

Contents

Glossary	xi
Acronyms	xiii
1 Introduction and Literature Review	1
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
2	Methods and Resources	38
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
3	Methods Developed During Thesis	57
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
4	Synthetic Lethal Analysis of Gene Expression Data	93
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
5	Synthetic Lethal Pathway Structure	125
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
6	Simulation and Modelling of Synthetic Lethal Pathways	147
6.1	Synthetic Lethal Detection Methods	148
6.1.1	Performance of SLIPT and χ^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
7	Discussion	189
7.1	Synthetic Lethality and <i>CDH1</i> Biology	189
7.1.1	Established Functions of <i>CDH1</i>	190
7.1.2	The Molecular Role of <i>CDH1</i> in Cancer	190
7.2	Significance	191
7.2.1	Synthetic Lethality in the Genomic Era	191
7.2.2	Clinical Interventions based on Synthetic Lethality	193
7.3	Future Directions	194
7.4	Conclusions	196
	Bibliography	198

A	Sample Quality	222
A.1	Sample Correlation	222
A.2	Replicate Samples in TCGA Breast Cancer Data	224
B	Software Used for Thesis	228
C	Mutation Analysis in Breast Cancer	237
C.1	Synthetic Lethal Genes and Pathways	237
C.2	Synthetic Lethal Expression Profiles	238
C.3	Comparison to Primary Screen	241
C.3.1	Resampling Analysis	243
C.4	Compare SLIPT genes	245
D	Metagene Analysis	247
D.1	Pathway Signature Expression	247
D.2	Synthetic Lethal Reactome Metagenes	251
E	Intrinsic Subtyping	252
F	Stomach Expression Analysis	254
F.1	Synthetic Lethal Genes and Pathways	254
F.2	Comparison to Primary Screen	258
F.2.1	Resampling Analysis	260
F.3	Metagene Analysis	262
G	Synthetic Lethal Genes in Pathways	263
H	Network Analysis for Mutation SLIPT	270
I	Pathway Structure for Mutation SLIPT	273
J	Performance of SLIPT and χ^2	277
J.1	Correlated Query Genes affects Specificity	283
K	Simulations on Graph Structures	289
K.0.1	Simulations from Inhibiting Graph Structures	290
K.1	Simulation across Graph Structures	293
K.2	Simulations from Complex Graph Structures	297
K.2.1	Simulations from Complex Inhibiting Graphs	300
K.3	Simulations from Pathway Graph Structures	306

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 χ^2 and SLIPT across quantiles	150
6.2	Performance of χ^2 and SLIPT across quantiles with more genes	151
6.3	Performance of χ^2 and SLIPT across quantiles with query correlation	152
6.4	Performance of χ^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	222
A.2	Correlation analysis and sample removal	223
A.3	Replicate excluded samples	224
A.4	Replicate samples with all remaining	225
A.5	Replicate samples with some excluded	226
C.1	Synthetic lethal expression profiles of analysed samples	239
C.2	Comparison of mtSLIPT to short interfering RNA (siRNA)	241
C.3	Compare mtSLIPT and siRNA genes with correlation	245
C.4	Compare mtSLIPT and siRNA genes with correlation	245
C.5	Compare mtSLIPT and siRNA genes with siRNA viability	246
D.1	Pathway metagene expression profiles	249

D.2	Expression profiles for estrogen receptor related genes	250
F.1	Synthetic lethal expression profiles of stomach samples	256
F.2	Comparison of SLIPT in stomach to siRNA	258
G.1	Synthetic lethality in the PI3K/AKT pathway	263
G.2	Synthetic lethality in the PI3K/AKT pathway in cancer	264
G.3	Synthetic lethality in the Extracellular Matrix	265
G.4	Synthetic lethality in the GPCR Downstream	266
G.5	Synthetic lethality in the Translation Elongation	267
G.6	Synthetic lethality in the Nonsense-mediated Decay	268
G.7	Synthetic lethality in the 3' UTR	269
H.1	Synthetic lethality and vertex degree	270
H.2	Synthetic lethality and centrality	271
H.3	Synthetic lethality and PageRank	271
I.1	Structure of synthetic lethality resampling	273
J.1	Performance of χ^2 and SLIPT across quantiles	277
J.2	Performance of χ^2 and SLIPT across quantiles	279
J.3	Performance of χ^2 and SLIPT across quantiles with more genes	281
J.4	Performance of χ^2 and SLIPT across quantiles with query correlation	283
J.5	Performance of χ^2 and SLIPT across quantiles with query correlation	285
J.6	Performance of χ^2 and SLIPT across quantiles with query correlation and more genes	287
K.1	Performance of simulations on a simple graph	289
K.2	Performance of simulations on an inhibiting graph	290
K.3	Performance of simulations on a constructed graph with inhibition	291
K.4	Performance of simulations on a constructed graph with inhibition	292
K.5	Detection of synthetic lethality within a graph structure	293
K.6	Detection of synthetic lethality within an inhibiting graph	295
K.7	Detection of synthetic lethality within an inhibiting graph	296
K.8	Performance of simulations on a branching graph	297
K.9	Performance of simulations on a complex graph	298
K.10	Performance of simulations on a large graph	299
K.11	Performance of simulations on a branching graph with inhibition	300
K.12	Performance of simulations on a branching graph with inhibition	301
K.13	Performance of simulations on a complex graph with inhibition	302
K.14	Performance of simulations on a complex graph with inhibition	303
K.15	Performance of simulations on a large constructed graph with inhibition	304
K.16	Performance of simulations on a large constructed graph with inhibition	305
K.17	Performance of simulations on the $G_{\alpha i}$ signalling pathway	306
K.18	Performance of simulations including the $G_{\alpha i}$ signalling pathway	307

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	228
C.1	Candidate synthetic lethal gene partners of <i>CDH1</i> from mtSLIPT . . .	237
C.2	Pathways for <i>CDH1</i> partners from mtSLIPT	238
C.3	Pathways for clusters of <i>CDH1</i> partners from mtSLIPT	240
C.4	Pathways for <i>CDH1</i> partners from mtSLIPT and siRNA	242
C.5	Pathways for <i>CDH1</i> partners from mtSLIPT	243
C.6	Pathways for <i>CDH1</i> partners from mtSLIPT and siRNA primary screen	244
D.1	Candidate synthetic lethal metagenes against <i>CDH1</i> from mtSLIPT . .	251

E.1	Comparison of intrinsic subtypes	252
F.1	Synthetic lethal gene partners of <i>CDH1</i> from SLIPT in stomach cancer	254
F.2	Pathways for <i>CDH1</i> partners from SLIPT in stomach cancer	255
F.3	Pathways for clusters of <i>CDH1</i> partners in stomach SLIPT	257
F.4	Pathways for <i>CDH1</i> partners from SLIPT and siRNA	259
F.5	Pathways for <i>CDH1</i> partners from SLIPT in stomach cancer	260
F.6	Pathways for <i>CDH1</i> partners from SLIPT in stomach and siRNA	261
F.7	Synthetic lethal metagenes against <i>CDH1</i> in stomach cancer	262
H.1	ANOVA for synthetic lethality and vertex degree	272
H.2	ANOVA for synthetic lethality and information centrality	272
H.3	ANOVA for synthetic lethality and PageRank centrality	272
I.1	Resampling for pathway structure of synthetic lethal detection methods	274

Glossary

centrality	A network metric which identifies important vertices .
edge or link	A relationship connecting a pair of elements of a graph structure or network, may be weighted or directional.
gene expression	A measure of the relative expression of each gene from the mRNA extracted from (pooled) cells.
graph or network	A mathematical structure modelling or depicting the relationships between elements.
information centrality	A network centrality metric which uses the impact of removing a vertex or node on connections in the network.
metagene	A consistent signal of expression for a collection of genes such as a biological pathway, derived from singular value decomposition.
mutation	A change in DNA sequence that disrupts gene function.
PageRank centrality	A network centrality metric which uses eigenvectors with a scaling factor (Brin and Page, 1998).
pathway	A series of biomolecules that produces a particular product or biological function.
shortest path	A path with the fewest possible edges which connects two particular vertices .
synthetic lethal	Genetic interactions where inactivation of multiple genes is inviable (or deleterious) which are viable if inactivated separately.

vertex degree	A network metric of connectivity of vertices which uses the number of edges connected to each vertex or node .
vertex or node	An element of a graph structure or network.

Acronyms

ANOVA	Analysis of Variance.
GPCR	G Protein Coupled Receptor.
mtSLIPT	Synthetic Lethal Interaction Prediction Tool (against mutation).
NMD	Nonsense-Mediated Decay.
PI3K	Phosphoinositide 3-kinase.
siRNA	Short Interfering RNA.
SLIPT	Synthetic Lethal Interaction Prediction Tool.
UTR	Untranslated Region (of mRNA).

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 wee1 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 polyadprribose 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 α -Dependent μ -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., Dollé, 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.
- Sørli, 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(adenosine diphosphate) 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) Usc 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.

Appendix G

Synthetic Lethal Genes in Pathways

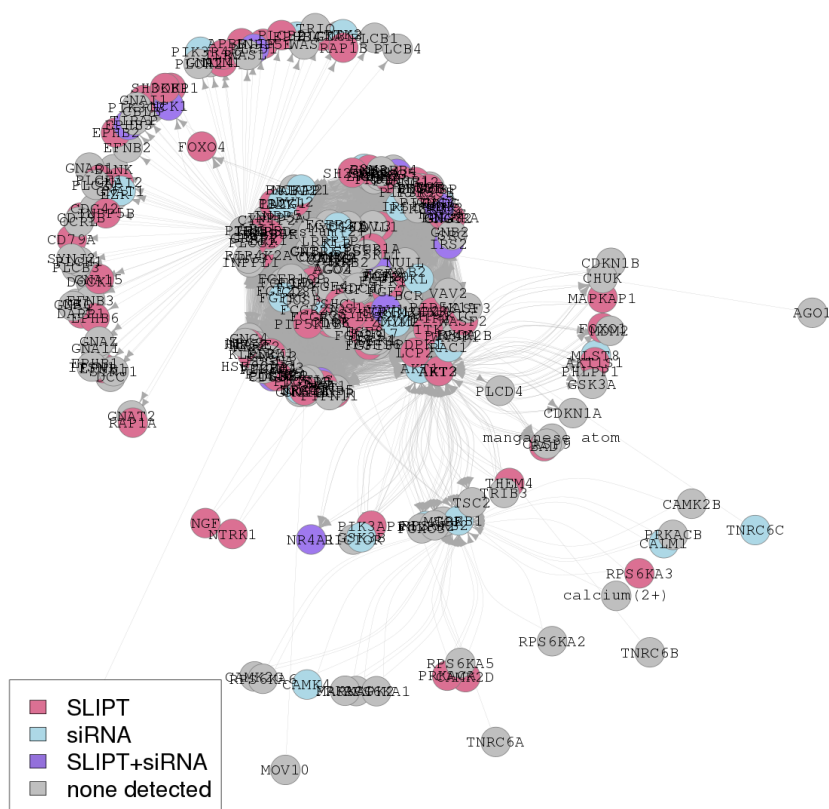


Figure G.1: **Synthetic lethality in the PI3K/AKT pathway.** The Reactome PI3K/AKT pathway with synthetic lethal candidates, coloured as shown in the legend.

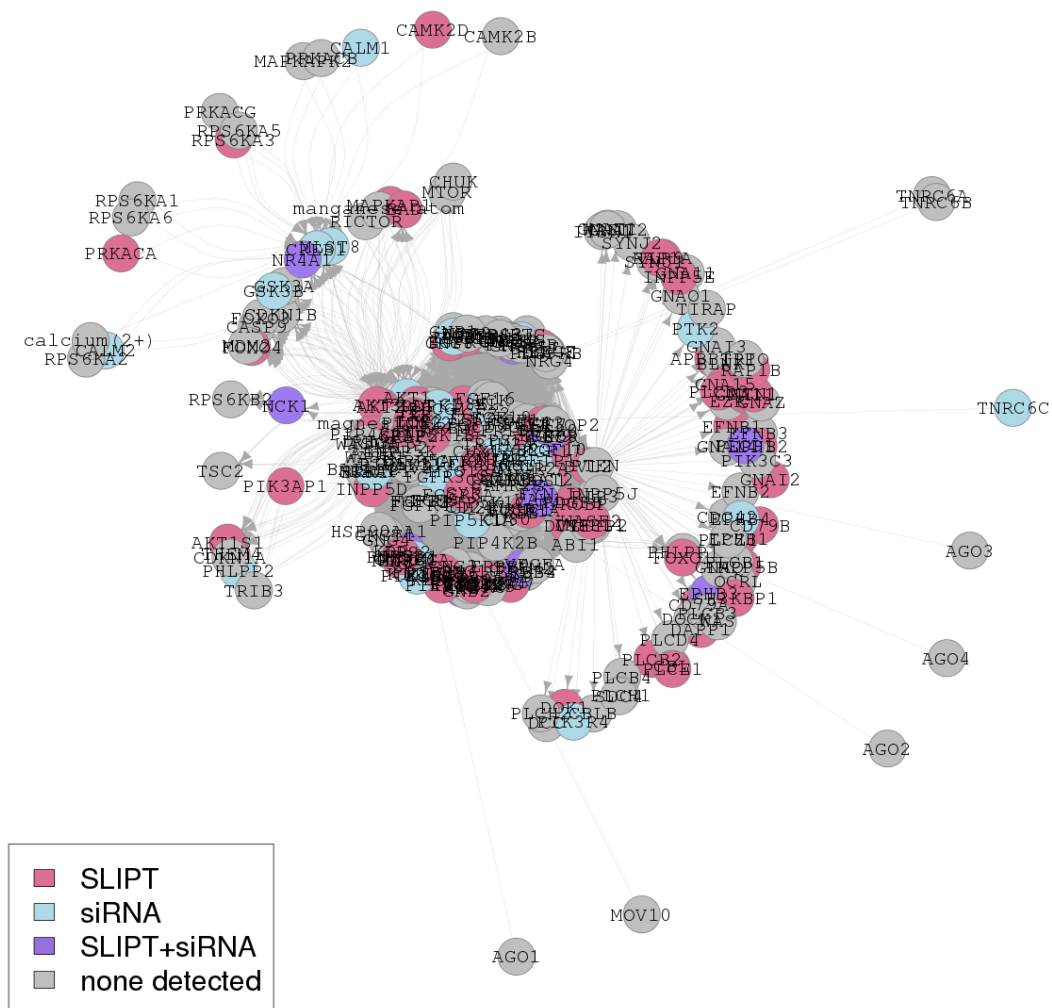


Figure G.2: **Synthetic lethality in the PI3K/AKT pathway in cancer.** The Reactome PI3K/AKT in cancer pathway with synthetic lethal candidates, coloured as shown in the legend.

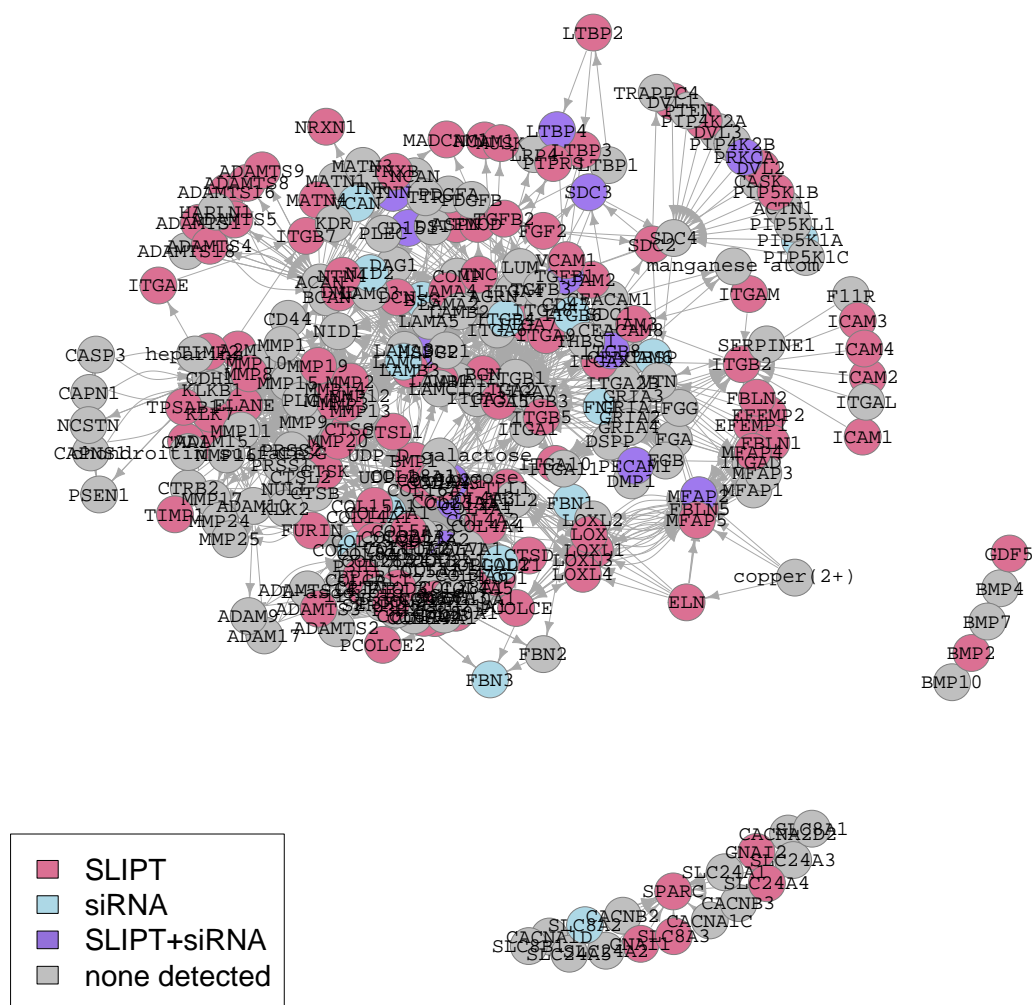


Figure G.3: **Synthetic lethality in the Extracellular Matrix.** The Reactome Extracellular Matrix pathway with synthetic lethal candidates, coloured as shown in the legend.

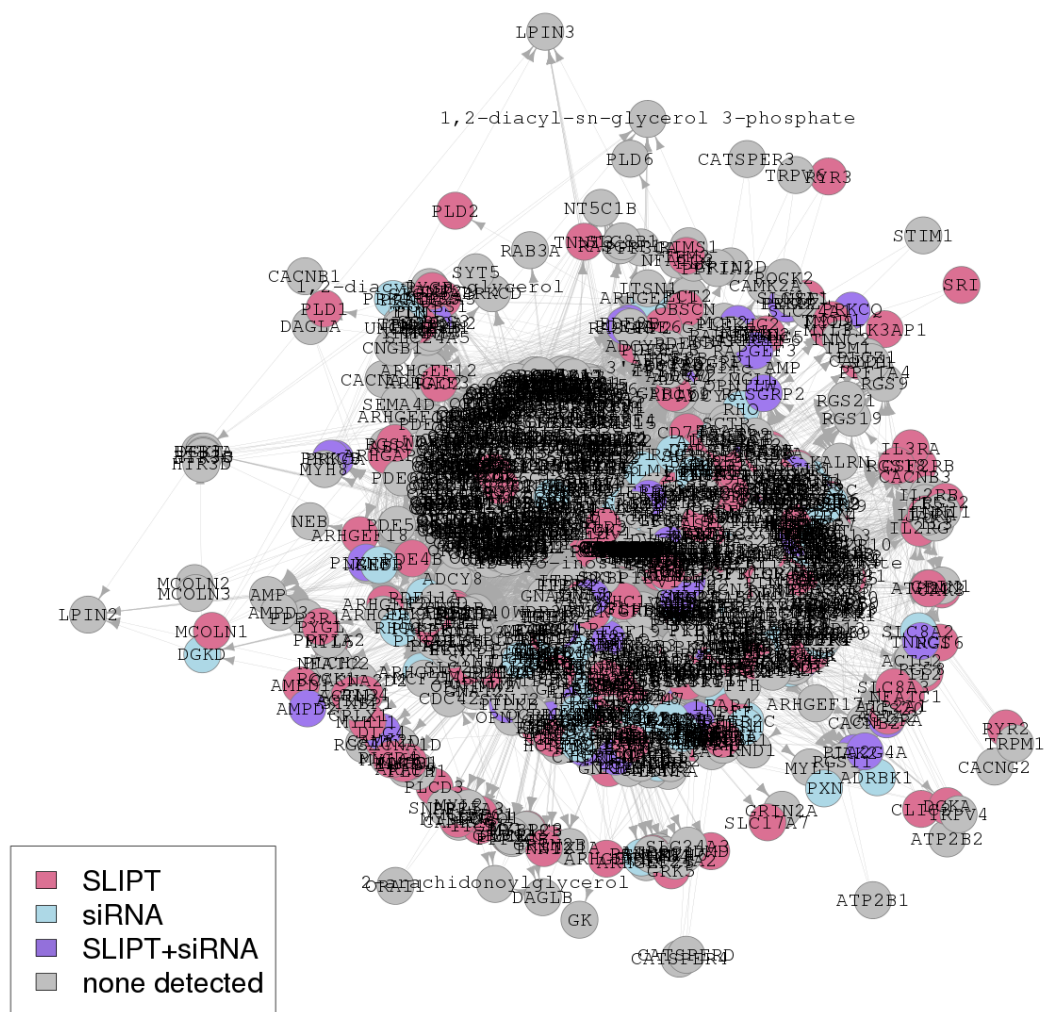


Figure G.4: **Synthetic lethality in the GPCR Downstream.** The Reactome **G protein coupled receptor (GPCR)** Downstream pathway with synthetic lethal candidates, coloured as shown in the legend.

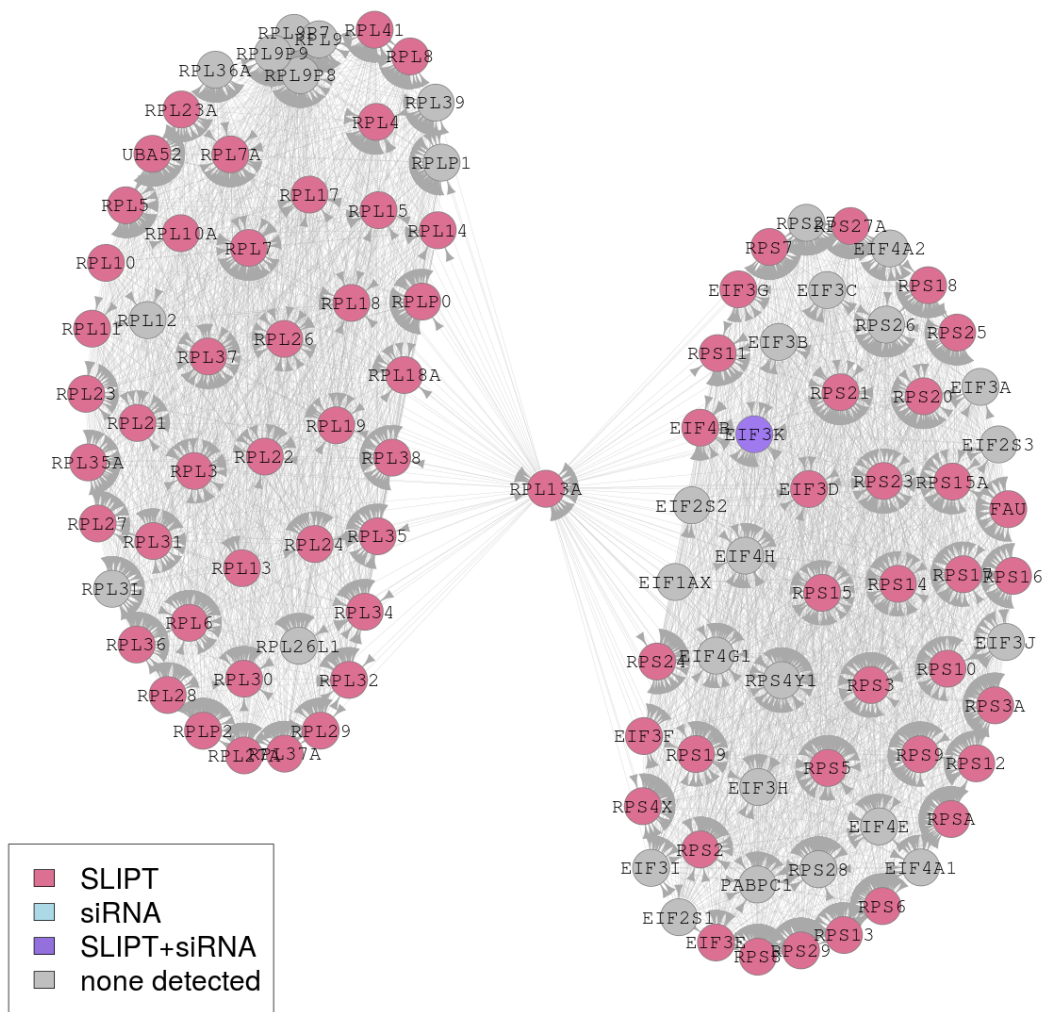


Figure G.7: **Synthetic lethality in the 3' UTR.** The Reactome 3' untranslated region (UTR) pathway with synthetic lethal candidates, coloured as shown in the legend.

Appendix H

Network Analysis for Mutation SLIPT

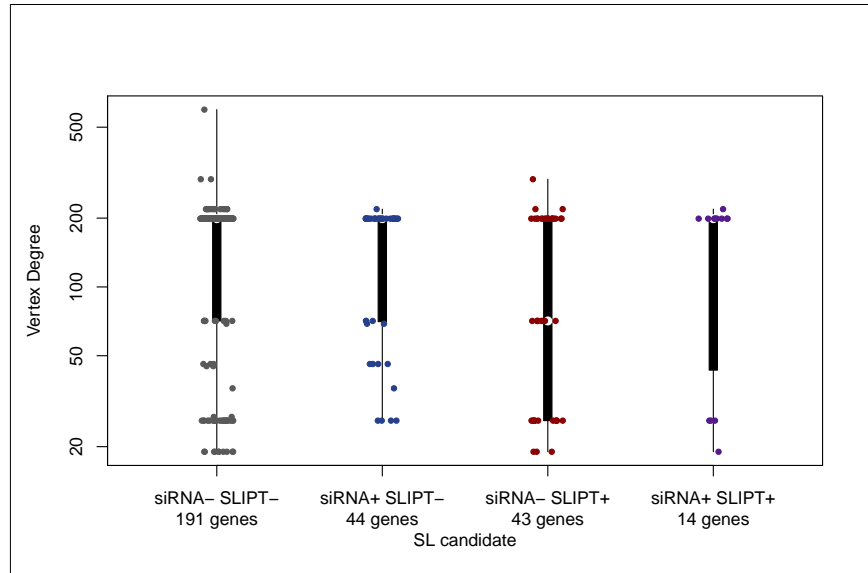


Figure H.1: **Synthetic lethality and vertex degree.** The number of connected genes (**vertex degree**) was compared (on a log-scale) across genes detected by **mtSLIPT** and **siRNA** screening in the Reactome G_{ai} pathway. There were no differences in **vertex degree** between the groups (shown in Table 5.1), although genes detected by **siRNA** included those with the fewest connections.

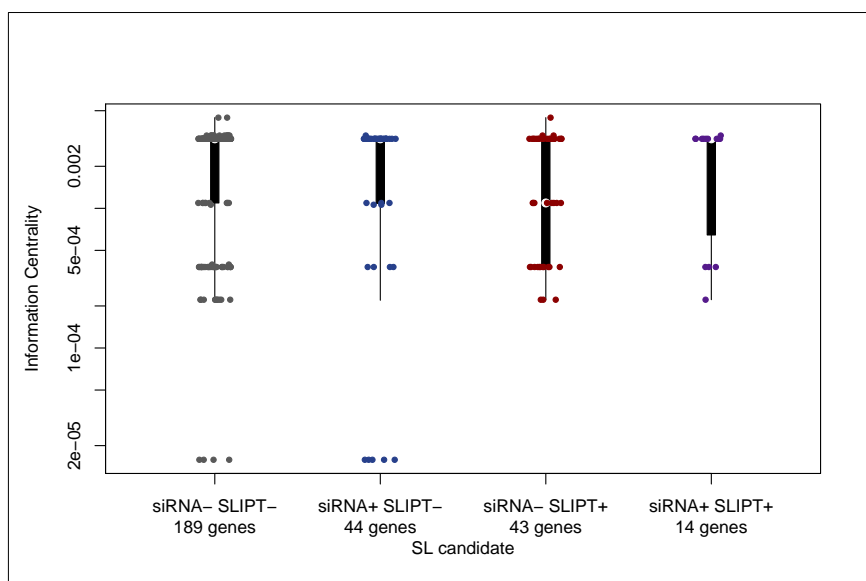


Figure H.2: **Synthetic lethality and centrality.** The [information centrality](#) was compared (on a log-scale) across genes detected by [SLIPT](#) and [siRNA](#) screening in the Reactome $G_{\alpha i}$ pathway. Genes detected by [SLIPT](#) or [siRNA](#) did not have higher centrality than other genes (shown in Table [H.2](#)). Genes detected by [SLIPT](#) spanned the range of [centrality](#) values.

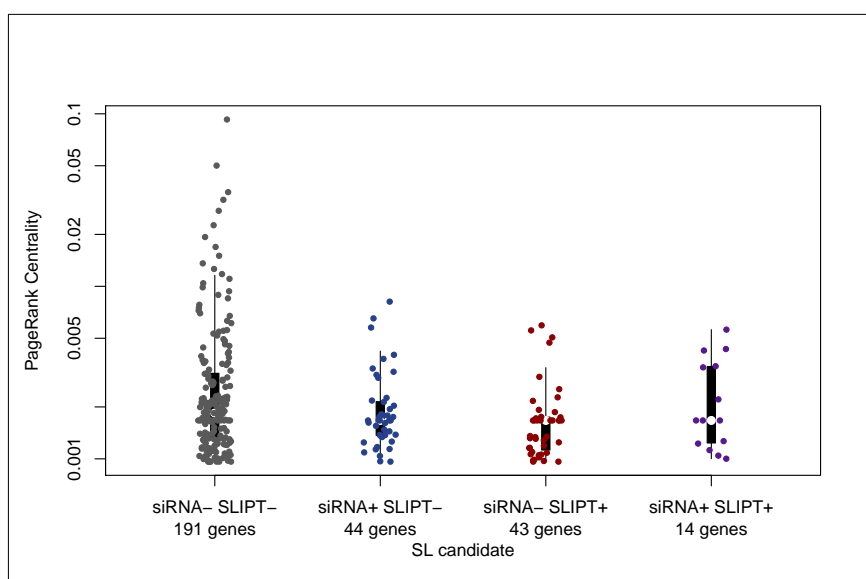


Figure H.3: **Synthetic lethality and PageRank.** The [PageRank centrality](#) was compared (on a log-scale) across genes detected by [mtSLIPT](#) and [siRNA](#) screening in the Reactome $G_{\alpha i}$ pathway. Genes detected by with either [synthetic lethal](#) detection approach had a more restricted range of [centrality](#) values neither of these had a significant association with centrality (shown in Table [H.3](#)).

Table H.1: ANOVA for synthetic lethality and vertex degree

	DF	Sum Squares	Mean Squares	F-value	p-value
siRNA	1	15	15.50	0.0134	0.9084
mtSLIPT	1	196	195.94	0.1689	0.6825
siRNA×mtSLIPT	1	9	9.17	0.0079	0.9294

Analysis of variance for **vertex degree** against **synthetic lethal** detection approaches (with an interaction term)

Table H.2: ANOVA for synthetic lethality and information centrality

	DF	Sum Squares	Mean Squares	F-value	p-value
siRNA	1	0.000256	0.0002561	0.1851	0.6685
mtSLIPT	1	0.003225	0.0032247	2.3308	0.1318
siRNA×mtSLIPT	1	0.001238	0.0012385	0.8952	0.3476

Analysis of variance for **information centrality** against **synthetic lethal** detection approaches (with an interaction term)

Table H.3: ANOVA for synthetic lethality and PageRank centrality

	DF	Sum Squares	Mean Squares	F-value	p-value
siRNA	1	0.0002038	2.0385×10^{-4}	1.1423	0.2892
mtSLIPT	1	0.0000208	2.0752×10^{-5}	0.1163	0.7342
siRNA×mtSLIPT	1	0.0000137	1.3743×10^{-5}	0.0770	0.7823

Analysis of variance for **PageRank centrality** against **synthetic lethal** detection approaches (with an interaction term)

Appendix I

Pathway Structure for Mutation SLIPT

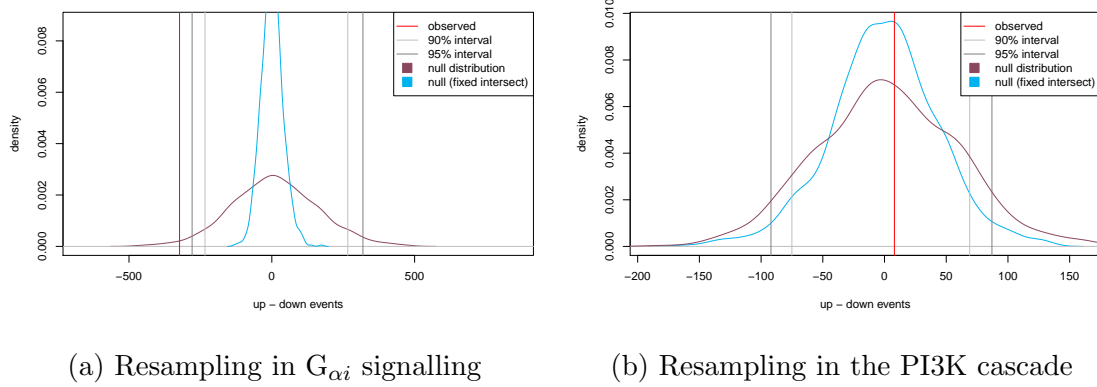


Figure I.1: **Structure of synthetic lethality resampling.** A null distribution with 10,000 iterations of the number of **siRNA** genes upstream or downstream of **mtSLIPT** genes (depicted as the difference of these) in each **pathway**. To assess significance, the observed events (with **shortest paths**) were compared to the 90% and 95% intervals for the null distribution (shown in blue). Genes detected by both methods were not fixed to the same number as observed for the alternative null distribution (shown in red), although the significance of the observed number of events (red) was changed in either case. The genes detected by both approaches were included in computing the number of **shortest paths** (in either direction) between **SLIPT** and **siRNA** genes. The permutations show (a) a significant pathway relationship for $G_{\alpha i}$ signalling and (b) and non-significant relationship for the **phosphoinositide 3-kinase (PI3K)** cascade.

Table I.1: Resampling for pathway structure of **synthetic lethal** detection methods

PPathway	Graph		Candidates		Observed				Permutation p-value		p-value (FDR)
	Nodes	Edges	SLIPT	siRNA	Up ¹	Down ²	Up–Down	Up/Down	Up–Down	Down–Up	Down–Up
PI3K Cascade	138	1495	42	25	131	123	8	1.065	0.4473	0.5466	0.7263
PI3K/AKT Signalling in Cancer	275	12882	56	44	478	440	38	1.086	0.4163	0.5810	0.7263
G_{ai} Signalling	292	22003	57	58	543	866	-323	0.627	0.9507	0.0488	0.488
GPCR downstream	1270	142071	218	160	7632	6500	1132	1.174	0.1707	0.8291	0.8751
Elastic fibre formation	42	175	16	7	6	7	-1	0.857	0.5512	0.3681	0.7263
Extracellular matrix	299	3677	81	29	313	347	-34	0.902	0.5762	0.4215	0.7263
Formation of Fibrin	52	243	11	5	8	19	-11	0.421	0.7993	0.1800	0.6000
Nonsense-Mediated Decay	103	102	56	2	0	0	0		0.197	0.1373	0.6000
3'-UTR-mediated translational regulation	107	2860	56	1	52	1	51	52	0.1210	0.8751	0.8751
Eukaryotic Translation Elongation	92	3746	57	0	0	0	0		0.4952	0.4892	0.7263

Pathways in the Reactome network tested for structural relationships between **mtSLIPT** and **siRNA** genes by resampling. The raw p-value (computed without adjusting for multiple comparisons over pathways) is given for the difference in upstream and downstream paths from **mtSLIPT** to **siRNA** gene candidate partners of CDH1 with significant pathways highlighted in bold. Sampling was performed only in the target pathway and shortest paths were computed within it. Loops or paths in either direction that could not be resolved were excluded from the analysis. The genes detected by both **mtSLIPT** and **siRNA** (or resampling for them) were included in the analysis and the number of these were fixed to the number observed.

¹ The number of paths where the **siRNA** candidate was upstream of a **mtSLIPT** candidate

² The number of paths where the **siRNA** candidate was downstream of a **mtSLIPT** candidate