

Contents

Glossary	xiii
Acronyms	xvi
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	Modeling 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	65
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	72
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	77
3.3.2.2	Specificity with Query-Correlated Pathways	84
3.3.2.3	Importance of Directional Testing	84
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	88
3.4.2	Simulating Gene Expression from Graph Structures	89
3.5	Customised Functions and Packages Developed	93
3.5.1	Synthetic Lethal Interaction Prediction Tool	93
3.5.2	Data Visualisation	94
3.5.3	Extensions to the iGraph Package	97
3.5.3.1	Sampling Simulated Data from Graph Structures	97
3.5.3.2	Plotting Directed Graph Structures	97
3.5.3.3	Computing Information Centrality	98
3.5.3.4	Testing Pathway Structure with Permutation Testing	98
3.5.3.5	Metapackage to Install iGraph Functions	99
4	Synthetic Lethal Analysis of Gene Expression Data	100
4.1	Synthetic Lethal Genes in Breast Cancer	101
4.1.1	Synthetic Lethal Pathways in Breast Cancer	103
4.1.2	Expression Profiles of Synthetic Lethal Partners	104
4.1.2.1	Subgroup Pathway Analysis	107
4.2	Comparing Synthetic Lethal Gene Candidates	110
4.2.1	Primary siRNA Screen Candidates	110
4.2.2	Comparison with Correlation	110
4.2.3	Comparison with Primary Screen Viability	112
4.2.4	Comparison with Secondary siRNA Screen Validation	114
4.2.5	Comparison to Primary Screen at Pathway Level	116
4.2.5.1	Resampling Genes for Pathway Enrichment	118
4.2.6	Integrating Synthetic Lethal Pathways and Screens	121
4.3	Metagene Analysis	123
4.3.1	Pathway Expression	124
4.3.2	Somatic Mutation	126
4.3.3	Synthetic Lethal Pathway Metagenes	130
4.3.4	Synthetic Lethality in Breast Cancer	131
4.4	Replication in Stomach Cancer	132
4.5	Discussion	133

4.5.1	Strengths of the SLIPT Methodology	133
4.5.2	Synthetic Lethal Pathways for E-cadherin	134
4.5.3	Replication and Validation	136
4.5.3.1	Integration with short interfering RNA (siRNA) Screen- ing	136
4.5.3.2	Replication across Tissues	137
4.6	Summary	137
5	Synthetic Lethal Pathway Structure	139
5.1	Synthetic Lethal Genes in Reactome Pathways	139
5.1.1	The PI3K/AKT Pathway	140
5.1.2	The Extracellular Matrix	142
5.1.3	G Protein Coupled Receptors	145
5.1.4	Gene Regulation and Translation	145
5.2	Network Analysis of Synthetic Lethal Genes	146
5.2.1	Gene Connectivity and Vertex Degree	147
5.2.2	Gene Importance and Centrality	148
5.2.2.1	Information Centrality	148
5.2.2.2	PageRank Centrality	150
5.3	Relationships between Synthetic Lethal Genes	152
5.3.1	Hierarchical Pathway Structure	152
5.3.1.1	Contextual Hierarchy of PI3K	152
5.3.1.2	Testing Contextual Hierarchy of Synthetic Lethal Genes	152
5.3.2	Upstream or Downstream Synthetic Lethality	156
5.3.2.1	Measuring Structure of Candidates within PI3K	156
5.3.2.2	Resampling for Synthetic Lethal Pathway Structure . .	158
5.4	Discussion	160
5.5	Summary	162
6	Simulation and Modeling of Synthetic Lethal Pathways	164
6.1	Synthetic Lethal Detection Methods	165
6.1.1	Performance of SLIPT and χ^2 across Quantiles	165
6.1.1.1	Correlated Query Genes affects Specificity	169
6.1.2	Alternative Synthetic Lethal Detection Strategies	171
6.1.2.1	Correlation for Synthetic Lethal Detection	171
6.1.2.2	Testing for Bimodality with BiSEp	173
6.2	Simulations with Graph Structures	174
6.2.1	Performance over Graph Structures	175
6.2.1.1	Simple Graph Structures	175
6.2.1.2	Constructed Graph Structures	177
6.2.2	Performance with Inhibitions	180
6.2.3	Synthetic Lethality across Graph Structures	185
6.2.4	Performance within a Simulated Human Genome	189
6.3	Simulations in More Complex Graph Structures	193
6.3.1	Simulations over Pathway-based Graphs	194
6.3.2	Pathway Structures in a Simulated Human Genome	197

6.4	Discussion	200
6.4.1	Simulation Procedure	200
6.4.2	Comparing Methods with Simulated Data	201
6.4.3	Design and Performance of SLIPT	202
6.4.4	Simulations from Graph Structures	204
6.5	Summary	205
7	Discussion	207
7.1	Synthetic Lethality and <i>CDH1</i> Biology	207
7.1.1	Established Functions of <i>CDH1</i>	208
7.1.2	The Molecular Role of <i>CDH1</i> in Cancer	208
7.2	Significance	209
7.2.1	Synthetic Lethality in the Genomic Era	209
7.2.2	Clinical Interventions based on Synthetic Lethality	211
7.3	Future Directions	212
7.4	Conclusions	214
	References	216
A	Sample Quality	240
A.1	Sample Correlation	240
A.2	Replicate Samples in The Cancer Genome Atlas (TCGA) Breast	243
B	Software Used for Thesis	247
C	Mutation Analysis in Breast Cancer	256
C.1	Synthetic Lethal Genes and Pathways	256
C.2	Synthetic Lethal Expression Profiles	257
C.3	Comparison to Primary Screen	260
C.3.1	Resampling Analysis	262
C.4	Compare Synthetic Lethal Interaction Prediction Tool (SLIPT) genes	264
C.5	Metagene Analysis	266
C.6	Expression of Somatic Mutations	267
C.7	Metagene Expression Profiles	270
D	Intrinsic Subtyping	273
E	Expression Analysis in Stomach Cancer	275
E.1	Synthetic Lethal Genes and Pathways	275
E.2	Comparison to Primary Screen	280
E.2.1	Resampling Analysis	282
E.3	Metagene Analysis	284
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	Graph Structures	315
	K.1 Simulations from Simple Graph Structures	315
	K.1.1 Simulations from Inhibiting Graph Structures	317
	K.2 Simulation across Graph Structures	320
	K.3 Simulations from Complex Graph Structures	324
	K.3.1 Simulations from Complex Inhibiting Graphs	327
	K.4 Simulations from Pathway Graph Structures	333

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	Modeling 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	78
3.13	Simulating expression with correlated gene blocks	79
3.14	Synthetic lethal prediction across simulations	80
3.15	Performance with correlations	81
3.16	Comparison of statistical performance with correlation structure	82
3.17	Performance with query correlations	83
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	91
3.22	Simulating expression from graph structure with inhibitions	92
3.23	Demonstration of violin plots with custom features	95
3.24	Demonstration of annotated heatmap	95
3.25	Simulating graph structures	98
4.1	Synthetic lethal expression profiles of analysed samples	106
4.2	Comparison of SLIPT to siRNA	110
4.3	Compare SLIPT and siRNA genes with correlation	111
4.4	Compare SLIPT and siRNA genes with correlation	112
4.5	Compare SLIPT and siRNA genes with viability	113

4.6	Compare SLIPT genes with siRNA viability	114
4.7	Resampled intersection of SLIPT and siRNA candidates	118
4.8	Pathway metagene expression profiles	125
4.9	Expression profiles for constituent genes of PI3K	127
4.10	Expression profiles for estrogen receptor related genes	128
4.11	Somatic mutation against the PI3K metagene	129
5.1	synthetic lethality in the PI3K cascade	141
5.2	synthetic lethality in Elastic Fibre Formation	143
5.3	Synthetic lethality in Fibrin Clot Formation	144
5.4	Synthetic lethality and vertex degree	147
5.5	Synthetic lethality and centrality	150
5.6	Synthetic lethality and PageRank	151
5.7	Hierarchical structure of PI3K	153
5.8	Hierarchy score in PI3K against synthetic lethality in PI3K	154
5.9	Structure of synthetic lethality in PI3K	156
5.10	Structure of synthetic lethality resampling in PI3K	157
6.1	Performance of χ^2 and SLIPT across quantiles	167
6.2	Performance of χ^2 and SLIPT across quantiles with more genes	168
6.3	Performance of χ^2 and SLIPT across quantiles with query correlation	169
6.4	Performance of χ^2 and SLIPT across quantiles with query correlation and more genes	170
6.5	Performance of negative correlation and SLIPT	172
6.6	Simple graph structures	175
6.7	Performance of simulations on a simple graph	176
6.8	Performance of simulations is similar in simple graphs	178
6.9	Performance of simulations on a pathway	179
6.10	Performance of simulations on a simple graph with inhibition	181
6.11	Performance is higher on a simple inhibiting graph	182
6.12	Performance of simulations on a constructed graph with inhibition	183
6.13	Performance is affected by inhibition in graphs	185
6.14	Detection of synthetic lethality within a graph structure	187
6.15	Performance of simulations including a simple graph	190
6.16	Performance on a simple graph improves with more genes	191
6.17	Performance on an inhibiting graph improves with more genes	193
6.18	Performance of simulations on the PI3K cascade	196
6.19	Performance of simulations including the PI3K cascade	198
6.20	Performance on pathways improves with more genes	199
A.1	Correlation profiles of removed samples	241
A.2	Correlation analysis and sample removal	242
A.3	Replicate excluded samples	243
A.4	Replicate samples with all remaining	244
A.5	Replicate samples with some excluded	245
C.1	Synthetic lethal expression profiles of analysed samples	258

C.2	Comparison of mtSLIPT to siRNA	260
C.3	Compare mtSLIPT and siRNA genes with correlation	264
C.4	Compare mtSLIPT and siRNA genes with correlation	264
C.5	Compare mtSLIPT and siRNA genes with siRNA viability	265
C.6	Somatic mutation against PIK3CA metagene	267
C.7	Somatic mutation against PI3K protein	268
C.8	Somatic mutation against AKT protein	269
C.9	Pathway metagene expression profiles	270
C.10	Expression profiles for p53 related genes	271
C.11	Expression profiles for BRCA related genes	272
E.1	Synthetic lethal expression profiles of stomach samples	278
E.2	Comparison of SLIPT in stomach to siRNA	280
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	316
K.2	Performance of simulations on an inhibiting graph	317
K.3	Performance of simulations on a constructed graph with inhibition	318
K.4	Performance of simulations on a constructed graph with inhibition	319
K.5	Detection of synthetic lethality within a graph structure	320
K.6	Detection of synthetic lethality within an inhibiting graph	322

K.7	Detection of synthetic lethality within an inhibiting graph	323
K.8	Performance of simulations on a branching graph	324
K.9	Performance of simulations on a complex graph	325
K.10	Performance of simulations on a large graph	326
K.11	Performance of simulations on a branching graph with inhibition	327
K.12	Performance of simulations on a branching graph with inhibition	328
K.13	Performance of simulations on a complex graph with inhibition	329
K.14	Performance of simulations on a complex graph with inhibition	330
K.15	Performance of simulations on a large constructed graph with inhibition	331
K.16	Performance of simulations on a large constructed graph with inhibition	332
K.17	Performance of simulations on the $G_{\alpha i}$ signalling pathway	333
K.18	Performance of simulations including the $G_{\alpha i}$ signalling pathway	334

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	102
4.2	Pathways for <i>CDH1</i> partners from SLIPT	104
4.3	Pathways for clusters of <i>CDH1</i> partners from SLIPT	108
4.4	ANOVA for synthetic lethality and correlation with <i>CDH1</i>	112
4.5	Comparing SLIPT genes against secondary siRNA screen	115
4.6	Pathways for <i>CDH1</i> partners from SLIPT and siRNA	117
4.7	Pathways for <i>CDH1</i> partners from SLIPT	120
4.8	Pathways for <i>CDH1</i> partners from SLIPT and siRNA primary screen .	122
4.9	Candidate synthetic lethal metagenes against <i>CDH1</i> from SLIPT . . .	131
5.1	ANOVA for synthetic lethality and vertex degree	148
5.2	ANOVA for synthetic lethality and information centrality	150
5.3	ANOVA for synthetic lethality and PageRank centrality	152
5.4	ANOVA for synthetic lethality and PI3K hierarchy	155
5.5	Resampling for pathway structure of synthetic lethal detection methods	159
B.1	R packages used during thesis	247
C.1	Candidate synthetic lethal gene partners of <i>CDH1</i> from mtSLIPT . . .	256
C.2	Pathways for <i>CDH1</i> partners from mtSLIPT	257
C.3	Pathways for clusters of <i>CDH1</i> partners from mtSLIPT	259
C.4	Pathways for <i>CDH1</i> partners from mtSLIPT and siRNA	261
C.5	Pathways for <i>CDH1</i> partners from mtSLIPT	262
C.6	Pathways for <i>CDH1</i> partners from mtSLIPT and siRNA primary screen	263
C.7	Candidate synthetic lethal metagenes against <i>CDH1</i> from mtSLIPT . .	266
D.1	Comparison of intrinsic subtypes	273

E.1	Synthetic lethal gene partners of <i>CDH1</i> from SLIPT in stomach cancer	276
E.2	Pathways for <i>CDH1</i> partners from SLIPT in stomach cancer	277
E.3	Pathways for clusters of <i>CDH1</i> partners in stomach SLIPT	279
E.4	Pathways for <i>CDH1</i> partners from SLIPT and siRNA	281
E.5	Pathways for <i>CDH1</i> partners from SLIPT in stomach cancer	282
E.6	Pathways for <i>CDH1</i> partners from SLIPT in stomach and siRNA . . .	283
E.7	Synthetic lethal metagenes against <i>CDH1</i> in stomach cancer	284
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

bioinformatics	Statistical or computational approaches to biological data or research tools.
driver mutation	A mutation which promotes cancer growth.
E-cadherin	Epithelial cadherin (calcium-dependent adhesion), a cell-adhesion protein encoded by <i>CDH1</i> .
essential	A gene which is required to be functional or expressed for a cell or organism to be viable, grow or develop.
functional redundancy	Genes which perform a common function, also known as genetic redundancy.
gene expression	A measure of the relative expression of each gene from the mRNA extracted from (pooled) cells.
genome	An analysis of all of the DNA sequence in the genome.
genomic	An approach or technology designed to generate or use data from all genes in the genome.
germline mutation	A mutation that occurred in germline cells and is passed between generation.
graph or network	A mathematical structure modelling or depicting the relationships between elements.
hereditary	A trait or disease which has a genetic cause and is inherited from family members.
<i>in vivo</i>	An investigation conducted using in the context of a biological cell or organism, including pre-clinical models and clinical trials.

intrinsic subtype	Distinguishing cancer by molecular and genetic features.
MCF10A cell line	A non-tumorigenic epithelial cell line derived from breast tissue.
metagene	A consistent signal of expression for a collection of genes such as a biological pathway, derived from singular value decomposition.
microarray	A high-throughput technique to measure presence or abundance of nucleic acid sequences from binding to probes.
mutant	A variant or dysfunctional phenotype arising from a mutation in a gene.
mutation	A change in DNA sequence that disrupts gene function.
non-oncogene addiction	The dependence of a cancer cell on functioning non-mutant genes.
oncogene	A gene that potentially causes cancer, typically by over-expression or mutant gene variants.
prognosis	The estimation of disease progression and patient outcome.
recurrent mutation	The repeated occurrence of mutations in a particular gene across cancers.
RNA-Seq	Transcriptome data from sequencing RNA.
scale-free	A property of a network which has a power law vertex degree distribution, that is several highly connected hub genes and many with very few connections.
somatic mutation	A mutation that occurs in somatic cells, during a patient's lifespan.
sporadic cancer	Cancers which do occur in patients with a family history or carry a high-risk genetic variant.
synthetic lethal	Genetic interactions where inactivation of multiple genes is inviable (or deleterious) which are viable if inactivated separately.

treatment	Medical procedures for a disease to improve patient outcomes.
tumour suppressor	A gene potentially causes cancer, typically by disruption of functions which protect the cell from cancer.

Acronyms

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

Chapter 4

Synthetic Lethal Analysis of Gene Expression Data

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

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

were used to replicate known relationships between clinical and molecular characteristics for breast cancer and to demonstrate application of SLIPT directly on metagenes to detect synthetic lethal pathways.

Together these results demonstrate the wide range of applications for SLIPT analysis and examine the synthetic lethal partners of *CDH1* in breast and stomach cancer. These synthetic lethal genes and pathways were identified both in the context of the functional implications of novel synthetic lethal relationships and as potential actionable targets against *CDH1* deficient tumours, in addition to replication of established functions of E-cadherin. In particular, these analyses focused on comparisons with experimental screening data to explore the potential for SLIPT to augment triage of candidate partners and support further experimental investigations. The key synthetic lethal partner pathways for *CDH1*, supported by both approaches, will be examined in more detail at the gene and pathway structure level in Chapter 5.

4.1 Synthetic Lethal Genes in Breast Cancer

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

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

Table 4.1: Candidate synthetic lethal gene partners of *CDH1* from SLIPT

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

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

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

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

supporting a interactions with extracellular proteins and the tumour microenvironment.

4.1.1 Synthetic Lethal Pathways in Breast Cancer

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

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

It is also notable that the pathways over-represented in SLIPT candidate genes have strongly significant over-representation of Reactome pathways from the hypergeometric test (as described in Section 2.3.2). Even after adjusting stringently for multiple tests, biologically related pathways were supported together. These pathways are further supported by testing for synthetic lethality against *CDH1* mutations (mtSLIPT) with many of these pathways also among the most strongly supported in this analysis (shown in Table C.2). This mutation-based analysis more closely represents the null *CDH1* mutations in HDGC (Guilford *et al.*, 1998) and the experimental MCF10A cell line cell model (Chen *et al.*, 2014). There was still support for translational and immune pathways not detected in the isolated experimental system. G protein coupled receptors also among the most strongly supported pathways, supporting the experimental findings of Telford *et al.* (2015) for these intracellular signalling pathways already being targeted for other diseases.

Table 4.2: Pathways for *CDH1* partners from SLIPT

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

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

4.1.2 Expression Profiles of Synthetic Lethal Partners

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

In these expression profiles, a gene with a moderate or high signal across samples exhibiting low *CDH1* expression would represent a potential drug target. However, it appears that several molecular subtypes of cancer have elevation of different clusters of synthetic lethal candidates in samples with low *CDH1*. This clustering suggests

that different targets (or combinations) could be effective in different patients, suggesting potential utility for stratification. In particular, estrogen receptor negative, basal-like subtype, and “normal-like” tumours (Dai *et al.*, 2015; Eroles *et al.*, 2012; Parker *et al.*, 2009) have elevation of genes specific to particular clusters, indicative of some synthetic lethal interactions being specific to a particular molecular subtype or genetic background. Thus synthetic lethal drug therapy against these subtypes may be ineffective if it were designed against genes in another cluster.

A similar correlation structure was observed among the candidates tested against *CDH1* mutation (mtSLIPT), as shown in Figure C.1. This clustering analysis similarly identified several major clusters of putative synthetic lethal partner genes. In this case, many partner genes had consistently high expression across most of the (predominantly lobular subtype) *CDH1* breast cancer samples. However, a major exception to this in the *CDH1* expression analysis were the normal samples which were excluded from the mutation data (as they were not tested for tumour-specific genotypes). This supports synthetic lethal interventions being more applicable to *CDH1* mutant tumours. There was still considerable correlation structure, particularly among *CDH1* wildtype samples, sufficient to distinguish gene clusters. In contrast to the expression analysis the (predominantly ductal *CDH1* wildtype) basal-like subtype and estrogen receptor negative samples had depleted expression among most candidate synthetic lethal partners. This is consistent with synthetic lethal interventions only being effective in lobular estrogen receptor positive breast cancers in which they are a more common, as recurrent (driver) mutation. However, the remaining samples are still informative for synthetic lethal analysis (by SLIPT) as it requires highly expressing *CDH1* samples for comparison.

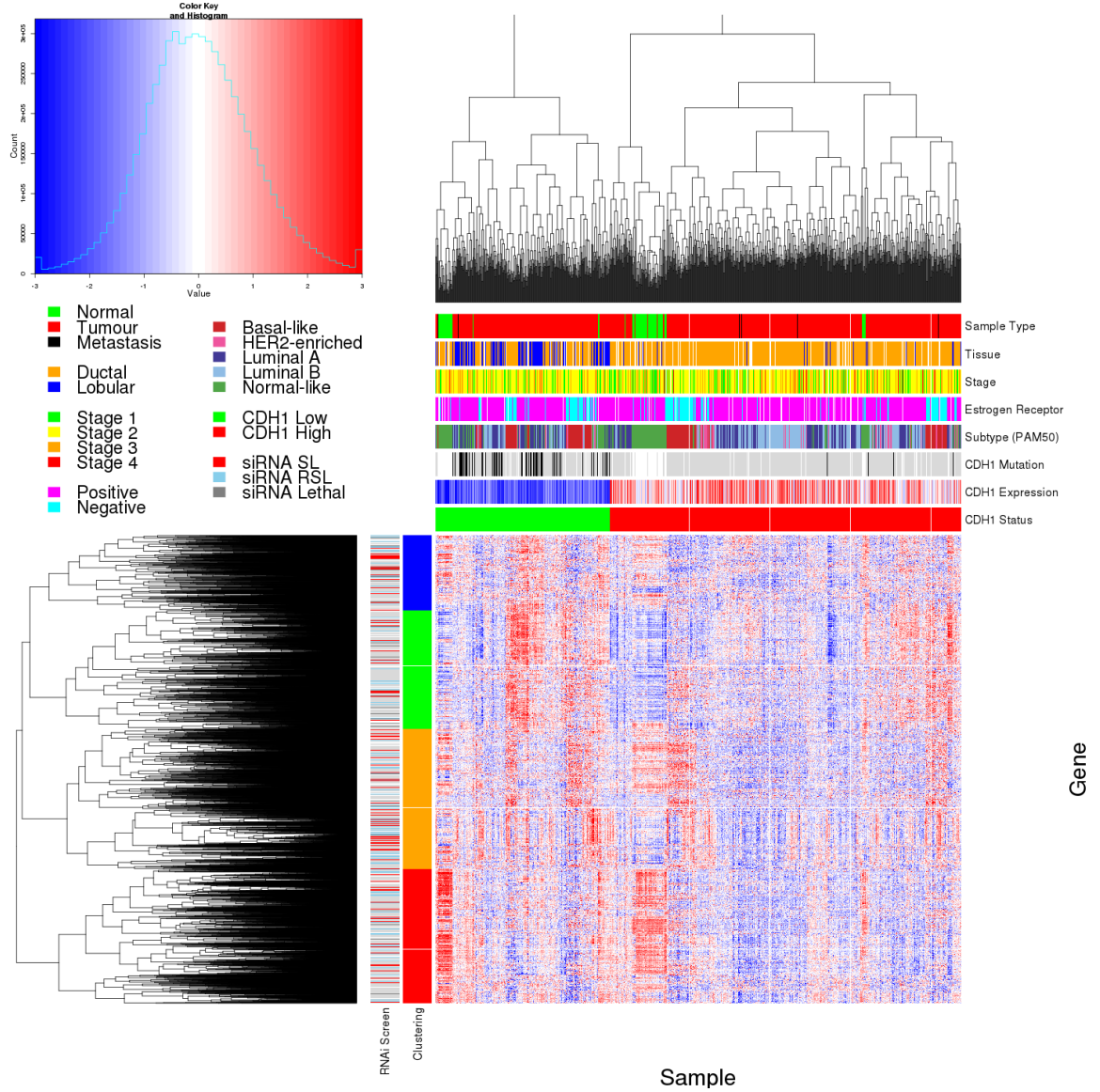


Figure 4.1: Synthetic lethal expression profiles of analysed samples. Gene expression profile heatmap (correlation distance, complete linkage) of all samples (separated by the $1/3$ quantile of *CDH1* expression) analysed in TCGA breast cancer dataset for gene expression of 5165 candidate partners of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

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

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

4.1.2.1 Subgroup Pathway Analysis

Synthetic lethal gene candidates for *CDH1* from SLIPT analysis of RNA-Seq gene expression data were also used for pathway over-representation analyses (as described in Section 2.3.2). The correlation structure in the expression of candidates synthetic lethal genes in *CDH1* low tumours (lowest $1/3^{\text{rd}}$ quantile of expression) was examined for distinct biological pathways in subgroups of genes elevated in different clusters of samples. These genes were highly expressed in different samples with their clinical

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

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

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.

factors including estrogen receptor status and intrinsic subtypes, from the Prediction Analysis of Microarray 50 (PAM50) procedure (Parker *et al.*, 2009) shown in Figure 4.1.

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

Based on these results, potential synthetic lethal partners of *CDH1* include processes known to be dysregulated in cancer, such as translational, cytoskeletal, and immune processes. Intracellular signalling cascades such as the GPCRs and extracellular stimuli for these pathways were also implicated in potential synthetic lethality with *CDH1*.

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

4.2 Comparing Synthetic Lethal Gene Candidates

4.2.1 Primary siRNA Screen Candidates

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

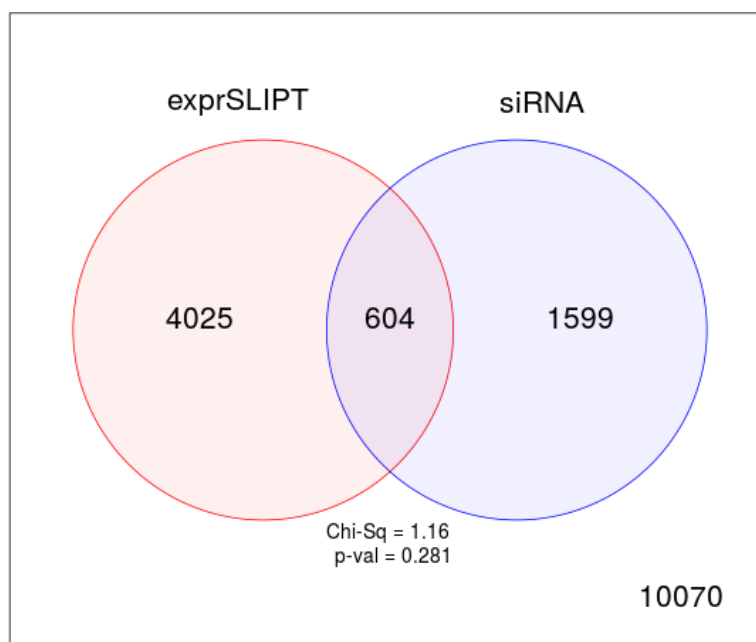


Figure 4.2: **Comparison of SLIPT to siRNA.** Testing the overlap of gene candidates for E-cadherin synthetic lethal partners between computational (SLIPT) and experimental screening (siRNA) approaches. The χ^2 test suggests that the overlap is no more than would be expected by chance ($p = 0.281$).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

4.2.2 Comparison with Correlation

Another potential means to triage drug target candidates is by correlation of expression profiles with *CDH1*. Correlation with *CDH1* was compared to SLIPT and siRNA

results in Figure 4.3. The genes not detected by SLIPT (including siRNA candidates) had included gene with insignificant SLIPT p-values. As expected, these genes were distributed around a correlation of zero and genes with higher correlation with *CDH1* (either direction) were more significant, although there were exceptions to this trend and larger positive correlations than negative correlations. The majority of SLIPT candidates had negative correlations, particularly genes detected by both approaches, although these were typically weak correlations and are unlikely to be sufficient to detect such genes on their own. This is supported by simulation results in Section 6.1.

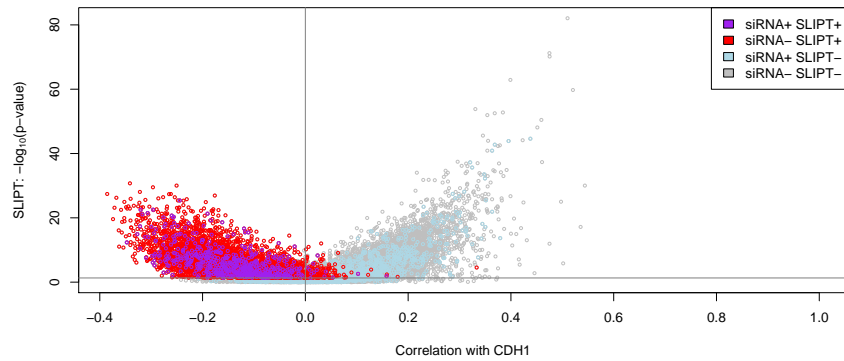


Figure 4.3: **Compare SLIPT and siRNA genes with correlation.** The χ^2 p-values for genes tested by SLIPT (in TCGA breast cancer) expression analysis were compared against Pearson’s correlation of gene expression with *CDH1*. Genes detected by SLIPT or siRNA are coloured according to the legend.

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

The apparent tendency for genes detected by SLIPT or siRNA to have negative correlations with *CDH1* expression is not due to the smaller number of genes in these groups. The distribution of *CDH1* correlations differed across these gene groups (as

shown by Figures 4.4 and C.4), specifically lower in SLIPT candidates (as supported by analysis of variance (ANOVA) in Table 4.4). However, these are relatively weak correlations and further triage of gene candidates by correlation is not suitable, nor is use of correlation itself to predict synthetic lethal partners in the first place.

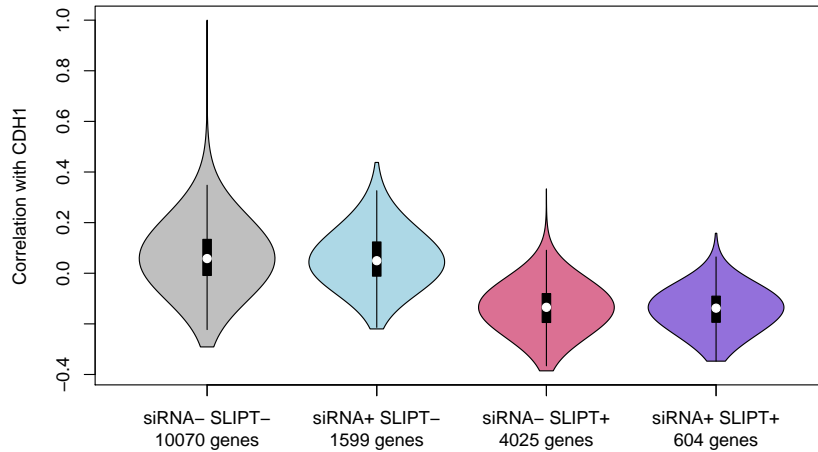


Figure 4.4: **Compare SLIPT and siRNA genes with correlation.** Genes detected as candidate synthetic lethal partners by SLIPT (in TCGA breast cancer) expression analysis and experimental screening (with siRNA) were compared against Pearson’s correlation of gene expression with *CDH1*. There were no differences in correlation between gene groups detected by either approach.

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

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

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

4.2.3 Comparison with Primary Screen Viability

A similar comparison of SLIPT results was made with the viability ratio (of *CDH1* mutant to wildtype) in the primary siRNA screen performed by Telford *et al.* (2015).

The significance and viability thresholds used for SLIPT and siRNA detection of synthetic lethal candidate partners of *CDH1* are shown in Figure 4.5. However, not all of the genes below the viability thresholds were necessarily selected to be candidate partners, as additional criteria were used in each case: directional criteria as for SLIPT (see Section 3.1) and minimum wildtype viability for siRNA (Telford *et al.*, 2015).

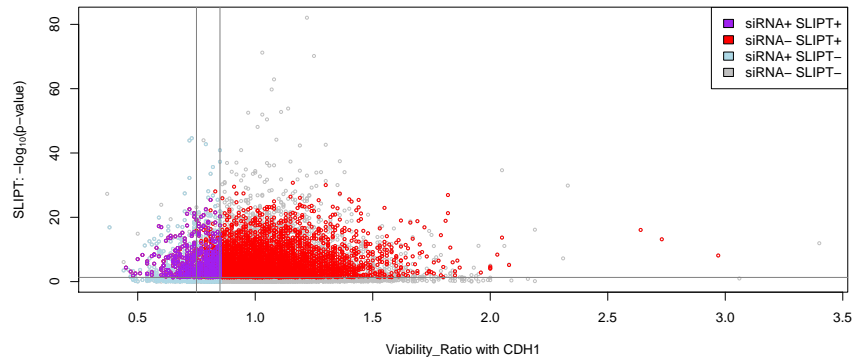


Figure 4.5: **Compare SLIPT and siRNA genes with viability.** The χ^2 p-values for genes tested by SLIPT (in TCGA breast cancer) expression analysis were compared (on a log-scale) against the viability ratio of *CDH1* mutant and wildtype cells in the primary siRNA screen. Genes detected by SLIPT or siRNA are coloured according to the legend.

There does not appear to be a clear relationship between SLIPT and siRNA candidates. Many genes not detected by both approaches were numerous in Figures 4.2 and C.2. These genes detected by either are not necessarily near the thresholds for the other. In this respect the SLIPT approach with patient data and cell line experiments are independent means to identify synthetic lethal candidates. While genes detected by both approaches were not necessarily more strongly supported by either, the genes with a viability closer to 1 (no synthetic lethal effect) in siRNA included those with more significant SLIPT p-values whereas more extreme viability ratios tended to be less significant (as shown by Figure 4.5). However, it should be noted that genes with more moderate viability ratios were more common and SLIPT was capable (despite adjusting for multiple testing) of detecting significant genes with extreme viability ratios, particularly those considerably lower than 1.

However, there was not little support for SLIPT candidates having considerably different viability ratios (as shown in Figures 4.6 and C.5). While the viability thresholds used by Telford *et al.* (2015) to detect synthetic lethal candidates in the primary

screen, the genes identified by SLIPT had a higher mean viability ratio (by t-test: $t = 2.1553$, $p = 0.03117$). However, the effect size was small (mean SLIPT- 1.029, mean SLIPT+ 1.037) and the vast majority of SLIPT candidate genes did not have different viability in the primary screen to genes not identified by SLIPT.

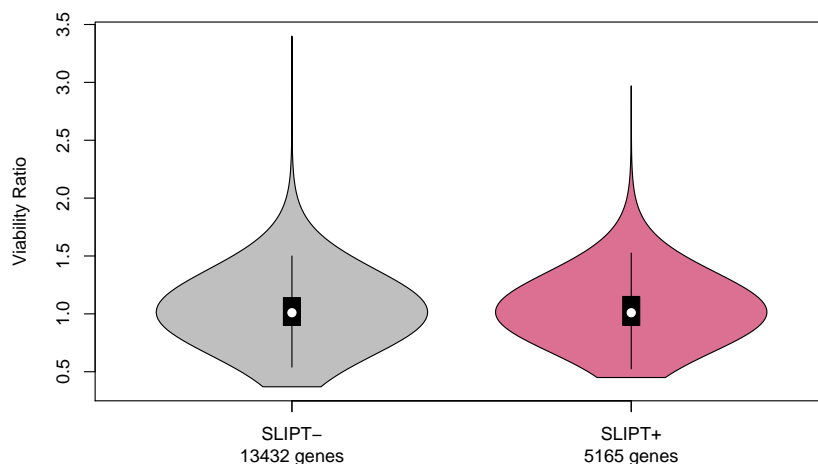


Figure 4.6: **Compare SLIPT genes with siRNA viability.** Genes detected as candidate synthetic lethal partners by SLIPT (in TCGA breast cancer) expression analysis were compared against the viability ratio of *CDH1* mutant and wildtype cells in the primary siRNA screen. There were clear no differences in viability between genes detected by SLIPT and those not with the differences being primarily due to viability thresholds being used to detect synthetic lethality by Telford *et al.* (2015).

4.2.4 Comparison with Secondary siRNA Screen Validation

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

dependent replication with 4 gene knockdowns with different targeting siRNAs. This was performed for the top 500 candidates (with the lowest viability ratio) from the primary screen and the 482 of these genes also tested by SLIPT in breast cancer.

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

Table 4.5: Comparing SLIPT genes against secondary siRNA screen

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

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

4.2.5 Comparison to Primary Screen at Pathway Level

These pathway over-representation analyses (performed as described in Section 2.3.2) correspond to genes separated into SLIPT or siRNA screen candidates unique to either method or detected by both (Table 4.6). The SLIPT-specific gene candidates were involved most strongly with translational and immune regulatory pathways, although extracellular matrix pathways were also supported. These pathways were largely consistent with those identified in Table 4.2 and in the clustering analysis (Table 4.3). The genes detected only by the siRNA screen had over-representation of cell signalling pathways, including many containing genes known to be involved in cancer (e.g., MAPK, PDGF, ERBB2, and FGFR), with the detection of Class A GPCRs supporting the independent analyses by Telford *et al.* (2015). The intersection of computational and experimental synthetic lethal partners of *CDH1* had stronger evidence for over-representation of GPCR pathways and more specific subclasses, such as visual phototransduction ($p = 6.9 \times 10^{-10}$) and $G_{\alpha s}$ signalling events ($p = 1.7 \times 10^{-7}$), than other signalling pathways.

The pathway analysis for mtSLIPT against *CDH1* mutations (in Table C.4) had concordant results for both mtSLIPT-specific and siRNA-specific pathways. While the specific pathway composition of the intersection of these analyses differed from SLIPT against low *CDH1* expression, signalling pathways including GPCRs, NOTCH, EERB2, PDGF, and SCF-KIT. These findings indicate the signalling pathways are among the most suitable vulnerability to exploit in targeting *CDH1* deficient tumours as they can be detected in both a patient cohort (with TCGA expression data) and tested in a laboratory system. However, it is possible that the isolated experimental system is set up to preferentially detect kinase singalling pathways (which are amenable to pharmacological inhibition and translation to the clinic) and the other pathways identified by SLIPT may still be informative of the role of *CDH1* loss of function in cancers or mechanisms by which further gene loss leads to specific inviability.

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

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

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

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

4.2.5.1 Resampling Genes for Pathway Enrichment

Comparisons of genes between experimental screen candidates and prediction from TCGA expression data were less consistent than comparisons of pathways. However, this is not unexpected, since synthetic lethal pathways are more robustly conserved (Dixon *et al.*, 2008) and the computational approach using patient samples from complex tumour microenvironment has considerably different strengths to an experimental screen (Telford *et al.*, 2015) based on genetically homogenous cell line models in an isolated laboratory environment. For instance, it is unlikely for immune signalling to be detected in an isolated cell culture system.

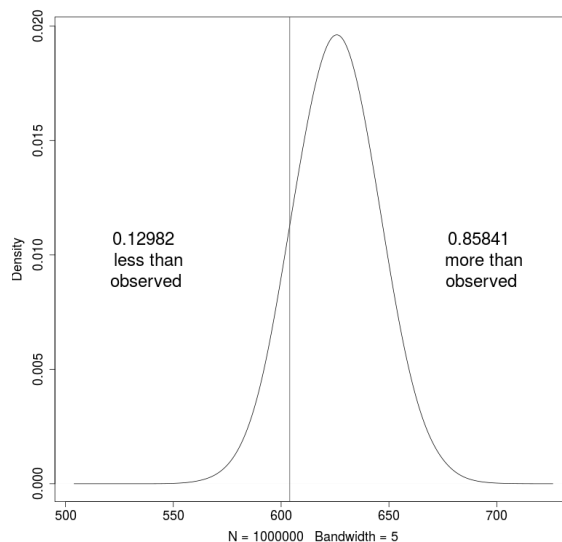


Figure 4.7: **Resampled intersection of SLIPT and siRNA candidates.** Resampling analysis of intersect size from genes detected by SLIPT and siRNA screening approaches over 1 million replicates. The proportion of expected intersection sizes for random samples below or above the observed intersection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

The overlap between synthetic lethal candidates from bioinformatics SLIPT predictions and siRNA screening has raised other questions, including whether the pathways over-represented would be expected by chance. This of particular concern since the siRNA candidate genes themselves are highly over-represented for particular pathways (e.g., GPCRs) so selecting any intersect with them could be enriched for these pathways. Another pathway-based approach is to test whether pathways are over-represented in randomly sampled genes, comparing many “resamplings” or “permutations” of these genes to the enrichment statistics observed for these pathways in the

SLIPT candidates and their intersection with the siRNA hits shows whether we detect these pathways more than we expect by chance (as described in Section 2.3.6).

Of particular concern are the over-represented pathways in genes detected by both methods. Pathway over-representation alone does not detect whether SLIPT predicted genes or siRNA candidates are enriched within each other. This resampling analysis therefore detects whether over-represented pathways were detected by SLIPT independently of their over-representation among siRNA candidates (without assuming an underlying test statistic distribution).

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

A permutation analysis was performed to resample the genes tested by both approaches to investigate whether the observed pathway over-representation could have occurred in a randomly selected sample of genes from the experimental candidates, that is, whether the pathway predictions from SLIPT could be expected by chance (as described in Sections 2.2.3.1 and 2.3.6). While the number of siRNA candidate genes also detected by SLIPT was not statistically significant ($p = 0.281$), this may be due to the vastly different limitations of the approaches and the correlation structure of gene expression not being independent (as assumed for multiple testing procedures). The intersection may still be functionally relevant to *CDH1*-deficient cancers, such as the pathway data in Table 4.6. The resampling analysis for pathways was compared to the pathway over-representation for SLIPT predicted synthetic lethal partners in Table 4.7. Similarly, the pathway resampling for intersection between SLIPT predictions and experimental screen candidates was compared to pathway over-representation in Table 4.8 for intersection with siRNA data.

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

were largely consistent for translation and post-transcriptional gene regulation, supporting gene set over-representation of the SLIPT-specific pathways in Table 4.7. In particular, some of the most significantly over-represented pathways had higher observed χ^2 values than any of the 1 million random permutations. Similar pathways were also replicated by permutation analysis for mtSLIPT candidate partners against *CDH1* mutation (shown in Table C.5). This shows that many of the pathways detected specifically by SLIPT are replicated by permutation procedures and that the permutation approach is capable of detecting many of the most strongly over-represented pathways.

Table 4.7: Pathways for *CDH1* partners from SLIPT

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

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

The permutation approach was then also applied to the intersection between computational and experimental candidates. The permutation analysis is testing for consistent detection of pathways was independent of their pre-existing status as experimental candidates. The pathway results for these candidate partners (in Table 4.8) differed between over-representation and resampling analyses.

Namely, many of the over-represented pathways were not significant in the resampling analysis, including visual phototransduction and retinoic acid signalling, and were likely over-represented in the intersection due to over-representation in the siRNA candidates rather than additional support from SLIPT. In contrast, pathways involving defective *EXT1* or *EXT2* genes approach significance after FDR adjustment for multiple tests in resampling. Of the highest over-represented pathways in the intersection, only $G_{\alpha s}$ signalling events were supported by both over-representation and resampling analyses. Other pathways supported by both analyses were cytoplasmic elastic fibre formation, associated HS-GAG protein modification pathways, energy metabolism, and the fibrin clotting cascade.

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

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

However, several pathways, including some immune functions and neurotransmitters, were supported by the resampling analysis (in Tables 4.8 and C.6) when the initial pathway over-representation test was not significant. These functions appear to have been detected by both approaches more than expected by chance but must be interpreted with caution since they were still not common enough to be detected in pathway over-representation analysis.

4.2.6 Integrating Synthetic Lethal Pathways and Screens

Based on these results, it appears that computational and experimental approaches to synthetic lethal screening for *CDH1* lead to a broader functional characterisation, and many candidate partners, when combined, despite different strengths and limitations. Compared to candidate gene approaches, experimental genomes-wide screens are an appealing unbiased strategy for identifying synthetic lethal interactions. Since these

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

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

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

screens are costly, laborious, and specific to genetic background, computational analysis can augment candidate triage to either reduce the initial panel of screened genes or prioritise validation.

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

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

The predicted synthetic lethal partners occurred across functionally distinct pathways, including characterised functions of *CDH1*. This diversity is consistent with the wide ranging role of *CDH1* in cell-cell adhesion, cell signalling, and the cytoskeletal structure of epithelial tissues. Pathway structure may be relevant to identifying potential drug targets from gene expression signatures, indicating downstream effector genes and mechanisms leading to cell inviability. These distinct synthetic lethal gene clusters and pathways may further lead to the elucidation of drug resistance mechanisms.

4.3 Metagene Analysis

The gene signatures (Gatza *et al.*, 2011, 2014) were used to demonstrate the utility of the metagene approach for use on a wider range of pathways as was performed with the Reactome (Croft *et al.*, 2014) pathways as an alternative approach to identification of synthetic lethal pathways. metagene serve as a summary of activity for each pathway. The direction of metagenes (derived by the singular value matrix decomposition) is

generally arbitrary but care has been taken to ensure that these occur in a direction which reflect overall activation of the pathway (as described in Section 2.2.3). meta-gene were derived for well characterised gene signatures in breast cancer (Gatza *et al.*, 2011, 2014) to verify that that these pathway signatures are consistent with expected molecular properties of each molecular subtype (Parker *et al.*, 2009; Perou *et al.*, 2000). This was performed by examining the pathway expression of these breast cancer gene signatures in TCGA expression data. These metagenes were also compared to somatic mutation mutation to evaluate mutation as a measure of gene activity in comparison to gene and protein expression.

The gene signatures (Gatza *et al.*, 2011, 2014) were used to demonstrate to utility of the metagene approach for use on a wider range of pathways. Having established that metagenes generated with this procedure reflect gene activity, the metagene procedure (in Section 2.2.3) was then applied to the Reactome pathways (Croft *et al.*, 2014). These Reactome metagenes were used for synthetic lethal analysis of pathways with SLIPT, directly using pathway activity for identifying synthetic lethal pathways with *CDH1*.

4.3.1 Pathway Expression

Pathway metagenes (generated as described in Section 2.2.3) for gene signatures of key processes in breast cancer (Gatza *et al.*, 2011) were used to check that metagenes were generated in the correct direction to indicate pathway activation. Some of these gene signatures are plotted in Figure 4.8 for comparison with clinical factors and somatic mutation mutations. The “intrinsic subtypes” was computed by performing the PAM50 procedure Parker *et al.* (2009) for RNA-Seq data which was highly concordant ($\chi^2 = 1305.9$, $p = 2.73 \times 10^{-268}$) with the subtypes provided by University of California, Santa Cruz (UCSC) (UCSC, 2012) for TCGA samples (TCGA, 2012) previously analysed by microarrays (as shown in Appendix D). Somatic mutation mutations were reported for gslinkrecurrent mutationrecurrently mutated genes in breast cancer, as reported by TCGA (TCGA, 2012), related genes, and those previously discussed to be important in hereditary breast cancers (*BRCA1*, *BRCA2*, and *CDH1*).

These gene signatures reflect intrinsic subtypes as expected. In particular, the estrogen and progesterone receptor signatures are low in the predominantly Estrogen receptor (ER)[−] and Progesterone receptor (PR)[−] basal-like subtype tumours. These tumours also had the highest frequency of *TP53* mutations and a corresponding reduction of p53 metagene activity, as expected for loss of a tumour suppressor. The

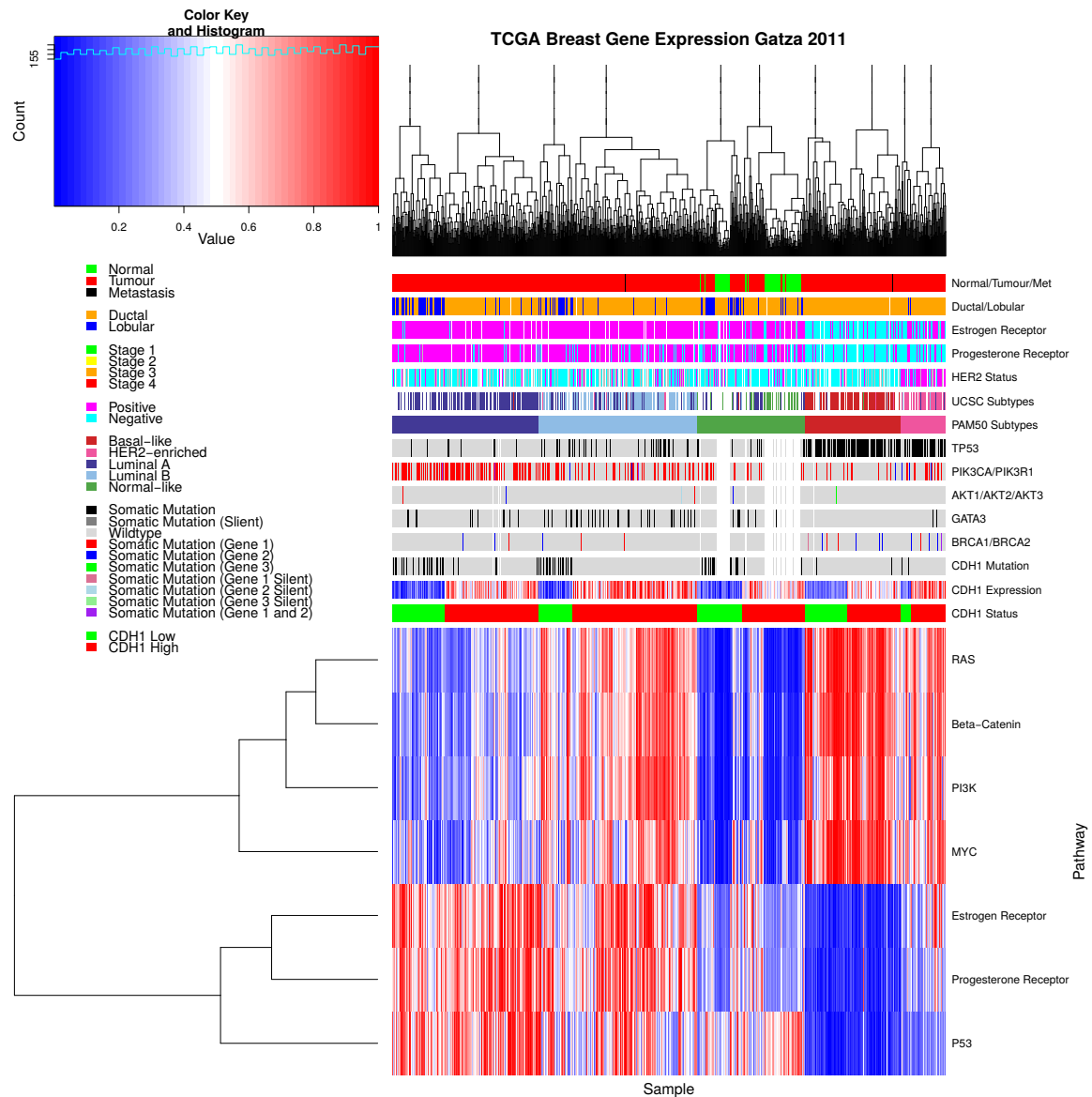


Figure 4.8: **Pathway metagene expression profiles.** Expression profiles for metagene signatures from Gatza *et al.* (2011) in TCGA breast data, annotated for clinical factors (with sample types and histological results coloured according to the legend) and cancer gene mutations (Negative values for mutation are light grey with missing data in white). Intrinsic subtypes are shown as derived from microarray (UCSC) and RNA-Seq (PAM50) data (Parker *et al.*, 2009; TCGA, 2012). Samples were clustered independently for each intrinsic subtypes and by *CDH1* expression status. Pathway expression signatures are consistent with mutations and clinical subgroups.

luminal A and luminal B tumour subtypes are the most similar, which is reflected in these metagenes signatures, although they are distinguishable molecular subtypes as shown by elevated phosphoinositide 3-kinase (PI3K), AKT, RAS, and β -catenin signalling in luminal B tumours. However, these pathways were also elevated in basal-like and HER2-enriched subtypes and lowly expressed in the “normal-like” subtype (which contained the normal samples). These intrinsic subtypes specific gene signature profiles were further supported with metagenes for an extended set of signatures (Gatza *et al.*, 2014), as shown in Figure C.9.

TP53 mutations were the most frequent and more common in the basal-like subtype. Similarly, *GATA3* mutations were more common in luminal subtype tumours. PI3K mutations were more frequent across breast tumours, although these were less common in the basal-like subtype despite an elevated metagene (this discrepancy will be discussed further in Section 4.3.2). *CDH1* mutations similarly occurred across molecular subtypes with the exception of the basal-like subtype (as observed in gene expression with Figure 4.1). *CDH1* low samples occurred in all subtypes but were predominantly of the lobular histological subtype. Apart from these genes, mutations did not show clear specificity to a particular subtype and the variation between samples reflects the range of molecular cascades that can result in tumours with similar molecular profiles, supporting the use of gene expression data for cancer diagnostics and identification of molecular targets.

The direction of each metagene was consistent with the clinical characteristics, which formed a consensus of gene activity as shown for the PI3K and ER signatures (Gatza *et al.*, 2011) in Figures 4.9 and 4.10, respectively. Supporting data for p53 and BRCA metagenes (Gatza *et al.*, 2011, 2014) are given in the Appendix (Figures C.10 and C.11). In each of the examples for gene signatures, the expression of the majority of the genes were highly concordant with the metagene, being either positively or negatively correlated. These were generally consistent with established clinical and molecular subtypes of breast cancer and the recurrent mutations shown. However, the *PIK3CA* and *PIK3R1* mutant samples did not necessarily have elevated PI3K pathway metagene activity (as shown in Figure 4.9).

4.3.2 Somatic Mutation

It should be noted that metagenes, while consistent with the consensus of constituent expressed genes, were not necessarily reflecting the somatic mutation status. The PI3K (Gatza *et al.*, 2011) metagene levels in particular, were not statistically signif-

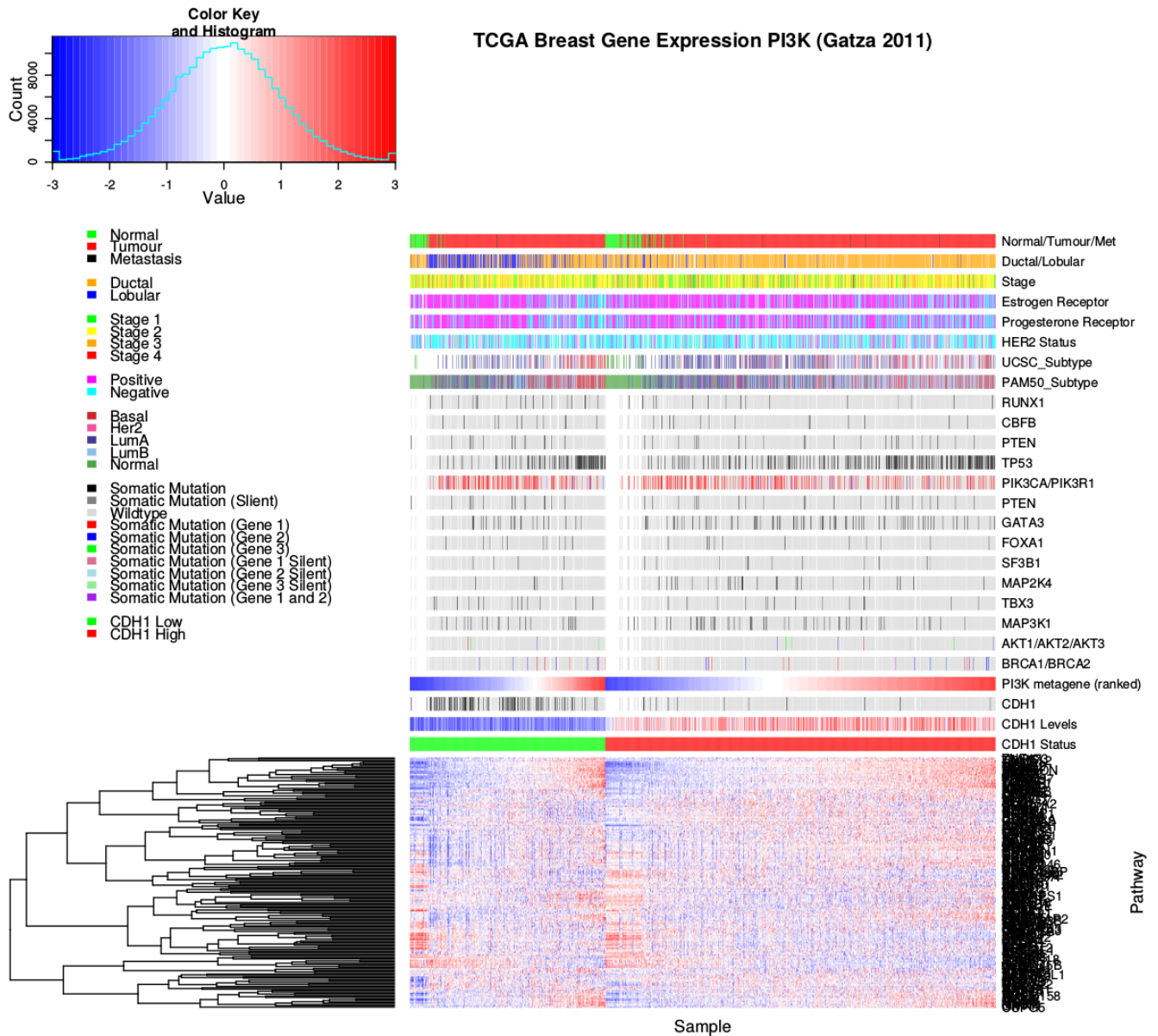


Figure 4.9: **Expression profiles for constituent genes of PI3K.** Expression profiles the genes contained in the PI3K gene signature from Gatza *et al.* (2011) in TCGA breast data, annotated for clinical factors and cancer gene mutations. Samples are separated by *CDH1* expression status and sorted by the metagene. In both cases, the majority of genes were consistent with the direction of the PI3K metagene, although considerable proportion were inversely correlated with the metagene. Normal samples had low PI3K metagene expression and *TP53* mutant samples had high PI3K expression. Although, oncogenic *PIK3CA* and tumour suppressor *PIK3R1* mutations across samples including those with low metagene response.

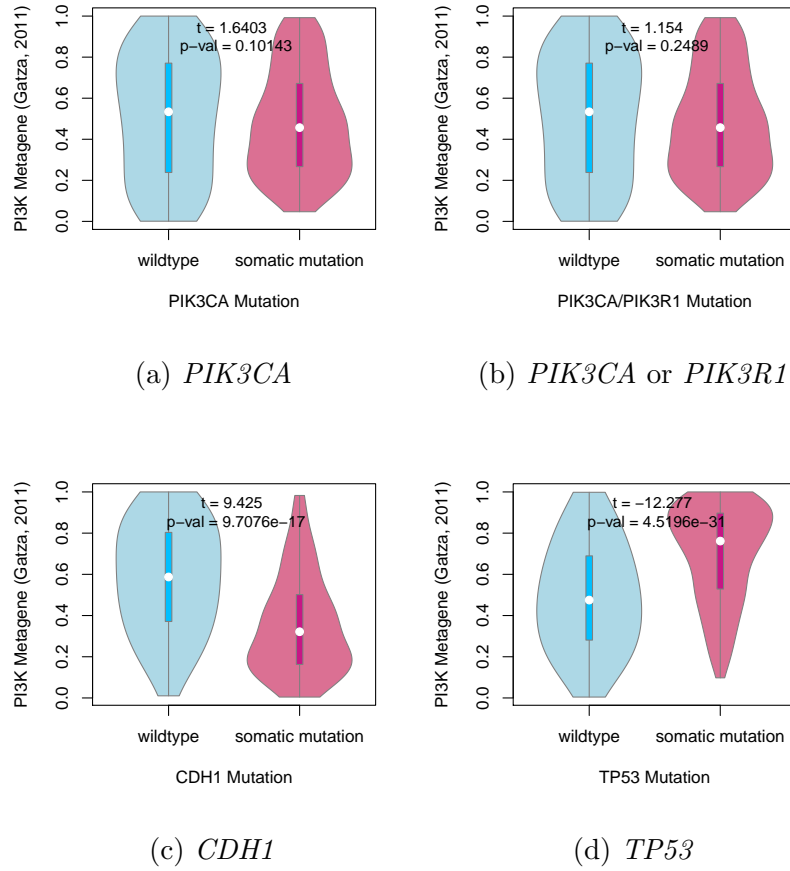


Figure 4.11: **Somatic mutation against the PI3K metagene.** Mutations in *PIK3CA*, *PIK3R1*, *CDH1*, and *TP53* were examined in TCGA breast cancer for their association with the PI3K (Gatza *et al.*, 2011) pathway metagene. The tumour suppressors *CDH1* and *TP53* showed an increase and decrease in the metagene respectively, whereas *PIK3CA* and *PIK3R1* mutations had little effect on the metagene levels.

icantly varying between mutant and wildtype *PIK3CA* samples (shown in Figure 4.11). However, the PI3K metagene differed across *CDH1* and *TP53* mutations, remarkably in opposite directions considering that PI3K is an oncogenic growth pathway and these are both most frequently tumour suppressors inactivated in cancers. This shows that *CDH1* and *TP53* deficient tumours have distinct molecular growth pathways and that synthetic lethal interventions against loss of *CDH1* function may not be applicable to other cancers with driver mutations such as *TP53*, although these were kept in the analysis for comparison. These differences may be related to these mutations being more frequent in tumours with difference clinical characteristics (as observed in Sec-

tion 4.3.1). Thus mutations do not necessarily have corresponding changes in pathway expression, particularly for oncogenes which may change in function rather than being upregulated.

While the more specific *PIK3CA* (Gatza *et al.*, 2014) metagene showed significant differences with *PIK3CA* and *PIK3R1* mutations (as shown in Figure C.6), this metagene replicated stronger differences for *CDH1* and *TP53*. These differences were less pronounced in the protein levels of p110 α (encoded by *PIK3CA*) and the downstream AKT gene (shown in Figures C.7 and C.8 respectively). However, this may be due to this regulatory cascade (kinases) being transmitted as a change in protein state (phosphorylation) rather than changes in expression levels. Another consideration is that mutations at different loci have different effects on protein function, particularly for oncogenes.

4.3.3 Synthetic Lethal Pathway Metagenes

Pathway metagenes for Reactome pathways (generated as described in Section 2.2.3) were also used for testing synthetic lethal partner pathways with *CDH1* by SLIPT. Since the metagenes have are higher when the pathway as a whole is activated, they are amenable to SLIPT analysis using low metagene levels for inactivated pathways. These synthetic lethal metagenes differed to the over-represented pathways among synthetic lethal gene candidates. However, there were some similarities to previous findings, as shown in Tables 4.9. In particular, translational pathways were replicated as observed in Table 4.2. While the specific pathways differ, immune pathways (e.g., NF- κ B) were also supported by metagene synthetic lethal analysis.

Signalling pathways were more strongly supported by mtSLIPT analysis of metagene pathway expression against *CDH1* mutation, as shown in Table C.7, although these results were generally less statistically significant than expression analyses. Signalling pathways detected as synthetic lethal metagenes include G $_{\alpha z}$, insulin-related growth factor (IGF), GABA receptor, G $_{\alpha s}$, S6K1 and various toxin responses mediated by GPCRs. Metabolic processes including processing of carbohydrates and fatty acids were also implicated across these analyses.

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

Table 4.9: Candidate synthetic lethal metagenes against *CDH1* from SLIPT

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

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

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

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

4.3.4 Synthetic Lethality in Breast Cancer

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

and metabolism necessary for the proliferation of cancer cells. This is consistent with the GPCR pathways such as $G_{\alpha s}$ signalling being supported by SLIPT gene candidates and the experimental primary siRNA screen, as shown by resampling in Section 4.2.5.1.

4.4 Replication in Stomach Cancer

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

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

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

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

4.5 Discussion

4.5.1 Strengths of the SLIPT Methodology

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

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

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

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

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

4.5.2 Synthetic Lethal Pathways for E-cadherin

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

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

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

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

This diversity in synthetic lethal functions is consistent with the wide ranging role of *CDH1* in cell-cell adhesion, cell signalling, and the cytoskeletal structure of epithelial tissues. Pathway structure may be relevant to identifying potential drug targets from gene expression signatures, indicating downstream effector genes and mechanisms leading to cell inviability. Identification of distinct synthetic lethal gene clusters may further lead to the elucidation of drug resistance mechanisms. While these pathways are indicative of the main functions of E-cadherin and synthetic lethal partners, it remains to identify the genes within these pathways that are the most actionable or supported across SLIPT analysis in patient samples and detected by experiments in preclinical models (Chen *et al.*, 2014; Telford *et al.*, 2015). The specific genes within key pathways will be discussed in Chapter 5, along with further investigations into

their relation to pathway structure. While these are important clinical implications, the synthetic lethal predictions lack enough confidence for direct translation into pre-clinical models or clinical applications leading to a need for statistical modelling and simulation of synthetic lethality in genomics expression data.

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

4.5.3 Replication and Validation

4.5.3.1 Integration with siRNA Screening

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

The pathways detected by both over-representation and resampling are the strongest candidates for further investigation and the pathway structure analyses in Chapter 5 will focus on these pathways detected by both over-representation and resampling. Particularly, those replicated across datasets or with pathway metagenes. In addition to GCPR pathways detected across these analyses, the PI3K cascade will also be investigated in Chapter 5, this signalling pathway is a well characterised mediator between GCPR receptors and regulation of translation (Gao and Roux, 2015) (both detected throughout this Chapter) and exhibited unexpected behaviour with pathway the metagenes (in Section 4.3). This pathway is activated by protein Phosphorylation states and thus inactivatino may not be detectable with expression.

However, the SLIPT approach was shown to be predictive of which siRNA primary screen candidate partners of *CDH1* were validated in a secondary screen (as shown in Section 4.2.4). These results further support SLIPT for identifying robust synthetic lethal candidates which can be validated and as a triage approach for interpreting screening experiments.

4.5.3.2 Replication across Tissues

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

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

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

4.6 Summary

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

This approach has been applied to robustly detect synthetic lethal pathways for the E-cadherin (*CDH1*) in TCGA breast cancer molecular profiles with comparisons to experimental screening (Telford *et al.*, 2015) in cell lines, and replication in TCGA stomach cancer molecular profiles and across cell types in the cancer cell line encyclopaedia. The pathway replicated across several analyses included extracellular matrix pathways (e.g., elastic fibres formation), cell signalling (including GPCRs), and core gene regulation and translation processes crucial for the growth and proliferation of cancer cells. These pathways show evidence of non-oncogene addiction for *CDH1* deficient cells and present vulnerabilities which may be exploited for specific treatment against *CDH1* mutations in HCGC and sporadic cancers. There was also support for

synthetic lethal pathways with *CDH1* in cell adhesion and cytoskeletal processes to which *CDH1* belongs, supporting the finding that synthetic lethality occurs within biological pathways (Boone *et al.*, 2007; Kelley and Ideker, 2005).

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

A characteristic of gene interaction networks is a scale-free topology leading to highly interacting hub genes, these represent important genes in a functional network. Cell surface interactions, the extracellular matrix, and cell signalling (particularly PI3K/AKT signalling) were also found to be synthetic lethal hubs with more interactions detected than other genes. This indicates that these pathways are functionally important to survival of cancer cells since they are subject to high functional redundancy, despite frequent disruptions in cancer. These pathways being involved in a disproportionate number of synthetic lethal interactions is also consistent with their detection for *CDH1*.

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

References

- Aarts, M., Bajrami, I., Herrera-Abreu, M.T., Elliott, R., Brough, R., Ashworth, A., Lord, C.J., and Turner, N.C. (2015) Functional genetic screen identifies increased sensitivity to wee1 inhibition in cells with defects in fanconi anemia and hr pathways. *Mol Cancer Ther*, **14**(4): 865–76.
- Abeshouse, A., Ahn, J., Akbani, R., Ally, A., Amin, S., Andry, C.D., Annala, M., Aprikian, A., Armenia, J., Arora, A., *et al.* (2015) The Molecular Taxonomy of Primary Prostate Cancer. *Cell*, **163**(4): 1011–1025.
- Adler, D. (2005) *vioplot: Violin plot*. R package version 0.2.
- Akbani, R., Akdemir, K.C., Aksoy, B.A., Albert, M., Ally, A., Amin, S.B., Arachchi, H., Arora, A., Auman, J.T., Ayala, B., *et al.* (2015) Genomic Classification of Cutaneous Melanoma. *Cell*, **161**(7): 1681–1696.
- Akobeng, A.K. (2007) Understanding diagnostic tests 3: receiver operating characteristic curves. *Acta Pdiatrica*, **96**(5): 644–647.
- American Cancer Society (2017) Genetics and cancer. <https://www.cancer.org/cancer/cancer-causes/genetics.html>. Accessed: 22/03/2017.
- Anjomshoaa, A., Lin, Y.H., Black, M.A., McCall, J.L., Humar, B., Song, S., Fukuzawa, R., Yoon, H.S., Holzmann, B., Friederichs, J., *et al.* (2008) Reduced expression of a gene proliferation signature is associated with enhanced malignancy in colon cancer. *Br J Cancer*, **99**(6): 966–973.
- Araki, H., Knapp, C., Tsai, P., and Print, C. (2012) GeneSetDB: A comprehensive meta-database, statistical and visualisation framework for gene set analysis. *FEBS Open Bio*, **2**: 76–82.

- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, **25**(1): 25–29.
- Ashworth, A. (2008) A synthetic lethal therapeutic approach: poly(adp) ribose polymerase inhibitors for the treatment of cancers deficient in dna double-strand break repair. *J Clin Oncol*, **26**(22): 3785–90.
- 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.
- Babiyak, 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., Gatz, M.L., Beck, A.H., Wilkerson, M.D., Rhie, S.K., Pastore, A., Zhang, H., McLellan, M., Yau, C., Kandoth, C., *et al.* (2015) Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell*, **163**(2): 506–519.

- Clark, M.J. (2004) Endogenous Regulator of G Protein Signaling Proteins Suppress G 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 re-activation of mapk signaling contributes to insensitivity of *BRAF*-mutant colorectal cancers to raf inhibition with vemurafenib. *Cancer Discovery*, **2**(3): 227–235.
- Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L., Toufighi, K., Mostafavi, S., *et al.* (2010) The genetic landscape of a cell. *Science*, **327**(5964): 425–31.
- Costanzo, M., Baryshnikova, A., Myers, C.L., Andrews, B., and Boone, C. (2011) Charting the genetic interaction map of a cell. *Curr Opin Biotechnol*, **22**(1): 66–74.
- 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 SNAIL1 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.
- Sørlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., *et al.* (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA*, **98**(19): 10869–10874.
- Stajich, J.E. and Lapp, H. (2006) Open source tools and toolkits for bioinformatics: significance, and where are we? *Brief Bioinformatics*, **7**(3): 287–296.
- Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009) The cancer genome. *Nature*, **458**(7239): 719–724.
- Ström, C. and Helleday, T. (2012) Strategies for the use of poly(adenosine diphosphate ribose) polymerase (parp) inhibitors in cancer therapy. *Biomolecules*, **2**(4): 635–649.
- Sun, C., Wang, L., Huang, S., Heynen, G.J.J.E., Prahallad, A., Robert, C., Haanen, J., Blank, C., Wesseling, J., Willems, S.M., *et al.* (2014) Reversible and adaptive resistance to *BRAF*(v600e) inhibition in melanoma. *Nature*, **508**(7494): 118–122.
- 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 gene-gene 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.