

Contents

Glossary	xiii
Acronyms	xiv
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 genomic Revolution in Cancer Research	4
1.1.2.1 High-Throughput Technologies	4
1.1.2.2 Bioinformatics and Genomic Data	6
1.1.3 Genomics Projects	6
1.1.3.1 The Cancer Genome Project	6
1.1.3.2 The Cancer Genome Atlas Project	7
1.1.4 Genomic Cancer Medicine	9
1.1.4.1 Cancer Genes and Driver Mutations	9
1.1.4.2 Precision Cancer Medicine	10
1.1.4.3 Molecular Diagnostics and Pan-Cancer Medicine	10
1.1.4.4 Targeted Therapeutics and Pharmacogenomics	10
1.1.5 Systems and Network Biology	11
1.1.5.1 Network Medicine and Polypharmacology	13
1.2 A Synthetic Lethal Approach to Cancer Medicine	14
1.2.1 Synthetic Lethal Genetic Interactions	14
1.2.2 Synthetic Lethal Concepts in Genetics	15
1.2.3 Synthetic Lethality in Model Systems	16
1.2.3.1 Synthetic Lethal Pathways and Networks	16
1.2.3.2 Evolution of Synthetic Lethality	17
1.2.4 Synthetic Lethality in Cancer	18
1.2.5 Clinical Impact of Synthetic Lethality in Cancer	19
1.2.6 High-throughput Screening for Synthetic Lethality	21
1.2.6.1 Synthetic Lethal Screens	22
1.2.7 Computational Prediction of Synthetic Lethality	25
1.2.7.1 Bioinformatics Approaches to Genetic Interactions	25
1.2.7.2 Comparative Genomics	26
1.2.7.3 Analysis and Modelling of Protein Data	29
1.2.7.4 Differential Gene Expression	31

1.2.7.5	Data Mining and Machine Learning	32
1.2.7.6	Mutually Exclusive Bimodality	35
1.2.7.7	Rationale for Further Development	36
1.3	E-cadherin as a Synthetic Lethal Target	36
1.3.1	The <i>CDH1</i> gene and its Biological Functions	36
1.3.1.1	Cytoskeleton	37
1.3.1.2	Extracellular and Tumour Micro-environment	37
1.3.1.3	Cell-Cell Adhesion and Signalling	37
1.3.2	<i>CDH1</i> as a Tumour (and Invasion) Suppressor	38
1.3.2.1	Breast Cancers and Invasion	38
1.3.3	Hereditary Diffuse Gastric Cancer and Lobular Breast Cancer	38
1.3.4	Cell Line Models of <i>CDH1</i> Null Mutations	40
1.4	Summary and Research Direction of Thesis	40
1.4.1	Thesis Aims	42
2	Methods and Resources	43
2.1	Bioinformatics Resources for Genomics Research	43
2.1.1	Public Data and Software Packages	43
2.1.1.1	Cancer Genome Atlas Data	44
2.1.1.2	Reactome and Annotation Data	45
2.2	Data Handling	45
2.2.1	Normalisation	45
2.2.2	Sample Triage	46
2.2.3	Metagenes and the Singular Value Decomposition	46
2.2.3.1	Candidate Triage and Integration with Screen Data	48
2.3	Techniques	49
2.3.1	Statistical Procedures and Tests	49
2.3.2	Gene Set Over-representation Analysis	50
2.3.3	Clustering	50
2.3.4	Heatmap	50
2.3.5	mMdelling and Simulations	51
2.3.5.1	Receiver Operating Characteristic (Performance)	52
2.3.6	Resampling Analysis	52
2.4	Pathway Structure Methods	53
2.4.1	Network and Graph Analysis	53
2.4.2	Sourcing Graph Structure Data	54
2.4.3	Constructing Pathway Subgraphs	54
2.4.4	Network Analysis Metrics	55
2.5	Implementation	56
2.5.1	Computational Resources and Linux Utilities	56
2.5.2	R Language and Packages	57
2.5.3	High Performance and Parallel Computing	60

3	Methods Developed During Thesis	62
3.1	A Synthetic Lethal Detection Methodology	62
3.2	Synthetic Lethal Simulation and Modelling	64
3.2.1	A Model of Synthetic Lethality in Expression Data	65
3.2.2	Simulation Procedure	69
3.3	Detecting Simulated Synthetic Lethal Partners	71
3.3.1	Binomial Simulation of Synthetic Lethality	72
3.3.2	Multivariate Normal Simulation of Synthetic Lethality	74
3.3.2.1	Multivariate Normal Simulation with Correlated Genes	76
3.3.2.2	Specificity with Query-Correlated Pathways	83
3.4	Graph Structure Methods	86
3.4.1	Upstream and Downstream Gene Detection	86
3.4.1.1	Permutation Analysis for Statistical Significance	87
3.4.1.2	Hierarchy Based on Biological Context	87
3.4.2	Simulating Gene Expression from Graph Structures	88
3.5	Customised Functions and Packages Developed	92
3.5.1	Synthetic Lethal Interaction Prediction Tool	93
3.5.2	Data Visualisation	93
3.5.3	Extensions to the iGraph Package	96
3.5.3.1	Sampling Simulated Data from Graph Structures	96
3.5.3.2	Plotting Directed Graph Structures	96
3.5.3.3	Computing Information Centrality	97
3.5.3.4	Testing Pathway Structure with Permutation Testing	97
3.5.3.5	Metapackage to Install iGraph Functions	98
4	Synthetic Lethal Analysis of Gene Expression Data	99
4.1	Synthetic Lethal Genes in Breast Cancer	100
4.1.1	Synthetic Lethal Pathways in Breast Cancer	102
4.1.2	Expression Profiles of Synthetic Lethal Partners	103
4.1.2.1	Subgroup Pathway Analysis	106
4.2	Comparing Synthetic Lethal Gene Candidates	109
4.2.1	Primary siRNA Screen Candidates	109
4.2.2	Comparison with Correlation	109
4.2.3	Comparison with Primary Screen Viability	111
4.2.4	Comparison with Secondary siRNA Screen Validation	113
4.2.5	Comparison to Primary Screen at Pathway Level	115
4.2.5.1	Resampling Genes for Pathway Enrichment	117
4.2.6	Integrating Synthetic Lethal Pathways and Screens	120
4.3	Metagene Analysis	122
4.3.1	Pathway Expression	123
4.3.2	Somatic Mutation	125
4.3.3	Synthetic Lethal Pathway Metagenes	129
4.3.4	Synthetic Lethality in Breast Cancer	130
4.4	Replication in Stomach Cancer	131
4.5	Discussion	132
4.5.1	Strengths of the SLIPT Methodology	132

4.5.2	Synthetic Lethal Pathways for E-cadherin	133
4.5.3	Replication and Validation	135
4.5.3.1	Integration with short interfering RNA (siRNA) Screen- ing	135
4.5.3.2	Replication across Tissues	136
4.6	Summary	136
5	Synthetic Lethal Pathway Structure	138
5.1	Synthetic Lethal Genes in Reactome Pathways	138
5.1.1	The PI3K/AKT Pathway	139
5.1.2	The Extracellular Matrix	141
5.1.3	G Protein Coupled Receptors	144
5.1.4	Gene Regulation and Translation	144
5.2	Network Analysis of Synthetic Lethal Genes	145
5.2.1	Gene Connectivity and Vertex Degree	146
5.2.2	Gene Importance and Centrality	147
5.2.2.1	Information Centrality	147
5.2.2.2	PageRank Centrality	149
5.3	Relationships between Synthetic Lethal Genes	151
5.3.1	Hierarchical Pathway Structure	151
5.3.1.1	Contextual Hierarchy of PI3K	151
5.3.1.2	Testing Contextual Hierarchy of Synthetic Lethal Genes	151
5.3.2	Upstream or Downstream Synthetic Lethality	155
5.3.2.1	Measuring Structure of Candidates within PI3K	155
5.3.2.2	Resampling for Synthetic Lethal Pathway Structure	157
5.4	Discussion	159
5.5	Summary	161
6	Simulation and mMdelling of Synthetic Lethal Pathways	163
6.1	Synthetic Lethal Detection Methods	164
6.1.1	Performance of SLIPT and χ^2 across Quantiles	164
6.1.1.1	Correlated Query Genes affects Specificity	168
6.1.2	Alternative Synthetic Lethal Detection Strategies	170
6.1.2.1	Correlation for Synthetic Lethal Detection	170
6.1.2.2	Testing for Bimodality with BiSEp	172
6.2	Simulations with Graph Structures	173
6.2.1	Performance over Graph Structures	174
6.2.1.1	Simple Graph Structures	174
6.2.1.2	Constructed Graph Structures	176
6.2.2	Performance with Inhibitions	179
6.2.3	Synthetic Lethality across Graph Structures	184
6.2.4	Performance within a Simulated Human Genome	188
6.3	Simulations in More Complex Graph Structures	192
6.3.1	Simulations over Pathway-based Graphs	193
6.3.2	Pathway Structures in a Simulated Human Genome	196
6.4	Discussion	199

6.4.1	Simulation Procedure	199
6.4.2	Comparing Methods with Simulated Data	200
6.4.3	Design and Performance of SLIPT	201
6.4.4	Simulations from Graph Structures	203
6.5	Summary	204
7	Discussion	206
7.1	Synthetic Lethality and <i>CDH1</i> Biology	206
7.1.1	Established Functions of <i>CDH1</i>	207
7.1.2	The Molecular Role of <i>CDH1</i> in Cancer	207
7.2	Significance	208
7.2.1	Synthetic Lethality in the Genomic Era	208
7.2.2	Clinical Interventions based on Synthetic Lethality	210
7.3	Future Directions	211
7.4	Conclusions	213
	Bibliography	215
A	Sample Quality	239
A.1	Sample Correlation	239
A.2	Replicate Samples in The Cancer Genome Atlas (TCGA) Breast	241
B	Software Used for Thesis	245
C	Mutation Analysis in Breast Cancer	254
C.1	Synthetic Lethal Genes and Pathways	254
C.2	Synthetic Lethal Expression Profiles	255
C.3	Comparison to Primary Screen	258
C.3.1	Resampling Analysis	260
C.4	Compare Synthetic Lethal Interaction Prediction Tool (SLIPT) genes	262
C.5	Metagene Analysis	264
C.6	Expression of Somatic Mutations	265
C.7	Metagene Expression Profiles	268
D	Intrinsic Subtyping	271
E	Stomach Expression Analysis	273
E.1	Synthetic Lethal Genes and Pathways	273
E.2	Comparison to Primary Screen	277
E.2.1	Resampling Analysis	279
E.3	Metagene Analysis	281
F	Synthetic Lethal Genes in Pathways	285
G	Pathway Connectivity for Mutation SLIPT	293
H	Information Centrality for Gene Essentiality	297

I	Pathway Structure for Mutation SLIPT	300
J	Performance of SLIPT and χ^2	303
J.1	Correlated Query Genes affects Specificity	309
K	Simulations on Graph Structures	315
K.0.1	Simulations from Inhibiting Graph Structures	316
K.1	Simulation across Graph Structures	319
K.2	Simulations from Complex Graph Structures	323
K.2.1	Simulations from Complex Inhibiting Graphs	326
K.3	Simulations from Pathway Graph Structures	332

List of Figures

1.1	Synthetic genetic interactions	15
1.2	Synthetic lethality in cancer	19
2.1	Read count density	47
2.2	Read count sample mean	47
3.1	Framework for synthetic lethal prediction	63
3.2	Synthetic lethal prediction adapted for mutation	64
3.3	A model of synthetic lethal gene expression	66
3.4	Modelling synthetic lethal gene expression	67
3.5	Synthetic lethality with multiple genes	68
3.6	Simulating gene function	70
3.7	Simulating synthetic lethal gene function	70
3.8	Simulating synthetic lethal gene expression	71
3.9	Performance of binomial simulations	73
3.10	Comparison of statistical performance	73
3.11	Performance of multivariate normal simulations	75
3.12	Simulating expression with correlated gene blocks	77
3.13	Simulating expression with correlated gene blocks	78
3.14	Synthetic lethal prediction across simulations	79
3.15	Performance with correlations	80
3.16	Comparison of statistical performance with correlation structure	81
3.17	Performance with query correlations	82
3.18	Statistical evaluation of directional criteria	84
3.19	Performance of directional criteria	85
3.20	Simulated graph structures	89
3.21	Simulating expression from a graph structure	90
3.22	Simulating expression from graph structure with inhibitions	91
3.23	Demonstration of violin plots with custom features	95
3.24	Demonstration of annotated heatmap	95
3.25	Simulating graph structures	97
4.1	Synthetic lethal expression profiles of analysed samples	105
4.2	Comparison of SLIPT to siRNA	109
4.3	Compare SLIPT and siRNA genes with correlation	110
4.4	Compare SLIPT and siRNA genes with correlation	111
4.5	Compare SLIPT and siRNA genes with viability	112

4.6	Compare SLIPT genes with siRNA viability	113
4.7	Resampled intersection of SLIPT and siRNA candidates	117
4.8	Pathway metagene expression profiles	124
4.9	Expression profiles for constituent genes of PI3K	126
4.10	Expression profiles for estrogen receptor related genes	127
4.11	Somatic mutation against the PI3K metagene	128
5.1	synthetic lethality in the PI3K cascade	140
5.2	synthetic lethality in Elastic Fibre Formation	142
5.3	Synthetic lethality in Fibrin Clot Formation	143
5.4	Synthetic lethality and vertex degree	146
5.5	Synthetic lethality and centrality	149
5.6	Synthetic lethality and PageRank	150
5.7	Hierarchical structure of PI3K	152
5.8	Hierarchy score in PI3K against synthetic lethality in PI3K	153
5.9	Structure of synthetic lethality in PI3K	155
5.10	Structure of synthetic lethality resampling in PI3K	156
6.1	Performance of χ^2 and SLIPT across quantiles	166
6.2	Performance of χ^2 and SLIPT across quantiles with more genes	167
6.3	Performance of χ^2 and SLIPT across quantiles with query correlation	168
6.4	Performance of χ^2 and SLIPT across quantiles with query correlation and more genes	169
6.5	Performance of negative correlation and SLIPT	171
6.6	Simple graph structures	174
6.7	Performance of simulations on a simple graph	175
6.8	Performance of simulations is similar in simple graphs	177
6.9	Performance of simulations on a pathway	178
6.10	Performance of simulations on a simple graph with inhibition	180
6.11	Performance is higher on a simple inhibiting graph	181
6.12	Performance of simulations on a constructed graph with inhibition	182
6.13	Performance is affected by inhibition in graphs	184
6.14	Detection of synthetic lethality within a graph structure	186
6.15	Performance of simulations including a simple graph	189
6.16	Performance on a simple graph improves with more genes	190
6.17	Performance on an inhibiting graph improves with more genes	192
6.18	Performance of simulations on the PI3K cascade	195
6.19	Performance of simulations including the PI3K cascade	197
6.20	Performance on pathways improves with more genes	198
A.1	Correlation profiles of removed samples	239
A.2	Correlation analysis and sample removal	240
A.3	Replicate excluded samples	241
A.4	Replicate samples with all remaining	242
A.5	Replicate samples with some excluded	243
C.1	Synthetic lethal expression profiles of analysed samples	256

C.2	Comparison of mtSLIPT to siRNA	258
C.3	Compare mtSLIPT and siRNA genes with correlation	262
C.4	Compare mtSLIPT and siRNA genes with correlation	262
C.5	Compare mtSLIPT and siRNA genes with siRNA viability	263
C.6	Somatic mutation against PIK3CA metagene	265
C.7	Somatic mutation against PI3K protein	266
C.8	Somatic mutation against AKT protein	267
C.9	Pathway metagene expression profiles	268
C.10	Expression profiles for p53 related genes	269
C.11	Expression profiles for BRCA related genes	270
E.1	Synthetic lethal expression profiles of stomach samples	275
E.2	Comparison of SLIPT in stomach to siRNA	277
F.1	Synthetic lethality in the PI3K/AKT pathway	285
F.2	Synthetic lethality in the PI3K/AKT pathway in cancer	286
F.3	Synthetic lethality in the Extracellular Matrix	287
F.4	Synthetic lethality in the GPCRs	288
F.5	Synthetic lethality in the GPCR Downstream	289
F.6	Synthetic lethality in the Translation Elongation	290
F.7	Synthetic lethality in the Nonsense-mediated Decay	291
F.8	Synthetic lethality in the 3' UTR	292
G.1	Synthetic lethality and vertex degree	293
G.2	Synthetic lethality and centrality	294
G.3	Synthetic lethality and PageRank	295
H.1	Information centrality distribution	299
I.1	Synthetic lethality and heirarchy score in PI3K	300
I.2	Heirarchy score in PI3K against synthetic lethality in PI3K	301
I.3	Structure of synthetic lethality in PI3K	301
I.4	Structure of synthetic lethality resampling	302
J.1	Performance of χ^2 and SLIPT across quantiles	303
J.2	Performance of χ^2 and SLIPT across quantiles	305
J.3	Performance of χ^2 and SLIPT across quantiles with more genes	307
J.4	Performance of χ^2 and SLIPT across quantiles with query correlation	309
J.5	Performance of χ^2 and SLIPT across quantiles with query correlation	311
J.6	Performance of χ^2 and SLIPT across quantiles with query correlation and more genes	313
K.1	Performance of simulations on a simple graph	315
K.2	Performance of simulations on an inhibiting graph	316
K.3	Performance of simulations on a constructed graph with inhibition	317
K.4	Performance of simulations on a constructed graph with inhibition	318
K.5	Detection of synthetic lethality within a graph structure	319
K.6	Detection of synthetic lethality within an inhibiting graph	321

K.7	Detection of synthetic lethality within an inhibiting graph	322
K.8	Performance of simulations on a branching graph	323
K.9	Performance of simulations on a complex graph	324
K.10	Performance of simulations on a large graph	325
K.11	Performance of simulations on a branching graph with inhibition	326
K.12	Performance of simulations on a branching graph with inhibition	327
K.13	Performance of simulations on a complex graph with inhibition	328
K.14	Performance of simulations on a complex graph with inhibition	329
K.15	Performance of simulations on a large constructed graph with inhibition	330
K.16	Performance of simulations on a large constructed graph with inhibition	331
K.17	Performance of simulations on the $G_{\alpha i}$ signalling pathway	332
K.18	Performance of simulations including the $G_{\alpha i}$ signalling pathway	333

List of Tables

1.1	Methods for predicting genetic interactions	26
1.2	Methods for predicting synthetic lethality in cancer	27
1.3	Methods used by Wu <i>et al.</i> (2014)	28
2.1	Excluded samples by batch and clinical characteristics.	46
2.2	Computers used during thesis	56
2.3	Linux utilities and applications used during thesis	57
2.4	R installations used during thesis	58
2.5	R Packages used during thesis	58
2.6	R packages developed during thesis	60
4.1	Candidate synthetic lethal gene partners of <i>CDH1</i> from SLIPT	101
4.2	Pathways for <i>CDH1</i> partners from SLIPT	103
4.3	Pathways for clusters of <i>CDH1</i> partners from SLIPT	107
4.4	ANOVA for synthetic lethality and correlation with <i>CDH1</i>	111
4.5	Comparing SLIPT genes against secondary siRNA screen	114
4.6	Pathways for <i>CDH1</i> partners from SLIPT and siRNA	116
4.7	Pathways for <i>CDH1</i> partners from SLIPT	119
4.8	Pathways for <i>CDH1</i> partners from SLIPT and siRNA primary screen .	121
4.9	Candidate synthetic lethal metagenes against <i>CDH1</i> from SLIPT . . .	130
5.1	ANOVA for synthetic lethality and vertex degree	147
5.2	ANOVA for synthetic lethality and information centrality	149
5.3	ANOVA for synthetic lethality and PageRank centrality	151
5.4	ANOVA for synthetic lethality and PI3K hierarchy	154
5.5	Resampling for pathway structure of synthetic lethal detection methods	158
B.1	Complete list of R packages used during this thesis	245
C.1	Candidate synthetic lethal gene partners of <i>CDH1</i> from mtSLIPT . . .	254
C.2	Pathways for <i>CDH1</i> partners from mtSLIPT	255
C.3	Pathways for clusters of <i>CDH1</i> partners from mtSLIPT	257
C.4	Pathways for <i>CDH1</i> partners from mtSLIPT and siRNA	259
C.5	Pathways for <i>CDH1</i> partners from mtSLIPT	260
C.6	Pathways for <i>CDH1</i> partners from mtSLIPT and siRNA primary screen	261
C.7	Candidate synthetic lethal metagenes against <i>CDH1</i> from mtSLIPT . .	264
D.1	Comparison of intrinsic subtypes	271

E.1	Synthetic lethal gene partners of <i>CDH1</i> from SLIPT in stomach cancer	273
E.2	Pathways for <i>CDH1</i> partners from SLIPT in stomach cancer	274
E.3	Pathways for clusters of <i>CDH1</i> partners in stomach SLIPT	276
E.4	Pathways for <i>CDH1</i> partners from SLIPT and siRNA	278
E.5	Pathways for <i>CDH1</i> partners from SLIPT in stomach cancer	279
E.6	Pathways for <i>CDH1</i> partners from SLIPT in stomach and siRNA	280
E.7	Synthetic lethal metagenes against <i>CDH1</i> in stomach cancer	281
G.1	ANOVA for synthetic lethality and vertex degree	296
G.2	ANOVA for synthetic lethality and information centrality	296
G.3	ANOVA for synthetic lethality and PageRank centrality	296
H.1	Information centrality for genes and molecules in the Reactome network	298
I.1	ANOVA for synthetic lethality and PI3K hierarchy	300
I.2	Resampling for pathway structure of synthetic lethal detection methods	302

Glossary

E-cadherin	Epithelial cadherin (calcium-dependent adhesion), a cell-adhesion protein encoded by <i>CDH1</i> .
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.
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.
RNA-Seq	Transcriptome data from sequencing RNA.
synthetic lethal	Genetic interactions where inactivation of multiple genes is inviable (or deleterious) which are viable if inactivated separately.

Acronyms

ANOVA	Analysis of Variance.
ER	Estrogen Receptor.
FDR	False Discovery Rate.
mRNA	Messenger RNA.
mtSLIPT	Synthetic Lethal Interaction Prediction Tool (against mutation).
PAM50	Prediction Analysis of Microarray 50.
RNA	Ribonucleic Acid.
siRNA	Short Interfering RNA.
SLIPT	Synthetic Lethal Interaction Prediction Tool.
TCGA	The Cancer Genome Atlas (genomics project).
UCSC	University of California, Santa Cruz.

Bibliography

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

- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, **25**(1): 25–29.
- Ashworth, A. (2008) A synthetic lethal therapeutic approach: poly(adp) ribose polymerase inhibitors for the treatment of cancers deficient in dna double-strand break repair. *J Clin Oncol*, **26**(22): 3785–90.
- 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., Kossenkov, 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.
- Brouxhon, S.M., Kyrkanides, S., Teng, X., Athar, M., Ghazizadeh, S., Simon, M., O’Banion, M.K., and Ma, L. (2014) Soluble E-cadherin: a critical oncogene modulating receptor tyrosine kinases, MAPK and PI3K/Akt/mTOR signaling. *Oncogene*, **33**(2): 225–235.

- Bryant, H.E., Schultz, N., Thomas, H.D., Parker, K.M., Flower, D., Lopez, E., Kyle, S., Meuth, M., Curtin, N.J., and Helleday, T. (2005) Specific killing of *BRCA2*-deficient tumours with inhibitors of polyadprbose polymerase. *Nature*, **434**(7035): 913–7.
- Bussey, H., Andrews, B., and Boone, C. (2006) From worm genetic networks to complex human diseases. *Nat Genet*, **38**(8): 862–3.
- Butland, G., Babu, M., Diaz-Mejia, J.J., Bohdana, F., Phanse, S., Gold, B., Yang, W., Li, J., Gagarinova, A.G., Pogoutse, O., *et al.* (2008) esga: *E. coli* synthetic genetic array analysis. *Nat Methods*, **5**(9): 789–95.
- 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.
- Chen, X. and Tompa, M. (2010) Comparative assessment of methods for aligning multiple genome sequences. *Nat Biotechnol*, **28**(6): 567–572.
- 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.
- 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.
- 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 reactivation 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.
- 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., Beroukhi, 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.
- 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., Fedyszyn, 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.
- Dong, L.L., Liu, L., Ma, C.H., Li, J.S., Du, C., Xu, S., Han, L.H., Li, L., and Wang, X.W. (2012) E-cadherin promotes proliferation of human ovarian cancer cells in vitro via activating MEK/ERK pathway. *Acta Pharmacol Sin*, **33**(6): 817–822.
- Dorogovtsev, S.N. and Mendes, J.F. (2003) *Evolution of networks: From biological nets to the Internet and WWW*. Oxford University Press, USA.
- 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. 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.
- 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.
- Fromental-Ramain, C., Warot, X., Lakkaraju, S., Favier, B., Haack, H., Birling, C., Dierich, A., Doll e, P., and Chambon, P. (1996) Specific and redundant functions of the paralogous Hoxa-9 and Hoxd-9 genes in forelimb and axial skeleton patterning. *Development*, **122**(2): 461–472.
- Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004) A census of human cancer genes. *Nat Rev Cancer*, **4**(3): 177–183.
- Futreal, P.A., Kasprzyk, A., Birney, E., Mullikin, J.C., Wooster, R., and Stratton, M.R. (2001) Cancer and genomics. *Nature*, **409**(6822): 850–852.
- Gao, B. and Roux, P.P. (2015) Translational control by oncogenic signaling pathways. *Biochimica et Biophysica Acta*, **1849**(7): 753–65.

- Gatza, M.L., Kung, H.N., Blackwell, K.L., Dewhirst, M.W., Marks, J.R., and Chi, J.T. (2011) Analysis of tumor environmental response and oncogenic pathway activation identifies distinct basal and luminal features in HER2-related breast tumor subtypes. *Breast Cancer Res*, **13**(3): R62.
- Gatza, M.L., Lucas, J.E., Barry, W.T., Kim, J.W., Wang, Q., Crawford, M.D., Datto, M.B., Kelley, M., Mathey-Prevot, B., Potti, A., *et al.* (2010) A pathway-based classification of human breast cancer. *Proc Natl Acad Sci USA*, **107**(15): 6994–6999.
- Gatza, M.L., Silva, G.O., Parker, J.S., Fan, C., and Perou, C.M. (2014) An integrated genomics approach identifies drivers of proliferation in luminal-subtype human breast cancer. *Nat Genet*, **46**(10): 1051–1059.
- Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*, **5**(10): R80.
- Genz, A. and Bretz, F. (2009) Computation of multivariate normal and t probabilities. In *Lecture Notes in Statistics*, volume 195. Springer-Verlag, Heidelberg.
- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2016) *mvtnorm: Multivariate Normal and t Distributions*. R package version 1.0-5. URL.
- Glaire, M.A., Brown, M., Church, D.N., and Tomlinson, I. (2017) Cancer predisposition syndromes: lessons for truly precision medicine. *J Pathol*, **241**(2): 226–235.
- Globus (Globus) (2017) Research data management simplified. <https://www.globus.org/>. Accessed: 25/03/2017.
- Goodwin, S., McPherson, J.D., and McCombie, W.R. (2016) Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*, **17**(6): 333–351.
- Grady, W.M., Willis, J., Guilford, P.J., Dunbier, A.K., Toro, T.T., Lynch, H., Wiesner, G., Ferguson, K., Eng, C., Park, J.G., *et al.* (2000) Methylation of the CDH1 promoter as the second genetic hit in hereditary diffuse gastric cancer. *Nat Genet*, **26**(1): 16–17.
- Graziano, F., Humar, B., and Guilford, P. (2003) The role of the E-cadherin gene (*CDH1*) in diffuse gastric cancer susceptibility: from the laboratory to clinical practice. *Annals of Oncology*, **14**(12): 1705–1713.

- 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.
- 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., 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.
- 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.
- 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.
- International HapMap 3 Consortium (HapMap) (2003) The International HapMap Project. *Nature*, **426**(6968): 789–796.
- 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, 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.

- 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.
- Kelly, S.T., Single, A.B., Telford, B.J., Beetham, H.G., Godwin, T.D., Chen, A., Black, M.A., and Guilford, P.J. (unpublished) 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*.
- 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.
- 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 SNAI1 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, 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.

- 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., Mastrogiannis, 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.
- 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.
- 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.

- 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.
- 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.
- 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.
- 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.
- 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.
- Schena, M. (1996) Genome analysis with gene expression microarrays. *Bioessays*, **18**(5): 427–431.
- Scheuer, L., Kauff, N., Robson, M., Kelly, B., Barakat, R., Satagopan, J., Ellis, N., Hensley, M., Boyd, J., Borgen, P., *et al.* (2002) Outcome of preventive surgery and screening for breast and ovarian cancer in BRCA mutation carriers. *J Clin Oncol*, **20**(5): 1260–1268.
- Semb, H. and Christofori, G. (1998) The tumor-suppressor function of E-cadherin. *Am J Hum Genet*, **63**(6): 1588–93.

- Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T. (2005) Rocr: visualizing classifier performance in r. *Bioinformatics*, **21**(20): 7881.
- Slurm development team (Slurm) (2017) Slurm workload manager. <https://slurm.schedmd.com/>. Accessed: 25/03/2017.
- Sørbye, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., *et al.* (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA*, **98**(19): 10869–10874.
- 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.
- 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) (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.

- 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.
- 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) triphosphate 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., Chandrashekhara, 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.
- 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 portal a 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 C

Mutation Analysis in Breast Cancer

C.1 Synthetic Lethal Genes and Pathways

SLIPT expression analysis (described in Section 3.1) on TCGA breast cancer data ($n = 969$) found the following genes and pathways, described in sections 4.1 and 4.1.1.

Table C.1: Candidate synthetic lethal gene partners of *CDH1* from mtSLIPT

Gene	Observed	Expected	χ^2 value	p-value	p-value (False discovery rate (FDR))
<i>TFAP2B</i>	8	36.7	89.5	3.60×10^{-20}	8.37×10^{-17}
<i>ZNF423</i>	15	36.7	78.8	7.89×10^{-18}	1.22×10^{-14}
<i>CALCOCO1</i>	11	36.7	76.8	2.09×10^{-17}	2.59×10^{-14}
<i>RBM5</i>	13	36.7	75.7	3.65×10^{-17}	4.00×10^{-14}
<i>BTG2</i>	7	36.7	71.7	2.72×10^{-16}	1.81×10^{-13}
<i>RXRA</i>	6	36.7	70.5	5.00×10^{-16}	2.97×10^{-13}
<i>SLC27A1</i>	11	36.7	70.3	5.42×10^{-16}	2.97×10^{-13}
<i>MEF2D</i>	12	36.7	69.6	7.86×10^{-16}	3.95×10^{-13}
<i>NISCH</i>	12	36.7	69.6	7.86×10^{-16}	3.95×10^{-13}
<i>AVPR2</i>	9	36.7	69.2	9.36×10^{-16}	4.58×10^{-13}
<i>CRY2</i>	13	36.7	68.9	1.07×10^{-15}	4.98×10^{-13}
<i>RAPGEF3</i>	13	36.7	68.9	1.07×10^{-15}	4.98×10^{-13}
<i>NRIP2</i>	10	36.7	68.2	1.58×10^{-15}	7.18×10^{-13}
<i>DARC</i>	12	36.7	66.4	3.76×10^{-15}	1.54×10^{-12}
<i>SFRS5</i>	12	36.7	66.4	3.76×10^{-15}	1.54×10^{-12}
<i>NOSTRIN</i>	5	36.7	65.1	7.40×10^{-15}	2.70×10^{-12}
<i>KIF13B</i>	12	36.7	63.4	1.69×10^{-14}	5.16×10^{-12}
<i>TENC1</i>	10	36.7	62.5	2.67×10^{-14}	7.40×10^{-12}
<i>MFAP4</i>	12	36.7	60.5	7.17×10^{-14}	1.67×10^{-11}
<i>ELN</i>	13	36.7	59.7	1.07×10^{-13}	2.32×10^{-11}
<i>SGK223</i>	14	36.7	59	1.51×10^{-13}	3.05×10^{-11}
<i>KIF12</i>	11	36.7	58.8	1.74×10^{-13}	3.34×10^{-11}
<i>SELP</i>	11	36.7	58.8	1.74×10^{-13}	3.34×10^{-11}
<i>CIRBP</i>	9	36.7	58.7	1.83×10^{-13}	3.41×10^{-11}
<i>CTDSP1</i>	9	36.7	58.7	1.83×10^{-13}	3.41×10^{-11}

Strongest candidate SL partners for *CDH1* by mtSLIPT with observed and expected numbers of *CDH1* mutant TCGA breast tumours with low expression of partner genes.

Table C.2: Pathways for *CDH1* partners from mtSLIPT

Pathways Over-represented	Pathway Size	SL Genes	p-value (FDR)
Eukaryotic Translation Elongation	86	60	2.0×10^{-128}
Peptide chain elongation	83	59	2.0×10^{-128}
Eukaryotic Translation Termination	83	58	2.3×10^{-125}
Viral mRNA Translation	81	57	2.5×10^{-124}
Nonsense Mediated Decay independent of the Exon Junction Complex	88	59	8.6×10^{-124}
Nonsense-Mediated Decay	103	61	5.2×10^{-117}
Nonsense Mediated Decay enhanced by the Exon Junction Complex	103	61	5.2×10^{-117}
Formation of a pool of free 40S subunits	93	58	1.6×10^{-116}
L13a-mediated translational silencing of Ceruloplasmin expression	103	59	1.3×10^{-111}
3' -UTR-mediated translational regulation	103	59	1.3×10^{-111}
GTP hydrolysis and joining of the 60S ribosomal subunit	104	59	6.2×10^{-111}
SRP-dependent cotranslational protein targeting to membrane	104	58	2.9×10^{-108}
Eukaryotic Translation Initiation	111	59	3.0×10^{-106}
Cap-dependent Translation Initiation	111	59	3.0×10^{-106}
Influenza Viral RNA Transcription and Replication	108	57	5.1×10^{-103}
Influenza Infection	117	59	1.5×10^{-102}
Translation	141	64	3.7×10^{-101}
Influenza Life Cycle	112	57	1.4×10^{-100}
GPCR downstream signalling	472	116	1.0×10^{-80}
Hemostasis	422	105	1.4×10^{-78}

Gene set over-representation analysis (hypergeometric test) for Reactome pathways in mtSLIPT partners for *CDH1*.

The genes and pathways identified in Tables C.1 and C.2 were derived from comparing the expression profiles of potential partners to the mutation status of *CDH1* (as shown in Figure 3.2). The following analysis was limited to the samples for which both expression and somatic mutation data were available from TCGA.

C.2 Synthetic Lethal Expression Profiles

Similar to the analysis of synthetic lethal partners against low *CDH1* expression in 4.1.2, the partners detected from *CDH1* mutation were also examined for their expression profiles and the pathway composition of gene clusters. Hierarchical clustering was performed on mtSLIPT partners for *CDH1* as showing in Figure C.1. Over-representation for Reactome pathways for each of the gene clusters identified is given in Table C.3.

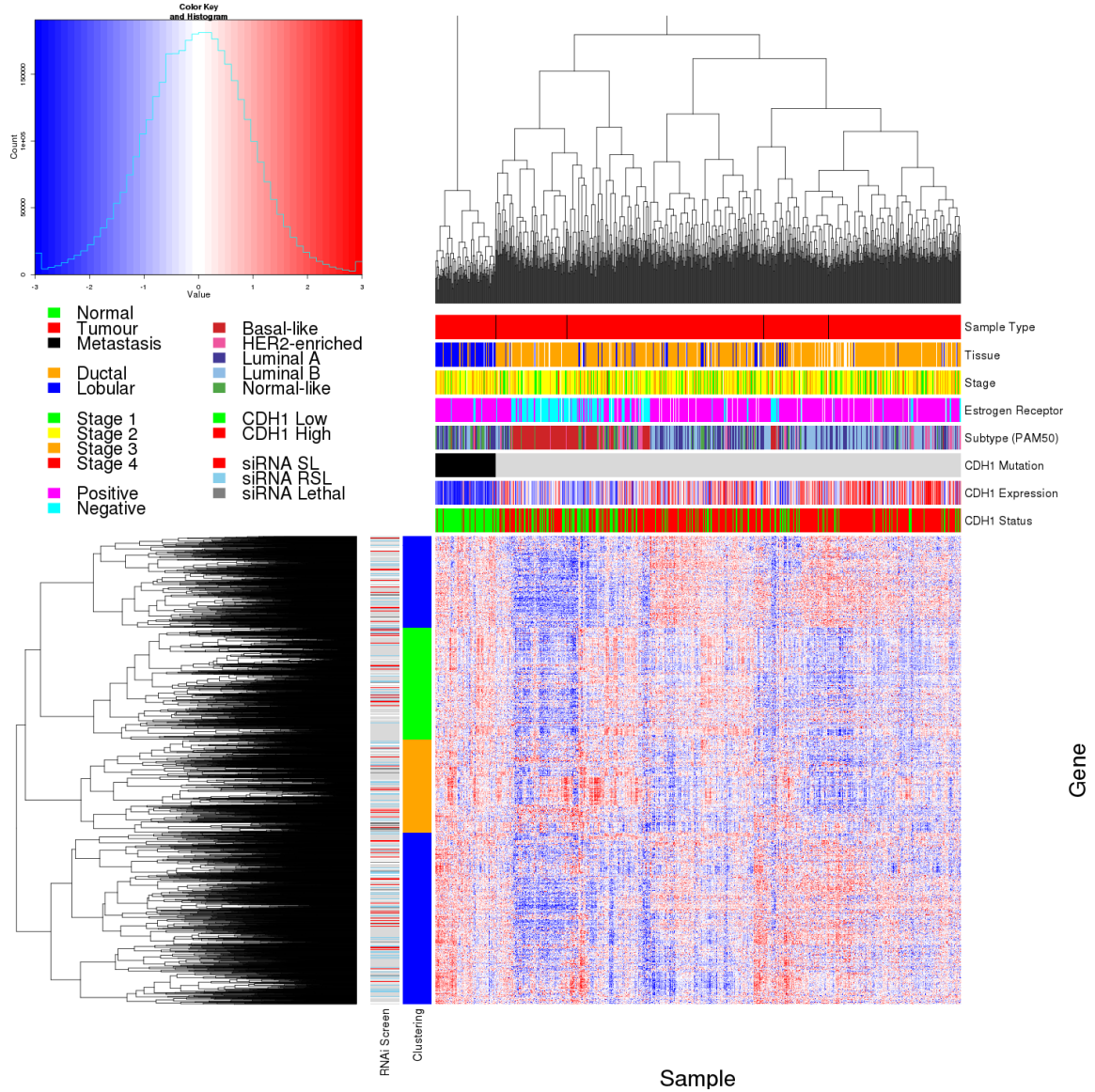


Figure C.1: **Synthetic lethal expression profiles of analysed samples.** Gene expression profile heatmap (correlation distance) of all samples (separated by *CDH1* somatic mutation status) analysed in TCGA breast cancer dataset for gene expression of 3743 candidate partners of E-cadherin (*CDH1*) from mtSLIPT prediction (with significant FDR adjusted $p < 0.05$). Deeply clustered, inter-correlated genes form several main groups, each containing genes that were SL candidates or toxic in an siRNA screen Telford *et al.* (2015). Clusters had different sample groups highly expressing the synthetic lethal candidates in *CDH1* mutant samples and often lowly expressing *CDH1* wildtype samples (which were not tested for), although many of the *CDH1* mutant samples had among the lowest *CDH1* expression. In contrast to the expression analysis the (predominantly *CDH1* wildtype) basal subtype and ER negative samples have depleted expression among most candidate synthetic lethal partners.

Table C.3: Pathways for clusters of *CDH1* partners from mtSLIPT

Pathways Over-represented in Cluster 1	Pathway Size	Cluster Genes	p-value (FDR)
Olfactory Signalling Pathway	57	8	7.1×10^{-9}
Assembly of the primary cilium	149	14	8.0×10^{-9}
Sphingolipid metabolism	62	8	9.6×10^{-9}
Signalling by ERBB4	133	12	5.1×10^{-8}
PI3K Cascade	65	7	4.9×10^{-7}
Circadian Clock	33	5	4.9×10^{-7}
Nuclear signalling by ERBB4	34	5	4.9×10^{-7}
Intraflagellar transport	35	5	4.9×10^{-7}
PI3K events in ERBB4 signalling	87	8	4.9×10^{-7}
PIP3 activates AKT signalling	87	8	4.9×10^{-7}
PI3K events in ERBB2 signalling	87	8	4.9×10^{-7}
PI-3K cascade:FGFR1	87	8	4.9×10^{-7}
PI-3K cascade:FGFR2	87	8	4.9×10^{-7}
PI-3K cascade:FGFR3	87	8	4.9×10^{-7}
PI-3K cascade:FGFR4	87	8	4.9×10^{-7}
Deadenylation of mRNA	22	4	5.6×10^{-7}
PI3K/AKT activation	90	8	5.6×10^{-7}
Cargo trafficking to the periciliary membrane	38	5	5.6×10^{-7}
Pathways Over-represented in Cluster 2	Pathway Size	Cluster Genes	p-value (FDR)
G _{αs} signalling events	83	19	5.1×10^{-25}
Extracellular matrix organization	238	30	1.4×10^{-18}
Hemostasis	422	46	2.7×10^{-16}
Aquaporin-mediated transport	32	9	2.7×10^{-16}
Transcriptional regulation of white adipocyte differentiation	56	11	1.7×10^{-15}
Degradation of the extracellular matrix	102	15	1.7×10^{-15}
Integration of energy metabolism	84	13	8.8×10^{-15}
GPCR downstream signalling	472	48	2.8×10^{-14}
G _{αz} signalling events	15	6	5.0×10^{-14}
Molecules associated with elastic fibres	33	8	5.4×10^{-14}
Phase 1 - Functionalization of compounds	67	11	5.6×10^{-14}
Platelet activation, signalling and aggregation	179	20	5.6×10^{-14}
Vasopressin regulates renal water homeostasis via Aquaporins	24	7	6.1×10^{-14}
Elastic fibre formation	37	8	$.03 \times 10^{-13}$
Calmodulin induced events	27	7	3.3×10^{-13}
CaM pathway	27	7	3.3×10^{-13}
cGMP effects	18	6	3.6×10^{-13}
G _{αi} signalling events	167	18	6.3×10^{-13}
Pathways Over-represented in Cluster 3	Pathway Size	Cluster Genes	p-value (FDR)
Eukaryotic Translation Elongation	86	55	1.1×10^{-112}
Peptide chain elongation	83	54	1.3×10^{-112}
Viral mRNA Translation	81	53	1.6×10^{-111}
Eukaryotic Translation Termination	83	53	7.1×10^{-110}
Nonsense Mediated Decay independent of the Exon Junction Complex	88	54	1.0×10^{-108}
Formation of a pool of free 40S subunits	93	53	4.1×10^{-102}
Nonsense-Mediated Decay	103	54	3.9×10^{-98}
Nonsense Mediated Decay enhanced by the Exon Junction Complex	103	54	3.9×10^{-98}
LI3a-mediated translational silencing of Ceruloplasmin expression	103	53	1.2×10^{-95}
3' -UTR-mediated translational regulation	103	53	1.2×10^{-95}
SRP-dependent cotranslational protein targeting to membrane	104	53	4.3×10^{-95}
GTP hydrolysis and joining of the 60S ribosomal subunit	104	53	4.3×10^{-95}
Influenza Viral RNA Transcription and Replication	108	53	9.6×10^{-93}
Eukaryotic Translation Initiation	111	53	4.2×10^{-91}
Cap-dependent Translation Initiation	111	53	4.2×10^{-91}
Influenza Life Cycle	112	53	1.4×10^{-90}
Influenza Infection	117	53	6.2×10^{-88}
Translation	141	55	3×10^{-81}
Pathways Over-represented in Cluster 4	Pathway Size	Cluster Genes	p-value (FDR)
ECM proteoglycans	66	10	2.9×10^{-11}
deactivation of the beta-catenin transactivating complex	38	7	5.1×10^{-10}
Arachidonic acid metabolism	41	7	1.1×10^{-9}
G _{αq} signalling events	149	14	4.0×10^{-9}
HS-GAG degradation	21	5	4.5×10^{-9}
Uptake and actions of bacterial toxins	22	5	6.1×10^{-9}
Gastrin-CREB signalling pathway via PKC and MAPK	170	15	6.1×10^{-9}
RNA Polymerase I, RNA Polymerase III, and Mitochondrial Transcription	64	8	6.1×10^{-9}
Non-integrin membrane-ECM interactions	53	7	1.5×10^{-8}
Syndecan interactions	25	5	1.5×10^{-8}
NOTCH1 Intracellular Domain Regulates Transcription	40	6	2.3×10^{-8}
Synthesis of Leukotrienes and Eoxins	15	4	3.2×10^{-8}
Signalling by NOTCH1	59	7	5.3×10^{-8}
Regulation of insulin secretion	44	6	6.0×10^{-8}
Metabolism of lipids and lipoproteins	471	37	8.2×10^{-8}
Signalling by NOTCH	80	8	1.2×10^{-7}
Platelet activation, signalling and aggregation	179	14	1.2×10^{-7}
Recruitment of mitotic centrosome proteins and complexes	64	7	1.2×10^{-7}

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

C.3 Comparison to Primary Screen

The mutation synthetic lethal partners with *CDH1* were also compared to siRNA primary screen data (Telford *et al.*, 2015), as performed in Section 4.2.1. These were expected to be more concordant with the experimental results performed on a null mutant, however this was not the case at the gene level: less genes overlapped with experimental candidates in Figure C.2. This discrepancy may be due to lower sample size for mutations in TCGA data or lower frequency (expected value) of *CDH1* mutations compared to low expression.

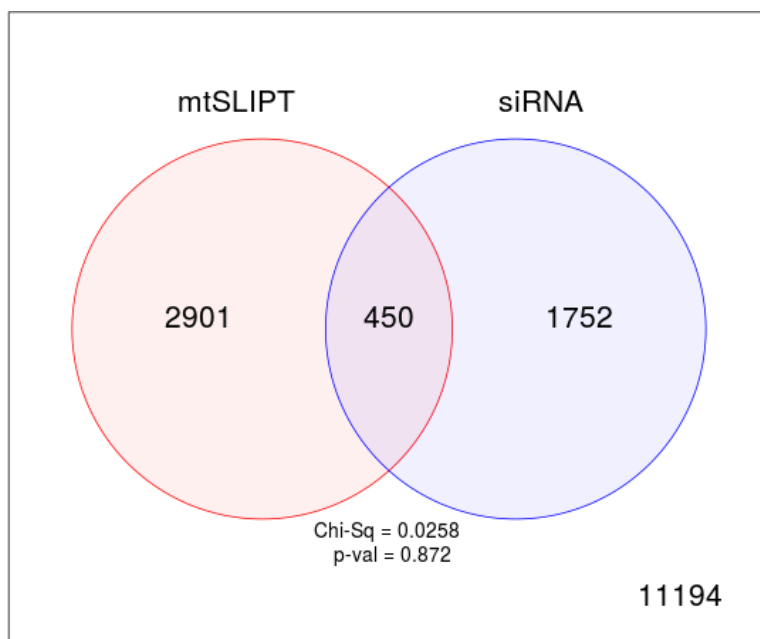


Figure C.2: **Comparison of mtSLIPT to siRNA.** Testing the overlap of gene candidates for *E-cadherin* synthetic lethal partners between computational (SLIPT) and experimental screening (siRNA) approaches. The χ^2 test suggests that the overlap is no more than would be expected by chance ($p = 0.281$).

Despite a lower sample size (and low number of a predicted partners) for mutation analysis, the pathway composition (Tables C.2 and C.4) was similar to expression analysis, as described in Section 4.2.5. In particular, the resampling analysis (Section C.3.1) supported many of the results of expression analysis (Section 4.2.5.1). Tables C.5 and C.6 detected many of the same or functionally-related pathways.

Table C.4: Pathways for *CDH1* partners from mtSLIPT and siRNA

Predicted only by SLIPT (2901 genes)	Pathway Size	Genes Identified	p-value (FDR)
Eukaryotic Translation Elongation	87	57	2.8×10^{-120}
Peptide chain elongation	84	56	3.1×10^{-120}
Eukaryotic Translation Termination	84	55	2.8×10^{-117}
Viral mRNA Translation	82	54	4.1×10^{-116}
Nonsense Mediated Decay independent of the Exon Junction Complex	89	55	3.7×10^{-113}
Formation of a pool of free 40S subunits	94	55	2.8×10^{-109}
Nonsense-Mediated Decay	104	57	8.4×10^{-108}
Nonsense Mediated Decay enhanced by the Exon Junction Complex	104	57	8.4×10^{-108}
L13a-mediated translational silencing of Ceruloplasmin expression	104	56	3.4×10^{-105}
3' -UTR-mediated translational regulation	104	56	3.4×10^{-105}
GTP hydrolysis and joining of the 60S ribosomal subunit	105	56	1.4×10^{-104}
Eukaryotic Translation Initiation	112	56	2.8×10^{-100}
Cap-dependent Translation Initiation	112	56	2.8×10^{-100}
SRP-dependent cotranslational protein targeting to membrane	105	54	2.2×10^{-99}
Influenza Viral RNA Transcription and Replication	109	54	5.3×10^{-97}
Influenza Life Cycle	113	54	9.6×10^{-95}
Influenza Infection	118	55	1.7×10^{-94}
Translation	142	60	3.5×10^{-94}
Infectious disease	349	77	5.9×10^{-62}
Extracellular matrix organization	241	54	3.0×10^{-52}

Detected only by siRNA screen (1752 genes)	Pathway Size	Genes Identified	p-value (FDR)
Class A/1 (Rhodopsin-like receptors)	282	69	1.9×10^{-59}
GPCR ligand binding	363	78	2.7×10^{-54}
Peptide ligand-binding receptors	175	41	1.5×10^{-42}
$G_{\alpha i}$ signalling events	184	41	1.1×10^{-40}
Gastrin-CREB signalling pathway via PKC and MAPK	180	37	1.5×10^{-35}
$G_{\alpha q}$ signalling events	159	34	3.7×10^{-35}
DAP12 interactions	159	27	1.1×10^{-24}
VEGFA-VEGFR2 Pathway	91	19	1.0×10^{-23}
Downstream signal transduction	146	24	1.9×10^{-22}
Signalling by VEGF	99	19	2.6×10^{-22}
DAP12 signalling	149	24	4.2×10^{-22}
Organelle biogenesis and maintenance	264	34	4.3×10^{-20}
Downstream signalling of activated FGFR1	134	21	4.3×10^{-20}
Downstream signalling of activated FGFR2	134	21	4.3×10^{-20}
Downstream signalling of activated FGFR3	134	21	4.3×10^{-20}
Downstream signalling of activated FGFR4	134	21	4.3×10^{-20}
Signalling by ERBB2	146	22	5.3×10^{-20}
Signalling by FGFR	146	22	5.3×10^{-20}
Signalling by FGFR1	146	22	5.3×10^{-20}
Signalling by FGFR2	146	22	5.3×10^{-20}

Intersection of SLIPT and siRNA screen (450 genes)	Pathway Size	Genes Identified	p-value (FDR)
HS-GAG degradation	21	4	4.9×10^{-6}
Retinoid metabolism and transport	39	5	4.9×10^{-6}
Platelet activation, signalling and aggregation	186	13	4.9×10^{-6}
Signalling by NOTCH4	11	3	4.9×10^{-6}
$G_{\alpha s}$ signalling events	100	8	5.0×10^{-6}
Defective EXT2 causes exostoses 2	12	3	5.0×10^{-6}
Defective EXT1 causes exostoses 1, TRPS2 and CHDS	12	3	5.0×10^{-6}
Class A/1 (Rhodopsin-like receptors)	289	18	2.2×10^{-5}
Signalling by PDGF	173	11	2.9×10^{-5}
Circadian Clock	34	4	2.9×10^{-5}
Signalling by ERBB4	139	9	4.3×10^{-5}
Role of LAT2/NTAL/LAB on calcium mobilization	99	7	4.4×10^{-5}
Peptide ligand-binding receptors	181	11	4.5×10^{-5}
Defective B4GALT7 causes EDS, progeroid type	19	3	4.5×10^{-5}
Defective B3GAT3 causes JDSSDHD	19	3	4.5×10^{-5}
Signalling by NOTCH	80	6	4.5×10^{-5}
$G_{\alpha q}$ signalling events	164	10	5.1×10^{-5}
Response to elevated platelet cytosolic Ca^{2+}	84	6	7.1×10^{-5}
Signalling by ERBB2	148	9	7.1×10^{-5}
Signalling by SCF-KIT	129	8	8.3×10^{-5}

C.3.1 Resampling Analysis

Table C.5: Pathways for *CDH1* partners from mtSLIPT

Reactome Pathway	Over-representation	Permutation
Eukaryotic Translation Elongation	3.2×10^{-128}	$< 7.035 \times 10^{-4}$
Peptide chain elongation	3.2×10^{-128}	$< 7.035 \times 10^{-4}$
Eukaryotic Translation Termination	3.7×10^{-125}	$< 7.035 \times 10^{-4}$
Viral mRNA Translation	4.1×10^{-124}	$< 7.035 \times 10^{-4}$
Nonsense Mediated Decay independent of the Exon Junction Complex	1.4×10^{-123}	$< 7.035 \times 10^{-4}$
Nonsense-Mediated Decay	8.4×10^{-117}	$< 7.035 \times 10^{-4}$
Nonsense Mediated Decay enhanced by the Exon Junction Complex	8.4×10^{-117}	$< 7.035 \times 10^{-4}$
Formation of a pool of free 40S subunits	2.6×10^{-116}	$< 7.035 \times 10^{-4}$
L13a-mediated translational silencing of Ceruloplasmin expression	2.0×10^{-111}	$< 7.035 \times 10^{-4}$
3' -UTR-mediated translational regulation	2.0×10^{-111}	$< 7.035 \times 10^{-4}$
GTP hydrolysis and joining of the 60S ribosomal subunit	9.9×10^{-111}	$< 7.035 \times 10^{-4}$
SRP-dependent cotranslational protein targeting to membrane	4.7×10^{-108}	$< 7.035 \times 10^{-4}$
Eukaryotic Translation Initiation	4.8×10^{-106}	$< 7.035 \times 10^{-4}$
Cap-dependent Translation Initiation	4.8×10^{-106}	$< 7.035 \times 10^{-4}$
Influenza Viral RNA Transcription and Replication	8.1×10^{-103}	$< 7.035 \times 10^{-4}$
Influenza Infection	2.4×10^{-102}	$< 7.035 \times 10^{-4}$
Translation	6.0×10^{-101}	$< 7.035 \times 10^{-4}$
Influenza Life Cycle	2.2×10^{-100}	$< 7.035 \times 10^{-4}$
Disease	2.1×10^{-90}	0.013347
GPCR downstream signalling	1.6×10^{-80}	0.095478
Hemostasis	2.1×10^{-78}	0.2671
Signalling by GPCR	1.2×10^{-73}	0.44939
<i>Extracellular matrix organization</i>	2.2×10^{-67}	0.054008
Metabolism of proteins	1.4×10^{-66}	0.9607
Signal Transduction	2.1×10^{-66}	0.48184
Developmental Biology	2.5×10^{-66}	0.54075
Innate Immune System	5.3×10^{-66}	0.9589
Infectious disease	9.6×10^{-66}	0.21075
Signalling by NGF	1.1×10^{-62}	0.43356
Immune System	2.8×10^{-62}	0.23052

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

Table C.6: Pathways for *CDH1* partners from mtSLIPT and siRNA primary screen

Reactome Pathway	Over-representation	Permutation
Visual phototransduction	1.2×10^{-9}	0.86279
G_{as} signalling events	2.9×10^{-7}	0.023066
Retinoid metabolism and transport	2.9×10^{-7}	0.299
Acyl chain remodelling of PS	1.1×10^{-5}	0.42584
Transcriptional regulation of white adipocyte differentiation	1.1×10^{-5}	0.53928
Chemokine receptors bind chemokines	1.1×10^{-5}	0.95259
<i>Signalling by NOTCH4</i>	1.2×10^{-5}	0.079229
Defective EXT2 causes exostoses 2	1.2×10^{-5}	0.22292
Defective EXT1 causes exostoses 1, TRPS2 and CHDS	1.2×10^{-5}	0.22292
Platelet activation, signalling and aggregation	1.2×10^{-5}	0.48853
Serotonin receptors	1.4×10^{-5}	0.34596
Nicotinamide salvaging	1.4×10^{-5}	0.70881
Phase 1 - Functionalization of compounds	2×10^{-5}	0.31142
Amine ligand-binding receptors	2.5×10^{-5}	0.34934
Acyl chain remodelling of PE	3.8×10^{-5}	0.42615
Signalling by GPCR	3.8×10^{-5}	0.93888
Molecules associated with elastic fibres	3.9×10^{-5}	0.017982
DAP12 interactions	3.9×10^{-5}	0.71983
Beta defensins	3.9×10^{-5}	0.91458
Cytochrome P ₄₅₀ - arranged by substrate type	4.7×10^{-5}	0.83493
GPCR ligand binding	5.7×10^{-5}	0.95258
Acyl chain remodelling of PC	6.1×10^{-5}	0.42584
Response to elevated platelet cytosolic Ca ²⁺	6.4×10^{-5}	0.54046
Arachidonic acid metabolism	6.7×10^{-5}	0.026696
Defective B4GALT7 causes EDS, progeroid type	7.3×10^{-5}	0.24921
Defective B3GAT3 causes JDSSDHD	7.3×10^{-5}	0.24921
Hydrolysis of LPC	7.3×10^{-5}	0.80663
Elastic fibre formation	7.4×10^{-5}	0.0058768
HS-GAG degradation	9.4×10^{-5}	0.0083179
<i>Bile acid and bile salt metabolism</i>	9.4×10^{-5}	0.079905
Netrin-1 signalling	0.00011	0.92216
Integration of energy metabolism	0.00011	0.011152
Dectin-2 family	0.00012	0.10385
Platelet sensitization by LDL	0.00012	0.34596
DAP12 signalling	0.00012	0.62787
Defensins	0.00012	0.77542
GPCR downstream signalling	0.00012	0.79454
<i>Diseases associated with glycosaminoglycan metabolism</i>	0.00013	0.065927
<i>Diseases of glycosylation</i>	0.00013	0.065927
Signalling by Retinoic Acid	0.00013	0.22292
Signalling by Leptin	0.00013	0.34596
Signalling by SCF-KIT	0.00013	0.70881
Opioid Signalling	0.00013	0.96053
Signalling by NOTCH	0.00015	0.26884
Platelet homeostasis	0.00015	0.4878
Signalling by NOTCH1	0.00016	0.13043
Class B/2 (Secretin family receptors)	0.00016	0.13994
<i>Diseases of Immune System</i>	0.0002	0.0795
<i>Diseases associated with the TLR signalling cascade</i>	0.0002	0.0795
A tetrasaccharide linker sequence is required for GAG synthesis	0.0002	0.42615

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

C.4 Compare SLIPT genes

The mutation synthetic lethal partners with *CDH1* were also compared to siRNA primary screen data (Telford *et al.*, 2015), by correlation and siRNA viability as described in sections 4.2.2 and 4.2.3.

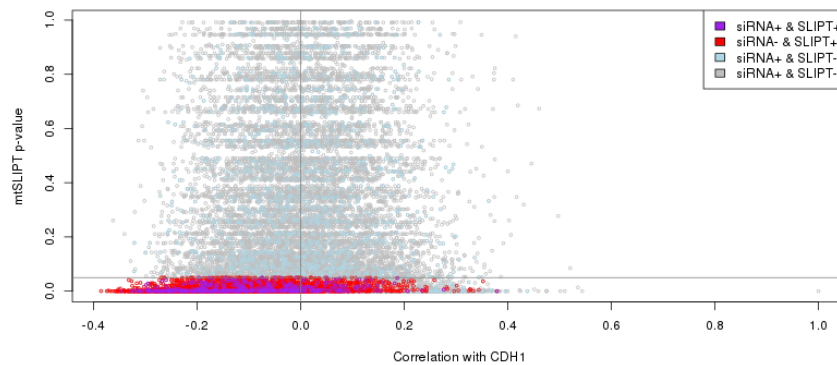


Figure C.3: **Compare mtSLIPT and siRNA genes with correlation.** The mtSLIPT p-values were compared against Pearson's correlation of expression with *CDH1*. Genes detected by SLIPT or siRNA were coloured according to the legend.

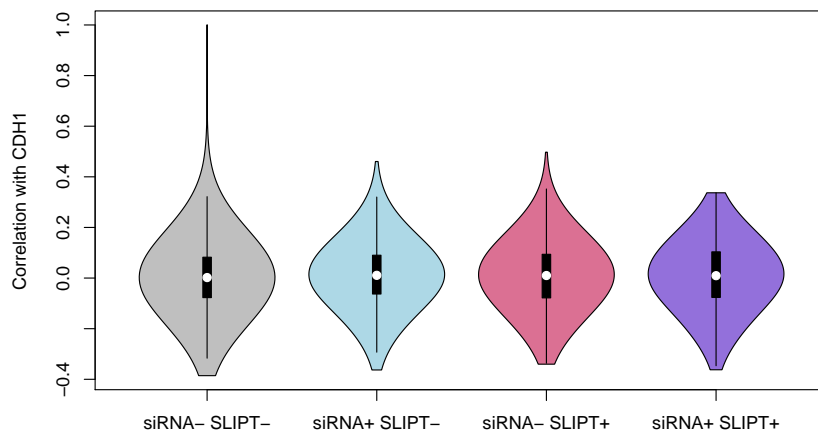


Figure C.4: **Compare mtSLIPT and siRNA genes with correlation.** Genes detected by mtSLIPT against *CDH1* mutation and siRNA screening were compared against Pearson's correlation of expression with *CDH1*. There were no differences in correlation between the gene groups.

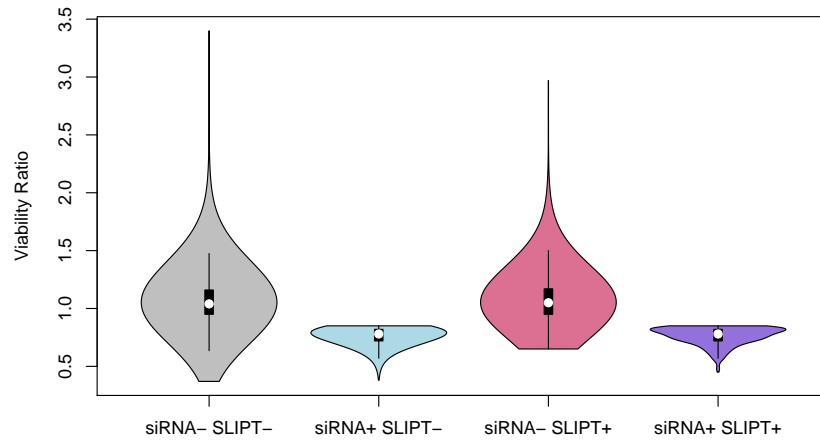


Figure C.5: **Compare *mtSLIPT* and *siRNA* genes with *siRNA* viability.** Genes detected as candidate synthetic lethal partners by *mtSLIPT* (in TCGA breast cancer) expression analysis against *CDH1* mutation and experimental screening (with *siRNA*) were compared against the viability ratio of *CDH1* mutant and wildtype cells in the primary *siRNA* screen. There were clear no differences in viability between genes detected by *mtSLIPT* and those not with the differences being primarily due to viability thresholds that were used to detect synthetic lethality by Telford *et al.* (2015).

C.5 Metagene Analysis

Metagene analysis was performed for synthetic lethal pathways against *CDH1* mutation. These were described and compared to expression analysis in Section 4.3.3.

Table C.7: Candidate synthetic lethal metagenes against *CDH1* from mtSLIPT

Pathway	ID	Observed	Expected	χ^2 value	p-value	p-value (FDR)
Neurotoxicity of clostridium toxins	168799	8	36.7	79.4	5.71×10^{-18}	3.14×10^{-15}
Aquaporin-mediated transport	445717	8	36.7	76.3	2.73×10^{-17}	9.01×10^{-15}
Toxicity of botulinum toxin type G (BoNT/G)	5250989	8	36.7	76.3	2.73×10^{-17}	9.01×10^{-15}
ABC-family proteins mediated transport	382556	10	36.7	68.2	1.58×10^{-15}	1.86×10^{-13}
G α_s signalling events	418597	10	36.7	59.9	9.97×10^{-14}	5.48×10^{-12}
Regulation of IGF transport and uptake by IGFBPs	381426	9	36.7	56.3	5.88×10^{-13}	2.11×10^{-11}
GP1b-IX-V activation signalling	430116	8	36.7	55.7	8.20×10^{-13}	2.76×10^{-11}
GABA receptor activation	977443	12	36.7	55.1	1.07×10^{-12}	3.26×10^{-11}
Vasopressin regulates renal water homeostasis via Aquaporins	432040	9	36.7	54.1	1.77×10^{-12}	4.88×10^{-11}
Toxicity of botulinum toxin type D (BoNT/D)	5250955	14	36.7	53.4	2.54×10^{-12}	6.64×10^{-11}
Toxicity of botulinum toxin type F (BoNT/F)	5250981	14	36.7	53.4	2.54×10^{-12}	6.64×10^{-11}
STAT6-mediated induction of chemokines	3249367	16	36.7	52.2	4.72×10^{-12}	1.13×10^{-10}
Toxicity of botulinum toxin type B (BoNT/B)	5250958	14	36.7	50.8	9.5×10^{-12}	1.98×10^{-10}
S6K1 signalling	165720	12	36.7	50.2	1.24×10^{-11}	2.5×10^{-10}
G α_s signalling events	418555	11	36.7	49.2	2.08×10^{-11}	3.85×10^{-10}
RHO GTPases activate CIT	5625900	14	36.7	48.2	3.34×10^{-11}	5.9×10^{-10}
NADE modulates death signalling	205025	15	36.7	47.4	5.00×10^{-11}	8.32×10^{-10}
Keratan sulfate degradation	2022857	10	36.7	46.6	7.5×10^{-11}	1.15×10^{-9}
Signalling by Retinoic Acid	5362517	10	36.7	46.6	7.5×10^{-11}	1.15×10^{-9}
Adenylate cyclase inhibitory pathway	170670	14	36.7	45.9	1.11×10^{-10}	1.59×10^{-9}
Inhibition of adenylate cyclase pathway	997269	14	36.7	45.9	1.11×10^{-10}	1.59×10^{-9}
Fatty acids	211935	6	36.7	45.7	1.21×10^{-10}	1.72×10^{-9}
Ionotropic activity of Kainate Receptors	451306	13	36.7	44.6	2.03×10^{-10}	2.58×10^{-9}
Activation of Ca-permeable Kainate Receptor	451308	13	36.7	44.6	2.03×10^{-10}	2.58×10^{-9}
RA biosynthesis pathway	5365859	13	36.7	44.6	2.03×10^{-10}	2.58×10^{-9}

Strongest candidate SL partners for *CDH1* by mtSLIPT with observed and expected numbers of mutant *CDH1* TCGA breast cancer tumours with low expression of partner metagenes.

C.6 Expression of Somatic Mutations

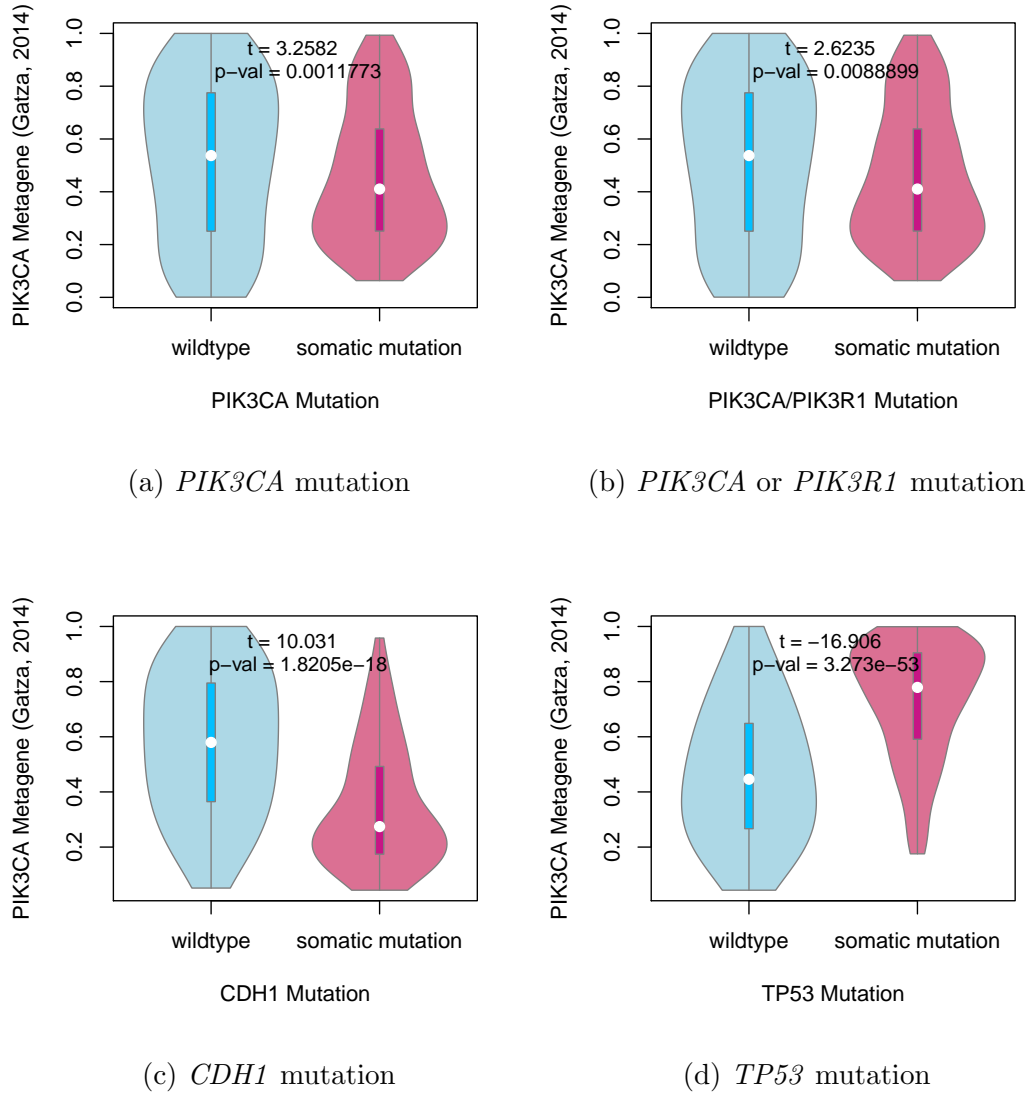
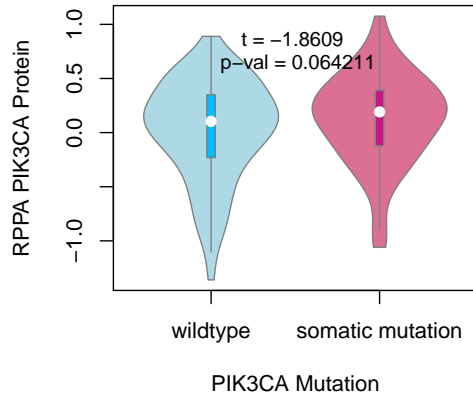
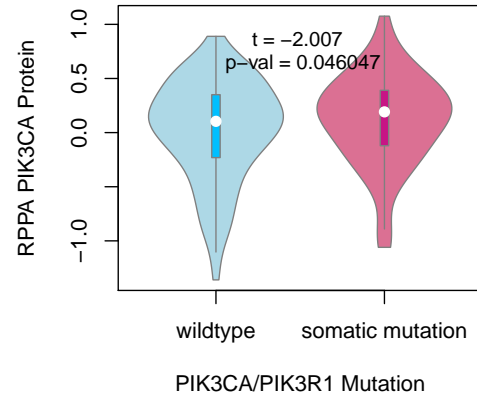


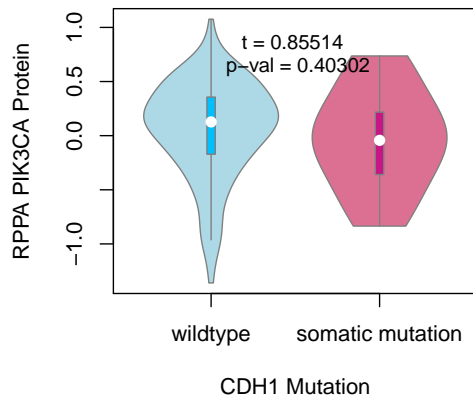
Figure C.6: **Somatic mutation against PIK3CA metagene.** Mutations in *PIK3CA*, *PIK3R1*, *CDH1*, and *TP53* were examined in TCGA breast cancer for their effect on the PIK3CA (Gatza *et al.*, 2014) pathway metagene. The tumour suppressors *CDH1* and *TP53* showed an increase and decrease in the metagene respectively, whereas *PIK3CA* and *PIK3R1* mutations weaker evidence of decrease in metagene levels.



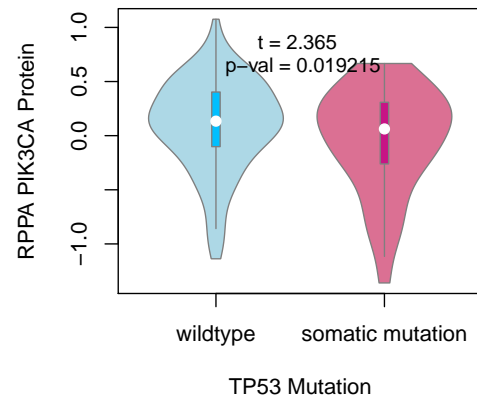
(a) *PIK3CA* mutation



(b) *PIK3CA* or *PIK3R1* mutation

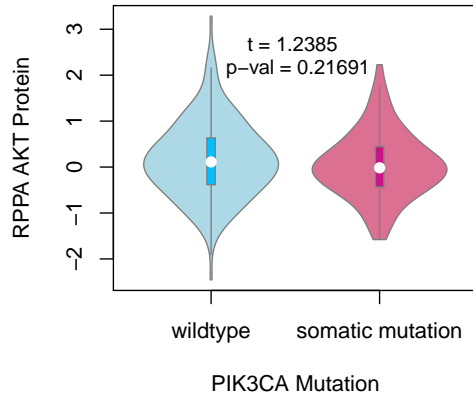


(c) *CDH1* mutation

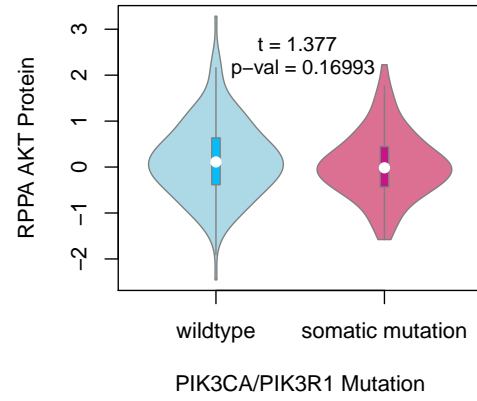


(d) *TP53* mutation

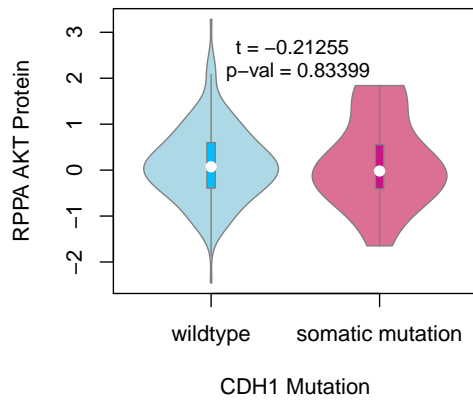
Figure C.7: **Somatic mutation against PI3K protein.** Mutations in *PIK3CA*, *PIK3R1*, *CDH1*, and *TP53* were examined in TCGA breast cancer for their effect on the expression of the p110 α protein (encoded by *PIK3CA*). Protein levels were significantly elevated in samples with *PIK3CA* or *PIK3R1* mutations and lower in samples with *TP53* mutations.



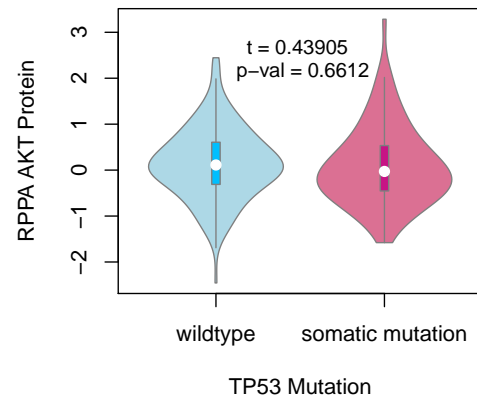
(a) *PIK3CA* mutation



(b) *PIK3CA* or *PIK3R1* mutation



(c) *CDH1* mutation



(d) *TP53* mutation

Figure C.8: **Somatic mutation against AKT protein.** Mutations in *PIK3CA*, *PIK3R1*, *CDH1*, and *TP53* were examined in TCGA breast cancer for their effect on the expression of the AKT protein (a downstream target of *PIK3CA*). Protein levels were not significantly different in samples mutations in any of these cancer genes.

C.7 Metagene Expression Profiles

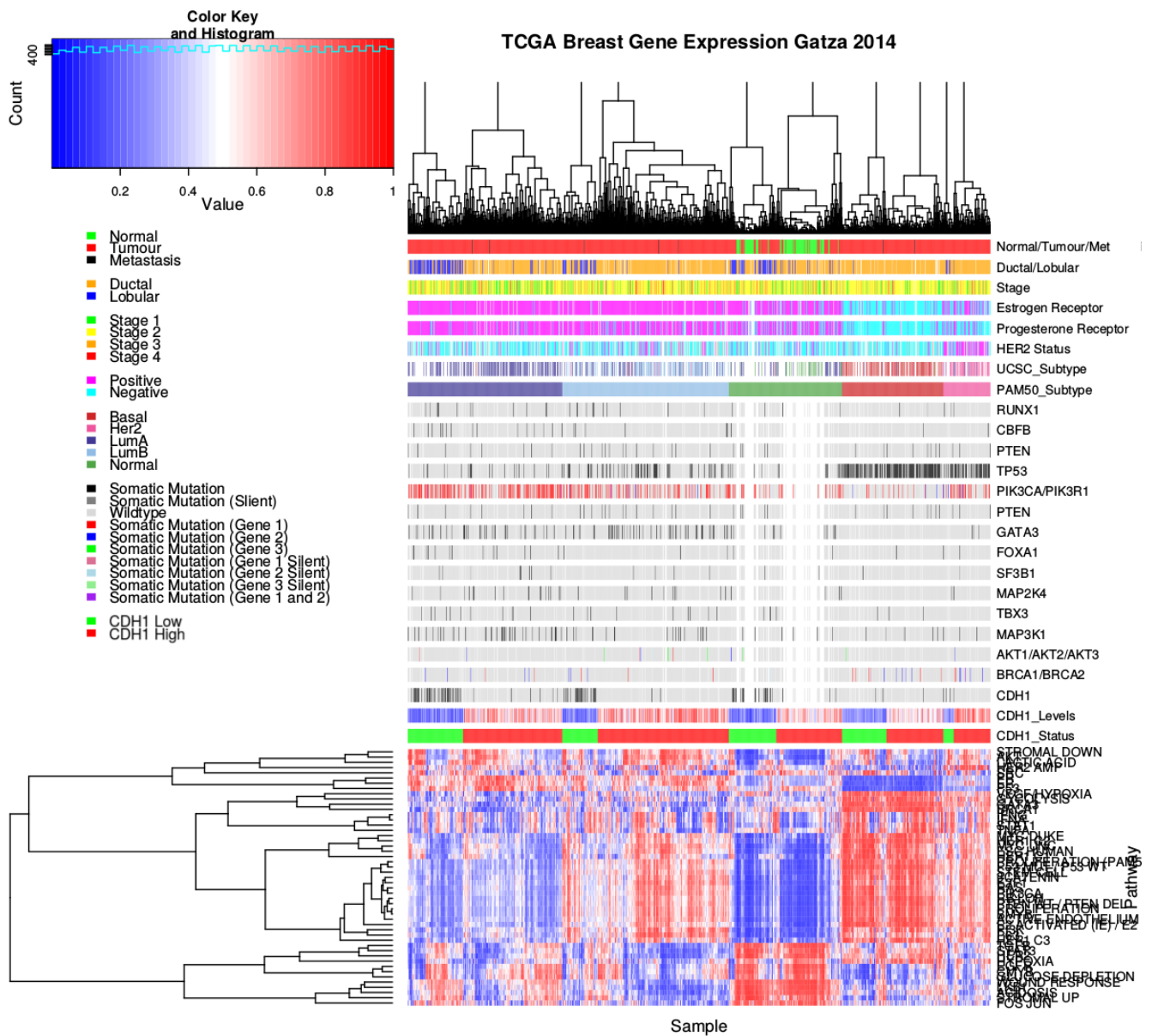


Figure C.9: **Pathway metagene expression profiles.** Expression profiles for metagene signatures from Gatz *et al.* (2014) in TCGA breast data, annotated for clinical factors and cancer gene mutations.

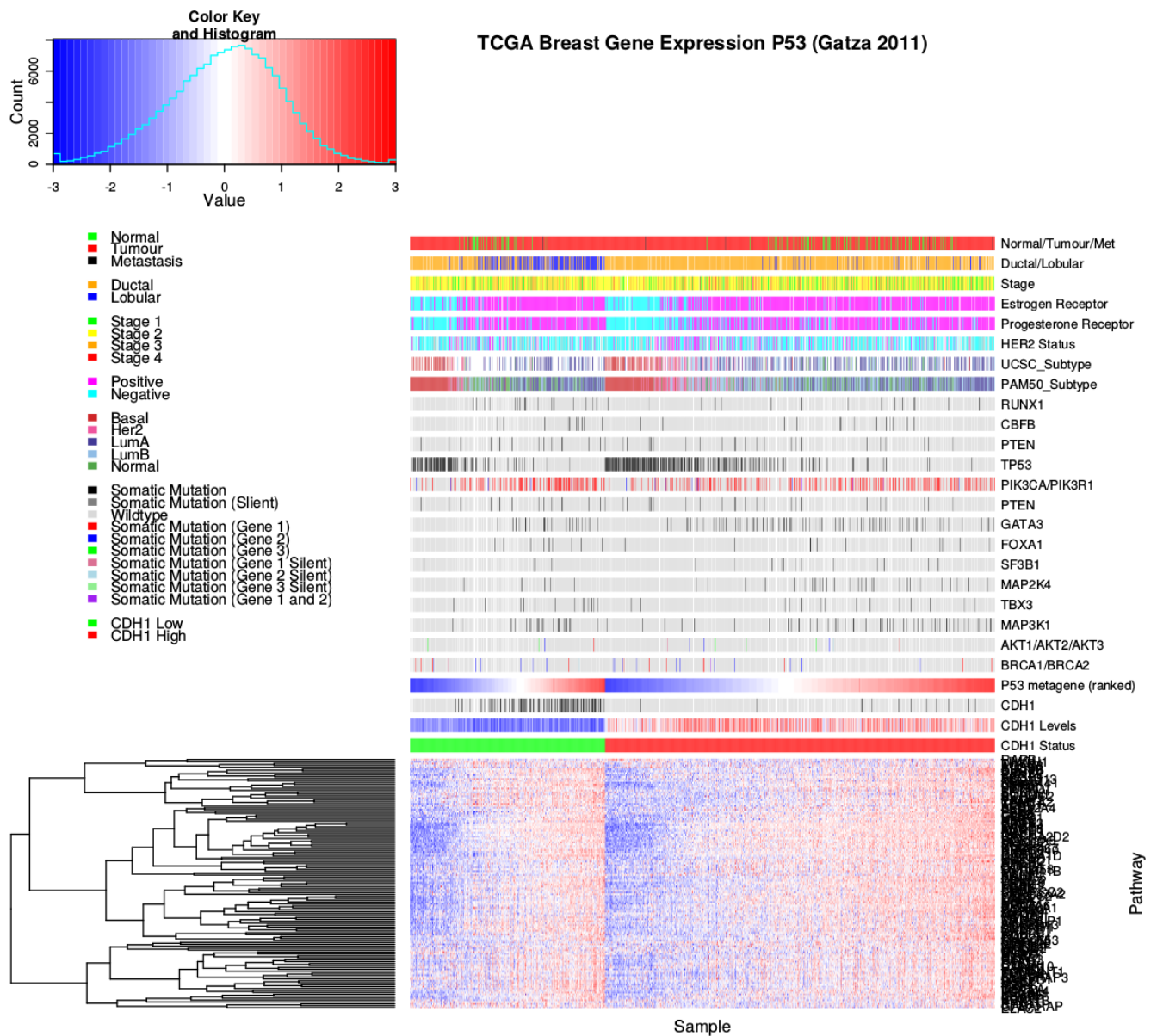


Figure C.10: **Expression profiles for p53 related genes.** Expression profiles the genes contained in the *TP53* gene signature from Gatza *et al.* (2011) in TCGA breast data, annotated for clinical factors and cancer gene mutations. Samples were separated by *CDH1* expression status and sorted by the metagene. In both cases, the majority of genes were consistent with the direction of the metagene, with few very exceptions. *TP53* mutant samples had low metagene expression, consistent with loss of tumour suppressor functions, and were less likely to have *CDH1* or *PIK3CA* mutations.

Appendix D

Intrinsic Subtyping

The intrinsic subtypes for [TCGA](#) breast cancer samples provided by [University of California, Santa Cruz \(UCSC\)](#) ([TCGA, 2012](#); [UCSC, 2012](#)) that were derived from microarray analysis have been compared to the [Prediction Analysis of Microarray 50 \(PAM50\)](#) results for performing subtyping from [RNA-Seq](#) data ([Parker *et al.*, 2009](#)). As shown in [Table D.1](#), these subtypes were highly concordant for samples which had both procedures performed upon them ($\chi^2 = 1305.9$, $p = 2.73 \times 10^{-268}$). The main exception were the luminal A samples some of which were reclassified as luminal B or “normal-like”.

Table D.1: Comparison of intrinsic subtypes

UCSC Subtype					
	Basal-like	HER2-enriched	Luminal A	Luminal B	Normal-like
	100	58	232	128	30

PAM50 Subtype					
	Basal-like	HER2-enriched	Luminal A	Luminal B	Normal-like
	208	94	314	334	227

UCSC Subtype					
PAM50 Subtype	Basal-like	HER2-enriched	Luminal A	Luminal B	Normal-like
Basal-like	96	4	2	2	1
HER2-enriched	0	47	5	3	0
Luminal A	1	0	141	1	0
Luminal B	2	7	49	121	0
Normal-like	1	0	35	1	29

The intrinsic subtypes of [TCGA](#) breast samples were compared between those provided by [UCSC](#) ([TCGA, 2012](#)) from microarray expression to those derived from [RNA-Seq](#) data ([Parker *et al.*, 2009](#)). Comparisons between these were limited to samples for which both data types were available.

The [PAM50](#) subtypes could be more accurate given similarity of these subtypes and that the remainder of the subtypes were accurately recapitulated with [RNA-Seq](#) data. Furthermore, [UCSC](#) subtypes correctly identified 22/22 normal samples as “normal-like” and [PAM50](#) subtyping in [RNA-Seq](#) data had a success rate of 112/113 (including all of those identified from microarrays). Therefore the [PAM50](#) subtypes (performed on a larger cohort of samples) are appropriate to use for further interpretation, superseding the [UCSC](#) subtypes available for a limited set of samples.

Appendix E

Stomach Expression Analysis

The following results are a replication of the [TCGA](#) results (in Chapter 4) with stomach cancer data, using synthetic lethality (SLIPT) against *CDH1*.

E.1 Synthetic Lethal Genes and Pathways

Table E.1: Synthetic lethal gene partners of *CDH1* from SLIPT in stomach cancer

Gene	Observed	Expected	χ^2 value	p-value	p-value (FDR)
<i>PRAF2</i>	17	50.4	121	3.54×10^{-25}	1.45×10^{-21}
<i>EMP3</i>	17	50.4	115	5.06×10^{-24}	1.48×10^{-20}
<i>PLEKHO1</i>	22	50.4	112	2.14×10^{-23}	4.75×10^{-20}
<i>SELM</i>	20	50.4	111	5.13×10^{-23}	8.09×10^{-20}
<i>GYPC</i>	20	50.4	110	5.77×10^{-23}	8.45×10^{-20}
<i>COX7A1</i>	18	50.4	109	1.15×10^{-22}	1.39×10^{-19}
<i>TNFSF12</i>	20	50.4	106	4.06×10^{-22}	4.38×10^{-19}
<i>SEPT4</i>	17	50.4	106	6.58×10^{-22}	5.91×10^{-19}
<i>LGALS1</i>	19	50.4	105	6.64×10^{-22}	5.91×10^{-19}
<i>RARRES2</i>	27	50.4	105	8.02×10^{-22}	6.85×10^{-19}
<i>VEGFB</i>	16	50.4	104	1.19×10^{-21}	9.74×10^{-19}
<i>PRR24</i>	22	50.4	102	2.96×10^{-21}	2.02×10^{-18}
<i>SYNC</i>	19	50.4	102	3.73×10^{-21}	2.39×10^{-18}
<i>MAGEH1</i>	17	50.4	100	9.52×10^{-21}	5.01×10^{-18}
<i>HSPB2</i>	23	50.4	99.6	1.19×10^{-20}	5.82×10^{-18}
<i>SMARCD3</i>	19	50.4	99	1.59×10^{-20}	7.57×10^{-18}
<i>CREM</i>	13	50.4	98.1	2.48×10^{-20}	1.13×10^{-17}
<i>GNG11</i>	20	50.4	97.3	3.68×10^{-20}	1.59×10^{-17}
<i>GNAI2</i>	17	50.4	96.4	5.75×10^{-20}	2.36×10^{-17}
<i>FUNDC2</i>	22	50.4	95.9	7.39×10^{-20}	2.91×10^{-17}
<i>CNRIP1</i>	21	50.4	95.3	1.0×10^{-19}	3.66×10^{-17}
<i>CALHM2</i>	22	50.4	93.1	2.94×10^{-19}	1.06×10^{-16}
<i>ARID5A</i>	18	50.4	92.7	3.47×10^{-19}	1.22×10^{-16}
<i>ST3GAL3</i>	27	50.4	92.2	4.49×10^{-19}	1.56×10^{-16}
<i>LOC339524</i>	21	50.4	92.1	4.8×10^{-19}	1.59×10^{-16}

SLIPT partners of *CDH1* with observed and expected numbers of [TCGA](#) stomach cancer samples with low expression of both genes.

Table E.2: Pathways for *CDH1* partners from SLIPT in stomach cancer

Pathways Over-represented	Pathway Size	SL Genes	p-value (FDR)
Extracellular matrix organization	241	104	7.5×10^{-140}
Hemostasis	445	138	1.8×10^{-121}
Developmental Biology	432	125	9.2×10^{-107}
Axon guidance	289	94	1.5×10^{-102}
Eukaryotic Translation Termination	84	49	1.9×10^{-99}
GPCR ligand binding	373	108	3.8×10^{-99}
Viral mRNA Translation	82	48	3.3×10^{-98}
Formation of a pool of free 40S subunits	94	51	3.3×10^{-98}
Eukaryotic Translation Elongation	87	49	1.6×10^{-97}
Peptide chain elongation	84	48	7.2×10^{-97}
Class A/1 (Rhodopsin-like receptors)	289	90	2.7×10^{-96}
Nonsense Mediated Decay independent of the Exon Junction Complex	89	49	3.0×10^{-96}
Infectious disease	349	100	2.6×10^{-94}
GTP hydrolysis and joining of the 60S ribosomal subunit	105	52	3.4×10^{-94}
L13a-mediated translational silencing of Ceruloplasmin expression	104	51	2.8×10^{-92}
3' -UTR-mediated translational regulation	104	51	2.8×10^{-92}
Neuronal System	272	84	8.4×10^{-92}
SRP-dependent cotranslational protein targeting to membrane	105	51	9.5×10^{-92}
Eukaryotic Translation Initiation	112	52	2.0×10^{-90}
Cap-dependent Translation Initiation	112	52	2.0×10^{-90}

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

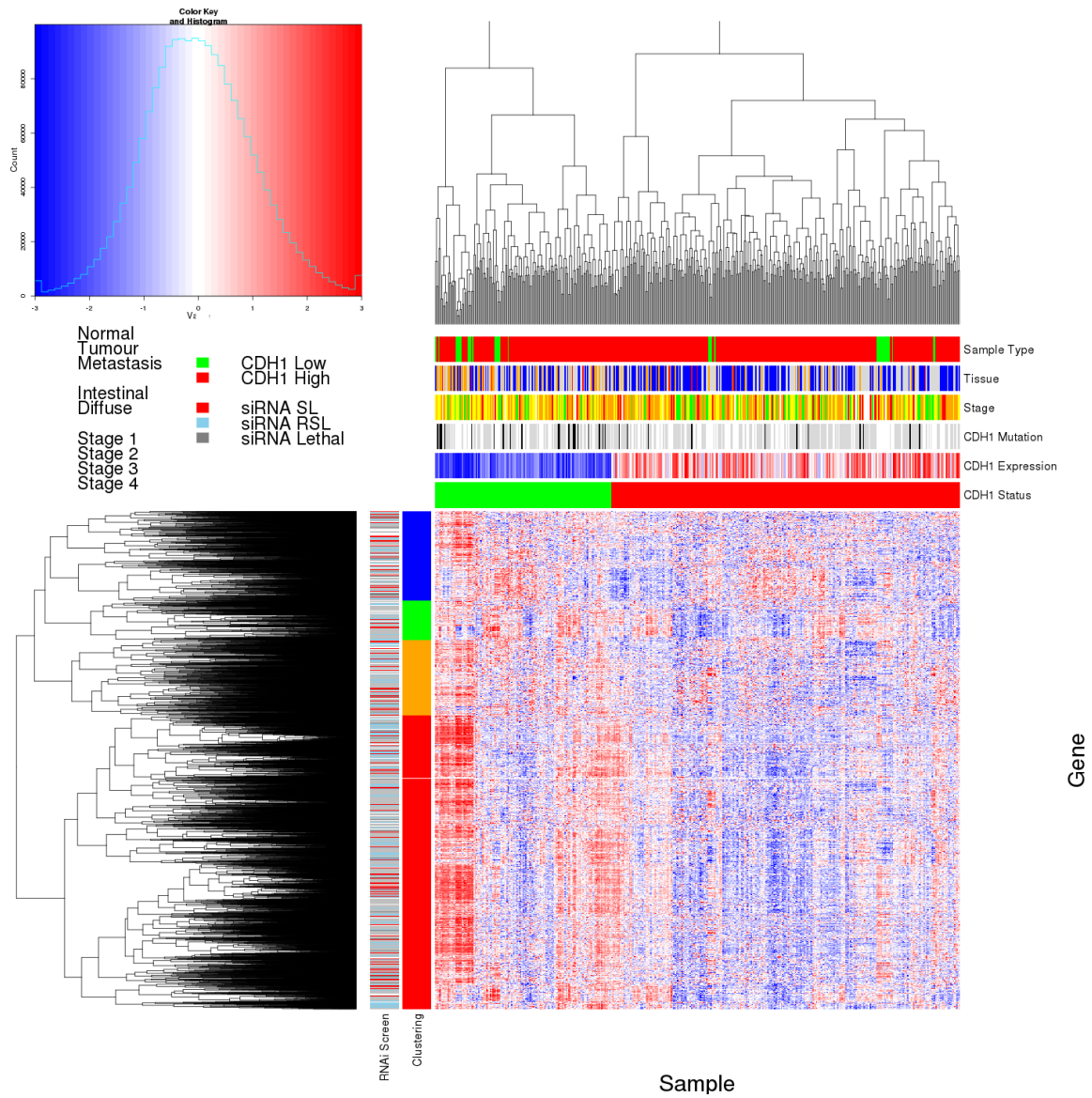


Figure E.1: **Synthetic lethal expression profiles of analysed samples.** Gene expression profile heatmap (correlation distance) of all samples (separated by the $1/3$ quantile of *CDH1* expression) analysed in TCGA stomach cancer dataset for gene expression of 4365 candidate partners of E-cadherin (*CDH1*) from SLIPT prediction (with significant FDR adjusted $p < 0.05$). Deeply clustered, inter-correlated genes form several main groups, each containing genes that were SL candidates or toxic in an siRNA screen Telford *et al.* (2015). Clusters had different sample groups highly expressing the synthetic lethal candidates in *CDH1* low samples. Notably, diffuse and *CDH1* mutant samples had elevated expression in one or more distinct clusters, although there was less complexity and variation among candidate synthetic lethal partners than in breast data. *CDH1* low samples also contained most of samples with *CDH1* mutations.

Table E.3: Pathways for clusters of *CDH1* partners in stomach SLIPT

Pathways Over-represented in Cluster 1	Pathway Size	Cluster Genes	p-value (FDR)
Viral mRNA Translation	82	48	1.3×10^{-97}
Formation of a pool of free 40S subunits	94	51	1.3×10^{-97}
Eukaryotic Translation Elongation	87	49	4.8×10^{-97}
Peptide chain elongation	84	48	1.4×10^{-96}
Eukaryotic Translation Termination	84	48	1.4×10^{-96}
GTP hydrolysis and joining of the 60S ribosomal subunit	105	52	7.9×10^{-94}
Nonsense Mediated Decay independent of the Exon Junction Complex	89	48	3.1×10^{-93}
L13a-mediated translational silencing of Ceruloplasmin expression	104	51	5.1×10^{-92}
3' -UTR-mediated translational regulation	104	51	5.1×10^{-92}
SRP-dependent cotranslational protein targeting to membrane	105	51	1.7×10^{-91}
Eukaryotic Translation Initiation	112	52	3.3×10^{-90}
Cap-dependent Translation Initiation	112	52	3.3×10^{-90}
Translation	142	56	3.6×10^{-85}
Nonsense-Mediated Decay	104	48	1.2×10^{-84}
Nonsense Mediated Decay enhanced by the Exon Junction Complex	104	48	1.2×10^{-84}
Influenza Viral RNA Transcription and Replication	109	48	4.1×10^{-82}
Influenza Life Cycle	113	48	3.4×10^{-80}
Influenza Infection	118	48	6.4×10^{-78}
Pathways Over-represented in Cluster 2	Pathway Size	Cluster Genes	p-value (FDR)
Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	65	12	1.3×10^{-15}
Phosphorylation of CD3 and TCR zeta chains	18	6	1.7×10^{-12}
Generation of second messenger molecules	29	7	2.7×10^{-12}
PD-1 signalling	21	6	7.4×10^{-12}
TCR signalling	62	9	4.3×10^{-11}
Translocation of ZAP-70 to Immunological synapse	16	5	1.1×10^{-10}
Interferon alpha/beta signalling	68	9	1.6×10^{-10}
Initial triggering of complement	17	5	1.6×10^{-10}
IKK complex recruitment mediated by RIP1	19	5	5.1×10^{-10}
TRIF-mediated programmed cell death	10	4	6.2×10^{-10}
Creation of C4 and C2 activators	11	4	1.3×10^{-9}
RHO GTPases Activate NADPH Oxidases	11	4	1.3×10^{-9}
Interferon Signalling	175	15	2.3×10^{-9}
Chemokine receptors bind chemokines	52	7	4.0×10^{-9}
Interferon gamma signalling	74	8	1.6×10^{-8}
TRAF6 mediated induction of TAK1 complex	15	4	1.6×10^{-8}
Activation of IRF3/IRF7 mediated by TBK1/IKK epsilon	16	4	2.7×10^{-8}
Downstream TCR signalling	45	6	3.5×10^{-8}
Pathways Over-represented in Cluster 3	Pathway Size	Cluster Genes	p-value (FDR)
Uptake and actions of bacterial toxins	22	4	3.5×10^{-6}
Neurotoxicity of clostridium toxins	10	3	3.5×10^{-6}
Activation of PPARGC1A (PGC-1alpha) by phosphorylation	10	3	3.5×10^{-6}
SMAD2/SMAD3:SMAD4 heterotrimer regulates transcription	28	4	1.4×10^{-5}
Assembly of the primary cilium	149	10	2.5×10^{-5}
Serotonin Neurotransmitter Release Cycle	15	3	2.5×10^{-5}
Glycosaminoglycan metabolism	114	8	3.3×10^{-5}
Platelet homeostasis	54	5	3.3×10^{-5}
Norepinephrine Neurotransmitter Release Cycle	17	3	3.3×10^{-5}
Acetylcholine Neurotransmitter Release Cycle	17	3	3.3×10^{-5}
G _{12s} signalling events	100	7	5.5×10^{-5}
GABA synthesis, release, reuptake and degradation	19	3	5.6×10^{-5}
deactivation of the beta-catenin transactivating complex	39	4	6.7×10^{-5}
Dopamine Neurotransmitter Release Cycle	20	3	6.7×10^{-5}
IRS-related events triggered by IGF1R	83	6	7.1×10^{-5}
Generic Transcription Pathway	186	11	7.1×10^{-5}
Termination of O-glycan biosynthesis	21	3	7.4×10^{-5}
Kinesins	22	3	8.5×10^{-5}
Pathways Over-represented in Cluster 4	Pathway Size	Cluster Genes	p-value (FDR)
Extracellular matrix organization	241	97	8.8×10^{-126}
Axon guidance	289	75	8.3×10^{-72}
Hemostasis	445	101	8.3×10^{-72}
Developmental Biology	432	95	3.0×10^{-67}
Response to elevated platelet cytosolic Ca ²⁺	84	37	5.8×10^{-67}
Platelet degranulation	79	36	5.8×10^{-67}
Degradation of the extracellular matrix	104	39	6.7×10^{-63}
Platelet activation, signalling and aggregation	186	52	6.6×10^{-62}
ECM proteoglycans	66	31	8.1×10^{-61}
Neuronal System	272	64	5.1×10^{-60}
Signalling by PDGF	173	47	9.7×10^{-57}
Integrin cell surface interactions	82	31	1.9×10^{-53}
Collagen biosynthesis and modifying enzymes	56	26	1.1×10^{-52}
Collagen formation	67	28	1.4×10^{-52}
Class A/1 (Rhodopsin-like receptors)	289	61	2.3×10^{-52}
GPCR ligand binding	373	73	2.8×10^{-52}
Elastic fibre formation	38	22	4.7×10^{-52}
Non-integrin membrane-ECM interactions	53	24	7.0×10^{-49}

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

E.2 Comparison to Primary Screen

The synthetic lethal partners with *CDH1* expression in stomach cancers were also compared to siRNA primary screen data (Telford *et al.*, 2015), as performed in Section 4.2.1. These were expected to be more concordant with the experimental results performed on a null mutant, however this was not the case at the gene level: less genes overlapped with experimental candidates in Figure E.2. This may be due to lower sample size for mutations in TCGA data or lower frequency (expected value) of *CDH1* mutations compared to low expression.

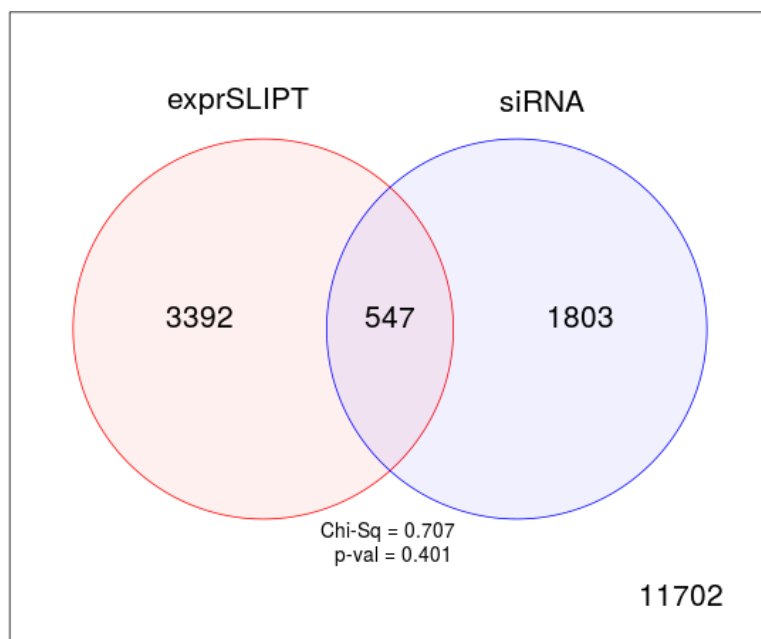


Figure E.2: **Comparison of SLIPT in stomach to siRNA.** The overlap of gene candidates for *E-cadherin* synthetic lethal partners between computational (SLIPT) and experimental screening (siRNA) approaches. The χ^2 test suggests that the overlap is no more than would be expected by chance ($p = 0.281$).

Table E.4: Pathways for *CDH1* partners from SLIPT and siRNA

Predicted only by SLIPT (3392 genes)	Pathway Size	Genes Identified	p-value (FDR)
Extracellular matrix organization	238	90	3.4×10^{-107}
Eukaryotic Translation Termination	79	46	7.6×10^{-91}
Viral mRNA Translation	77	45	1.2×10^{-89}
Eukaryotic Translation Elongation	82	46	5.8×10^{-89}
Peptide chain elongation	79	45	2.1×10^{-88}
Nonsense Mediated Decay independent of the Exon Junction Complex	84	46	9.4×10^{-88}
Formation of a pool of free 40S subunits	89	47	3.3×10^{-87}
GTP hydrolysis and joining of the 60S ribosomal subunit	100	48	3.2×10^{-83}
Axon guidance	284	84	3.9×10^{-82}
Developmental Biology	426	111	4.2×10^{-82}
L13a-mediated translational silencing of Ceruloplasmin expression	99	47	1.4×10^{-81}
3' -UTR-mediated translational regulation	99	47	1.4×10^{-81}
SRP-dependent cotranslational protein targeting to membrane	99	47	1.4×10^{-81}
Nonsense-Mediated Decay	99	47	1.4×10^{-81}
Nonsense Mediated Decay enhanced by the Exon Junction Complex	99	47	1.4×10^{-81}
Hemostasis	438	112	1.2×10^{-80}
Eukaryotic Translation Initiation	107	48	8.0×10^{-80}
Cap-dependent Translation Initiation	107	48	8.0×10^{-80}
Infectious disease	338	90	1.6×10^{-76}
Neuronal System	267	77	1.6×10^{-76}

Detected only by siRNA screen (1803 genes)	Pathway Size	Genes Identified	p-value (FDR)
Class A/1 (Rhodopsin-like receptors)	282	62	8.1×10^{-50}
GPCR ligand binding	363	71	4.9×10^{-46}
Peptide ligand-binding receptors	175	38	7.9×10^{-38}
G _{αi} signalling events	184	37	1.1×10^{-34}
Gastrin-CREB signalling pathway via PKC and MAPK	180	35	1.4×10^{-32}
G _{αq} signalling events	159	32	4.8×10^{-32}
DAP12 interactions	159	29	1.4×10^{-27}
Downstream signal transduction	146	26	2.4×10^{-25}
DAP12 signalling	149	26	6.4×10^{-25}
VEGFA-VEGFR2 Pathway	91	19	8.1×10^{-24}
Signalling by PDGF	172	27	5.7×10^{-23}
Signalling by ERBB2	146	24	1.4×10^{-22}
Signalling by VEGF	99	19	2.0×10^{-22}
Visual phototransduction	85	17	1.3×10^{-21}
Downstream signalling of activated FGFR1	134	22	1.3×10^{-21}
Downstream signalling of activated FGFR2	134	22	1.3×10^{-21}
Downstream signalling of activated FGFR3	134	22	1.3×10^{-21}
Downstream signalling of activated FGFR4	134	22	1.3×10^{-21}
Signalling by FGFR	146	23	2.0×10^{-21}
Signalling by FGFR1	146	23	2.0×10^{-21}

Intersection of SLIPT and siRNA screen (547 genes)	Pathway Size	Genes Identified	p-value (FDR)
Class A/1 (Rhodopsin-like receptors)	282	25	3.9×10^{-9}
Platelet activation, signalling and aggregation	182	17	3.9×10^{-9}
Response to elevated platelet cytosolic Ca ²⁺	82	9	5.5×10^{-8}
Platelet homeostasis	53	7	5.7×10^{-8}
Nucleotide-like (purinergic) receptors	16	4	1.8×10^{-7}
Platelet degranulation	77	8	2.8×10^{-7}
Peptide ligand-binding receptors	175	14	3.8×10^{-7}
Molecules associated with elastic fibres	34	5	7.1×10^{-7}
Amine ligand-binding receptors	35	5	8.6×10^{-7}
G _{αi} signalling events	184	14	9.8×10^{-7}
GPCR ligand binding	363	27	1.1×10^{-6}
Elastic fibre formation	38	5	1.5×10^{-6}
G _{αq} signalling events	159	12	1.9×10^{-6}
Serotonin receptors	12	3	3.8×10^{-6}
P2Y receptors	12	3	3.8×10^{-6}
Signal amplification	16	3	2.3×10^{-5}
Gastrin-CREB signalling pathway via PKC and MAPK	180	12	2.3×10^{-5}
Complement cascade	33	4	2.4×10^{-5}
Glycosaminoglycan metabolism	110	8	2.5×10^{-5}
Glycogen breakdown (glycogenolysis)	17	3	2.7×10^{-5}

E.2.1 Resampling Analysis

Table E.5: Pathways for *CDH1* partners from SLIPT in stomach cancer

Reactome Pathway	Over-representation	Permutation
<i>Extracellular matrix organization</i>	7.5×10^{-140}	0.070215
Hemostasis	1.8×10^{-121}	0.25804
Developmental Biology	9.2×10^{-107}	0.53032
Axon guidance	1.5×10^{-102}	0.6704
Eukaryotic Translation Termination	1.9×10^{-99}	$> 1.031 \times 10^{-5}$
GPCR ligand binding	3.8×10^{-99}	0.54914
Viral mRNA Translation	3.3×10^{-98}	$> 1.031 \times 10^{-5}$
Formation of a pool of free 40S subunits	3.3×10^{-98}	$> 1.031 \times 10^{-5}$
Eukaryotic Translation Elongation	1.6×10^{-97}	$> 1.031 \times 10^{-5}$
Peptide chain elongation	7.2×10^{-97}	$> 1.031 \times 10^{-5}$
Class A/1 (Rhodopsin-like receptors)	2.7×10^{-96}	0.58174
Nonsense Mediated Decay independent of the Exon Junction Complex	3×10^{-96}	$> 1.031 \times 10^{-5}$
Infectious disease	2.6×10^{-94}	0.25484
GTP hydrolysis and joining of the 60S ribosomal subunit	3.4×10^{-94}	$> 1.031 \times 10^{-5}$
L13a-mediated translational silencing of Ceruloplasmin expression	2.8×10^{-92}	$> 1.031 \times 10^{-5}$
3' -UTR-mediated translational regulation	2.8×10^{-92}	$> 1.031 \times 10^{-5}$
Neuronal System	8.4×10^{-92}	0.53433
SRP-dependent cotranslational protein targeting to membrane	9.5×10^{-92}	$> 1.031 \times 10^{-5}$
Eukaryotic Translation Initiation	2.0×10^{-90}	$> 1.031 \times 10^{-5}$
Cap-dependent Translation Initiation	2.0×10^{-90}	$> 1.031 \times 10^{-5}$
Nonsense-Mediated Decay	7.4×10^{-90}	$> 1.031 \times 10^{-5}$
Nonsense Mediated Decay enhanced by the Exon Junction Complex	7.4×10^{-90}	$> 1.031 \times 10^{-5}$
Adaptive Immune System	8.1×10^{-88}	0.14116
Translation	1.3×10^{-87}	$> 1.031 \times 10^{-5}$
Platelet activation, signalling and aggregation	1.3×10^{-86}	0.28959
Influenza Infection	1×10^{-82}	$> 1.031 \times 10^{-5}$
Influenza Viral RNA Transcription and Replication	2.4×10^{-82}	$> 1.031 \times 10^{-5}$
Influenza Life Cycle	2×10^{-80}	$> 1.031 \times 10^{-5}$
Response to elevated platelet cytosolic Ca^{2+}	4.9×10^{-78}	0.50817
Signalling by NGF	1.6×10^{-75}	0.38518
Rho GTPase cycle	5.1×10^{-75}	0.14864
Signalling by PDGF	7.4×10^{-74}	0.40493
<i>Signalling by Rho GTPases</i>	5.1×10^{-73}	0.077217
Glycosaminoglycan metabolism	1.4×10^{-68}	0.52984
$\text{G}_{\alpha i}$ signalling events	1.8×10^{-66}	0.9254
Metabolism of carbohydrates	1.1×10^{-65}	0.39501
$\text{G}_{\alpha s}$ signalling events	2.7×10^{-65}	0.0050293
Potassium Channels	2.7×10^{-65}	0.53359
Transmission across Chemical Synapses	1.8×10^{-64}	0.81833
ECM proteoglycans	3.4×10^{-64}	0.083482
Peptide ligand-binding receptors	4.8×10^{-64}	0.62817
Degradation of the extracellular matrix	1.1×10^{-63}	0.80879
Platelet homeostasis	5.3×10^{-63}	0.53134
NGF signalling via TRKA from the plasma membrane	6.1×10^{-63}	0.5717
Integration of energy metabolism	4.5×10^{-61}	0.10889
Collagen formation	5.4×10^{-61}	0.29896
Integrin cell surface interactions	7×10^{-59}	0.18167
Collagen biosynthesis and modifying enzymes	7×10^{-59}	0.30208
Neurotransmitter Receptor Binding And Downstream Transmission	8.7×10^{-57}	0.82522
In The Postsynaptic Cell	8.7×10^{-57}	0.25468
Signalling by Wnt	8.7×10^{-57}	0.25468

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

Table E.6: Pathways for *CDH1* partners from SLIPT in stomach and siRNA

Reactome Pathway	Over-representation	Permutation
Platelet activation, signalling and aggregation	3.9×10^{-9}	0.49557
Class A/1 (Rhodopsin-like receptors)	3.9×10^{-9}	0.98432
Response to elevated platelet cytosolic Ca^{2+}	5.5×10^{-8}	0.54349
Platelet homeostasis	5.7×10^{-8}	0.45017
Nucleotide-like (purinergic) receptors	1.8×10^{-7}	0.36966
Peptide ligand-binding receptors	3.8×10^{-7}	0.91294
Molecules associated with elastic fibres	7.1×10^{-7}	0.0025868
Amine ligand-binding receptors	8.6×10^{-7}	0.43303
$G_{\alpha i}$ signalling events	9.8×10^{-7}	0.99626
GPCR ligand binding	1.1×10^{-6}	0.97733
Elastic fibre formation	1.5×10^{-6}	0.0025868
$G_{\alpha q}$ signalling events	1.9×10^{-6}	0.86089
P2Y receptors	3.8×10^{-6}	0.18795
Serotonin receptors	3.8×10^{-6}	0.37853
Signal amplification	2.3×10^{-5}	0.47856
Gastrin-CREB signalling pathway via PKC and MAPK	2.3×10^{-5}	0.98567
Complement cascade	2.4×10^{-5}	$> 3.4628 \times 10^{-6}$
Glycosaminoglycan metabolism	2.5×10^{-5}	0.38953
Glycogen breakdown (glycogenolysis)	2.7×10^{-5}	0.83772
Defective B4GALT7 causes EDS, progeroid type	4.9×10^{-5}	0.10792
Defective B3GAT3 causes JDSSDHD	4.9×10^{-5}	0.10792
Role of LAT2/NTAL/LAB on calcium mobilization	5.6×10^{-5}	0.35373
Cell surface interactions at the vascular wall	5.6×10^{-5}	0.47642
$G_{\alpha s}$ signalling events	6×10^{-5}	0.019858
Signalling by NOTCH	6×10^{-5}	0.19008
A tetrasaccharide linker sequence is required for GAG synthesis	0.00017	0.47642
Extracellular matrix organization	0.00018	0.0047308
Collagen formation	0.00018	0.19245
Effects of PIP2 hydrolysis	0.0002	0.37779
Syndecan interactions	0.0002	0.37779
Diseases associated with glycosaminoglycan metabolism	0.00023	0.01028
Diseases of glycosylation	0.00023	0.01028
<i>Chondroitin sulfate/dermatan sulfate metabolism</i>	0.00023	0.085541
Integrin $\alpha\text{IIb}\beta 3$ signalling	0.00028	0.76936
Keratan sulfate biosynthesis	0.00034	0.68744
Rho GTPase cycle	0.00034	0.15675
Creation of C4 and C2 activators	0.00035	0.12275
Abacavir transport and metabolism	0.00035	0.12443
Amine compound SLC transporters	0.00037	0.69773
FCERI mediated NF- κ B activation	0.00037	0.69846
Fc epsilon receptor (FCERI) signalling	0.00056	0.43303
Defective EXT2 causes exostoses 2	0.00067	0.16053
Defective EXT1 causes exostoses 1, TRPS2 and CHDS	0.00067	0.16053
<i>Collagen biosynthesis and modifying enzymes</i>	0.00071	0.052911
Keratan sulfate/keratin metabolism	0.00073	0.46533
G α (12/13) signalling events	0.00078	0.59164
SEMA3A-Plexin repulsion signalling by inhibiting Integrin adhesion	0.00084	0.038504
Signal attenuation	0.00084	0.37779
Eicosanoid ligand-binding receptors	0.0011	0.11117
SOS-mediated signalling	0.0011	0.25387

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

E.3 Metagene Analysis

Metagenes used to detect synthetic lethal pathways with *CDH1* in stomach cancer.

Table E.7: Synthetic lethal metagenes against *CDH1* in stomach cancer

Pathway	ID	Observed	Expected	χ^2 value	p-value	p-value (FDR)
Cell-Cell communication	1500931	18	50.4	110	7.43×10^{-23}	1.53×10^{-20}
VEGFR2 mediated vascular permeability	5218920	19	50.4	109	1.36×10^{-22}	2.49×10^{-20}
Sema4D in semaphorin signalling	400685	20	50.4	104	1.62×10^{-21}	2.12×10^{-19}
Ion transport by P-type ATPases	936837	17	50.4	100	8.29×10^{-21}	8.06×10^{-19}
Sialic acid metabolism	4085001	19	50.4	95.3	9.95×10^{-20}	7.82×10^{-18}
Synthesis of pyrophosphates in the cytosol	1855167	26	50.4	94	1.86×10^{-19}	1.23×10^{-17}
Keratan sulfate/keratin metabolism	1638074	25	50.4	93.5	2.36×10^{-19}	1.44×10^{-17}
Ion channel transport	983712	19	50.4	92.8	3.37×10^{-19}	1.99×10^{-17}
Keratan sulfate biosynthesis	2022854	26	50.4	91.4	6.79×10^{-19}	3.62×10^{-17}
Arachidonic acid metabolism	2142753	22	50.4	90.6	9.81×10^{-19}	5.07×10^{-17}
RHO GTPases activate CIT	5625900	22	50.4	87	5.80×10^{-18}	2.66×10^{-16}
Stimuli-sensing channels	2672351	25	50.4	85.8	1.03×10^{-17}	4.58×10^{-16}
Synthesis of PI	1483226	19	50.4	85.6	1.15×10^{-17}	4.89×10^{-16}
G-protein activation	202040	19	50.4	85.3	1.34×10^{-17}	5.53×10^{-16}
NrCAM interactions	447038	22	50.4	84.3	2.1×10^{-17}	8.27×10^{-16}
Inwardly rectifying K^+ channels	1296065	24	50.4	83.5	3.19×10^{-17}	1.22×10^{-15}
Calcitonin-like ligand receptors	419812	20	50.4	82.2	6.07×10^{-17}	2.13×10^{-15}
Prostacyclin signalling through prostacyclin receptor	392851	24	50.4	81.8	7.27×10^{-17}	2.5×10^{-15}
Presynaptic function of Kainate receptors	500657	26	50.4	79.7	2.00×10^{-16}	6.34×10^{-15}
ADP signalling through P2Y purinoceptor 12	392170	23	50.4	79.2	2.57×10^{-16}	7.71×10^{-15}
regulation of FZD by ubiquitination	4641263	22	50.4	78.8	3.15×10^{-16}	9.3×10^{-15}
Toxicity of tetanus toxin (TeNT)	5250982	27	50.4	78.7	3.36×10^{-16}	9.75×10^{-15}
Gap junction degradation	190873	21	50.4	78.5	3.66×10^{-16}	1.04×10^{-14}
Nephrin interactions	373753	25	50.4	78.2	4.21×10^{-16}	1.14×10^{-14}
GABA synthesis, release, reuptake and degradation	888590	26	50.4	77	7.69×10^{-16}	1.95×10^{-14}

Strongest candidate SL partners for *CDH1* by SLIPT with observed and expected numbers of TCGA stomach cancer samples with low expression of both genes.