

Contents

Glossary	xii
Acronyms	xiii
1 Introduction	1
1.1 Cancer Research in the Post-Genomic Era	1
1.1.1 Cancer as a Global Health Concern	2
1.1.1.1 Genetics and Molecular Biology in Cancers	3
1.1.2 The Human Genome Revolution	5
1.1.2.1 The First Human Genome Sequence	6
1.1.2.2 Impact of Genomics	6
1.1.3 Technologies to Enable Genetics Research	7
1.1.3.1 DNA Sequencing and Genotyping Technologies	7
1.1.3.2 Microarrays and Quantitative Technologies	7
1.1.3.3 Massively Parallel “Next Generation” Sequencing	8
1.1.3.3.1 Molecular Profiling with Genomics Technology	10
1.1.3.3.2 Sequencing Technologies	10
1.1.3.4 Bioinformatics as Interdisciplinary Genomic Analysis	11
1.1.4 Follow-up Large-Scale Genomics Projects	12
1.1.5 Cancer Genomes	13
1.1.5.1 The Cancer Genome Atlas Project	14
1.1.5.1.1 Findings from Cancer Genomes	14
1.1.5.1.2 Genomic Comparisons Across Cancer Tissues	16
1.1.5.1.3 Cancer Genomic Data Resources	17
1.1.6 Genomic Cancer Medicine	17
1.1.6.1 Cancer Genes and Driver Mutations	18
1.1.6.2 Personalised or Precision Cancer Medicine	18
1.1.6.2.1 Molecular Diagnostics and Pan-Cancer Medicine	19
1.1.6.3 Targeted Therapeutics and Pharmacogenomics	20
1.1.6.3.1 Targeting Oncogenic Driver Mutations	20
1.1.6.4 Systems and Network Biology	21
1.1.6.4.1 Network Medicine, and Polypharmacology	23
1.2 A Synthetic Lethal Approach to Cancer Medicine	24
1.2.1 Synthetic Lethal Genetic Interactions	25
1.2.2 Synthetic Lethal Concepts in Genetics	25
1.2.3 Studies of Synthetic Lethality	26

1.2.3.1	Synthetic Lethal Pathways and Networks	27
1.2.3.1.1	Evolution of Synthetic Lethality	28
1.2.4	Synthetic Lethal Concepts in Cancer	28
1.2.5	Clinical Impact of Synthetic Lethality in Cancer	30
1.2.6	High-throughput Screening for Synthetic Lethality	32
1.2.6.1	Synthetic Lethal Screens	33
1.2.7	Computational Prediction of Synthetic Lethality	36
1.2.7.1	Bioinformatics Approaches to Genetic Interactions . .	36
1.2.7.2	Comparative Genomics	37
1.2.7.3	Analysis and Modelling of Protein Data	40
1.2.7.4	Differential Gene Expression	42
1.2.7.5	Data Mining and Machine Learning	43
1.2.7.6	Bimodality	46
1.2.7.7	Rationale for Further Development	47
1.3	E-cadherin as a Synthetic Lethal Target	47
1.3.1	The <i>CDH1</i> gene and it's Biological Functions	47
1.3.1.1	Cytoskeleton	48
1.3.1.2	Extracellular and Tumour Micro-Environment	48
1.3.1.3	Cell-Cell Adhesion and Signalling	48
1.3.2	<i>CDH1</i> as a Tumour (and Invasion) Suppressor	49
1.3.2.1	Breast Cancers and Invasion	49
1.3.3	Hereditary Diffuse Gastric Cancer and Lobular Breast Cancer .	49
1.3.4	Somatic Mutations	51
1.3.4.1	Mutation Rate	51
1.3.4.2	Co-occurring Mutations	51
1.3.5	Models of <i>CDH1</i> loss in cell lines	52
1.4	Summary and Research Direction of Thesis	53
2	Methods and Resources	57
2.1	Bioinformatics Resources for Genomics Research	57
2.1.1	Public Data and Software Packages	57
2.1.1.1	Cancer Genome Atlas Data	58
2.1.1.2	Reactome and Annotation Data	59
2.2	Data Handling	60
2.2.1	Normalisation	60
2.2.2	Sample Triage	60
2.2.3	Metagenes and the Singular Value Decomposition	62
2.2.3.1	Candidate Triage and Integration with Screen Data . .	62
2.3	Techniques	63
2.3.1	Statistical Procedures and Tests	63
2.3.2	Gene Set Over-representation Analysis	64
2.3.3	Clustering	65
2.3.4	Heatmap	65
2.3.5	Modeling and Simulations	65
2.3.5.1	Receiver Operating Characteristic (Performance) . . .	66
2.3.6	Resampling Analysis	67

2.4	Pathway Structure Methods	68
2.4.1	Network and Graph Analysis	68
2.4.2	Sourcing Graph Structure Data	69
2.4.3	Constructing Pathway Subgraphs	69
2.4.4	Network Analysis Metrics	69
2.5	Implementation	70
2.5.1	Computational Resources and Linux Utilities	70
2.5.2	R Language and Packages	72
2.5.3	High Performance and Parallel Computing	74
3	Methods Developed During Thesis	76
3.1	A Synthetic Lethal Detection Methodology	76
3.2	Synthetic Lethal Simulation and Modelling	79
3.2.1	A Model of Synthetic Lethality in Expression Data	79
3.2.2	Simulation Procedure	83
3.3	Detecting Simulated Synthetic Lethal Partners	86
3.3.1	Binomial Simulation of Synthetic lethality	86
3.3.2	Multivariate Normal Simulation of Synthetic lethality	88
3.3.2.1	Multivariate Normal Simulation with Correlated Genes	91
3.3.2.2	Specificity with Query-Correlated Pathways	98
3.3.2.2.1	Importance of Directional Testing	98
3.4	Graph Structure Methods	100
3.4.1	Upstream and Downstream Gene Detection	100
3.4.1.1	Permutation Analysis for Statistical Significance	101
3.4.1.2	Ranking Based on Biological Context	102
3.4.2	Simulating Gene Expression from Graph Structures	103
3.5	Customised Functions and Packages Developed	107
3.5.1	Synthetic Lethal Interaction Prediction Tool	107
3.5.2	Data Visualisation	108
3.5.3	Extensions to the iGraph Package	110
3.5.3.1	Sampling Simulated Data from Graph Structures	110
3.5.3.2	Plotting Directed Graph Structures	110
3.5.3.3	Computing Information Centrality	111
3.5.3.4	Testing Pathway Structure with Permutation Testing	111
3.5.3.5	Metapackage to Install iGraph Functions	112
4	Synthetic Lethal Analysis of Gene Expression Data	113
4.1	Synthetic lethal genes in breast cancer	114
4.1.1	Synthetic lethal pathways in breast cancer	116
4.1.2	Expression profiles of synthetic lethal partners	117
4.1.2.1	Subgroup pathway analysis	120
4.2	Comparison of synthetic lethal gene candidates	123
4.2.1	Comparison with siRNA screen candidates	123
4.2.1.1	Comparison with correlation	124
4.2.1.2	Comparison with viability	125
4.2.1.3	Comparison with secondary siRNA screen candidates	129

4.2.1.4	Comparison of screen at pathway level	129
4.2.1.4.1	Resampling of genes for pathway enrichment . .	131
4.3	Metagene Analysis	137
4.3.1	Pathway expression	137
4.3.2	Somatic mutation	140
4.3.3	Mutation locus	141
4.3.4	Synthetic lethal metagenes	143
4.4	Replication in stomach cancer	145
4.4.1	Synthetic Lethal Genes and Pathways	145
4.4.2	Synthetic Lethal Expression Profiles	147
4.4.3	Comparison to Primary Screen	149
4.4.3.1	Resampling Analysis	150
4.4.4	Metagene Analysis	150
4.5	Global Synthetic Lethality	151
4.5.1	Hub Genes	152
4.5.2	Hub Pathways	154
4.6	Replication in cell line encyclopaedia	155
4.7	Discussion	157
4.7.1	Strengths of the SLIPT Methodology	157
4.7.2	Synthetic Lethal Pathways for E-cadherin	158
4.7.3	Replication and Validation	160
4.7.3.1	Integration with siRNA Screening	160
4.7.3.2	Replication across Tissues and Cell lines	161
4.8	Summary	162
5	Synthetic Lethal Pathway Structure	165
5.1	Synthetic Lethal Genes in Reactome Pathways	167
5.1.1	The PI3K/AKT Pathway	167
5.1.2	The Extracellular Matrix	169
5.1.3	G Protein Coupled Receptors	171
5.1.4	Gene Regulation and Translation	172
5.2	Network Analysis of Synthetic Lethal Genes	173
5.2.1	Gene Connectivity and Vertex Degree	173
5.2.2	Gene Importance and Centrality	173
5.2.2.1	Information Centrality	173
5.2.2.2	PageRank Centrality	174
5.3	Testing Pathway Structure of Synthetic Lethal Genes	175
5.3.1	Hierarchical Pathway Structure	175
5.3.1.1	Contextual Ranking of PI3K	175
5.3.1.2	Testing Contextual Ranking of Synthetic Lethal Genes	175
5.3.2	Upstream or Downstream Synthetic Lethality	177
5.3.2.1	Measuring Structure of Candidates within PI3K	177
5.3.2.2	Resampling for Synthetic Lethal Pathway Structure . .	177
5.4	Discussion	178
5.5	Conclusion	178

6	Simulation and Modeling of Synthetic Lethal Pathways	179
6.1	Simulations and Modelling Synthetic Lethality in Expression Data . . .	182
6.2	Simulations over simple graph structures	183
6.2.1	Performance	183
6.2.2	Synthetic lethality across graph structures	183
6.2.3	Performance with inhibition links	183
6.2.4	Performance with 20,000 genes	183
6.3	Simulations over pathway-based graphs	183
6.4	Comparing methods	183
6.4.1	SLIPT and Chi-Squared	183
6.4.1.1	Correlated query genes	183
6.4.2	Correlation	183
6.4.3	Bimodality with BiSEp	183
7	Discussion	184
7.1	Significance	186
7.2	Future Directions	187
7.3	Conclusion	188
8	Conclusion	190
	References	191
A	Sample Quality	216
A.1	Sample Correlation	216
A.2	Replicate Samples in TCGA Breast	219
B	Software Used for Thesis	223
C	Secondary Screen Data	232
D	Mutation Analysis in Breast Cancer	234
D.1	Synthetic Lethal Genes and Pathways	234
D.2	Synthetic Lethal Expression Profiles	237
J	Synthetic Lethal Genes in Pathways	285
K	Pathway Connectivity for Mutation SLIPT	293
L	Pathway Structure for Mutation SLIPT	295
M	Information Centrality for Gene Essentiality	296

List of Figures

1.1	Synthetic genetic interactions	26
1.2	Synthetic lethality in cancer	29
2.1	Read count density	61
2.2	Read count sample mean	61
3.1	Framework for synthetic lethal prediction	77
3.2	Synthetic lethal prediction adapted for mutation	78
3.3	A model of synthetic lethal gene expression	80
3.4	Modeling synthetic lethal gene expression	81
3.5	Synthetic lethality with multiple genes	82
3.6	Simulating gene function	84
3.7	Simulating synthetic lethal gene function	84
3.8	Simulating synthetic lethal gene expression	85
3.9	Performance of binomial simulations	87
3.10	Comparison of statistical performance	87
3.11	Performance of multivariate normal simulations	89
3.12	Simulating expression with correlated gene blocks	92
3.13	Simulating expression with correlated gene blocks	93
3.14	Synthetic lethal prediction across simulations	94
3.15	Performance with correlations	95
3.16	Comparison of statistical performance with correlation structure	96
3.17	Performance with query correlations	97
3.18	Statistical evaluation of directional criteria	98
3.19	Performance of directional criteria	99
3.20	Simulated graph structures	103
3.21	Simulating expression from a graph structure	105
3.22	Simulating expression from graph structure with inhibitions	106
3.23	Demonstration of violin plots with custom features	109
3.24	Demonstration of annotated heatmap	109
3.25	Simulating graph structures	111
4.1	Synthetic lethal expression profiles of analysed samples	119
4.2	Comparison of SLIPT to siRNA	123
4.3	Compare SLIPT and siRNA genes with correlation	124
4.4	Compare SLIPT and siRNA genes with correlation	124
4.5	Compare SLIPT and siRNA genes with siRNA viability	126

4.6	Compare SLIPT and siRNA genes with viability	126
4.7	Compare SLIPT and siRNA genes with siRNA viability	128
4.8	Resampled intersection of SLIPT and siRNA candidates	132
4.9	Pathway metagene expression profiles	138
4.10	Somatic mutation against PI3K metagene	140
4.11	Somatic mutation locus against expression	142
4.12	Synthetic lethal expression profiles of stomach samples	148
4.13	Synthetic lethal partners across query genes	152
5.1	Synthetic Lethality in the PI3K Cascade	168
5.2	Synthetic Lethality in the Elastic Fibre Formation Pathway	170
5.3	Synthetic Lethality in the Fibrin Clot Formation	171
5.4	Synthetic Lethality and Vertex Degree	173
5.5	Synthetic Lethality and Centrality	174
5.6	Synthetic Lethality and PageRank	174
5.7	Structure of PI3K Ranking	175
5.8	Structure of Synthetic Lethality in PI3K	176
5.9	Structure of Synthetic Lethality Resampling	177
A.1	Correlation profiles of removed samples	217
A.2	Correlation analysis and sample removal	218
A.3	Replicate excluded samples	219
A.4	Replicate samples with all remaining	220
A.5	Replicate samples with some excluded	221
A.5	Replicate samples with some excluded	222
D.1	Synthetic lethal expression profiles of analysed samples	238
J.1	Synthetic Lethality in the PI3K/AKT Pathway	285
J.2	Synthetic Lethality in the PI3K/AKT Pathway in Cancer	286
J.3	Synthetic Lethality in the Extracellular Matrix	287
J.4	Synthetic Lethality in the GPCRs	288
J.5	Synthetic Lethality in the GPCR Downstream	289
J.6	Synthetic Lethality in the Translation Elongation	290
J.7	Synthetic Lethality in the Nonsense-mediated Decay	291
J.8	Synthetic Lethality in the 3' UTR	292
K.1	Synthetic Lethality and Vertex Degree	293
K.2	Synthetic Lethality and Centrality	294
K.3	Synthetic Lethality and PageRank	294
M.1	Information centrality distribution	298

List of Tables

1.1	Methods for Predicting Genetic Interactions	37
1.2	Methods for Predicting Synthetic Lethality in Cancer	38
1.3	Methods used by Wu <i>et al.</i> (2014)	39
2.1	Excluded Samples by Batch and Clinical Characteristics.	62
2.2	Computers used during Thesis	71
2.3	Linux Utilities and Applications used during Thesis	71
2.4	R Installations used during Thesis	72
2.5	R Packages used during Thesis	72
2.6	R Packages Developed during Thesis	74
4.1	Candidate synthetic lethal gene partners of <i>CDH1</i> from SLIPT	115
4.2	Pathways for <i>CDH1</i> partners from SLIPT	117
4.3	Pathway composition for clusters of <i>CDH1</i> partners from SLIPT	121
4.4	Pathway composition for <i>CDH1</i> partners from SLIPT and siRNA screen- ing	130
4.5	Pathways for <i>CDH1</i> partners from SLIPT	134
4.6	Pathways for <i>CDH1</i> partners from SLIPT and siRNA primary screen .	135
4.7	Candidate synthetic lethal metagenes against <i>CDH1</i> from SLIPT	144
4.8	Pathways for <i>CDH1</i> partners from SLIPT in stomach cancer	146
4.9	Query synthetic lethal genes with the most SLIPT partners	153
4.10	Pathways for genes with the most SLIPT partners	154
4.11	Pathways for <i>CDH1</i> partners from SLIPT in CCLE	155
4.12	Pathways for <i>CDH1</i> partners from SLIPT in breast CCLE	157
5.1	Information centrality for genes and molecules in the Reactome network	178
B.1	R Packages used during Thesis	223
C.1	Comparing SLIPT genes against Secondary siRNA Screen in breast cancer	232
C.2	Comparing mtSLIPT genes against Secondary siRNA Screen in breast cancer	233
C.3	Comparing SLIPT genes against Secondary siRNA Screen in stomach cancer	233
D.1	Candidate synthetic lethal gene partners of <i>CDH1</i> from mtSLIPT . . .	235
D.2	Pathways for <i>CDH1</i> partners from mtSLIPT	236
D.3	Pathway composition for clusters of <i>CDH1</i> partners from mtSLIPT . .	239

L.1	Information centrality for genes and molecules in the Reactome network	295
M.1	Information centrality for genes and molecules in the Reactome network	297

Glossary

synthetic lethal Genetic interactions where inactivation of multiple genes is inviable (or deleterious) when they are viable if inactivated separately.

Acronyms

AMPK	AMP-activated protein kinase.
BMP	Bone morphogenic protein.
EMT	Epithelial-mesenchymal transition.
GPCR	G protein coupled receptor.
PDE	Phosphodiesterase.
PI3K	Phosphoinositide 3-kinase.
siRNA	Short interfering ribonucleic acid.
SLIPT	Synthetic lethal interaction prediction tool.
TGF β	Transforming growth factor β .

Chapter 5

Synthetic Lethal Pathway Structure

Aims

- Synthetic Lethal Genes within a Biological Pathway Structure
- Importance and Connectivity of Synthetic Lethal Genes within Pathway Networks
- Upstream and Downstream Relationships between SLIPT and siRNA Candidates

Summary

- Synthetic Lethal genes were explored within a graph structures for key pathways identified previously
- In some cases these graph structures appeared to have relationships between synthetic lethal genes
- However, no existing network metrics of importance and connectivity with the networks were elevated significantly for Synthetic Lethal genes
- Nor was there significant evidence of upstream and downstream relationships between SLIPT and siRNA Candidates in a shortest path permutation analysis

Having identified key pathways implicated in synthetic lethal genetic interactions with *CDH1*, these were investigated for the underlying synthetic lethal genes within them and their relationships to pathway structure in Reactome pathways. This chapter will focus on the pathway structure of biological pathways detected across analyses in Chapter 4. The synthetic lethal genes considered here are those candidates detected by SLIPT (as described in Section 3.1) in TCGA breast cancer expression and mutation data (TCGA, 2012) in comparison to the candidate gene partners from the siRNA screening in breast cell lines (Telford *et al.*, 2015).

The graph structure for Reactome pathways was obtained from Pathway Commons via BioPaX (as described in Section 2.4.2). The pathways describe the (directional) relationships between biomolecules, including proteins (encoded by genes), in biological pathways. These relationships include cell signalling (such as kinase phosphorylation cascades), gene regulation (such as transcription factors, chromatin modifiers, RNA binding proteins), and metabolism (such as the product of an enzyme being the substrate of another). Together these relationships describe the known functional pathways in a human cell with a reasonable resolution, from a curated database supported by publications documenting pathway relationships. While this functional pathway network encapsulates protein complexes and functional modules, protein binding or text-mining alone are not used to determine relationships between genes. The Reactome network is sufficient to test pathway relationships with directional information, although it is not documented whether these relationships are activating or inhibitory.

Pathway structures were derived from the Reactome network (as described in Section 2.4.3) for the graph structure of each biological pathway. The synthetic lethal candidate genes for notable pathways discussed in Chapter 4, including candidate synthetic lethal pathways of *CDH1*, were examined to show the SLIPT and siRNA candidates within these pathways. Thus synthetic lethal genes were identified within a biological context and with further investigations into their relationship with pathway structure and between synthetic lethal candidates detect by each approach. Synthetic lethal gene candidates in the context of pathway structures and additional support for belonging to a synthetic lethal pathway are ideal for triage of targets specific to *CDH1* deficient tumours and for further experimental validation in preclinical models.

Network analysis metrics (as described in Sections 2.4.4 and 3.5.3) were applied to test whether gene detected by SLIPT, siRNA, or both approaches varied according to these network analysis metrics (of connectivity and importance in the network) to test whether they differed between synthetic lethal genes or approaches to detect them.

Another consideration is the relationships between synthetic lethal candidates detected by either approach: these were tested by both a resampling approach (as described in Sections 3.4.1 and 3.4.1.1) and compared across a ranking based on biological context (Section 3.4.1.2). Together these approaches serve to test the pathway relationships between SLIPT and siRNA synthetic lethal gene candidate partners for *CDH1* within the biological pathways identified and demonstrate a combination of network biology and statistical investigations into structural relationships between genes identified by a Bioinformatics analysis.

5.1 Synthetic Lethal Genes in Reactome Pathways

5.1.1 The PI3K/AKT Pathway

The phosphoinositide 3-kinase (PI3K) cascade signalling pathway exhibited unexpected results with metagene analyses (as discussed in Section 4.3). This pathway is also of interest because mediating signals between the G protein coupled receptors and regulation of protein translation which have both been strongly implicated to be synthetic lethal pathways with loss of *CDH1* function. All three of these pathways have are also subject to dysregulation in cancer and other diseases. Thus the PI3K cascade will be examined along with the most supported synthetic lethal pathways (as identified in Chapter 4).

The phosphoinositide 3-kinase (PI3K) pathway is also an ideal pathway to test pathway structure since it has an established direction of signal transduction from extracellular stimuli (and membrane bound receptors) to the inner mechanisms of the cell, namely the regulation of protein translation. The production of proteins is necessary for the growth of the cell so it is reasonable to suggest that these processes may be subject to (non-oncogene) addiction in some cancer cells which rely upon them for sustained protein production and cell growth. This is also supported by the oncogenes *PIK3CA* and *AKT1* being involved with the PI3K cascade and related PI3K/AKT pathway which may be subject to oncogene addiction when these proto-oncogenes are activated.

Despite the PI3K cascade not being supported across SLIPT and siRNA analysis by over-representation (in Section 4.2.1.4) or resampling (in Section 4.2.1.4.1), numerous genes were detected by either Synthetic Lethal Interaction Prediction Tool (SLIPT) in TCGA breast expression data or the short interfering ribonucleic acid (siRNA) primary screen (as shown in Figure 5.1). It is also notable, that of the few genes that

were identified by both approaches, these include genes that are highly connected in the PI3K cascade and are hubs to information transmission such as *FGF9*, *PDE3B*, and *PDE4A*. The key upstream genes *PIK3CA* and *PIK3CG* were detected by siRNA whereas the downstream *PIK3R1* and *AKT2* genes were detected by SLIPT. Gene detected by either method were also prevalent in the PI3K, phosphodiesterase (PDE), and AMP-activated protein kinase (AMPK) modules, in addition to the downstream translation factors and ribosomal genes (*EIF4B*, *EEF2K*, and *RPS6*). Together these suggest that there may further be structure between the SLIPT and siRNA candidates partners of *CDH1* in pathways such as this example. As such, pathway structure will be tested to detect differences in the upstream and downstream gene candidates of those detected by either method. This may further explain the disparity between SLIPT and siRNA genes, even in pathways such as PI3K where they did not significantly intersect.

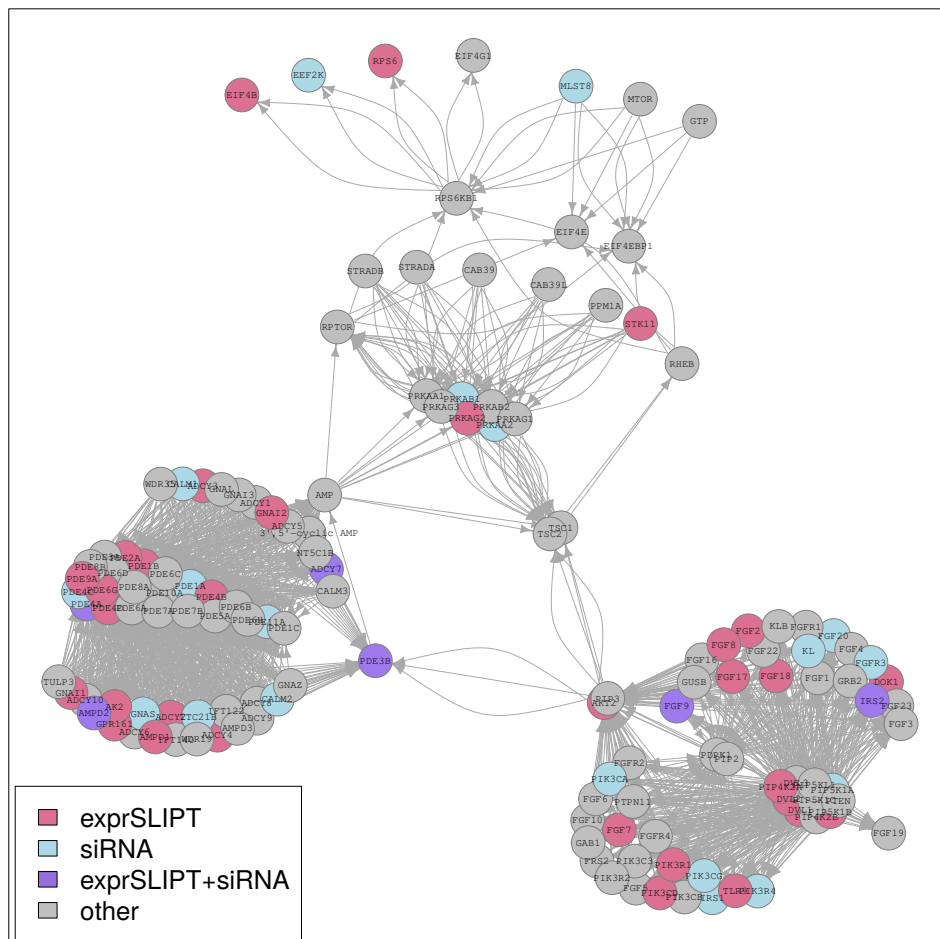


Figure 5.1: **Synthetic Lethality in the PI3K Cascade.** The Reactome PI3K Cascade pathway with synthetic lethal candidates coloured as shown in the Legend.

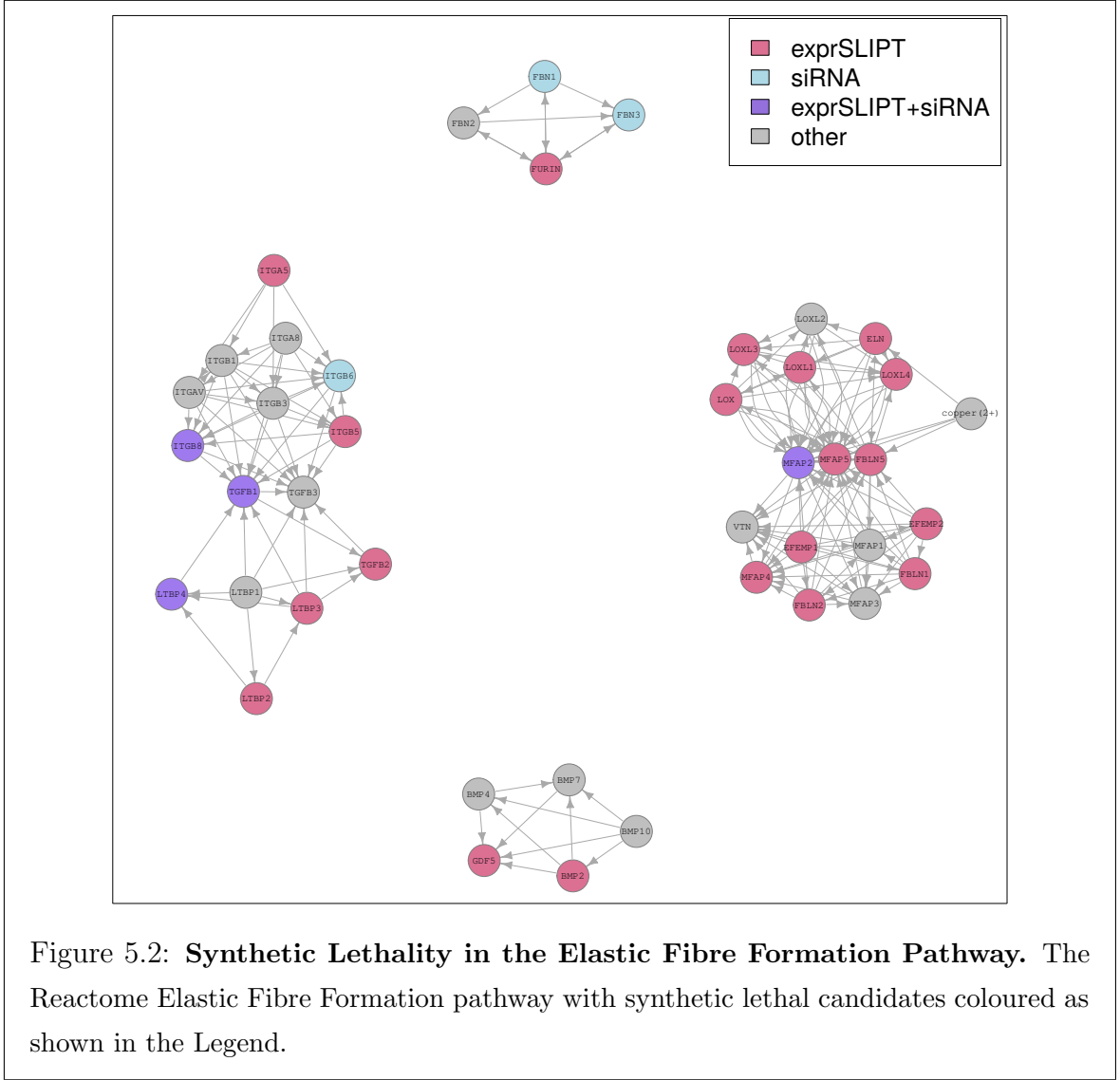
This disparity between SLIPT and siRNA gene candidate synthetic lethal partners of *CDH1*, that is a high number of genes detected by either approach with few detected by both, was replicated the related PI3K/AKT pathway and the “PI3K/AKT in cancer” pathway (shown in Figures J.1 and J.2). With many synthetic lethal candidates at the upstream core of these pathway networks and the downstream extremities. It is particularly notable that the many genes important in cell signalling and gene regulation were detected by either sytnhetic lethal detection approach. These include *AKT1*, *AKT2*, and *AKT3*, the Calmodulin signalling genes *CALM1* and *CAMK4*, and the forkhead family transcription factors *FOXO1* (a tumour suppressor) and *FOXO4* and inhibitor of epithelial-mesenchymal transition (EMT).

5.1.2 The Extracellular Matrix

The extracellular pathways elastic fibre formation and fibrin clot formation (shown in Figures 5.2 and 5.3 respectively) were both supported across analyses (in Chapter 4). This includes a significant over-representation and resampling the interaction between SLIPT (for TCGA breast cancer) and siRNA gene candidates showing that SLIPT has identified these pathways in addition to their over-representation in the siRNA screen.

Particularly for elastic fibres (in Figure 5.2), the vast majority of genes were detected by either approach in addition to a significant proportion of genes detected by both approaches (as determined in Section 4.2.1.4). The genes detected by both approaches also appeared to have a non-random distribution in the network with *TFGB1*, *ITGB8*, and *MFAP2* exhibiting high connectivity and a cental role in their respective pathway modules. In addition to a structural role in the extracellular matrix and connective tissue (including the tumour microenvironment), these proteins including Furin, transforming growth factor β (TGF β), and the bone morphogenic proteins (BMPs), are also involved in responses to endocrine signals and interacting with the cellular receptors for signalling pathways. Therefore it is plausible that *CDH1* deficient tumours will be subject to non-oncogene addiction to the extracellular environment and growth signals arising from this pathway. The pathway structure is also worth further investigation into whether the genes detected by siRNA or both approaches are downstream of those detected by SLIPT in addition to whether they have higher connectivity or centrality than other genes in the pathway.

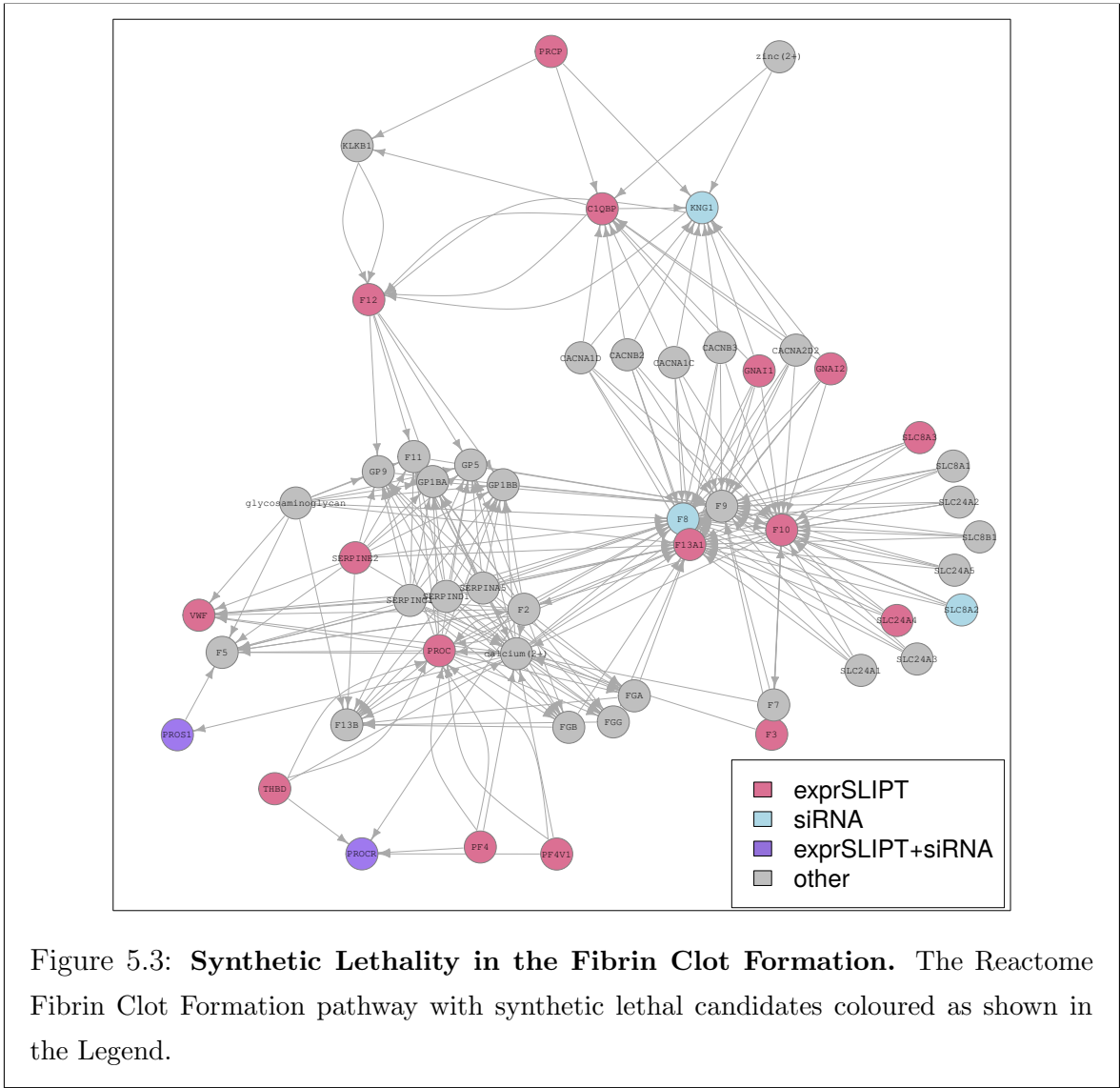
Genes detected as synthetic lethal partners of *CDH1* by SLIPT or siRNA screening were also common in the Fibrin clot formation pathway (shown in Figure 5.3). This is consistent with the established pleiotropic role of *CDH1* in regulating fibrin clotting. It



is also notable that the genes detected by either method appear to be highly connected such as *C1QBP*, *KNG1*, *F8*, *F10*, *F12*, *F13A*, and *PROC* (including many of the coagulation factors). Synthetic lethal candidates also include *SERPINE2* and *PRCP*, which only affect downstream genes, in addition to *PROCR* and *VWF*, which are only affected by upstream genes.

Many of these genes are involved in the larger Extracellular Matrix pathway (shown in Figure J.3), including many of the synthetic lethal candidates discussed for elastic fibres. The number of SLIPT candidate genes outnumbers those identified by siRNA as expected from an isolated cell model. However, the endocrine response genes (such as *TGFB1* and *LTBP4*) which are potentially artifacts of the cell line growth process were replicated with SLIPT analysis in patient tumours (TCGA breast cancer data).

There is also additional support for synthetic lethal genes such as *ITGB2*, *MFAP2*, and *SPARC* being highly connected networks hubs of the pathway. Although the complexity of extracellular matrix pathway lends credence to the need for formal network analysis approaches to aid interpretation of the structure and relationships among synthetic lethal candidates in a pathway network, in addition to statistical approaches to determine whether such relationships are unlikely to be observed by sampling error.



5.1.3 G Protein Coupled Receptors

Similarly, G protein coupled receptor (GPCR) pathways are highly complex (as shown in Figures J.4 and J.5). Many of these were synthetic lethal candidates by either SLIPT or siRNA screening with many detected with both approaches, consistent with these path-

ways being supported by prior analyses (in Sections 4.2.1.4 and 4.2.1.4.1). Synthetic lethal candidates include the PDE and Calmodulin genes (as discussed in Section 5.1.3) in addition to others such as the regulators of G-protein signaling (RGS), chemokine receptors (CXCR), Janus kinase (JAK), and the Ras homolog family (RHO) genes. These are important regulatory signalling pathways necessary for cellular growth and cancer proliferation. Thus the GPCR pathways (and downstream PI3K/AKT signals) are a potentially actionable vulnerability against *CDH1* deficient cancers, particularly since many existing drug targets exist among these signalling pathways, some of which have been experimentally validated (Kelly *et al.*, 2017b; Telford *et al.*, 2015). However, the complexity of GPCR networks containing hundreds of genes requires the relationships between SLIPT and experimental candidates to be tested with a network based statistical approach, although a statistically significant intersection of these approaches has been established (in Sections 4.2.1.4 and 4.2.1.4.1).

5.1.4 Gene Regulation and Translation

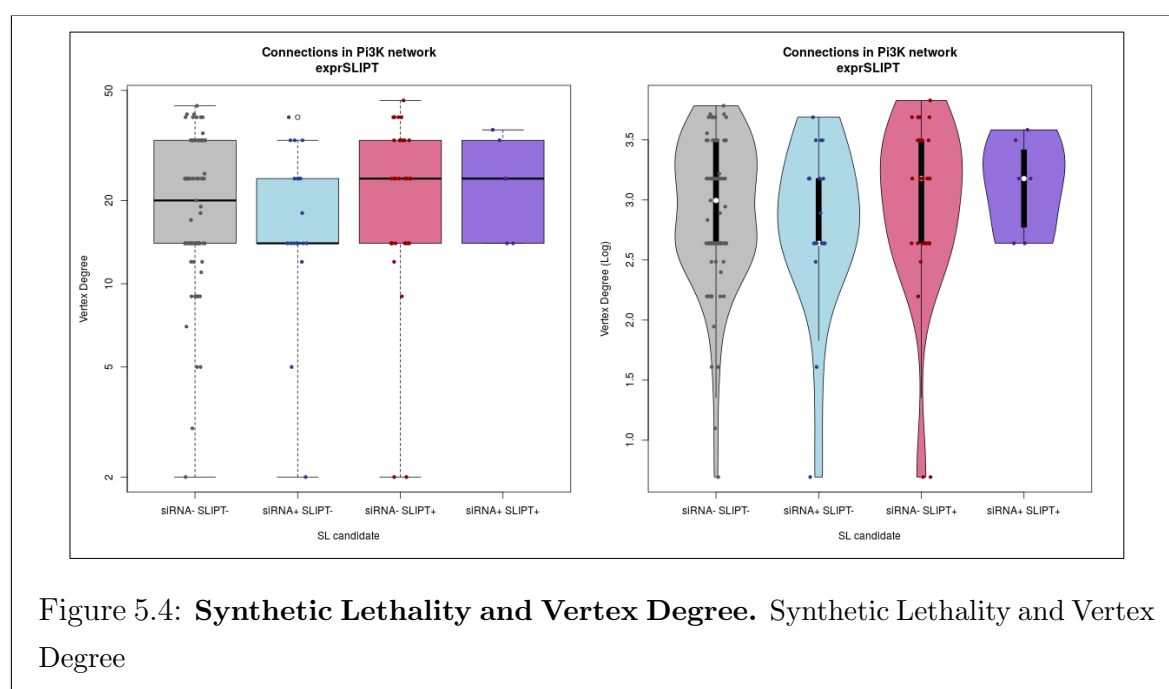
While very few synthetic lethal genes were detected in translational pathways in an experimental screen against *CDH1* Telford *et al.* (2015), these were highly over-represented in translational elongation (as shown in Figure J.6). These SLIPT genes include many ribosomal proteins and the regulatory “elongation factors” which may be subject to responses in the upstream signalling pathways. This observation lends support the notion of pathway structure among synthetic lethal candidates detected by SLIPT in comparison with siRNA as the computational approach with SLIPT has demonstrated the ability to detect downstream genes in the core translational processes which experimental screening did not identify. Although it is possible the experimental screening may detect upstream regulatory genes that are less sensitive inactivation, that is less likely to be indiscriminately lethal to both genotypes at high doses of inactivation.

Many of these SLIPT candidate genes are also among the nonsense-mediated decay (NMD) pathway (shown in Figure J.7) or 3' untranslated region (UTR) mediated translational regulation (shown in Figure J.8). While genes in these pathways were also supported by experimental screening with siRNA, there was clear pathway structure. In particular, *UPF1* was detected in the siRNA screen and is the focal downstream gene for the entire NMD pathway showing that (in this case) siRNA genes are downstream effectors of those detected by SLIPT. 3' UTR mediated translational regulation has a similar structure with two modules connected solely by *RPL13A*, giving an example of SLIPT candidates genes with high connectivity, although there were many

ribosomal proteins detected by SLIPT. However, *EIF3K* a regulatory elongation factor (not essential to ribosomal function) that was detected by SLIPT was replicated with siRNA screening while the majority of the elongation factors were not detected by either approach. Regulatory genes being more amenable to experimental validation also support further investigation into pathway structure as the SLIPT candidates may support them by structural relationships and the downstream genes not being detectable by experimental screening with high dose inhibitors may explain the greater number of SLIPT candidate partners of *CDH1* than those experimentally identified.

5.2 Network Analysis of Synthetic Lethal Genes

5.2.1 Gene Connectivity and Vertex Degree



5.2.2 Gene Importance and Centrality

5.2.2.1 Information Centrality

Information centrality is a measure of the importance of nodes in a network but how vital they are to the transmission of information throughout the network. This naturally applies well to biological pathways, particularly gene regulation and cell signalling. The nodes with the highest information centrality are not necessarily the most connected

as they may also include nodes which pass signals between highly connected network hubs.

Information centrality has also been suggested to indicate essentiality of genes or proteins (Kranthi *et al.*, 2013). See also Appendix M on gene essentiality.

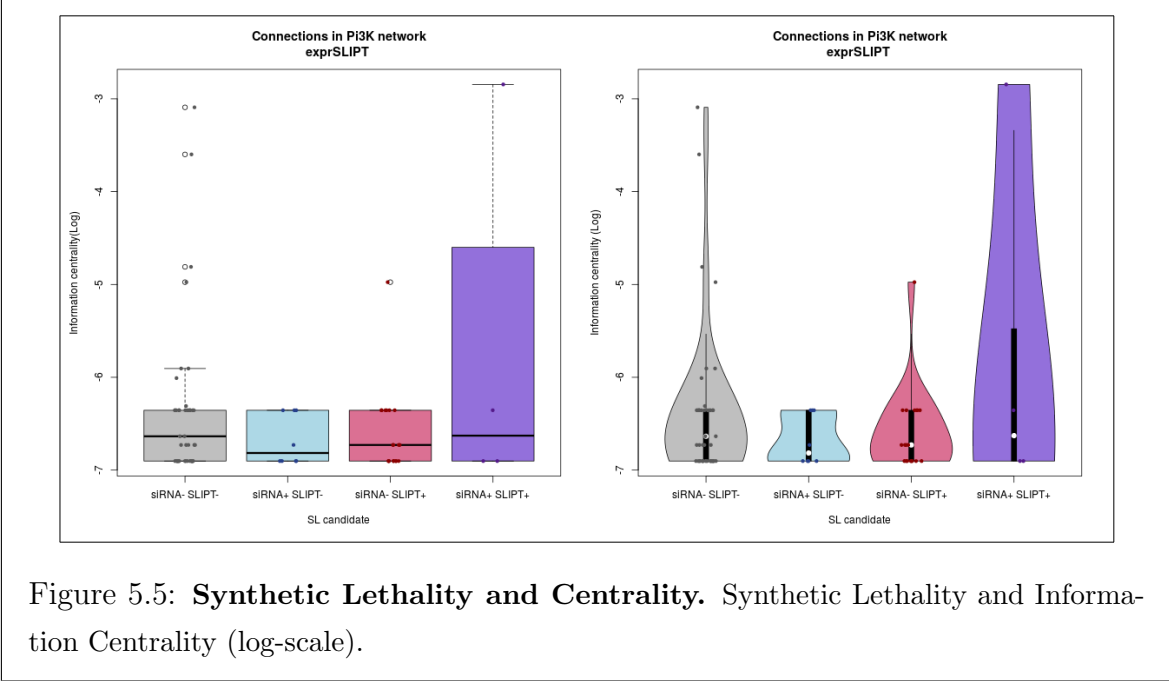


Figure 5.5: **Synthetic Lethality and Centrality.** Synthetic Lethality and Information Centrality (log-scale).

5.2.2.2 PageRank Centrality

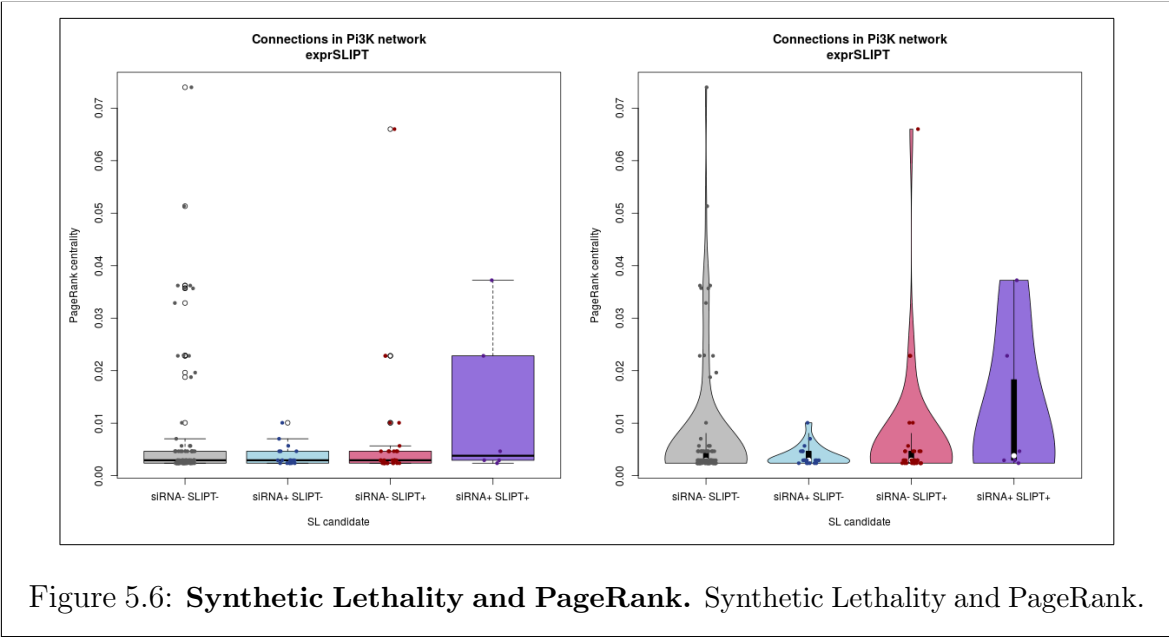
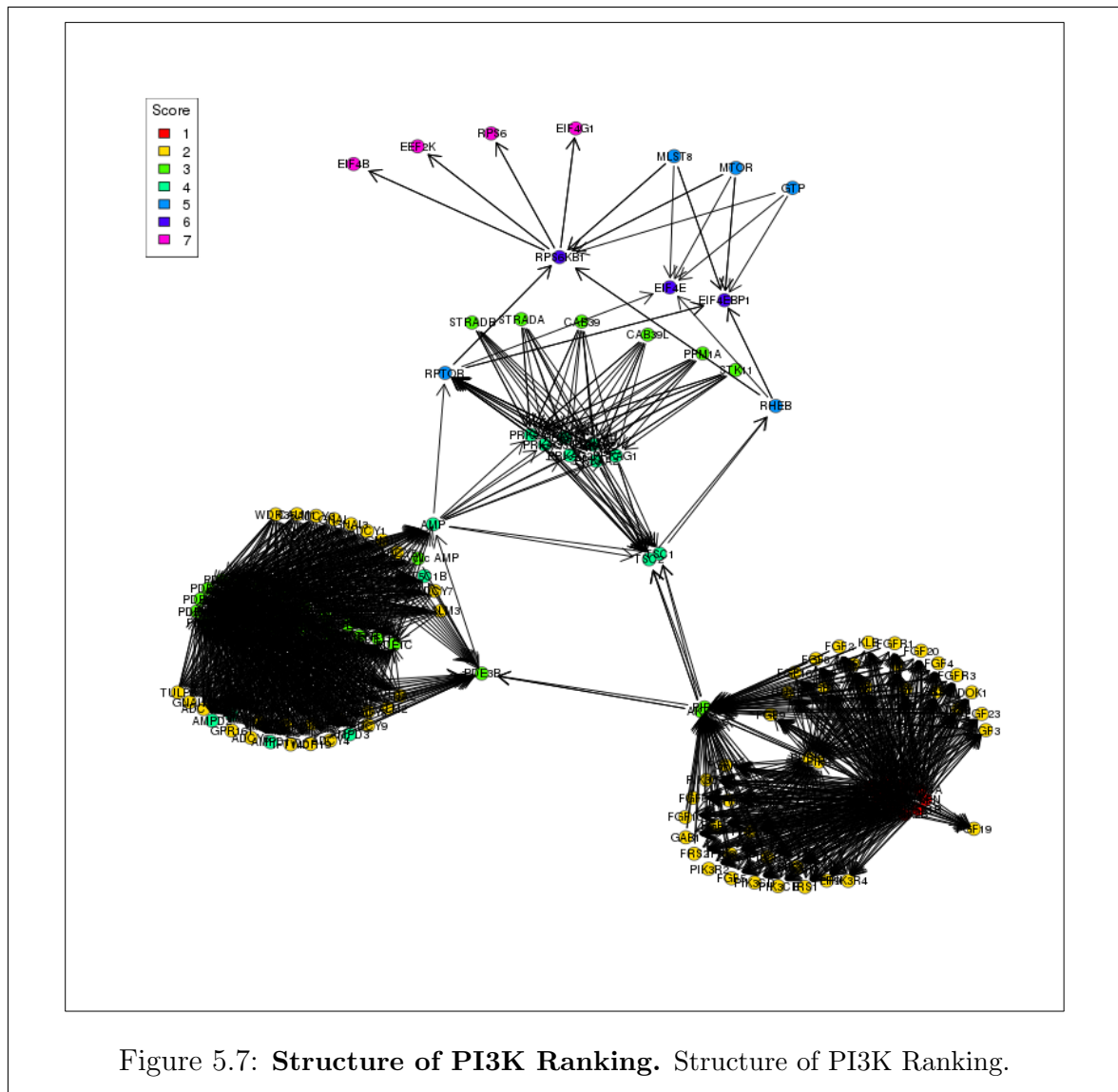


Figure 5.6: **Synthetic Lethality and PageRank.** Synthetic Lethality and PageRank.

5.3 Testing Pathway Structure of Synthetic Lethal Genes

5.3.1 Hierarchical Pathway Structure

5.3.1.1 Contextual Ranking of PI3K



5.3.1.2 Testing Contextual Ranking of Synthetic Lethal Genes

- Are there more SL genes of a particular rank?
- Is there an association with SLIPT (Chi-sq) or siRNA (viability) score?

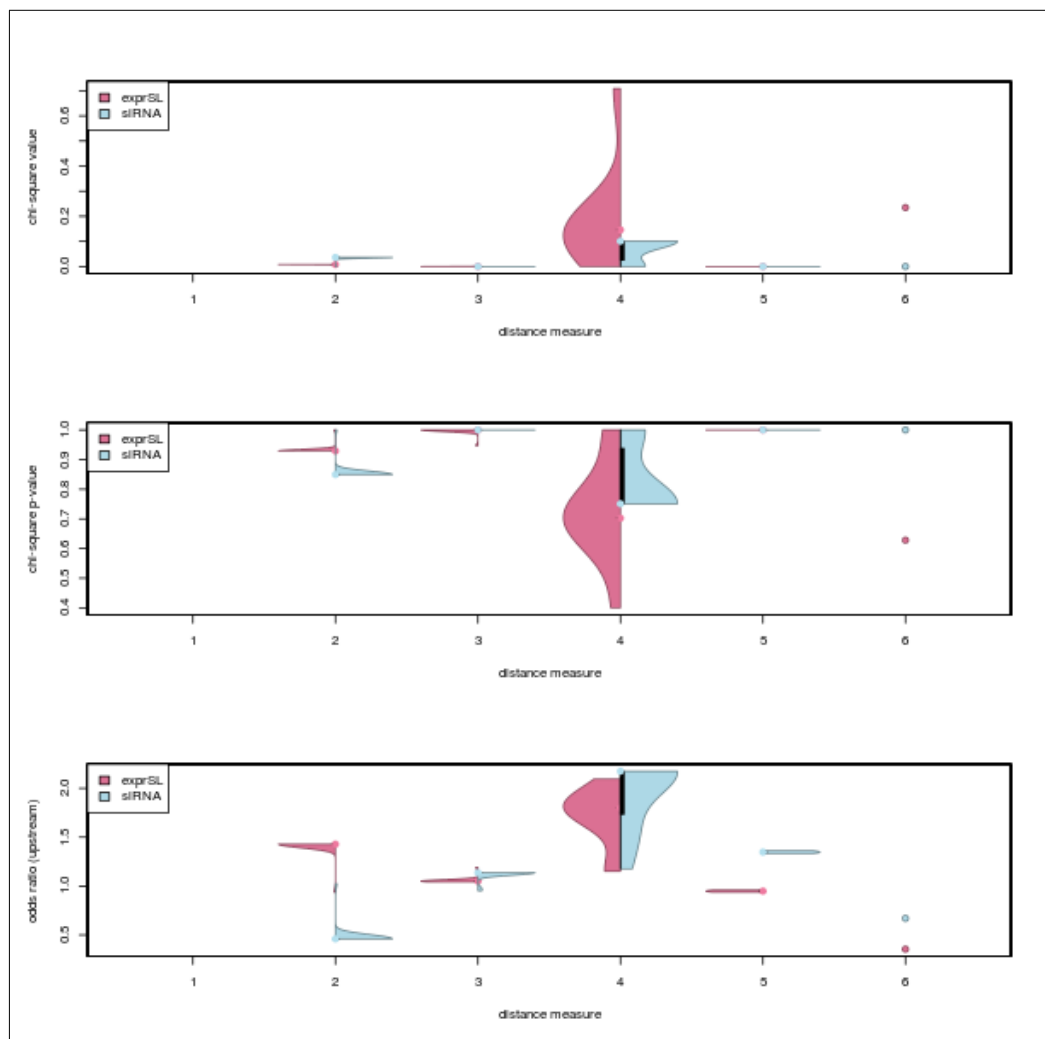
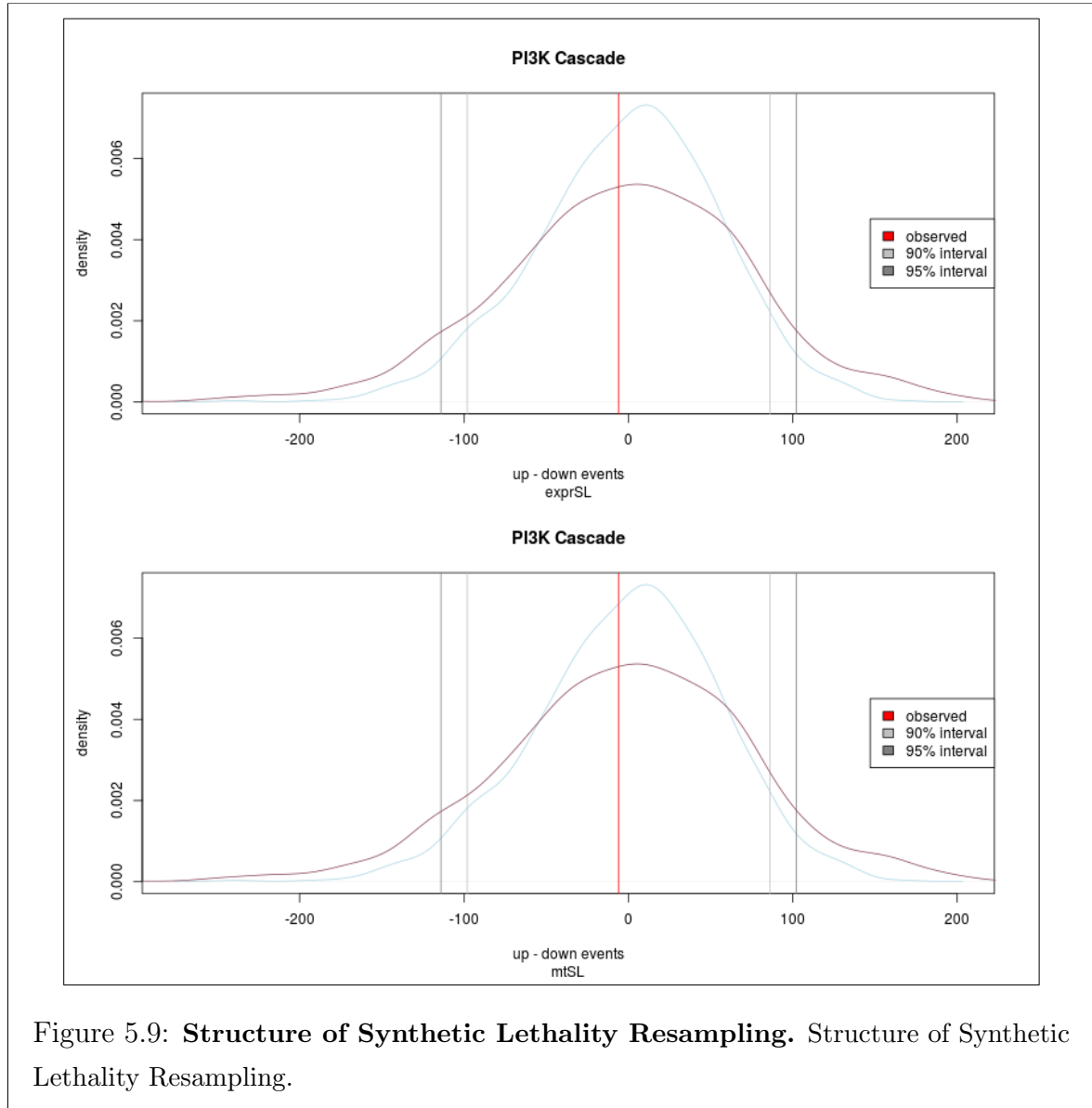


Figure 5.8: **Structure of Synthetic Lethality in PI3K.** Structure of Synthetic Lethality in PI3K.

5.3.2 Upstream or Downstream Synthetic Lethality

5.3.2.1 Measuring Structure of Candidates within PI3K



5.3.2.2 Resampling for Synthetic Lethal Pathway Structure

Table 5.1: Information centrality for genes and molecules in the Reactome network

Pathway	Graph:		States:		Observed:				Permutation p-value:	
	Nodes	Edges	SLIPT	siRNA	Up	Down	Up-Down	Up/Down	Up-Down	Down-Up
PI3K Cascade	138	1495	38	25	122	128	-6	0.953	0.5326	0.4606
PI3K/AKT Signaling in Cancer	275	12882	98	44	779	679	100	1.147	0.3255	0.6734
G _{αi} Signaling	292	22003	95	58	836	1546	-710	0.541	0.9971	0.0029
GPCR downstream	1270	142071	312	160	9755	9261	494	1.053	0.3692	0.6305
Elastic fibre formation	42	175	24	7	1	2	-1	0.500	0.5461	0.3865
Extracellular matrix	299	3677	127	29	547	455	92	1.202	0.3351	0.6636
Formation of Fibrin	52	243	18	5	12	17	-5	0.706	0.6198	0.3564
Nonsense-Mediated Decay	103	102	74	2	0	74	-74	0	1.0000	0.0000
3' -UTR-mediated translational regulation	107	2860	77	1	0	0	0	NaN	0.4902	0.5027
Eukaryotic Translation Elongation	92	3746	76	0	0	0	0	NaN	0.4943	0.4933

Pathways in the Reactome network tested for structural relationships between SLIPT and siRNA genes by resampling (raw p-value)

Significant resampling in bold

Sampling only within target pathway

Number of siRNA+SLIPT matched to observed

siRNA+SLIPT kept for up/down evaluation

5.4 Discussion

5.5 Conclusion

References

- 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.
- Adamski, M.G., Gumann, P., and Baird, A.E. (2014) A method for quantitative analysis of standard and high-throughput qPCR expression data based on input sample quantity. *PLoS ONE*, **9**(8): e103917.
- Adler, D. (2005) *vioplot: Violin plot*. R package version 0.2.
- Agarwal, S., Deane, C.M., Porter, M.A., and Jones, N.S. (2010) Revisiting date and party hubs: Novel approaches to role assignment in protein interaction networks. *PLoS Comput Biol*, **6**(6): e1000817.
- Agrawal, N., Akbani, R., Aksoy, B.A., Ally, A., Arachchi, H., Asa, S.L., Auman, J.T., Balasundaram, M., Balu, S., Baylin, S.B., *et al.* (2014) Integrated genomic characterization of papillary thyroid carcinoma. *Cell*, **159**(3): 676–690.
- 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 Paediatrica*, **96**(5): 644–647.
- American Cancer Society (2017) Genetics and cancer. <https://www.cancer.org/cancer/cancer-causes/genetics.html>. Accessed: 22/03/2017.

- American Society for Clinical Oncology (ASCO) (2017) The genetics of cancer. <http://www.cancer.net/navigating-cancer-care/cancer-basics/genetics/genetics-cancer>. Accessed: 22/03/2017.
- 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.
- 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. 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.].
- Beck, T.F., Mullikin, J.C., and Biesecker, L.G. (2016) Systematic Evaluation of Sanger Validation of Next-Generation Sequencing Variants. *Clin Chem*, **62**(4): 647–654.
- 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.
- 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.

- 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.
- Burk, R.D., Chen, Z., Saller, C., Tarvin, K., Carvalho, A.L., Scapulatempo-Neto, C., Silveira, H.C., Fregnani, J.H., Creighton, C.J., Anderson, M.L., *et al.* (2017) Integrated genomic and molecular characterization of cervical cancer. *Nature*, **543**(7645): 378–384.
- 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.
- Cancer Research UK (2017) Family history and cancer genes. <http://www.cancerresearchuk.org/about-cancer/causes-of-cancer/inherited-cancer-genes-and-increased-cancer-risk/family-history-and-inherited-cancer-genes>. Accessed: 22/03/2017.
- Cancer Cell Line Encyclopedia (CCLE) (2014) Broad-Novartis Cancer Cell Line Encyclopedia. <http://www.broadinstitute.org/ccle>. Accessed: 07/11/2014.
- 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, K., Yang, D., Li, X., Sun, B., Song, F., Cao, W., Brat, D.J., Gao, Z., Li, H., Liang, H., *et al.* (2015) Mutational landscape of gastric adenocarcinoma in Chinese: implications for prognosis and therapy. *Proc Natl Acad Sci USA*, **112**(4): 1107–1112.

- Chen, S. and Parmigiani, G. (2007) Meta-analysis of BRCA1 and BRCA2 penetrance. *J Clin Oncol*, **25**(11): 1329–1333.
- Chen, X. and Tompa, M. (2010) Comparative assessment of methods for aligning multiple genome sequences. *Nat Biotechnol*, **28**(6): 567–572.
- Cherniack, A.D., Shen, H., Walter, V., Stewart, C., Murray, B.A., Bowlby, R., Hu, X., Ling, S., Soslow, R.A., Broaddus, R.R., *et al.* (2017) Integrated Molecular Characterization of Uterine Carcinosarcoma. *Cancer Cell*, **31**(3): 411–423.
- 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., Gatz, 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 o-Dependent μ -Opioid Agonist-Mediated Adenylyl Cyclase Supersensitization. *Journal of Pharmacology and Experimental Therapeutics*, **310**(1): 215–222.
- Clough, E. and Barrett, T. (2016) The Gene Expression Omnibus Database. *Methods Mol Biol*, **1418**: 93–110.
- 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.
- Collins, F.S., Morgan, M., and Patrinos, A. (2003) The Human Genome Project: lessons from large-scale biology. *Science*, **300**(5617): 286–290.
- 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.

- Corcoran, R.B., Ebi, H., Turke, A.B., Coffee, E.M., Nishino, M., Cogdill, A.P., Brown, R.D., Della Pelle, P., Dias-Santagata, D., Hung, K.E., *et al.* (2012) Egfr-mediated re-activation of mapk signaling contributes to insensitivity of *BRAF*-mutant colorectal cancers to raf inhibition with vemurafenib. *Cancer Discovery*, **2**(3): 227–235.
- 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.
- 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.
- Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., *et al.* (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, **486**(7403): 346–352.
- 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.
- 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.
- Dienstmann, R. and Tabernero, J. (2011) *BRAF* as a target for cancer therapy. *Anti-cancer Agents Med Chem*, **11**(3): 285–95.
- 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., Fedysyn, 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.
- Dorogovtsev, S.N. and Mendes, J.F. (2003) *Evolution of networks: From biological nets to the Internet and WWW*. Oxford University Press, USA.
- 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. In *Publ. Math. Inst. Hung. Acad. Sci*, volume 5, 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.

- Ezkurdia, I., Juan, D., Rodriguez, J.M., Frankish, A., Diekhans, M., Harrow, J., Vazquez, J., Valencia, A., and Tress, M.L. (2014) Multiple evidence strands suggest that there may be as few as 19 000 human protein-coding genes. *Human Molecular Genetics*, **23**(22): 5866.
- 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.
- 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., 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.
- Gilbert, W. and Maxam, A. (1973) The nucleotide sequence of the lac operator. *Proceedings of the National Academy of Sciences*, **70**(12): 3581–3584.
- Git, A., Dvinge, H., Salmon-Divon, M., Osborne, M., Kutter, C., Hadfield, J., Bertone, P., and Caldas, C. (2010) Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. *RNA*, **16**(5): 991–1006.

- Globus (Globus) (2017) Research data management simplified. <https://www.globus.org/>. Accessed: 25/03/2017.
- 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.
- 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. (1999) E-cadherin downregulation in cancer: fuel on the fire? *Molecular Medicine Today*, **5**(4): 172 – 177.
- 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.
- Han, J.D.J., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J.M., Cusick, M.E., Roth, F.P., *et al.* (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*, **430**(6995): 88–93.
- 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.
- Heiskanen, M., Bian, X., Swan, D., and Basu, A. (2014) caArray microarray database in the cancer biomedical informatics gridTM (caBIGTM). *Cancer Research*, **67**(9 Supplement): 3712–3712.
- 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.
- Herschkowitz, J.I., Simin, K., Weigman, V.J., Mikaelian, I., Usary, J., Hu, Z., Rasmussen, K.E., Jones, L.P., Assefnia, S., Chandrasekharan, S., *et al.* (2007) Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol*, **8**(5): R76.
- 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.
- Holme, P. and Kim, B.J. (2002) Growing scale-free networks with tunable clustering. *Physical Review E*, **65**(2): 026107.
- 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.
- Illumina, Inc (Illumina) (2017) Sequencing and array-based solutions for genetic research. <https://www.illumina.com/>. Accessed: 26/03/2017.
- International HapMap 3 Consortium (HapMap) (2003) The International HapMap Project. *Nature*, **426**(6968): 789–796.
- International Human Genome Sequencing Consortium (IHGSC) (2004) Finishing the euchromatic sequence of the human genome. *Nature*, **431**(7011): 931–945.
- 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, Ikonf, J.C.B. Christopher, and J.S. Alexander (editors), *Advances in kernel methods*, 169–184. MIT Press.
- Ju, Z., Liu, W., Roebuck, P.L., Siwak, D.R., Zhang, N., Lu, Y., Davies, M.A., Akbani, R., Weinstein, J.N., Mills, G.B., *et al.* (2015) Development of a robust classifier for quality control of reverse-phase protein arrays. *Bioinformatics*, **31**(6): 912.
- 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.
- Kakiuchi, M., Nishizawa, T., Ueda, H., Gotoh, K., Tanaka, A., Hayashi, A., Yamamoto, S., Tatsuno, K., Katoh, H., Watanabe, Y., *et al.* (2014) Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. *Nat Genet*, **46**(6): 583–587.
- Kamada, T. and Kawai, S. (1989) An algorithm for drawing general undirected graphs. *Information Processing Letters*, **31**(1): 7–15.
- Kandoth, C., Schultz, N., Cherniack, A.D., Akbani, R., Liu, Y., Shen, H., Robertson, A.G., Pashtan, I., Shen, R., Benz, C.C., *et al.* (2013) Integrated genomic characterization of endometrial carcinoma. *Nature*, **497**(7447): 67–73.
- 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., Chen, A., Guilford, P., and Black, M. (2017a) Synthetic lethal interaction prediction of target pathways in E-cadherin deficient breast cancers. Submitted to *BMC Genomics*.
- Kelly, S.T. (2013) *Statistical Predictions of Synthetic Lethal Interactions in Cancer*. Dissertation, University of Otago.
- Kelly, S.T., Single, A.B., Telford, B.J., Beetham, H.G., Godwin, T.D., Chen, A., Black, M.A., and Guilford, P.J. (2017b) Towards HDGC chemoprevention: vulnerabilities

- in E-cadherin-negative cells identified by genome-wide interrogation of isogenic cell lines and whole tumors. Submitted to *Cancer Prev Res*.
- 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.
- 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.
- Lawrence, M.S., Sougnez, C., Lichtenstein, L., Cibulskis, K., Lander, E., Gabriel, S.B., Getz, G., Ally, A., Balasundaram, M., Birol, I., *et al.* (2015) Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*, **517**(7536): 576–582.
- 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, 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 anticancer 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.
- 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.
- Maxam, A.M. and Gilbert, W. (1977) A new method for sequencing DNA. *Proceedings of the National Academy of Science*, **74**(2): 560–564.
- McCourt, C.M., McArt, D.G., Mills, K., Catherwood, M.A., Maxwell, P., Waugh, D.J., Hamilton, P., O’Sullivan, J.M., and Salto-Tellez, M. (2013) Validation of next generation sequencing technologies in comparison to current diagnostic gold standards for BRAF, EGFR and KRAS mutational analysis. *PLoS ONE*, **8**(7): e69604.
- 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.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*, **5**(7): 621–628.
- 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.
- Neeley, E.S., Kornblau, S.M., Coombes, K.R., and Baggerly, K.A. (2009) Variable slope normalization of reverse phase protein arrays. *Bioinformatics*, **25**(11): 1384.
- Novomestky, F. (2012) *matrixcalc: Collection of functions for matrix calculations*. R package version 1.0-3.
- 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.
- Peltonen, L. and McKusick, V.A. (2001) Genomics and medicine. Dissecting human disease in the postgenomic era. *Science*, **291**(5507): 1224–1229.
- 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.
- Pleasance, E.D., Cheetham, R.K., Stephens, P.J., McBride, D.J., Humphray, S.J., Greenman, C.D., Varela, I., Lin, M.L., Ordóñez, G.R., Bignell, G.R., *et al.* (2010) A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature*, **463**(7278): 191–196.
- Polyak, K. and Weinberg, R.A. (2009) Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nat Rev Cancer*, **9**(4): 265–73.
- Prahalad, A., Sun, C., Huang, S., Di Nicolantonio, F., Salazar, R., Zecchin, D., Beijersbergen, R.L., Bardelli, A., and Bernards, R. (2012) Unresponsiveness of colon cancer to *BRAF*(v600e) inhibition through feedback activation of egfr. *Nature*, **483**(7387): 100–3.
- R Core Team (2016) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. R version 3.3.2.
- Ravnan, M.C. and Matalaka, M.S. (2012) Vemurafenib in patients with *BRAF* v600e mutation-positive advanced melanoma. *Clin Ther*, **34**(7): 1474–86.
- 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.
- Robin, J.D., Ludlow, A.T., LaRanger, R., Wright, W.E., and Shay, J.W. (2016) Comparison of DNA Quantification Methods for Next Generation Sequencing. *Sci Rep*, **6**: 24067.
- Robinson, M.D. and Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*, **11**(3): R25.
- 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.

- Rung, J. and Brazma, A. (2013) Reuse of public genome-wide gene expression data. *Nat Rev Genet*, **14**(2): 89–99.
- Rustici, G., Kolesnikov, N., Brandizi, M., Burdett, T., Dylag, M., Emam, I., Farne, A., Hastings, E., Ison, J., Keays, M., *et al.* (2013) ArrayExpress update—trends in database growth and links to data analysis tools. *Nucleic Acids Res*, **41**(Database issue): D987–990.
- Ryan, C., Lord, C., and Ashworth, A. (2014) Daisy: Picking synthetic lethals from cancer genomes. *Cancer Cell*, **26**(3): 306–308.
- Sander, J.D. and Joung, J.K. (2014) Crispr-cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*, **32**(4): 347–55.
- Sanger, F. and Coulson, A. (1975) A rapid method for determining sequences in dna by primed synthesis with dna polymerase. *Journal of Molecular Biology*, **94**(3): 441 – 448.
- 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.
- Sørlie, 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.
- 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.
- Sun, C., Wang, L., Huang, S., Heynen, G.J.J.E., Prahallad, A., Robert, C., Haanen, J., Blank, C., Wesseling, J., Willems, S.M., *et al.* (2014) Reversible and adaptive resistance to *BRAF*(v600e) inhibition in melanoma. *Nature*, **508**(7494): 118–122.
- Taylor, I.W., Linding, R., Warde-Farley, D., Liu, Y., Pesquita, C., Faria, D., Bull, S., Pawson, T., Morris, Q., and Wrana, J.L. (2009) Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nat Biotechnol*, **27**(2): 199–204.
- 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) (2012) Comprehensive molecular portraits of human breast tumours. *Nature*, **490**(7418): 61–70.
- The Cancer Genome Atlas Research Network (TCGA) (2017a) The Cancer Genome Atlas Project. <https://cancergenome.nih.gov/>. Accessed: 26/03/2017.
- The Cancer Genome Atlas Research Network (TCGA) (2017b) The Cancer Genome Atlas Project Data Portal. <https://tcga-data.nci.nih.gov/>. Accessed: 06/02/2017 (via cBioPortal).
- The Cancer Society of New Zealand (Cancer Society of NZ) (2017) What is cancer? <https://otago-southland.cancernz.org.nz/en/cancer-information/other-links/what-is-cancer-3/>. Accessed: 22/03/2017.
- 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 International Cancer Genome Consortium (ICGC) (2017) ICGC Data Portal. <https://dcc.icgc.org/>. Accessed: 06/02/2017.
- 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.
- The Pharmaceutical Management Agency (PHARMAC) (2016) Approval of multi-product funding proposal with roche.
- 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.

- Travers, J. and Milgram, S. (1969) An experimental study of the small world problem. *Sociometry*, **32**(4): 425–443.
- Tsai, H.C., Li, H., Van Neste, L., Cai, Y., Robert, C., Rassool, F.V., Shin, J.J., Harbom, K.M., Beaty, R., Pappou, E., *et al.* (2012) Transient low doses of dna-demethylating agents exert durable antitumor effects on hematological and epithelial tumor cells. *Cancer Cell*, **21**(3): 430–46.
- 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.
- 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 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.
- Vargas, J.J., Gusella, G., Najfeld, V., Klotman, M., and Cara, A. (2004) Novel integrase-defective lentiviral episomal vectors for gene transfer. *Hum Gene Ther*, **15**: 361–372.
- 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.
- 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, K., Yuen, S.T., Xu, J., Lee, S.P., Yan, H.H., Shi, S.T., Siu, H.C., Deng, S., Chu, K.M., Law, S., *et al.* (2014) Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat Genet*, **46**(6): 573–582.
- 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.
- 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.