

Contents

1	Introduction	1
1.1	Cancer Research in the Post-Genomic Era	1
1.1.1	Cancer as a Global Health Concern	2
1.1.1.1	Genetics and Molecular Biology in Cancers	3
1.1.2	The Human Genome Revolution	5
1.1.2.1	The First Human Genome Sequence	5
1.1.2.2	Impact of Genomics	6
1.1.3	Technologies to Enable Genetics Research	7
1.1.3.1	DNA Sequencing and Genotyping Technologies	7
1.1.3.2	Microarrays and Quantitative Technologies	8
1.1.3.3	Massively Parallel “Next Generation” Sequencing	9
1.1.3.3.1	Molecular Profiling with Genomics Technology	10
1.1.3.3.2	Established Sequencing Technologies	11
1.1.3.3.3	Emerging Sequencing Technologies	12
1.1.3.4	Bioinformatics as Interdisciplinary Genomic Analysis	14
1.1.4	Follow-up Large-Scale Genomics Projects	14
1.1.5	Cancer Genomes	15
1.1.5.1	The Cancer Genome Atlas Project	16
1.1.5.2	The International Cancer Genome Consortium	17
1.1.5.2.1	Findings from Cancer Genomes	17
1.1.5.2.2	Genomic Comparisons Across Cancer Tissues	19
1.1.5.2.3	Cancer Genomic Data Resouces	20
1.1.6	Genomic Cancer Medicine	20
1.1.6.1	Cancer Genes and Driver Mutations	21
1.1.6.2	Personalised or Precision Cancer Medicine	22
1.1.6.2.1	Molecular Diagnostics and Pan-Cancer Medicine	22
1.1.6.3	Targeted Therapeutics and Pharmacogenomics	23
1.1.6.3.1	Targeting Oncogenic Driver Mutations	23
1.1.6.4	Systems and Network Biology	24
1.1.6.4.1	Network Medicine, and Polypharmacology	27
1.2	A Synthetic Lethal Approach to Cancer Medicine	28
1.2.1	Synthetic Lethal Genetic Interactions	28
1.2.2	Synthetic Lethal Concepts in Genetics	29
1.2.3	Studies of Synthetic Lethality	30
1.2.3.1	Synthetic Lethal Pathways and Networks	30
1.2.3.1.1	Evolution of Synthetic Lethality	31
1.2.4	Synthetic Lethal Concepts in Cancer	32
1.2.5	Clinical Impact of Synthetic Lethality in Cancer	33
1.2.6	High-throughput Screening for Synthetic Lethality	35
1.2.6.1	Synthetic Lethal Screens	37

1.2.7	Computational Prediction of Synthetic Lethality	40
1.2.7.1	Bioinformatics Approaches to Genetic Interactions . .	40
1.2.7.2	Comparative Genomics	41
1.2.7.3	Analysis and Modelling of Protein Data	44
1.2.7.4	Differential Gene Expression	46
1.2.7.5	Data Mining and Machine Learning	47
1.2.7.6	Bimodality	50
1.2.7.7	Rationale for Further Development	50
1.3	E-cadherin as a Synthetic Lethal Target	51
1.3.1	The <i>CDH1</i> gene and it's Biological Functions	51
1.3.1.1	Cytoskeleton	51
1.3.1.2	Extracellular and Tumour Micro-Environment	52
1.3.1.3	Cell-Cell Adhesion and Signalling	52
1.3.2	<i>CDH1</i> as a Tumour (and Invasion) Suppressor	52
1.3.2.1	Breast Cancers and Invasion	53
1.3.3	Hereditary Diffuse Gastric Cancer and Lobular Breast Cancer .	53
1.3.4	Somatic Mutations	54
1.3.4.1	Mutation Rate	54
1.3.4.2	Co-occurring Mutations	55
1.3.5	Models of <i>CDH1</i> loss in cell lines	56
1.4	Summary and Research Direction of Thesis	56
2	Methods and Resources	61
2.1	Bioinformatics Resources for Genomics Research	61
2.1.1	Public Data and Software Packages	61
2.1.1.1	Cancer Genome Atlas Data	62
2.1.1.2	Reactome and Annotation Data	63
2.2	Data Handling	63
2.2.1	Normalisation	63
2.2.2	Sample Triage	65
2.2.3	Metagenes and the Singular Value Decomposition	65
2.2.3.1	Candidate Triage and Integration with Screen Data . .	66
2.3	Techniques	67
2.3.1	Statistical Procedures and Tests	67
2.3.2	Gene Set Over-representation Analysis	68
2.3.3	Clustering	68
2.3.4	Heatmap	68
2.3.5	Modeling and Simulations	69
2.3.5.1	Receiver Operating Characteristic (Performance) . . .	70
2.3.6	Resampling Analysis	70
2.4	Pathway Structure Methods	71
2.4.1	Network and Graph Analysis	71
2.4.2	Sourcing Graph Structure Data	72
2.4.3	Constructing Pathway Subgraphs	72
2.4.4	Network Analysis Metrics	72
2.5	Implementation	74

2.5.1	Computational Resources and Linux Utilities	74
2.5.2	R Language and Packages	74
2.5.3	High Performance and Parallel Computing	77
3	Methods Developed During Thesis	79
3.1	A Synthetic Lethal Detection Methodology	79
3.2	Synthetic Lethal Simulation and Modelling	81
3.2.1	A Model of Synthetic Lethality in Expression Data	82
3.2.2	Simulation Procedure	86
3.3	Detecting Simulated Synthetic Lethal Partners	87
3.3.1	Binomial Simulation of Synthetic lethality	89
3.3.2	Multivariate Normal Simulation of Synthetic lethality	91
3.3.2.1	Multivariate Normal Simulation with Correlated Genes	93
3.3.2.2	Specificity with Query-Correlated Pathways	99
3.3.2.2.1	Importance of Directional Testing	101
3.4	Graph Structure Methods	103
3.4.1	Upstream and Downstream Gene Detection	103
3.4.1.1	Permutation Analysis for Statistical Significance	104
3.4.1.2	Ranking Based on Biological Context	104
3.4.2	Simulating Gene Expression from Graph Structures	105
3.5	Customised Functions and Packages Developed	110
3.5.1	Synthetic Lethal Interaction Prediction Tool	110
3.5.2	Data Visualisation	111
3.5.3	Extensions to the iGraph Package	111
3.5.3.1	Sampling Simulated Data from Graph Structures	113
3.5.3.2	Plotting Directed Graph Structures	113
3.5.3.3	Computing Information Centrality	113
3.5.3.4	Testing Pathway Structure with Permutation Testing	114
3.5.3.5	Metapackage to Install iGraph Functions	114
4	Synthetic Lethal Analysis of Gene Expression Data	115
4.1	Synthetic lethal genes in breast cancer	117
4.1.1	Synthetic lethal pathways in breast cancer	119
4.1.2	Expression profiles of synthetic lethal partners	119
4.1.2.1	Subgroup pathway analysis	120
4.2	Comparison of synthetic lethal gene candidates	122
4.2.1	Comparison with differential expression	122
4.2.2	Comparison with primary siRNA screen candidates	122
4.2.2.1	Comparison with correlation	125
4.2.2.2	Comparison with viability	126
4.2.2.3	Comparison of screen at pathway level	126
4.2.2.3.1	Resampling of genes for pathway enrichment	127
4.2.3	Comparison with secondary screen siRNA screen candidates	131
4.2.3.1	Comparison of candidate SL Pathways	131
4.3	Mutation, Copy Number, and Methylation	131
4.3.1	Synthetic lethality by DNA copy number	133

4.3.2	Synthetic lethality by somatic mutation	133
4.3.2.1	Mutation analysis	133
4.3.3	ANOVA of Expression Predictors	133
4.4	Global Synthetic Lethality	134
4.4.1	Hub Genes	134
4.5	Metagene Analysis	134
4.5.1	Pathway expression	134
4.5.2	Somatic mutation	134
4.5.3	Synthetic lethal metagenes	134
4.6	Replication in stomach cancer	134
4.7	Replication in cell line encyclopaedia	135
4.8	Summary	135
5	Synthetic Lethal Pathway Structure	142
5.1	Reactome Network structure and Information Centrality as a measure of gene essentiality	143
5.2	Synthetic lethal genes in synthetic lethal pathways	143
5.3	Centrality and connectivity of synthetic lethal genes	143
5.4	Upstream or downstream synthetic lethal candidates	143
5.5	Hierachical approach	143
5.6	Discussion	143
5.7	Conclusion	143
6	Simulation and Modeling of Synthetic Lethal Pathways	144
6.1	Simulations and Modelling Synthetic Lethality in Expression Data . . .	146
6.2	Simulations over simple graph structures	147
6.2.1	Performance	147
6.2.2	Synthetic lethality across graph stuctures	147
6.2.3	Performance with inhibition links	147
6.2.4	Performance with 20,000 genes	147
6.3	Simulations over pathway-based graphs	147
6.4	Comparing methods	147
6.4.1	SLIPT and Chi-Squared	147
6.4.1.1	Correlated query genes	147
6.4.2	Correlation	147
6.4.3	Bimodality with BiSEp	147
7	Discussion	148
7.1	Significance	150
7.2	Future Directions	151
7.3	Conclusion	152
8	Conclusion	153
	References	154

A	Sample Quality	154
A.1	Sample Correlation	154
A.2	Replicate Samples in TCGA Breast	156
B	Software Used for Thesis	160
C	Secondary Screen Data	169
D	Mutation Analysis in Breast Cancer	171
D.1	Synthetic Lethal Genes and Pathways	171
D.2	Comparison to Primary Screen	175
D.3	Resampling Analysis	177
D.4	Metagene Analysis	179
D.5	Mutation Variation	180
D.5.1	Mutation Frequency	180
D.5.2	PI3K Mutation Expression	181
D.6	Compare SLIPT genes	186
E	Expression Analysis in Stomach Cancer	187
F	Mutation Analysis in Stomach Cancer	188

List of Figures

1.1	Synthetic genetic interactions	29
1.2	Synthetic lethality in cancer	33
2.1	Read count density	64
2.2	Read count sample mean	64
3.1	Framework for synthetic lethal prediction	80
3.2	Synthetic lethal prediction adapted for mutation	81
3.3	A model of synthetic lethal gene expression	83
3.4	Modeling synthetic lethal gene expression	84
3.5	Synthetic lethality with multiple genes	85
3.6	Simulating gene function	87
3.7	Simulating synthetic lethal gene function	88
3.8	Simulating synthetic lethal gene expression	88
3.9	Performance of binomial simulations	90
3.10	Comparison of statistical performance	90
3.11	Performance of multivariate normal simulations	92
3.12	Simulating expression with correlated gene blocks	95
3.13	Simulating expression with correlated gene blocks	96
3.14	Synthetic lethal prediction across simulations	97
3.15	Performance with correlations	98
3.16	Comparison of statistical performance with correlation structure	99
3.17	Performance with query correlations	100
3.18	Statistical evaluation of directional criteria	101
3.19	Performance of directional criteria	102
3.20	Simulated graph structures	106
3.21	Simulating expression from a graph structure	107
3.22	Simulating expression from graph structure with inhibitions	109
3.23	Demonstration of violin plots with custom features	112
3.24	Demonstration of annotated heatmap	112
3.25	Simulating graph structures	114
4.1	Synthetic lethal expression profiles of analysed samples	121
4.2	Comparison of SLIPT to siRNA	124
4.3	Compare exprSLIPT and siRNA genes with correlation	125
4.4	Compare exprSLIPT and siRNA genes with correlation	125
4.5	Compare exprSLIPT and siRNA genes with siRNA viability	126
4.6	Compare exprSLIPT and siRNA genes with siRNA viability	126
4.7	Compare exprSLIPT and siRNA genes with viability	127
4.8	Resampled intersection of SLIPT and siRNA candidates	129
4.9	Synthetic lethal expression profiles of stomach samples	142

4.10	Comparison of SLIPT in stomach to siRNA	143
A.1	Correlation profiles of removed samples	154
A.2	Correlation analysis and sample removal	155
A.3	Replicate excluded samples	156
A.4	Replicate samples with all remaining	157
A.5	Replicate samples with some excluded	158
A.5	Replicate samples with some excluded	159
D.1	Synthetic lethal expression profiles of analysed samples	173
D.2	Comparison of mtSLIPT to siRNA	175
D.3	Somatic mutation locus	180
D.4	Somatic mutation against expression	181
D.5	Somatic mutation against PI3K protein	182
D.6	Somatic mutation against AKT protein	183
D.7	Somatic mutation against PI3K metagene	184
D.8	Somatic mutation against PIK3CA metagene	185
D.9	Compare mtSLIPT and siRNA genes with correlation	186
D.10	Compare mtSLIPT and siRNA genes with siRNA viability	186
F.1	Synthetic lethal expression profiles of stomach samples	190
F.2	Comparison of mtSLIPT in stomach to siRNA	192

List of Tables

1.1	Methods for Predicting Genetic Interactions	40
1.2	Methods for Predicting Synthetic Lethality in Cancer	41
1.3	Methods used by Wu <i>et al.</i> (2014)	43
2.1	Excluded Samples by Batch and Clinical Characteristics.	65
2.2	Computers used during Thesis	74
2.3	Linux Utilities and Applications used during Thesis	75
2.4	R Installations used during Thesis	75
2.5	R Packages Developed during Thesis	76
2.6	R Packages used during Thesis	76
4.1	Candidate synthetic lethal genes against E-cadherin from SLIPT	118
4.2	Pathways for <i>CDH1</i> partners from SLIPT	119
4.3	Pathway composition for clusters of <i>CDH1</i> partners from SLIPT	123
4.4	Pathway composition for <i>CDH1</i> partners from SLIPT and siRNA screen- ing	128
4.5	Pathways for <i>CDH1</i> partners from SLIPT	130
4.6	Pathways for <i>CDH1</i> partners from SLIPT and siRNA primary screen .	132
4.7	Candidate synthetic lethal metagenes against <i>CDH1</i> from SLIPT	135
4.8	Candidate synthetic lethal genes against E-cadherin from SLIPT in stomach cancer	136
4.9	Pathways for <i>CDH1</i> partners from SLIPT in stomach cancer	137
4.10	Pathway composition for clusters of <i>CDH1</i> partners in stomach SLIPT	138
4.11	Pathway composition for <i>CDH1</i> partners from SLIPT and siRNA screen- ing	139
4.12	Pathways for <i>CDH1</i> partners from SLIPT in stomach cancer	140
4.13	Pathways for <i>CDH1</i> partners from SLIPT in stomach and siRNA screen	141
4.14	Candidate synthetic lethal metagenes against <i>CDH1</i> from SLIPT in stomach cancer	144
4.15	Candidate synthetic lethal genes against E-cadherin from SLIPT in CCLE	145
4.16	Pathways for <i>CDH1</i> partners from SLIPT in CCLE	146
4.17	Candidate synthetic lethal genes against E-cadherin from SLIPT in breast CCLE	147
4.18	Pathways for <i>CDH1</i> partners from SLIPT in breast CCLE	148
4.19	Candidate synthetic lethal genes against E-cadherin from SLIPT in stomach CCLE	149
4.20	Pathways for <i>CDH1</i> partners from SLIPT in stomach CCLE	150
B.1	R Packages used during Thesis	160
C.1	Comparing SLIPT genes against Secondary siRNA Screen in breast cancer	169

C.2	Comparing mtSLIPT genes against Secondary siRNA Screen in breast cancer	170
C.3	Comparing SLIPT genes against Secondary siRNA Screen in stomach cancer	170
D.1	Candidate synthetic lethal genes against E-cadherin from mtSLIPT . .	171
D.2	Pathways for <i>CDH1</i> partners from mtSLIPT	172
D.3	Pathway composition for clusters of <i>CDH1</i> partners from mtSLIPT . .	174
D.4	Pathway composition for <i>CDH1</i> partners from mtSLIPT and siRNA . .	176
D.5	Pathways for <i>CDH1</i> partners from mtSLIPT	177
D.6	Pathways for <i>CDH1</i> partners from mtSLIPT and siRNA primary screen	178
D.7	Candidate synthetic lethal metagenes against <i>CDH1</i> from mtSLIPT . .	179
F.1	Candidate synthetic lethal genes against E-cadherin from mtSLIPT in stomach cancer	188
F.2	Pathways for <i>CDH1</i> partners from mtSLIPT in stomach cancer	189
F.3	Pathway composition for clusters of <i>CDH1</i> partners in stomach mtSLIPT	191
F.4	Pathway composition for <i>CDH1</i> partners from mtSLIPT and siRNA . .	193
F.5	Pathways for <i>CDH1</i> partners from mtSLIPT in stomach cancer	194
F.6	Pathways for <i>CDH1</i> partners from mtSLIPT in stomach and siRNA screen	195
F.7	Candidate synthetic lethal metagenes against <i>CDH1</i> from mtSLIPT in stomach cancer	196

Chapter 4

Synthetic Lethal Analysis of Gene Expression Data

Aims

- Pathway Structure of Candidate Synthetic Lethal Genes for *CDH1* from TCGA breast data
- Comparisons to Experimental siRNA Screen Candidates
- Replication of Pathways across in TCGA Stomach data

Summary

- We have developed a Synthetic Lethal detection method that generates a high number of synthetic lethal candidates
- Pathways in cell signalling, extracellular matrix, and cytoskeletal functions were supported with experimental candidates and the known functions of E-cadherin
- Several candidate pathways were supported by mutation analysis and replicated across breast and stomach cancer
- Translation and immune functions were uniquely detected by the computational approach which may be explained by differences between patient samples and cell line models
- There remains the need to identify actionable genes within these pathways, relationships with experimental candidates, and how these pathways may affect viability when lost

Having developed a statistical synthetic lethal detection methodology (SLIPT), it was applied to empirical (publicly available) cancer gene expression datasets in this chapter. The analysis largely focuses findings from the TCGA breast cancer data which covers a range of clinical subtypes and is more closely modelled by siRNA data (Telford *et al.*, 2015) generated from screening experiments conducted in MCF10A breast cells. Although stomach cancer data will also be considered to replicate findings in an independent dataset and for its relevance to syndromic hereditary diffuse gastric cancer. The TCGA data also has the advantages of other clinical and molecular profiles (e.g., somatic mutation and DNA copy number) for many of the same samples, in addition to a considerable sample size for RNASeq expression data, treated with a rigorous procedure to minimise batch effects. Some findings will be replicated in the Cancer Cell Line Encyclopaedia (CCLE) which may be more comparable to the cell line experiments.

Synthetic lethal candidate partners for *CDH1* will be described at both the gene and pathway level. SLIPT gene candidates will be 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, an alternative SLIPT methodology which uses mutation data for *CDH1* against expression of candidate partners will also be presented which may better represent the null mutations in HDGC patients and the experiment cell model (Chen *et al.*, 2014). Pathways will be analysed by over-representation analysis (with resampling for comparisons with siRNA data) and supported by a metagene analysis of pathway gene signatures. The pathway metagene expression profiles will be 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 will 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 will be described in both 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, the focus of these analysis will be in comparisons with experimental screening data to explore the potential for SLIPT to augment such triage of candidate partners and support further experimental inves-

tigations. 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.

Some of the findings presented in this Chapter have also been included in manuscripts submitted for publication (Kelly *et al.*, 2017a,b) and may bear similarity to them, although the results in this thesis have been edited to cohesively fit with additional findings (including consistent data versions). These findings are the result of investigations conducted throughout this thesis project and only these contributions to the articles are included in this chapter, not that conducted by co-authors.

4.1 Synthetic lethal genes in breast cancer

- exprSL
- mtSL
- heatmap

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 gene were among the highest 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 signaling 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 identify synthetic lethal pathways.

Table 4.1: Candidate synthetic lethal genes against E-cadherin from SLIPT

Gene	Observed	Expected	χ^2 value	p-value	p-value (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 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 loss of function mutations in *CDH1*. As shown in Table D.1, the most significant genes also had strong evidence of expression associated with *CDH1* mutations (high χ^2 values) with fewer samples exhibiting both low expression and mutations of each gene than expected statistically. Although, these were not as strongly supported as the expression analysis (in Table 4.1) nor were as many genes detected. This is unsurprising due to the lower sample size with matching somatic mutation data and the lower frequency of *CDH1* mutations compared to low expression 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 include 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

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 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, there were several large clusters of genes among the expression profiles of the *CDH1* synthetic lethal candidate partners, which suggests co-regulation of genes or pathway correlation. A number of candidates from an experimental RNAi screen study performed by Telford *et al.* Telford *et al.* (2015) were also identified by this approach. In addition, we identified novel gene candidates, which had little effect on 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 subtype, and “normal-like” samples Dai *et al.* (2015); Eroles *et al.* (2012); Parker *et al.* (2009) have elevation of genes specific to particular clusters which is 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.

Table 5. Gene set enrichment results for subgroups of *CDH1* SL partners shows functional variation.

Figure 3. Heatmap of RNASeq gene expression in predicted SL partners of *CDH1* showing distinct subgroups of SL partners and links between SL partner expression and clinical variables.

4.1.2.1 Subgroup pathway analysis

Synthetic lethal gene candidates for *CDH1* from SLIPT performed on RNA-Seq 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 including some by clinical factors such as estrogen receptor status, intrinsic (PAM50) subtype (Parker *et al.*, 2009), and somatic mutation (of highest impact genes) against these gene clusters.

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

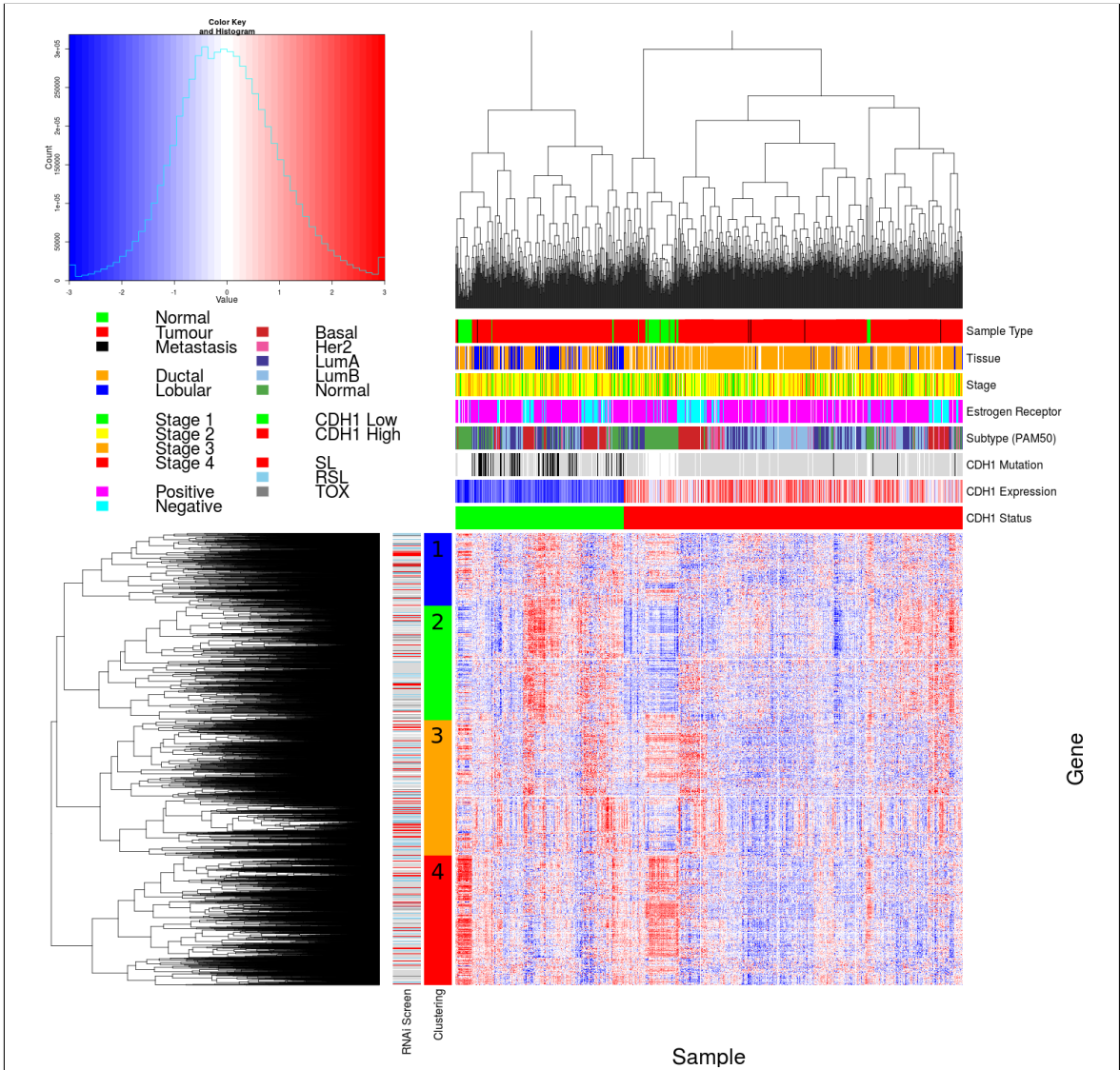


Figure 4.1: Synthetic lethal expression profiles of analysed samples. Gene expression profile heatmap (correlation distance) of all samples (separated by the $1/3$ quantile of *CDH1* expression) analysed in TCGA breast cancer dataset for gene expression of 4,629 candidate partners of E-cadherin (*CDH1*) from SLIPT prediction (with significant FDR adjusted $p < 0.05$). Deeply clustered, inter-correlated genes form several main groups, each containing genes that were SL candidates or toxic in an siRNA screen Telford *et al.* (2015). Clusters had different sample groups highly expressing the synthetic lethal candidates in *CDH1* low samples, notably 'normal-like', basal, 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.

processes. Similarly, Cluster 3 has over-representation of immune signalling pathways (including chemokines, secondary messenger, and TCR signaling) and downstream intracellular signalling cascades such as G protein coupled receptor (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*.

4.2 Comparison of synthetic lethal gene candidates

4.2.1 Comparison with differential expression

4.2.2 Comparison with primary siRNA screen candidates

Secondary screen results in Appendix C

Gene candidates were compared between computational and experimental screening 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. These intersecting genes could be useful in 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).

The overlap between synthetic lethal from bioinformatics SLIPT predictions and siRNA screening has raised other questions including whether the number of genes and pathways enriched would be expected by chance. This of particular concern since the siRNA candidate genes themselves are highly enriched for particular pathways so selecting any intersect with them would be enriched for these pathways. The siRNA data is also based on cell line models which have limitations in application to a genetically variable patient population with a complex tumour microenvironment interacting with immune cells. One approach is to compare the candidate genes is to exclude genes that were not tested in both systems, such as those not expressed in cell lines or those with more than $1/3$ of TCGA patients without any RNA Seq reads so the lowest quantile

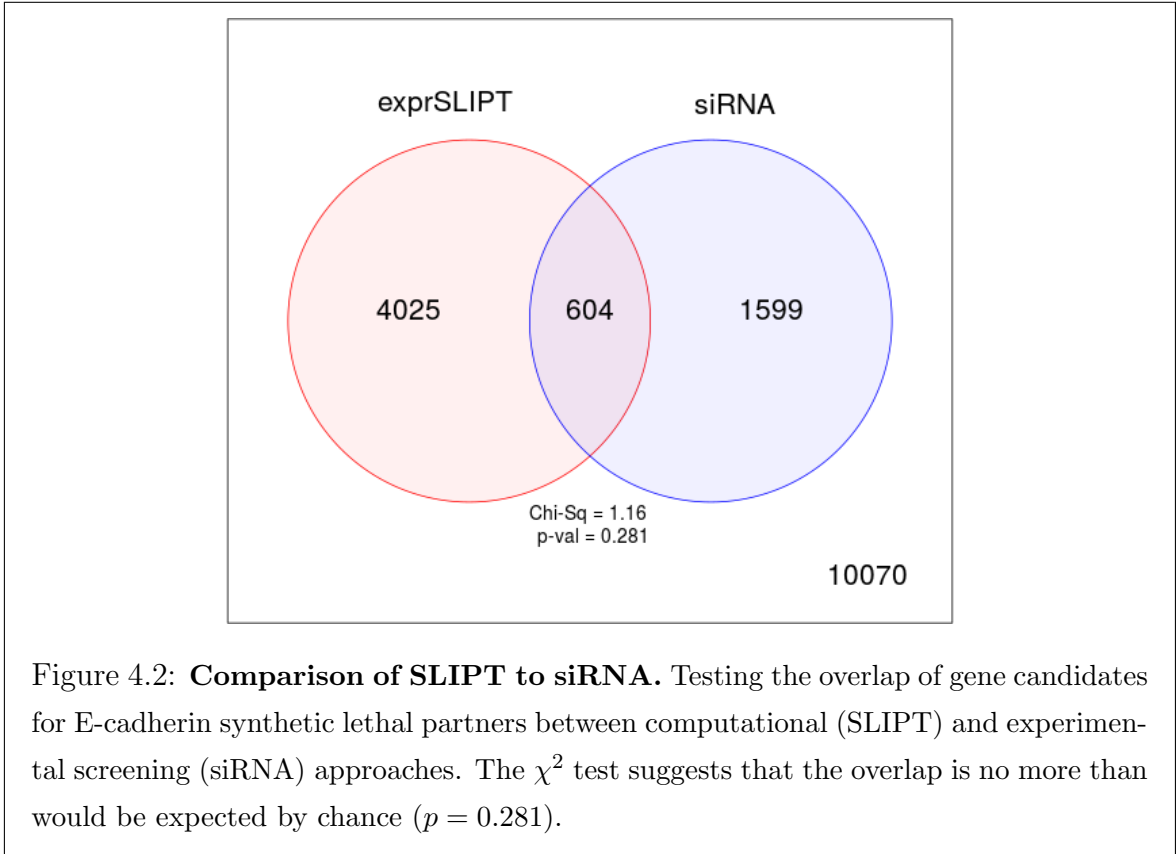
Table 4.3: Pathway composition for clusters of *CDH1* partners from SLIPT

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}
Signaling 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}
Smooth Muscle Contraction	28	4	4.4×10^{-6}
ECM proteoglycans	66	6	6.3×10^{-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}
LI3a-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}
Infectious disease	347	103	2.2×10^{-35}
Formation of the ternary complex, and subsequently, the 43S complex	47	33	6.8×10^{-80}

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 Signaling in Immune system	268	48	8.9×10^{-37}
Downstream TCR signalling	45	18	1.8×10^{-35}
G _α 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}
Extracellular matrix organisation	238	35	2.7×10^{-25}
Antigen activates B Cell Receptor leading to generation of second messengers	32	12	7.2×10^{-25}

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 _α 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 _α 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 _α 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}
Signalling by Retinoic Acid	42	13	6.7×10^{-24}
Degradation of the extracellular matrix	102	19	1.4×10^{-22}



cannot be defined for SLIPT analysis. Another approach is to test whether pathways are enriched in randomly sampled genes, comparing many resampled 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.

Both of these are being applied with developing a method and overcoming technical challenges for the latter being the focus of recent work. The main challenge at the moment is to compare SLIPT results to experimental candidates and explain why so few genes (and so many pathways) overlap.

As discussed above, comparing genes between experimental screen candidates and prediction from TCGA expression data has been difficult. Figure 3 summarises the approaches to comparing genes accounting for some of the differences between the datasets. Of particular concern are the over-represented pathways in genes detected by both methods. There is no statistical evidence that SLIPT predicted genes or siRNA candidates are enriched in with each other. The siRNA candidates themselves are over-represented with many pathways including GPCRs so any intersection with

these would contain some of these pathways. Whether these pathways are contained in the intersection more than expected by chance is the problem the two approaches below were designed to tackle.

4.2.2.1 Comparison with correlation

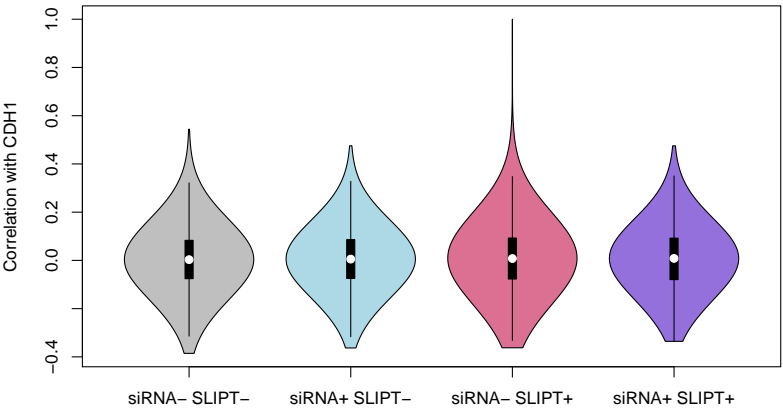


Figure 4.3: Compare exprSLIPT and siRNA genes with correlation. Caption, Caption, Caption...

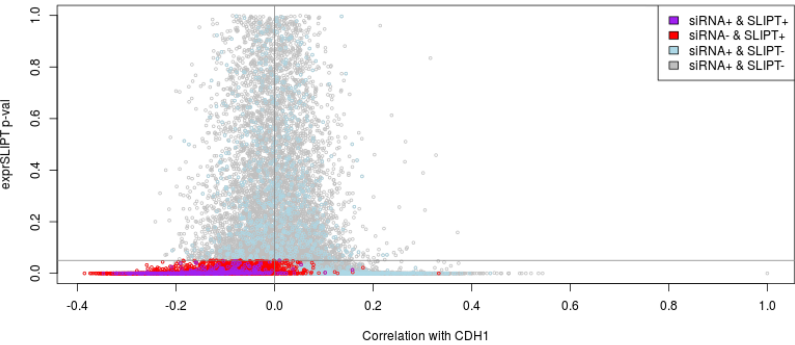


Figure 4.4: Compare exprSLIPT and siRNA genes with correlation. Caption, Caption, Caption...

4.2.2.2 Comparison with viability

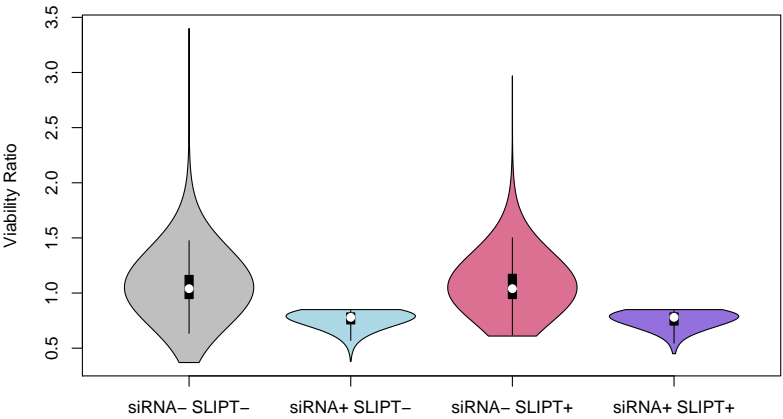


Figure 4.5: Compare exprSLIPT and siRNA genes with siRNA viability. Caption, Caption, Caption...

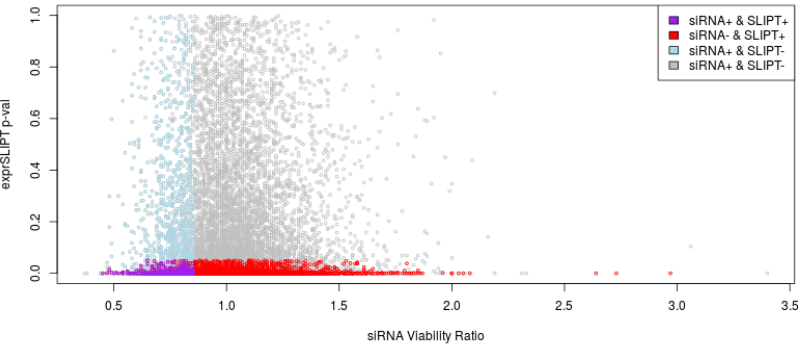


Figure 4.6: Compare exprSLIPT and siRNA genes with siRNA viability. Caption, Caption, Caption...

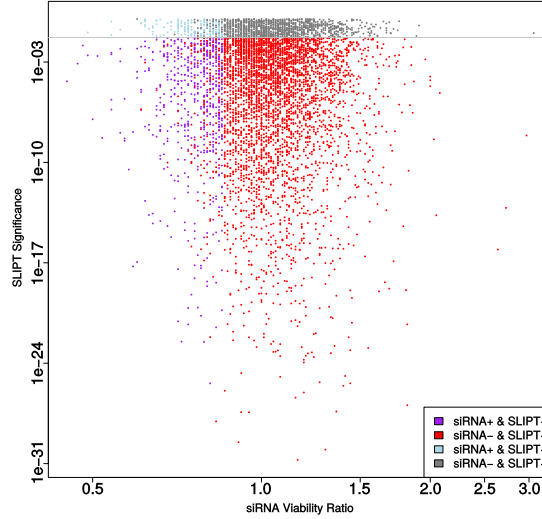


Figure 4.7: Compare exprSLIPT and siRNA genes with viability. Caption, Caption, Caption...

4.2.2.3 Comparison of screen at pathway level

These pathway analyses correspond to genes separated into SLIPT or siRNA screen candidates unique to either method or detected by both (Table 4.4). The SLIPT-specific gene candidates were involved most strongly with translational and immune regulatory pathways, which were also identified 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.* Telford *et al.* (2015).

The intersection of computational and experimental synthetic lethal partners of *CDH1* has 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

4.2.2.3.1 Resampling of genes for pathway enrichment

As shown in Figure 4.8, resampling did not find evidence of significant depletion or over-representation for experimental synthetic lethal candidates in the computationally predicted synthetic lethal partners of *CDH1*, which suggested that the overlap across the two methods was no better than by chance.

A permutation analysis was performed to resample the genes tested by both ap-

Table 4.4: Pathway composition for *CDH1* partners from SLIPT and siRNA screening

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}
DAPI2 interactions	159	23	2.6×10^{-19}
DAPI2 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}
DAPI2 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}

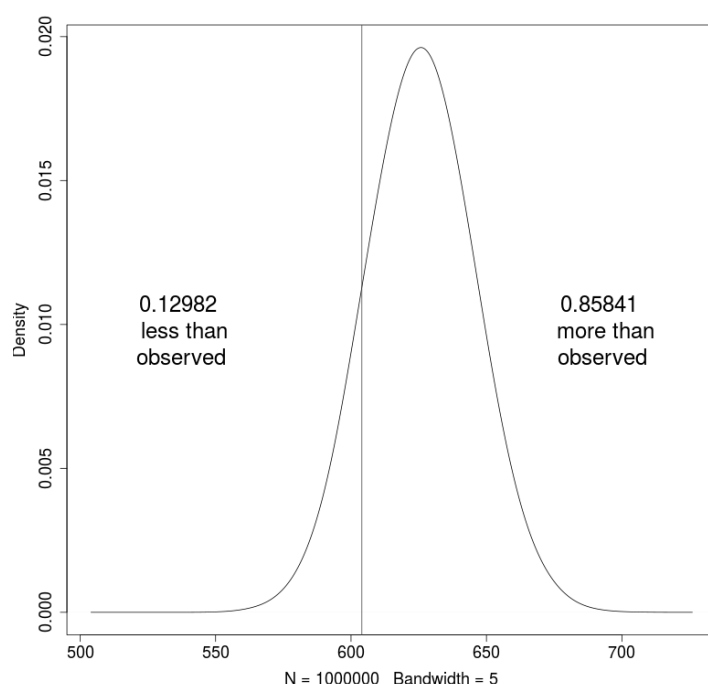


Figure 4.8: **Resampled intersection of SLIPT and siRNA candidates.** Re-sampling analysis of intersect size from genes detected by SLIPT and siRNA screening approaches over 1 million replicates. The proportion of expected intersection sizes for random samples below or above the observed intersection size respectively, lacking significant over-representation or depletion of siRNA screen candidates within the SLIPT predictions for *CDH1*.

proaches 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. While the number of siRNA candidate genes 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 4.4. The re-sampling analysis for pathways was compared to the pathway over-representation for SLIPT predicted synthetic lethal partners in Table 4.5. Similarly, the pathway resampling for intersection between SLIPT predictions and experimental screen candidates was compared to pathway over-representation in Table 4.6 for intersection with siRNA

Table 4.5: 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}	$< 1.241 \times 10^{-5}$
Extracellular matrix organisation	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
GPCR downstream signalling	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$_{\alpha i}$ 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).

data.

The pathway resampling approach for SLIPT-specific gene candidates (Table 4.5) 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 4.5. In particular, some of the most significantly over-represented pathways had higher observed χ^2 values than any of the 1 million random permutations.

The intersection between computational and experimental candidates (in Table 4.6) 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, although pathways involving

defective *EXT1* or *EXT2* genes approach significance after FDR adjustment for multiple tests. 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 protein modification pathways, energy metabolism, and the fibrin clotting cascade.

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, we note that several pathways, including some immune functions and neurotransmitters, were supported by the resampling analysis (in Table 4.6) 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.

4.2.3 Comparison with secondary screen siRNA screen candidates

4.2.3.1 Comparison of candidate SL Pathways

Thus we have identified candidate synthetic lethal pathways by gene set over-representation, metagene synthetic lethality, and re-sampled empirical pathway over-representation. The challenge currently under consideration is whether these methods can be compared and which may lead to biologically meaningful or clinically relevant synthetic lethal candidate pathways.

4.3 Mutation, Copy Number, and Methylation

Due to promising synthetic lethal data on mutation and DNA copy number analyses (Jerby-Arnon *et al.*, 2014; Lu *et al.*, 2015), these were also investigated to compare genes for synthetic lethality in an analogous manner to expression analyses in the TCGA data. Due to the low somatic mutation rate (and lack of available) germline mutations for many genes, it was not possible to detect many double mutations with significantly under-representation in cancers. There were also concerns about using rare mutations

Table 4.6: 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).

with unknown significance or excluding functional mutations by only using those in the exons. It was possible to compare deletion and duplication of DNA copy number in a manner analogous to expression quantiles. However, these overlapped poorly with candidate interacting partners from expression analyses and concerns were raised that they may not be relevant to *CDH1* which is typically inactivated in tumours by loss of function mutations or DNA methylation (PJ Guilford, personal communication).

DNA methylation data was also prepared for synthetic lethal analysis but was discontinued due to computational challenges, expected similarity to expression results, difficulty defining loss of function methylation at a gene level across CpG sites, and the concerns raised in the next section.

4.3.1 Synthetic lethality by DNA copy number

4.3.2 Synthetic lethality by somatic mutation

4.3.2.1 Mutation analysis

4.3.3 ANOVA of Expression Predictors

[include?]

Another approach was to only use copy number, mutation, or hyper-methylation data for genes in which they would impact on gene function and occur frequently in tumours. Before investigating whether these impact on gene function, they were investigated as predictors of variation in gene expression. If these are not giving variation independent of gene expression, expression would be a more suitable measure of gene function as it is widely generated in studies and useful as a clinical biomarker.

Globally predicting gene expression across all genes from DNA copy number and somatic mutation was attempted by ANOVA. However, this was computationally challenging and gene-specific analyses would be more informative. Gene specific ANOVA and linear regression was performed but was raised more issues than it addressed. There were issues with interaction terms and mutation data, many genes were not tested for these since there were so few mutations for these genes in the dataset. It was possible to include DNA methylation in gene-specific analyses (despite the concerns raised above) but the R^2 values for each gene were still generally very low and issues with insufficient mutant samples for interaction terms became worse. This means that the approach used differs for each gene making it difficult to compare them. The challenges raised here suggested that expression is very difficult to predict with other factors but including these other factors would be difficult and plagued by multiple-testing, particularly

comparing between them with the current synthetic lethal prediction method. This led to investigations into the simulation of synthetic lethality.

4.4 Global Synthetic Lethality

[include?]

Global levels of synthetic lethality were analysed as part of my Honours project (Kelly, 2013) to address concerns of high numbers of synthetic lethal candidates for *CDH1*. This turned out to be typical for most genes in the microarray dataset. Due to newer samples and concerns about sample quality in TCGA microarrays, RNA-Seq datasets were used here. The focus of this thesis is gene expression data generated by RNA-Seq, this was replicated using the TCGA breast cancer RNA-Seq dataset on the New Zealand eScience Infrastructure Intel Pan supercomputer.

4.4.1 Hub Genes

Table 1. Hub gene function in TCGA breast cancer microarray expression SL predictions (n=600).

Table 2 Hub gene function in TCGA breast cancer RNA-Seq expression SL predictions (n=878). [revise for n=1168]

Table 3. Hub gene function in BC2116 breast cancer microarray expression SL predictions (n=2116).

4.5 Metagene Analysis

[include?]

4.5.1 Pathway expression

4.5.2 Somatic mutation

4.5.3 Synthetic lethal metagenes

4.6 Replication in stomach cancer

- exprSL
- mtSL
- heatmap
- Venn

Table 4.7: Candidate synthetic lethal metagenes against *CDH1* from SLIPT

Pathway	ID	Observed	Expected	χ^2 value	p-value	p-value (FDR)
Activation of BMF and translocation to mitochondria	139910	213	130.22	205.32	2.6909×10^{-43}	4.4373×10^{-40}
Downregulation of ERBB2:ERBB3 signaling	1358803	197	130.22	189.57	6.5577×10^{-40}	5.4069×10^{-370}
Activation of PKB	165158	209	130.22	188.57	1.0771×10^{-39}	5.9203×10^{-370}
Glycogen storage diseases	3229121	68	130.22	175.58	6.6178×10^{-37}	1.8188×10^{-340}
Myoclonic epilepsy of Lafora	3785653	68	130.22	175.58	6.6178×10^{-37}	1.8188×10^{-340}
Diseases of carbohydrate metabolism	5663084	68	130.22	175.58	6.6178×10^{-37}	1.8188×10^{-340}
HSF1 activation	3371511	212	130.22	171.21	5.7399×10^{-36}	1.3522×10^{-330}
Downregulation of ERBB4 signaling	1253288	192	130.22	161.77	6.0875×10^{-34}	1.2548×10^{-310}
Arachidonic acid metabolism	2142753	81	130.22	156.53	8.1254×10^{-33}	1.4888×10^{-300}
Translation initiation complex formation	72649	70	130.22	152.14	7.0837×10^{-32}	1.1681×10^{-290}
Synthesis of 5-eicosatetraenoic acids	2142688	68	130.22	150.98	1.2533×10^{-31}	1.8787×10^{-290}
SRP-dependent cotranslational protein targeting to membrane	1799339	69	130.22	150.03	2.0095×10^{-31}	2.7613×10^{-290}
L13a-mediated translational silencing of Ceruloplasmin expression	156827	72	130.22	147.84	5.9094×10^{-31}	6.4389×10^{-290}
3' -UTR-mediated translational regulation	157279	72	130.22	147.84	5.9094×10^{-31}	6.4389×10^{-290}
Trafficking of AMPA receptors	399719	198	130.22	147.73	6.2476×10^{-31}	6.4389×10^{-290}
Glutamate Binding, Activation of AMPA Receptors and Synaptic Plasticity	399721	198	130.22	147.73	6.2476×10^{-31}	6.4389×10^{-290}
Scavenging by Class F Receptors	3000484	202	130.22	146.85	9.6215×10^{-31}	9.2823×10^{-290}
Activation of the mRNA upon binding of the cap-binding complex and eIFs, and subsequent binding to 43S	72662	70	130.22	146.51	1.1365×10^{-30}	9.2823×10^{-290}
Formation of the ternary complex, and subsequently, the 43S complex	72695	70	130.22	146.51	1.1365×10^{-30}	9.2823×10^{-290}
Ribosomal scanning and start codon recognition	72702	70	130.22	146.51	1.1365×10^{-30}	9.2823×10^{-290}
Eukaryotic Translation Elongation	156842	72	130.22	146.42	1.192×10^{-30}	9.2823×10^{-290}
Nonsense Mediated Decay independent of the Exon Junction Complex	975956	71	130.22	146.34	1.2384×10^{-30}	9.2823×10^{-290}
Viral mRNA Translation	192823	70	130.22	145.93	1.5135×10^{-30}	1.0399×10^{-280}
Eukaryotic Translation Termination	72764	70	130.22	145.93	1.5135×10^{-30}	1.0399×10^{-280}
NF-kB is activated and signals survival	209560	71	130.22	145.48	1.8975×10^{-30}	1.1857×10^{-280}

Strongest candidate SL partners for *CDH1* by SLIPT with observed and expected samples with low expression of both genes

- Pathway enrichment
- Permutations

4.7 Replication in cell line encyclopaedia

As breast cancer cell lines are the experimental system in which many cancer genetics and drug targets are investigated, these were analysed in addition to patient samples from TCGA. The cancer cell line encyclopaedia (CCLE) is a resource for genomics profiles across a range of cell lines. These have also been used to generate synthetic lethal candidates for comparison to those in experimental screen and predictions from TCGA expression data. A transcriptome experiment has been conducted by the Cancer Genetics Laboratory to test their *CDH1*^{-/-} null MCF10A cell lines compared to an otherwise isogenic wildtype (Chen *et al.*, 2014). While differential expression analysis was inconclusive due to few technical replicates, this data was also useful to determine genes which were not detectable in MCF10A cell lines which would not be expected to detect synthetic lethality in siRNA screen data even if they were predicted to be synthetic lethal in expression data.

Table 4.8: Candidate synthetic lethal genes against E-cadherin from SLIPT in stomach cancer

Gene	Observed	Expected	χ^2 value	p-value	p-value (FDR)
<i>PRAF2</i>	17	50.4	121	3.54×10^{-25}	1.45×10^{-21}
<i>EMP3</i>	17	50.4	115	5.06×10^{-24}	1.48×10^{-20}
<i>PLEKHO1</i>	22	50.4	112	2.14×10^{-23}	4.75×10^{-20}
<i>SELM</i>	20	50.4	111	5.13×10^{-23}	8.09×10^{-20}
<i>GYPC</i>	20	50.4	110	5.77×10^{-23}	8.45×10^{-20}
<i>COX7A1</i>	18	50.4	109	1.15×10^{-22}	1.39×10^{-19}
<i>TNFSF12</i>	20	50.4	106	4.06×10^{-22}	4.38×10^{-19}
<i>SEPT4</i>	17	50.4	106	6.58×10^{-22}	5.91×10^{-19}
<i>LGALS1</i>	19	50.4	105	6.64×10^{-22}	5.91×10^{-19}
<i>RARRES2</i>	27	50.4	105	8.02×10^{-22}	6.85×10^{-19}
<i>VEGFB</i>	16	50.4	104	1.19×10^{-21}	9.74×10^{-19}
<i>PRR24</i>	22	50.4	102	2.96×10^{-21}	2.02×10^{-18}
<i>SYNC</i>	19	50.4	102	3.73×10^{-21}	2.39×10^{-18}
<i>MAGEH1</i>	17	50.4	100	9.52×10^{-21}	5.01×10^{-18}
<i>HSPB2</i>	23	50.4	99.6	1.19×10^{-20}	5.82×10^{-18}
<i>SMARCD3</i>	19	50.4	99	1.59×10^{-20}	7.57×10^{-18}
<i>CREM</i>	13	50.4	98.1	2.48×10^{-20}	1.13×10^{-17}
<i>GNG11</i>	20	50.4	97.3	3.68×10^{-20}	1.59×10^{-17}
<i>GNAI2</i>	17	50.4	96.4	5.75×10^{-20}	2.36×10^{-17}
<i>FUNDC2</i>	22	50.4	95.9	7.39×10^{-20}	2.91×10^{-17}
<i>CNRIP1</i>	21	50.4	95.3	1×10^{-19}	3.66×10^{-17}
<i>CALHM2</i>	22	50.4	93.1	2.94×10^{-19}	1.06×10^{-16}
<i>ARID5A</i>	18	50.4	92.7	3.47×10^{-19}	1.22×10^{-16}
<i>ST3GAL3</i>	27	50.4	92.2	4.49×10^{-19}	1.56×10^{-16}
<i>LOC339524</i>	21	50.4	92.1	4.8×10^{-19}	1.59×10^{-16}

Strongest candidate SL partners for *CDH1* by SLIPT with observed and expected samples with low expression of both genes

4.8 Summary

We have developed a simple, interpretable, computational approach to predict synthetic lethal partners from genomics data. Originally developed for microarray gene expression data, it has been expanded to test DNA copy number, or RNA-Seq gene expression data which are both also supported by the TCGA dataset. DNA copy number was included for comparison with the DAISY tool of Jerby-Arnon *et al.* (2014). Predictions based on microarray data were inconclusive when compared with an RNAi screen for *CDH1* in MCF10A breast cells as performed by Telford *et al.* (2015), few predictions replicated between BC2116, CCLE, or TCGA microarray datasets, results

Table 4.9: Pathways for *CDH1* partners from SLIPT in stomach cancer

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

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

with gene expression and DNA copy number were vastly different, and predictions from TCGA microarray and RNA-Seq datasets for the same samples differed were inconsistent. The Aligent TCGA microarray data in particular is difficult to compare to other datasets and will in the future use Affymetrix microarrays or RNA-Seq platforms for predictions from gene expression data. The analyses focus on gene expression data as it is widely available for applications in other cancers and current attempts to use gene expression data for synthetic lethal discovery vary widely (Jerby-Arnon *et al.*, 2014; Lu *et al.*, 2015; Tiong *et al.*, 2014). There is no consensus for which approach is more appropriate since they lack much a basis on biological experimental data or statistical modelling and often use difficult to interpret machine learning methodology.

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 SGA and RNAi screens for synthetic lethality are also error-prone, especially with false-positives, raising the need for understanding the expected behaviour and number of functional relationships and genetic interactions in the genome, or in discovery of synthetic lethal partners of a particular query gene. A characteristic of gene interaction networks is a scale-free topology lead-

Table 4.10: Pathway composition for clusters of *CDH1* partners in stomach SLIPT

Pathways Over-represented in Cluster 1	Pathway Size	Cluster Genes	p-value (FDR)
Viral mRNA Translation	82	48	1.3×10^{-97}
Formation of a pool of free 40S subunits	94	51	1.3×10^{-97}
Eukaryotic Translation Elongation	87	49	4.8×10^{-97}
Peptide chain elongation	84	48	1.4×10^{-96}
Eukaryotic Translation Termination	84	48	1.4×10^{-96}
GTP hydrolysis and joining of the 60S ribosomal subunit	105	52	7.9×10^{-94}
Nonsense Mediated Decay independent of the Exon Junction Complex	89	48	3.1×10^{-93}
L13a-mediated translational silencing of Ceruloplasmin expression	104	51	5.1×10^{-92}
3' -UTR-mediated translational regulation	104	51	5.1×10^{-92}
SRP-dependent cotranslational protein targeting to membrane	105	51	1.7×10^{-91}
Eukaryotic Translation Initiation	112	52	3.3×10^{-90}
Cap-dependent Translation Initiation	112	52	3.3×10^{-90}
Translation	142	56	3.6×10^{-85}
Nonsense-Mediated Decay	104	48	1.2×10^{-84}
Nonsense Mediated Decay enhanced by the Exon Junction Complex	104	48	1.2×10^{-84}
Influenza Viral RNA Transcription and Replication	109	48	4.1×10^{-82}
Influenza Life Cycle	113	48	3.4×10^{-80}
Influenza Infection	118	48	6.4×10^{-78}
Infectious disease	349	68	1.8×10^{-50}
Formation of the ternary complex, and subsequently, the 43S complex	48	21	3.7×10^{-43}

Pathways Over-represented in Cluster 2	Pathway Size	Cluster Genes	p-value (FDR)
Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	65	12	1.3×10^{-15}
Phosphorylation of CD3 and TCR zeta chains	18	6	1.7×10^{-12}
Generation of second messenger molecules	29	7	2.7×10^{-12}
PD-1 signaling	21	6	7.4×10^{-12}
TCR signaling	62	9	4.3×10^{-11}
Translocation of ZAP-70 to Immunological synapse	16	5	1.1×10^{-10}
Interferon alpha/beta signaling	68	9	1.6×10^{-10}
Initial triggering of complement	17	5	1.6×10^{-10}
IKK complex recruitment mediated by RIP1	19	5	5.1×10^{-10}
TRIF-mediated programmed cell death	10	4	6.2×10^{-10}
Creation of C4 and C2 activators	11	4	1.3×10^{-9}
RHO GTPases Activate NADPH Oxidases	11	4	1.3×10^{-9}
Interferon Signaling	175	15	2.3×10^{-9}
Chemokine receptors bind chemokines	52	7	4.0×10^{-9}
Interferon gamma signaling	74	8	1.6×10^{-8}
TRAF6 mediated induction of TAK1 complex	15	4	1.6×10^{-8}
Activation of IRF3/IRF7 mediated by TBK1/IKK epsilon	16	4	2.7×10^{-8}
Downstream TCR signaling	45	6	3.5×10^{-8}
Ligand-dependent caspase activation	17	4	4.2×10^{-8}
Complement cascade	34	5	1.3×10^{-7}

Pathways Over-represented in Cluster 3	Pathway Size	Cluster Genes	p-value (FDR)
Uptake and actions of bacterial toxins	22	4	3.5×10^{-6}
Neurotoxicity of clostridium toxins	10	3	3.5×10^{-6}
Activation of PPARGC1A (PGC-1alpha) by phosphorylation	10	3	3.5×10^{-6}
SMAD2/SMAD3/SMAD4 heterotrimer regulates transcription	28	4	1.4×10^{-5}
Assembly of the primary cilium	149	10	2.5×10^{-5}
Serotonin Neurotransmitter Release Cycle	15	3	2.5×10^{-5}
Glycosaminoglycan metabolism	114	8	3.3×10^{-5}
Platelet homeostasis	54	5	3.3×10^{-5}
Norepinephrine Neurotransmitter Release Cycle	17	3	3.3×10^{-5}
Acetylcholine Neurotransmitter Release Cycle	17	3	3.3×10^{-5}
Gos signalling events	100	7	5.5×10^{-5}
GABA synthesis, release, reuptake and degradation	19	3	5.6×10^{-5}
deactivation of the beta-catenin transactivating complex	39	4	6.7×10^{-5}
Dopamine Neurotransmitter Release Cycle	20	3	6.7×10^{-5}
IRS-related events triggered by IGF1R	83	6	7.1×10^{-5}
Generic Transcription Pathway	186	11	7.1×10^{-5}
Termination of O-glycan biosynthesis	21	3	7.4×10^{-5}
Kinesins	22	3	8.5×10^{-5}
Signaling by Type 1 Insulin-like Growth Factor 1 Receptor (IGF1R)	86	6	8.5×10^{-5}
IGF1R signaling cascade	86	6	8.5×10^{-5}

Pathways Over-represented in Cluster 4	Pathway Size	Cluster Genes	p-value (FDR)
Extracellular matrix organization	241	97	8.8×10^{-126}
Axon guidance	289	75	8.3×10^{-72}
Hemostasis	445	101	8.3×10^{-72}
Developmental Biology	432	95	3.0×10^{-67}
Response to elevated platelet cytosolic Ca^{2+}	84	37	5.8×10^{-67}
Platelet degranulation	79	36	5.8×10^{-67}
Degradation of the extracellular matrix	104	39	6.7×10^{-63}
Platelet activation, signaling and aggregation	186	52	6.6×10^{-62}
ECM proteoglycans	66	31	8.1×10^{-61}
Neuronal System	272	64	5.1×10^{-60}
Signaling by PDGF	173	47	9.7×10^{-57}
Integrin cell surface interactions	82	31	1.9×10^{-53}
Collagen biosynthesis and modifying enzymes	56	26	1.1×10^{-52}
Collagen formation	67	28	1.4×10^{-52}
Class A/1 (Rhodopsin-like receptors)	289	61	2.3×10^{-52}
GPCR ligand binding	373	73	2.8×10^{-52}
Elastic fibre formation	38	22	4.7×10^{-52}
Non-integrin membrane-ECM interactions	53	24	7.0×10^{-49}
Glycosaminoglycan metabolism	114	33	4.7×10^{-47}
Platelet homeostasis	54	23	1.0×10^{-45}

Table 4.11: Pathway composition for *CDH1* partners from SLIPT and siRNA screening

Predicted only by SLIPT (3392 genes)	Pathway Size	Genes Identified	p-value (FDR)
Eukaryotic Translation Elongation	87	76	3.5×10^{-187}
Peptide chain elongation	84	73	1.6×10^{-180}
Eukaryotic Translation Termination	84	72	1.1×10^{-176}
Viral mRNA Translation	82	71	3.6×10^{-176}
Formation of a pool of free 40S subunits	94	75	3.1×10^{-173}
Nonsense Mediated Decay independent of the Exon Junction Complex	89	72	2.4×10^{-169}
L13a-mediated translational silencing of Ceruloplasmin expression	104	76	1.8×10^{-164}
3' -UTR-mediated translational regulation	104	76	1.8×10^{-164}
GTP hydrolysis and joining of the 60S ribosomal subunit	105	76	2×10^{-163}
Nonsense-Mediated Decay	104	75	2.4×10^{-161}
Nonsense Mediated Decay enhanced by the Exon Junction Complex	104	75	2.4×10^{-161}
SRP-dependent cotranslational protein targeting to membrane	105	74	4.2×10^{-157}
Eukaryotic Translation Initiation	112	76	2.4×10^{-156}
Cap-dependent Translation Initiation	112	76	2.4×10^{-156}
Translation	142	85	3.5×10^{-156}
Influenza Infection	118	75	6.8×10^{-148}
Influenza Viral RNA Transcription and Replication	109	72	4.2×10^{-147}
Infectious disease	349	131	7.9×10^{-145}
Influenza Life Cycle	113	72	1.5×10^{-143}
Adaptive Immune System	418	144	1.6×10^{-140}

Detected only by siRNA screen (1803 genes)	Pathway Size	Genes Identified	p-value (FDR)
Class A/1 (Rhodopsin-like receptors)	282	58	1.5×10^{-44}
GPCR ligand binding	363	66	2×10^{-40}
G _{ai} signalling events	184	36	4.2×10^{-33}
Peptide ligand-binding receptors	175	33	1.8×10^{-30}
G _{aq} signalling events	159	29	1.9×10^{-27}
Gastrin-CREB signalling pathway via PKC and MAPK	180	30	1.3×10^{-25}
Olfactory Signaling Pathway	348	46	1.6×10^{-22}
Downstream signal transduction	146	24	2.1×10^{-22}
Signaling by PDGF	172	26	1.5×10^{-21}
Signaling by ERBB2	146	23	4.6×10^{-21}
DAP12 interactions	159	24	1.0×10^{-20}
DAP12 signaling	149	23	1.0×10^{-20}
Downstream signaling of activated FGFR1	134	21	4.3×10^{-20}
Downstream signaling of activated FGFR2	134	21	4.3×10^{-20}
Downstream signaling of activated FGFR3	134	21	4.3×10^{-20}
Downstream signaling of activated FGFR4	134	21	4.3×10^{-20}
Signalling by NGF	266	34	5.3×10^{-20}
Signaling by FGFR	146	22	5.3×10^{-20}
Signaling by FGFR1	146	22	5.3×10^{-20}
Signaling by FGFR2	146	22	5.3×10^{-20}

Intersection of SLIPT and siRNA screen (547 genes)	Pathway Size	Genes Identified	p-value (FDR)
Chemokine receptors bind chemokines	52	11	5.2×10^{-16}
Class A/1 (Rhodopsin-like receptors)	289	29	6.4×10^{-14}
Peptide ligand-binding receptors	181	19	8.8×10^{-13}
Visual phototransduction	86	11	1.8×10^{-11}
GPCR ligand binding	373	32	8.1×10^{-11}
Retinoid metabolism and transport	39	7	1.3×10^{-10}
Gastrin-CREB signalling pathway via PKC and MAPK	185	17	1.5×10^{-10}
G _{aq} signalling events	164	15	5.6×10^{-10}
Platelet activation, signaling and aggregation	186	16	1.7×10^{-9}
G _{ai} signalling events	191	15	3.5×10^{-8}
Response to elevated platelet cytosolic Ca ²⁺	84	8	1.8×10^{-7}
HS-GAG degradation	21	4	4.2×10^{-7}
Platelet homeostasis	54	6	4.7×10^{-7}
VEGFA-VEGFR2 Pathway	91	8	5.1×10^{-7}
Transcriptional regulation of white adipocyte differentiation	56	6	6.4×10^{-7}
Signaling by NOTCH4	11	3	1.2×10^{-6}
Signaling by VEGF	99	8	1.5×10^{-6}
Signaling by NOTCH	80	7	1.5×10^{-6}
G _{oa} signalling events	100	8	1.7×10^{-6}
Defective EXT2 causes exostoses 2	12	3	1.7×10^{-6}

Table 4.12: Pathways for *CDH1* partners from SLIPT in stomach cancer

Reactome Pathway	Over-representation	Permutation
Eukaryotic Translation Elongation	1.3×10^{-207}	$< 1.001 \times 10^{-3}$
Peptide chain elongation	5.6×10^{-201}	$< 1.001 \times 10^{-3}$
Viral mRNA Translation	1.2×10^{-196}	$< 1.001 \times 10^{-3}$
Eukaryotic Translation Termination	1.2×10^{-196}	$< 1.001 \times 10^{-3}$
Formation of a pool of free 40S subunits	3.7×10^{-194}	$< 1.001 \times 10^{-3}$
Nonsense Mediated Decay independent of the Exon Junction Complex	5.3×10^{-187}	$< 1.001 \times 10^{-3}$
L13a-mediated translational silencing of Ceruloplasmin expression	9.6×10^{-183}	$< 1.001 \times 10^{-3}$
3' -UTR-mediated translational regulation	9.6×10^{-183}	$< 1.001 \times 10^{-3}$
GTP hydrolysis and joining of the 60S ribosomal subunit	1.9×10^{-181}	$< 1.001 \times 10^{-3}$
Nonsense-Mediated Decay	6.2×10^{-176}	$< 1.001 \times 10^{-3}$
Nonsense Mediated Decay enhanced by the Exon Junction Complex	6.2×10^{-176}	$< 1.001 \times 10^{-3}$
Adaptive Immune System	6.5×10^{-174}	0.11122
Eukaryotic Translation Initiation	5.7×10^{-173}	$< 1.001 \times 10^{-3}$
Cap-dependent Translation Initiation	5.7×10^{-173}	$< 1.001 \times 10^{-3}$
SRP-dependent cotranslational protein targeting to membrane	2×10^{-171}	$< 1.001 \times 10^{-3}$
Translation	6.1×10^{-170}	$< 1.001 \times 10^{-3}$
Infectious disease	1.6×10^{-166}	0.1467
Influenza Infection	1.9×10^{-163}	$< 1.001 \times 10^{-3}$
Influenza Viral RNA Transcription and Replication	1.9×10^{-160}	$< 1.001 \times 10^{-3}$
Influenza Life Cycle	2.5×10^{-156}	$< 1.001 \times 10^{-3}$
Extracellular matrix organization	1.1×10^{-152}	0.054712
GPCR ligand binding	1.1×10^{-143}	0.50343
Class A/1 (Rhodopsin-like receptors)	1.5×10^{-142}	0.51419
GPCR downstream signaling	7.6×10^{-140}	0.087065
Hemostasis	1.9×10^{-134}	0.18151
Developmental Biology	2×10^{-123}	0.42551
Metabolism of lipids and lipoproteins	3.3×10^{-120}	0.6772
Cytokine Signaling in Immune system	2.6×10^{-119}	0.27238
Peptide ligand-binding receptors	3.7×10^{-109}	0.46952
G <i>αi</i> signalling events	8.9×10^{-100}	$< 1.001 \times 10^{-3}$
Axon guidance	1.4×10^{-96}	0.63789
Platelet activation, signaling and aggregation	3.7×10^{-94}	0.17679
Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	1.4×10^{-93}	$< 1.001 \times 10^{-3}$
Formation of the ternary complex, and subsequently, the 43S complex	7×10^{-91}	$< 1.001 \times 10^{-3}$
Translation initiation complex formation	9.6×10^{-87}	0.001001
Ribosomal scanning and start codon recognition	9.6×10^{-87}	0.001001
Activation of the mRNA upon binding of the cap-binding complex and eIFs, and subsequent binding to 43S	8.7×10^{-86}	0.001001
Chemokine receptors bind chemokines	5.1×10^{-82}	0.77614
Signalling by NGF	1.2×10^{-81}	0.25326
Toll-Like Receptors Cascades	5.3×10^{-80}	0.52118
Interferon gamma signaling	6.3×10^{-80}	0.45042
Transmembrane transport of small molecules	5.3×10^{-78}	0.13759
Signaling by Rho GTPases	1.1×10^{-77}	0.055108
Degradation of the extracellular matrix	7.3×10^{-77}	0.63362
Interferon Signaling	1.1×10^{-76}	0.12689
NGF signalling via TRKA from the plasma membrane	1.4×10^{-74}	0.53792
Gastrin-CREB signalling pathway via PKC and MAPK	3.1×10^{-74}	$< 1.001 \times 10^{-3}$
Rho GTPase cycle	3.2×10^{-73}	0.091991
DAP12 interactions	2×10^{-71}	0.44074
Cell surface interactions at the vascular wall	3.3×10^{-71}	0.63362

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

Table 4.13: Pathways for *CDH1* partners from SLIPT in stomach and siRNA screen

Reactome Pathway	Over-representation	Permutation
Chemokine receptors bind chemokines	5.2×10^{-16}	0.0026524
Class A/1 (Rhodopsin-like receptors)	6.4×10^{-14}	0.05974
Peptide ligand-binding receptors	8.8×10^{-13}	0.10988
Visual phototransduction	1.8×10^{-11}	0.30639
GPCR ligand binding	8.1×10^{-11}	0.17895
Retinoid metabolism and transport	1.3×10^{-10}	0.17481
Gastrin-CREB signalling pathway via PKC and MAPK	1.5×10^{-10}	0.52377
Gαq signalling events	5.6×10^{-10}	0.57601
Platelet activation, signaling and aggregation	1.7×10^{-9}	0.34977
Gαi signalling events	3.5×10^{-8}	0.23131
Response to elevated platelet cytosolic Ca ²⁺	1.8×10^{-7}	0.18637
HS-GAG degradation	4.2×10^{-7}	0.24605
Platelet homeostasis	4.7×10^{-7}	0.18996
VEGFA-VEGFR2 Pathway	5.1×10^{-7}	0.87816
Transcriptional regulation of white adipocyte differentiation	6.4×10^{-7}	0.18505
Signaling by NOTCH4	1.2×10^{-6}	0.36495
Signaling by NOTCH	1.5×10^{-6}	0.76112
Signaling by VEGF	1.5×10^{-6}	0.52553
Defective EXT2 causes exostoses 2	1.7×10^{-6}	0.24605
Defective EXT1 causes exostoses 1, TRPS2 and CHDS	1.7×10^{-6}	0.24605
Gαs signalling events	1.7×10^{-6}	0.31637
Generation of second messenger molecules	3.5×10^{-6}	0.032952
DAP12 interactions	3.5×10^{-6}	0.8492
Mitochondrial Fatty Acid Beta-Oxidation	4×10^{-6}	0.033295
Acyl chain remodelling of PS	6×10^{-6}	0.46799
Phase 1 - Functionalization of compounds	6.5×10^{-6}	0.068729
Costimulation by the CD28 family	6.5×10^{-6}	0.031427
Translocation of ZAP-70 to Immunological synapse	8.1×10^{-6}	$< 2.299 \times 10^{-4}$
Complement cascade	9.8×10^{-6}	$< 2.299 \times 10^{-4}$
Molecules associated with elastic fibres	9.8×10^{-6}	0.025491
Signal amplification	1.1×10^{-5}	0.36204
Phosphorylation of CD3 and TCR zeta chains	1.5×10^{-5}	$< 2.299 \times 10^{-4}$
Cell surface interactions at the vascular wall	1.6×10^{-5}	0.039572
Hemostasis	1.7×10^{-5}	0.22035
FCERI mediated MAPK activation	1.7×10^{-5}	0.35433
Defective B4GALT7 causes EDS, progeroid type	1.8×10^{-5}	0.36204
Defective B3GAT3 causes JDSSDHD	1.8×10^{-5}	0.36204
Elastic fibre formation	1.9×10^{-5}	0.0026524
Signaling by NOTCH1	1.9×10^{-5}	0.52553
Acyl chain remodelling of PE	2.9×10^{-5}	0.46799
TCR signaling	2.9×10^{-5}	0.1269
Signaling by Leptin	2.9×10^{-5}	0.36091
PD-1 signaling	2.9×10^{-5}	$< 2.299 \times 10^{-4}$
Opioid Signalling	3.3×10^{-5}	0.81326
Signaling by SCF-KIT	3.4×10^{-5}	0.79924
Arachidonic acid metabolism	3.4×10^{-5}	0.0033013
DAP12 signaling	3.4×10^{-5}	0.9366
Netrin-1 signaling	3.4×10^{-5}	0.76768
Signaling by Retinoic Acid	3.4×10^{-5}	0.011724
Respiratory electron transport	4×10^{-5}	0.28245

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).

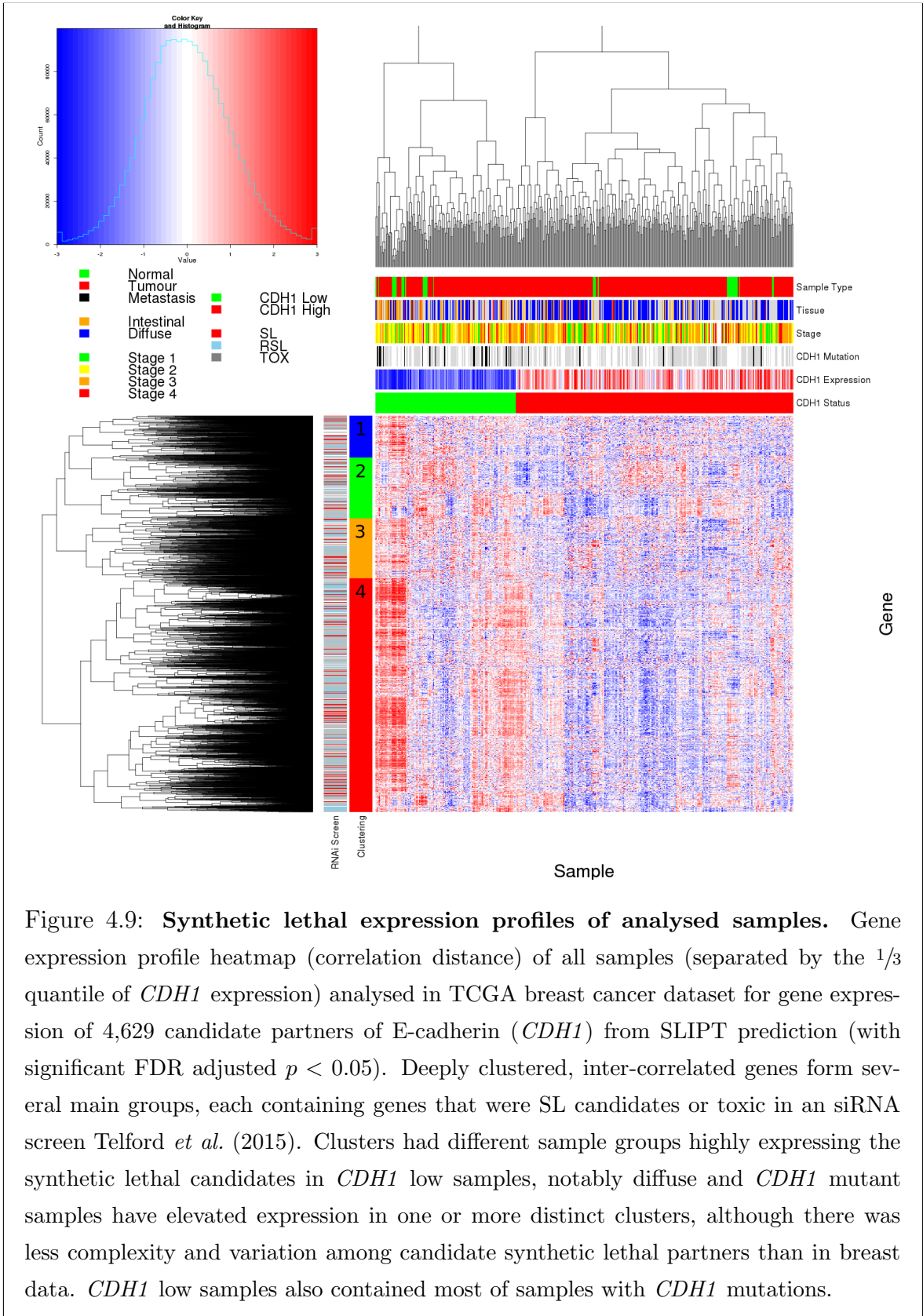
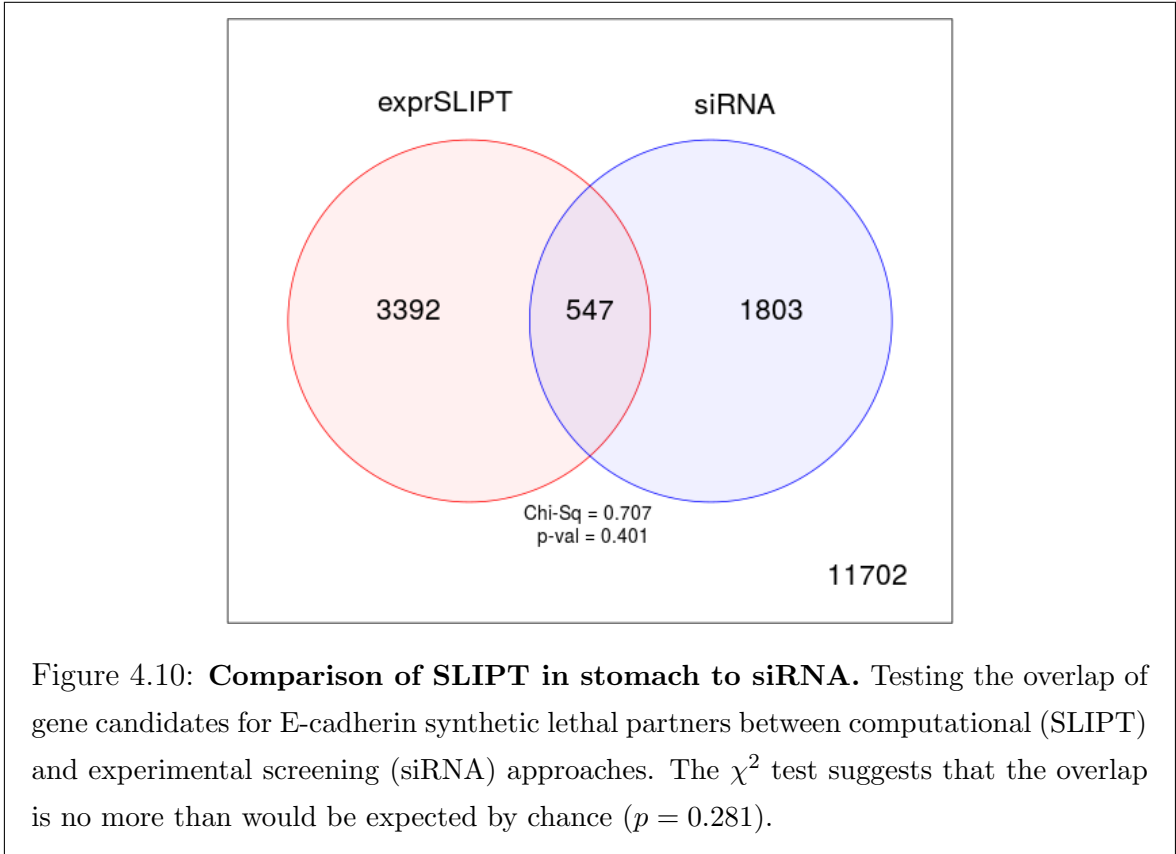


Figure 4.9: **Synthetic lethal expression profiles of analysed samples.** Gene expression profile heatmap (correlation distance) of all samples (separated by the $1/3$ quantile of *CDH1* expression) analysed in TCGA breast cancer dataset for gene expression of 4,629 candidate partners of E-cadherin (*CDH1*) from SLIPT prediction (with significant FDR adjusted $p < 0.05$). Deeply clustered, inter-correlated genes form several main groups, each containing genes that were SL candidates or toxic in an siRNA screen Telford *et al.* (2015). Clusters had different sample groups highly expressing the synthetic lethal candidates in *CDH1* low samples, notably diffuse and *CDH1* mutant samples have elevated expression in one or more distinct clusters, although there was less complexity and variation among candidate synthetic lethal partners than in breast data. *CDH1* low samples also contained most of samples with *CDH1* mutations.



ing to highly interacting hub genes, these represent important genes in a functional network. As shown in Tables 1-3, Gene Ontology terms for genes important in cancer proliferation, progression, and drug response were enriched in hub genes, showing that synthetic lethal interactions are among important genes in cancer cells. Gene functions replicated across the breast cancer datasets are highlighted in bold, despite differences in particular hits, gene expression platforms, and only correcting for multiple tests for each gene query separately, there are many gene functions replicated across breast cancer gene expression analyses. TCGA microarray data was less consistent with the other datasets, as expected from lower sample size, lower concordance of particular hits for the example query of *CDH1*, and suspected lower quality of data on the Aligent microarray platform.

As specific genes were difficult to replicate across experiments, gene expression profiles for synthetic lethal partners must be more complex than originally expected to directly compensate for loss of query gene or completely lack (or clearly under-represent) co-loss (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

Table 4.14: Candidate synthetic lethal metagenes against *CDH1* from SLIPT in stomach cancer

Pathway	ID	Observed	Expected	χ^2 value	p-value	p-value (FDR)
Apoptotic cleavage of cell adhesion proteins	351906	106	50.45	160.39	1.205×10^{-33}	1.9906×10^{-300}
Nef mediated downregulation of MHC class I complex cell surface expression	164940	100	50.45	128.73	7.2777×10^{-27}	6.0114×10^{-240}
Cell-cell junction organization	421270	94	50.45	125.26	4.0251×10^{-26}	2.2165×10^{-230}
Cytochrome P450 - arranged by substrate type	211897	96	50.45	116.16	3.5335×10^{-24}	1.2741×10^{-210}
Cell junction organization	446728	93	50.45	115.98	3.8563×10^{-24}	1.2741×10^{-210}
N-Glycan antennae elongation	975577	98	50.45	115.26	5.5032×10^{-24}	1.5152×10^{-210}
N-glycan antennae elongation in the medialtrans-Golgi	975576	95	50.45	113.42	1.3541×10^{-23}	3.1958×10^{-210}
Cell-Cell communication	1500931	18	50.45	109.96	7.426×10^{-23}	1.5335×10^{-200}
VEGFR2 mediated vascular permeability	5218920	19	50.45	108.73	1.3555×10^{-22}	2.4882×10^{-200}
Synthesis of PE	1483213	93	50.45	108.33	1.6505×10^{-22}	2.7266×10^{-200}
Lysosome Vesicle Biogenesis	432720	92	50.45	105.43	6.8635×10^{-22}	1.0308×10^{-190}
Sema4D in semaphorin signaling	400685	20	50.45	103.68	1.6182×10^{-21}	2.1167×10^{-190}
Transport of glucose and other sugars, bile salts and organic acids, metal ions and amine compounds	425366	83	50.45	103.62	1.6657×10^{-21}	2.1167×10^{-190}
Phase 1 - Functionalization of compounds	211945	93	50.45	102.76	2.5461×10^{-21}	3.0044×10^{-190}
Sphingolipid de novo biosynthesis	1660661	94	50.45	102.39	3.0471×10^{-21}	3.3558×10^{-190}
Transport of nucleotide sugars	727802	91	50.45	101.47	4.7818×10^{-21}	4.9372×10^{-190}
Ion transport by P-type ATPases	936837	17	50.45	100.35	8.2923×10^{-21}	8.0581×10^{-190}
PPARA activates gene expression	1989781	93	50.45	99.78	1.0972×10^{-20}	1.007×10^{-180}
Adherens junctions interactions	418990	93	50.45	99.09	1.5361×10^{-20}	1.3356×10^{-180}
Tight junction interactions	420029	92	50.45	98.35	2.2075×10^{-20}	1.8234×10^{-180}
Sialic acid metabolism	4085001	19	50.45	95.28	9.947×10^{-20}	7.8249×10^{-180}
Transport of inorganic cationsanions and amino acidsoligopeptides	425393	89	50.45	94.10	1.7698×10^{-19}	1.2268×10^{-170}
Biological oxidations	211859	87	50.45	94.05	1.8182×10^{-19}	1.2268×10^{-170}
GRB7 events in ERBB2 signaling	1306955	92	50.45	94.01	1.8492×10^{-19}	1.2268×10^{-170}
Synthesis of pyrophosphates in the cytosol	1855167	26	50.45	94.00	1.8566×10^{-19}	1.2268×10^{-170}

Strongest candidate SL partners for *CDH1* by SLIPT with observed and expected samples with low expression of both genes

profiles and clinical variables to find relationships in gene expression, gene function, and clinical characteristics. The large number of hits indicate that synthetic lethality is error-prone and identifying genes or pathways relevant for clinical application will be difficult.

The expression profiles of the SL partners of *CDH1* predicted from the TCGA breast cancer RNA-Seq data in *CDH1* low tumours (where synthetic lethal partners are expected to have compensating high or stable expression) are shown in Figure 7 and their corresponding functional enrichment is given below in Table 5, computed as WikiPathways in GeneSetDB (Araki *et al.*, 2012). The 3 subgroups of genes are showed functional organisation of expression profiles in *CDH1* low breast tumours. The first group is enriched for G protein coupled receptors, an established drug target and supported in cell line experiments (Telford *et al.*, 2015). The second group contains genes involved in development and metabolism consistent with cancer cells showing stem cell properties and the Warburg hypothesis (Merlos-Suarez *et al.*, 2011; Warburg, 1956). The third group contains cell signalling and focal adhesion functions, including path-

Table 4.15: Candidate synthetic lethal genes against E-cadherin from SLIPT in CCLE

Gene	Observed	Expected	χ^2 value	p-value	p-value (FDR)
<i>ZEB1</i>	24	115	555	7.84×10^{-119}	3.62×10^{-116}
<i>RP11-620J15.3</i>	17	115	471	1.54×10^{-100}	3.68×10^{-98}
<i>AP1S2</i>	20	115	462	1.38×10^{-98}	3.07×10^{-96}
<i>VIM</i>	24	115	424	1.73×10^{-90}	3.06×10^{-88}
<i>CCDC88A</i>	24	115	418	3.94×10^{-89}	6.86×10^{-87}
<i>RECK</i>	28	115	416	8.23×10^{-89}	1.42×10^{-86}
<i>AP1M1</i>	16	115	414	2.42×10^{-88}	4.06×10^{-86}
<i>ZEB2</i>	23	115	396	2.32×10^{-84}	3.4×10^{-82}
<i>WIPF1</i>	25	115	390	4.9×10^{-83}	6.74×10^{-81}
<i>SLC35B4</i>	29	115	386	3.2×10^{-82}	4.38×10^{-80}
<i>SACS</i>	28	115	373	2.13×10^{-79}	2.7×10^{-77}
<i>ST3GAL2</i>	25	115	351	9.7×10^{-75}	1.08×10^{-72}
<i>ATP8B2</i>	38	115	341	1.53×10^{-72}	1.61×10^{-70}
<i>IFFO1</i>	39	115	332	1.66×10^{-70}	1.65×10^{-68}
<i>EMP3</i>	38	115	329	5.04×10^{-70}	4.95×10^{-68}
<i>LEPRE1</i>	40	115	325	5.4×10^{-69}	5.22×10^{-67}
<i>STARD9</i>	39	115	311	4.52×10^{-66}	3.96×10^{-64}
<i>DENND5A</i>	48	115	304	1.89×10^{-64}	1.59×10^{-62}
<i>SYT11</i>	38	115	300	1.21×10^{-63}	9.89×10^{-62}
<i>EID2B</i>	38	115	299	1.99×10^{-63}	1.61×10^{-61}
<i>NXPE3</i>	35	115	294	1.71×10^{-62}	1.35×10^{-60}
<i>STX2</i>	49	115	293	3.83×10^{-62}	3×10^{-60}
<i>ARHGEF6</i>	43	115	289	2.2×10^{-61}	1.71×10^{-59}
<i>KATNAL1</i>	50	115	283	4.45×10^{-60}	3.38×10^{-58}
<i>ANXA6</i>	37	115	282	8.92×10^{-60}	6.67×10^{-58}

Strongest candidate SL partners for *CDH1* by SLIPT with observed and expected samples with low expression of both genes

ways involved in cancer proliferation, metastasis, and consistent with internal synthetic lethality within the pathways containing *CDH1* (Barabási and Oltvai, 2004).

Ductal breast cancers show higher expression of synthetic lethal partners suggesting treatment would be more effective in this tumour subtype. However, there is consistently low expression of SL partners in ER 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

Table 4.16: Pathways for *CDH1* partners from SLIPT in CCLE

Pathways Over-represented	Pathway Size	SL Genes	p-value (FDR)
Cell Cycle	442	207	1.2×10^{-215}
Cell Cycle, Mitotic	365	180	2.9×10^{-209}
Signaling by Rho GTPases	311	136	9.4×10^{-156}
M Phase	212	104	8.8×10^{-145}
Infectious disease	289	123	1.3×10^{-142}
RHO GTPase Effectors	207	98	5.3×10^{-135}
HIV Infection	200	94	2×10^{-130}
Separation of Sister Chromatids	140	77	5.6×10^{-128}
Organelle biogenesis and maintenance	258	107	1.4×10^{-127}
Chromatin modifying enzymes	181	87	4.7×10^{-126}
Chromatin organization	181	87	4.7×10^{-126}
Mitotic Metaphase and Anaphase	149	78	1.2×10^{-124}
Mitotic Anaphase	148	77	6.3×10^{-123}
Developmental Biology	421	142	1.6×10^{-121}
RHO GTPases Activate Formins	94	60	5.3×10^{-118}
Mitotic Prometaphase	93	59	5.4×10^{-116}
Hemostasis	421	138	7.2×10^{-116}
Adaptive Immune System	397	132	3.2×10^{-115}
Assembly of the primary cilium	143	72	2.4×10^{-114}
Transcription	133	68	6.2×10^{-111}

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

CDH1 gene loss; and the expression of synthetic partners could be a clinically important biomarker. 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.

Table 4.17: Candidate synthetic lethal genes against E-cadherin from SLIPT in breast CCLE

Gene	Observed	Expected	χ^2 value	p-value	p-value (FDR)
<i>MIR155HG</i>	1	6.78	31.5	2.41×10^{-6}	0.00371
<i>ENPP2</i>	1	6.78	30.7	3.47×10^{-6}	0.00383
<i>DCLK2</i>	3	6.78	28.3	1.08×10^{-5}	0.0071
<i>PID1</i>	1	6.78	27.8	1.34×10^{-5}	0.00791
<i>SCFD2</i>	5	6.78	27.7	1.42×10^{-5}	0.00791
<i>FAT4</i>	4	6.78	27.3	1.69×10^{-5}	0.00865
<i>ILK</i>	1	6.78	26.9	2.04×10^{-5}	0.00884
<i>RWDD1</i>	0	6.78	26.8	2.15×10^{-5}	0.00884
<i>RIC8A</i>	2	6.78	26.8	2.2×10^{-5}	0.00884
<i>F2RL2</i>	1	6.78	26.6	2.34×10^{-5}	0.00901
<i>SDCBP</i>	5	6.78	25.9	3.26×10^{-5}	0.0108
<i>PPM1F</i>	4	6.78	25.8	3.41×10^{-5}	0.0108
<i>IKBIP</i>	5	6.78	25.8	3.49×10^{-5}	0.0108
<i>SPRED1</i>	3	6.78	25.5	3.97×10^{-5}	0.0108
<i>RNH1</i>	1	6.78	25.4	4.22×10^{-5}	0.0108
<i>SYDE1</i>	3	6.78	25.4	4.22×10^{-5}	0.0108
<i>LINC00968</i>	1	6.78	25.2	4.63×10^{-5}	0.0109
<i>ARHGEF10</i>	5	6.78	24.5	6.22×10^{-5}	0.0116
<i>P4HA1</i>	0	6.78	24.5	6.34×10^{-5}	0.0116
<i>AZI2</i>	2	6.78	24.5	6.34×10^{-5}	0.0116
<i>TNFAIP6</i>	2	6.78	24.5	6.34×10^{-5}	0.0116
<i>CD200</i>	4	6.78	24.5	6.37×10^{-5}	0.0116
<i>SMPD1</i>	1	6.78	24.4	6.67×10^{-5}	0.0116
<i>ATP6V1G2</i>	3	6.78	24.2	7.33×10^{-5}	0.0123
<i>FGF2</i>	4	6.78	24.1	7.49×10^{-5}	0.0123

Strongest candidate SL partners for *CDH1* by SLIPT with observed and expected samples with low expression of both genes

Table 4.18: Pathways for *CDH1* partners from SLIPT in breast CCLE

Pathways Over-represented	Pathway Size	SL Genes	p-value (FDR)
Cell junction organization	71	5	0.006
Adherens junctions interactions	29	3	0.006
Dermatan sulfate biosynthesis	11	2	0.006
Non-integrin membrane-ECM interactions	52	4	0.006
Regulation of pyruvate dehydrogenase (PDH) complex	12	2	0.0069
Cell-extracellular matrix interactions	17	2	0.021
Pyruvate metabolism	17	2	0.021
Cell-cell junction organization	46	3	0.039
Synthesis of substrates in N-glycan biosynthesis	50	3	0.057
Detoxification of Reactive Oxygen Species	26	2	0.082
Keratan sulfate biosynthesis	28	2	0.092
Laminin interactions	28	2	0.092
Cell-Cell communication	118	5	0.12
Keratan sulfate/keratin metabolism	32	2	0.12
Opioid Signalling	63	3	0.12
Biosynthesis of the N-glycan precursor (dolichol lipid-linked oligosaccharide) and transfer to a nascent protein	63	3	0.12
Intraflagellar transport	34	2	0.14
Signaling by Retinoic Acid	36	2	0.16
Pyruvate metabolism and Citric Acid (TCA) cycle	36	2	0.16
Nef mediated downregulation of MHC class I complex cell surface expression	10	1	0.22

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

Table 4.19: Candidate synthetic lethal genes against E-cadherin from SLIPT in stomach CCLE

Gene	Observed	Expected	χ^2 value	p - value	p-value (FDR)
<i>ZEB1</i>	1	4.45	36	2.84×10^{-07}	0.00175
<i>WDR47</i>	0	4.45	26.7	2.3×10^{-05}	0.013
<i>KANK2</i>	1	4.45	25.1	4.81×10^{-05}	0.0222
<i>LEPRE1</i>	0	4.45	24.5	6.26×10^{-05}	0.0228
<i>KATNAL1</i>	0	4.45	24.3	6.88×10^{-05}	0.0231
<i>TET1</i>	0	4.45	23.9	8.23×10^{-05}	0.0249
<i>AP1S2</i>	1	4.45	23.1	0.00012	0.0273
<i>CDKN2C</i>	1	4.45	22.8	0.000136	0.0292
<i>ARMC4</i>	1	4.45	22.4	0.000164	0.0315
<i>CSTF3</i>	1	4.45	22.4	0.000166	0.0315
<i>FAM216A</i>	1	4.45	22.4	0.000166	0.0315
<i>ANKRD32</i>	1	4.45	22.4	0.000166	0.0315
<i>WDR35</i>	1	4.45	22.4	0.000169	0.0315
<i>ECI2</i>	0	4.45	21.7	0.000232	0.0378
<i>SAMD8</i>	0	4.45	21.7	0.000232	0.0378
<i>CHST12</i>	0	4.45	21.7	0.000232	0.0378
<i>RPL23AP32</i>	0	4.45	21.7	0.000232	0.0378
<i>STARD9</i>	1	4.45	21.7	0.000232	0.0378
<i>MCM8</i>	0	4.45	21.5	0.000255	0.0379

Strongest candidate SL partners for *CDH1* by SLIPT with observed and expected samples with low expression of both genes

Table 4.20: Pathways for *CDH1* partners from SLIPT in stomach CCLE

Pathways Over-represented	Pathway Size	SL Genes	p-value (FDR)
Nef mediated downregulation of MHC class I complex cell surface expression	10	1	1
Unwinding of DNA	11	1	1
Processing of Intronless Pre-mRNAs	13	1	1
E2F mediated regulation of DNA replication	20	1	1
Chondroitin sulfate biosynthesis	20	1	1
Post-Elongation Processing of Intronless pre-mRNA	21	1	1
Nef-mediates down modulation of cell surface receptors by recruiting them to clathrin adapters	21	1	1
Processing of Capped Intronless Pre-mRNA	21	1	1
Post-Elongation Processing of Intron-Containing pre-mRNA	23	1	1
Activation of the pre-replicative complex	23	1	1
mRNA 3'-end processing	23	1	1
Golgi Associated Vesicle Biogenesis	24	1	1
Lysosome Vesicle Biogenesis	25	1	1
Oncogene Induced Senescence	27	1	1
The role of Nef in HIV-1 replication and disease pathogenesis	28	1	1
Cyclin D associated events in G1	29	1	1
G1 Phase	29	1	1
Cleavage of Growing Transcript in the Termination Region	31	1	1
Activation of ATR in response to replication stress	31	1	1
DNA strand elongation	31	1	1

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

References

- Aarts, M., Bajrami, I., Herrera-Abreu, M.T., Elliott, R., Brough, R., Ashworth, A., Lord, C.J., and Turner, N.C. (2015) Functional genetic screen identifies increased sensitivity to weel inhibition in cells with defects in fanconi anemia and hr pathways. *Mol Cancer Ther*, **14**(4): 865–76.
- Abeshouse, A., Ahn, J., Akbani, R., Ally, A., Amin, S., Andry, C.D., Annala, M., Aprikian, A., Armenia, J., Arora, A., *et al.* (2015) The Molecular Taxonomy of Primary Prostate Cancer. *Cell*, **163**(4): 1011–1025.
- Adamski, M.G., Gumann, P., and Baird, A.E. (2014) A method for quantitative analysis of standard and high-throughput qPCR expression data based on input sample quantity. *PLoS ONE*, **9**(8): e103917.
- Adler, D. (2005) *vioplot: Violin plot*. R package version 0.2.
- Agarwal, S., Deane, C.M., Porter, M.A., and Jones, N.S. (2010) Revisiting date and party hubs: Novel approaches to role assignment in protein interaction networks. *PLoS Comput Biol*, **6**(6): e1000817.
- Agrawal, N., Akbani, R., Aksoy, B.A., Ally, A., Arachchi, H., Asa, S.L., Auman, J.T., Balasundaram, M., Balu, S., Baylin, S.B., *et al.* (2014) Integrated genomic characterization of papillary thyroid carcinoma. *Cell*, **159**(3): 676–690.
- Akbani, R., Akdemir, K.C., Aksoy, B.A., Albert, M., Ally, A., Amin, S.B., Arachchi, H., Arora, A., Auman, J.T., Ayala, B., *et al.* (2015) Genomic Classification of Cutaneous Melanoma. *Cell*, **161**(7): 1681–1696.
- Akobeng, A.K. (2007) Understanding diagnostic tests 3: receiver operating characteristic curves. *Acta Paediatrica*, **96**(5): 644–647.
- American Cancer Society (2017) Genetics and cancer. <https://www.cancer.org/cancer/cancer-causes/genetics.html>. Accessed: 22/03/2017.
- American Society for Clinical Oncology (ASCO) (2017) The genetics of cancer. <http://www.cancer.net/navigating-cancer-care/cancer-basics/genetics/genetics-cancer>. Accessed: 22/03/2017.

- Araki, H., Knapp, C., Tsai, P., and Print, C. (2012) GeneSetDB: A comprehensive meta-database, statistical and visualisation framework for gene set analysis. *FEBS Open Bio*, **2**: 76–82.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, **25**(1): 25–29.
- Ashworth, A. (2008) A synthetic lethal therapeutic approach: poly(adp) ribose polymerase inhibitors for the treatment of cancers deficient in dna double-strand break repair. *J Clin Oncol*, **26**(22): 3785–90.
- Audeh, M.W., Carmichael, J., Penson, R.T., Friedlander, M., Powell, B., Bell-McGuinn, K.M., Scott, C., Weitzel, J.N., Oaknin, A., Loman, N., *et al.* (2010) Oral poly(adp-ribose) polymerase inhibitor olaparib in patients with brca1 or brca2 mutations and recurrent ovarian cancer: a proof-of-concept trial. *Lancet*, **376**(9737): 245–51.
- Babyak, M.A. (2004) What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom Med*, **66**(3): 411–21.
- Bamford, S., Dawson, E., Forbes, S., Clements, J., Pettett, R., Dogan, A., Flanagan, A., Teague, J., Futreal, P.A., Stratton, M.R., *et al.* (2004) The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer*, **91**(2): 355–358.
- Barabási, A.L. and Albert, R. (1999) Emergence of scaling in random networks. *Science*, **286**(5439): 509–12.
- Barabási, A.L. and Oltvai, Z.N. (2004) Network biology: understanding the cell’s functional organization. *Nat Rev Genet*, **5**(2): 101–13.
- Barrat, A. and Weigt, M. (2000) On the properties of small-world network models. *The European Physical Journal B - Condensed Matter and Complex Systems*, **13**(3): 547–560.
- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehar, J., Kryukov, G.V., Sonkin, D., *et al.* (2012) The Cancer

- Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, **483**(7391): 603–607.
- Barry, W.T. (2016) *safe: Significance Analysis of Function and Expression*. R package version 3.14.0.
- Baryshnikova, A., Costanzo, M., Dixon, S., Vizeacoumar, F.J., Myers, C.L., Andrews, B., and Boone, C. (2010a) Synthetic genetic array (sga) analysis in *saccharomyces cerevisiae* and *schizosaccharomyces pombe*. *Methods Enzymol*, **470**: 145–79.
- Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K., Youn, J.Y., Ou, J., San Luis, B.J., Bandyopadhyay, S., *et al.* (2010b) Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat Meth*, **7**(12): 1017–1024.
- Bass, A.J., Thorsson, V., Shmulevich, I., Reynolds, S.M., Miller, M., Bernard, B., Hinoue, T., Laird, P.W., Curtis, C., Shen, H., *et al.* (2014) Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*, **513**(7517): 202–209.
- Bates, D. and Maechler, M. (2016) *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version 1.2-7.1.
- Bateson, W. and Mendel, G. (1909) *Mendel's principles of heredity, by W. Bateson*. University Press, Cambridge [Eng.].
- Beck, T.F., Mullikin, J.C., and Biesecker, L.G. (2016) Systematic Evaluation of Sanger Validation of Next-Generation Sequencing Variants. *Clin Chem*, **62**(4): 647–654.
- Becker, K.F., Atkinson, M.J., Reich, U., Becker, I., Nekarda, H., Siewert, J.R., and Hfler, H. (1994) E-cadherin gene mutations provide clues to diffuse type gastric carcinomas. *Cancer Research*, **54**(14): 3845–3852.
- Bell, D., Berchuck, A., Birrer, M., Chien, J., Cramer, D., Dao, F., Dhir, R., DiSaia, P., Gabra, H., Glenn, P., *et al.* (2011) Integrated genomic analyses of ovarian carcinoma. *Nature*, **474**(7353): 609–615.
- Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)*, **57**(1): 289–300.

- Berx, G., Cleton-Jansen, A.M., Nollet, F., de Leeuw, W.J., van de Vijver, M., Cornelisse, C., and van Roy, F. (1995) E-cadherin is a tumour/invasion suppressor gene mutated in human lobular breast cancers. *EMBO J*, **14**(24): 6107–15.
- Berx, G., Cleton-Jansen, A.M., Strumane, K., de Leeuw, W.J., Nollet, F., van Roy, F., and Cornelisse, C. (1996) E-cadherin is inactivated in a majority of invasive human lobular breast cancers by truncation mutations throughout its extracellular domain. *Oncogene*, **13**(9): 1919–25.
- Berx, G. and van Roy, F. (2009) Involvement of members of the cadherin superfamily in cancer. *Cold Spring Harb Perspect Biol*, **1**: a003129.
- Bitler, B.G., Aird, K.M., Garipov, A., Li, H., Amatangelo, M., 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.
- Breiman, L. (2001) Random forests. *Machine Learning*, **45**(1): 5–32.
- Brin, S. and Page, L. (1998) The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, **30**(1): 107 – 117.

- Bryant, H.E., Schultz, N., Thomas, H.D., Parker, K.M., Flower, D., Lopez, E., Kyle, S., Meuth, M., Curtin, N.J., and Helleday, T. (2005) Specific killing of *brca2*deficient tumours with inhibitors of polyadprbose polymerase. *Nature*, **434**(7035): 913–7.
- Burk, R.D., Chen, Z., Saller, C., Tarvin, K., Carvalho, A.L., Scapulatempo-Neto, C., Silveira, H.C., Fregnani, J.H., Creighton, C.J., Anderson, M.L., *et al.* (2017) Integrated genomic and molecular characterization of cervical cancer. *Nature*, **543**(7645): 378–384.
- Bussey, H., Andrews, B., and Boone, C. (2006) From worm genetic networks to complex human diseases. *Nat Genet*, **38**(8): 862–3.
- Butland, G., Babu, M., Diaz-Mejia, J.J., Bohdana, F., Phanse, S., Gold, B., Yang, W., Li, J., Gagarinova, A.G., Pogoutse, O., *et al.* (2008) esga: E. coli synthetic genetic array analysis. *Nat Methods*, **5**(9): 789–95.
- Cancer Research UK (2017) Family history and cancer genes. <http://www.cancerresearchuk.org/about-cancer/causes-of-cancer/inherited-cancer-genes-and-increased-cancer-risk/family-history-and-inherited-cancer-genes>. Accessed: 22/03/2017.
- Cancer Cell Line Encyclopedia (CCLE) (2014) Broad-Novartis Cancer Cell Line Encyclopedia. <http://www.broadinstitute.org/ccle>. Accessed: 07/11/2014.
- cBioPortal for Cancer Genomics (cBioPortal) (2017) cBioPortal for Cancer Genomics. <http://www.cbioportal.org/>. Accessed: 26/03/2017.
- Cerami, E.G., Gross, B.E., Demir, E., Rodchenkov, I., Babur, O., Anwar, N., Schultz, N., Bader, G.D., and Sander, C. (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res*, **39**(Database issue): D685–690.
- Chen, A., Beetham, H., Black, M.A., Priya, R., Telford, B.J., Guest, J., Wiggins, G.A.R., Godwin, T.D., Yap, A.S., and Guilford, P.J. (2014) E-cadherin loss alters cytoskeletal organization and adhesion in non-malignant breast cells but is insufficient to induce an epithelial-mesenchymal transition. *BMC Cancer*, **14**(1): 552.
- Chen, S. and Parmigiani, G. (2007) Meta-analysis of BRCA1 and BRCA2 penetrance. *J Clin Oncol*, **25**(11): 1329–1333.
- Chen, X. and Tompa, M. (2010) Comparative assessment of methods for aligning multiple genome sequences. *Nat Biotechnol*, **28**(6): 567–572.

- Cherniack, A.D., Shen, H., Walter, V., Stewart, C., Murray, B.A., Bowlby, R., Hu, X., Ling, S., Soslow, R.A., Broaddus, R.R., *et al.* (2017) Integrated Molecular Characterization of Uterine Carcinosarcoma. *Cancer Cell*, **31**(3): 411–423.
- Chipman, K. and Singh, A. (2009) Predicting genetic interactions with random walks on biological networks. *BMC Bioinformatics*, **10**(1): 17.
- Christofori, G. and Semb, H. (1999) The role of the cell-adhesion molecule e-cadherin as a tumour-suppressor gene. *Trends in Biochemical Sciences*, **24**(2): 73 – 76.
- Ciriello, G., Gatz, M.L., Beck, A.H., Wilkerson, M.D., Rhie, S.K., Pastore, A., Zhang, H., McLellan, M., Yau, C., Kandoth, C., *et al.* (2015) Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell*, **163**(2): 506–519.
- Clark, M.J. (2004) Endogenous Regulator of G Protein Signaling Proteins Suppress G o-Dependent μ -Opioid Agonist-Mediated Adenylyl Cyclase Supersensitization. *Journal of Pharmacology and Experimental Therapeutics*, **310**(1): 215–222.
- Clough, E. and Barrett, T. (2016) The Gene Expression Omnibus Database. *Methods Mol Biol*, **1418**: 93–110.
- Collingridge, D.S. (2013) A primer on quantitized data analysis and permutation testing. *Journal of Mixed Methods Research*, **7**(1): 81–97.
- Collins, F.S. and Barker, A.D. (2007) Mapping the cancer genome. Pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies. *Sci Am*, **296**(3): 50–57.
- Collins, F.S., Morgan, M., and Patrinos, A. (2003) The Human Genome Project: lessons from large-scale biology. *Science*, **300**(5617): 286–290.
- Collisson, E., Campbell, J., Brooks, A., Berger, A., Lee, W., Chmielecki, J., Beer, D., Cope, L., Creighton, C., Danilova, L., *et al.* (2014) Comprehensive molecular profiling of lung adenocarcinoma. *Nature*, **511**(7511): 543–550.
- Corcoran, R.B., Ebi, H., Turke, A.B., Coffee, E.M., Nishino, M., Cogdill, A.P., Brown, R.D., Della Pelle, P., Dias-Santagata, D., Hung, K.E., *et al.* (2012) Egfr-mediated reactivation of mapk signaling contributes to insensitivity of braf-mutant colorectal cancers to raf inhibition with vemurafenib. *Cancer Discovery*, **2**(3): 227–235.

- Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L., Toufighi, K., Mostafavi, S., *et al.* (2010) The genetic landscape of a cell. *Science*, **327**(5964): 425–31.
- Costanzo, M., Baryshnikova, A., Myers, C.L., Andrews, B., and Boone, C. (2011) Charting the genetic interaction map of a cell. *Curr Opin Biotechnol*, **22**(1): 66–74.
- Creighton, C.J., Morgan, M., Gunaratne, P.H., Wheeler, D.A., Gibbs, R.A., Robertson, A., Chu, A., Beroukhim, R., Cibulskis, K., Signoretti, S., *et al.* (2013) Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature*, **499**(7456): 43–49.
- Croft, D., Mundo, A.F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., Kamdar, M.R., *et al.* (2014) The Reactome pathway knowledge-base. *Nucleic Acids Res*, **42**(database issue): D472D477.
- Crunkhorn, S. (2014) Cancer: Predicting synthetic lethal interactions. *Nat Rev Drug Discov*, **13**(11): 812.
- Csardi, G. and Nepusz, T. (2006) The igraph software package for complex network research. *InterJournal*, **Complex Systems**: 1695.
- Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., *et al.* (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, **486**(7403): 346–352.
- Dai, X., Li, T., Bai, Z., Yang, Y., Liu, X., Zhan, J., and Shi, B. (2015) Breast cancer intrinsic subtype classification, clinical use and future trends. *Am J Cancer Res*, **5**(10): 2929–2943.
- Davierwala, A.P., Haynes, J., Li, Z., Brost, R.L., Robinson, M.D., Yu, L., Mnaimneh, S., Ding, H., Zhu, H., Chen, Y., *et al.* (2005) The synthetic genetic interaction spectrum of essential genes. *Nat Genet*, **37**(10): 1147–1152.
- De Leeuw, W.J., Berx, G., Vos, C.B., Peterse, J.L., Van de Vijver, M.J., Litvinov, S., Van Roy, F., Cornelisse, C.J., and Cleton-Jansen, A.M. (1997) Simultaneous loss of e-cadherin and catenins in invasive lobular breast cancer and lobular carcinoma in situ. *J Pathol*, **183**(4): 404–11.

- Demir, E., Babur, O., Rodchenkov, I., Aksoy, B.A., Fukuda, K.I., Gross, B., Sumer, O.S., Bader, G.D., and Sander, C. (2013) Using biological pathway data with Paxtools. *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. *Anticancer 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., Fedyshyn, Y., Koh, J.L., Prasad, T.S., Chahwan, C., Chua, G., Toufighi, K., Baryshnikova, A., Hayles, J., Hoe, K.L., *et al.* (2008) Significant conservation of synthetic lethal genetic interaction networks between distantly related eukaryotes. *Proc Natl Acad Sci U S A*, **105**(43): 16653–8.
- Dorogovtsev, S.N. and Mendes, J.F. (2003) *Evolution of networks: From biological nets to the Internet and WWW*. Oxford University Press, USA.
- Erdős, P. and Rényi, A. (1959) On random graphs I. *Publ Math Debrecen*, **6**: 290–297.
- Erdős, P. and Rényi, A. (1960) On the evolution of random graphs. In *Publ. Math. Inst. Hung. Acad. Sci*, volume 5, 17–61.
- Eroles, P., Bosch, A., Perez-Fidalgo, J.A., and Lluch, A. (2012) Molecular biology in breast cancer: intrinsic subtypes and signaling pathways. *Cancer Treat Rev*, **38**(6): 698–707.
- Ezkurdia, I., Juan, D., Rodriguez, J.M., Frankish, A., Diekhans, M., Harrow, J., Vazquez, J., Valencia, A., and Tress, M.L. (2014) Multiple evidence strands suggest that there may be as few as 19 000 human protein-coding genes. *Human Molecular Genetics*, **23**(22): 5866.

- Farmer, H., McCabe, N., Lord, C.J., Tutt, A.N., Johnson, D.A., Richardson, T.B., Santarosa, M., Dillon, K.J., Hickson, I., Knights, C., *et al.* (2005) Targeting the dna repair defect in brca mutant cells as a therapeutic strategy. *Nature*, **434**(7035): 917–21.
- Fawcett, T. (2006) An introduction to ROC analysis. *Pattern Recognition Letters*, **27**(8): 861 – 874. {ROC} Analysis in Pattern Recognition.
- Fece de la Cruz, F., Gapp, B.V., and Nijman, S.M. (2015) Synthetic lethal vulnerabilities of cancer. *Annu Rev Pharmacol Toxicol*, **55**: 513–531.
- Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F. (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*, **136**(5): E359–386.
- Fisher, R.A. (1919) Xv.the correlation between relatives on the supposition of mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, **52**(02): 399–433.
- Fong, P.C., Boss, D.S., Yap, T.A., Tutt, A., Wu, P., Mergui-Roelvink, M., Mortimer, P., Swaisland, H., Lau, A., O’Connor, M.J., *et al.* (2009) Inhibition of poly(adenosine) 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(adenosine) polymerase inhibition: frequent durable responses in brca carrier ovarian cancer correlating with platinum-free interval. *J Clin Oncol*, **28**(15): 2512–9.
- Forbes, S.A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S., *et al.* (2015) COSMIC: exploring the world’s knowledge of somatic mutations in human cancer. *Nucleic Acids Res*, **43**(Database issue): D805–811.
- Fraser, A. (2004) Towards full employment: using rnai to find roles for the redundant. *Oncogene*, **23**(51): 8346–52.

- Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004) A census of human cancer genes. *Nat Rev Cancer*, **4**(3): 177–183.
- Futreal, P.A., Kasprzyk, A., Birney, E., Mullikin, J.C., Wooster, R., and Stratton, M.R. (2001) Cancer and genomics. *Nature*, **409**(6822): 850–852.
- Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*, **5**(10): R80.
- Genz, A. and Bretz, F. (2009) Computation of Multivariate Normal and t Probabilities. In *Lecture Notes in Statistics*, volume 195. Springer-Verlag, Heidelberg.
- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2016) *mvtnorm: Multivariate Normal and t Distributions*. R package version 1.0-5. URL.
- Gilbert, W. and Maxam, A. (1973) The nucleotide sequence of the lac operator. *Proceedings of the National Academy of Sciences*, **70**(12): 3581–3584.
- Git, A., Dvinge, H., Salmon-Divon, M., Osborne, M., Kutter, C., Hadfield, J., Bertone, P., and Caldas, C. (2010) Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. *RNA*, **16**(5): 991–1006.
- Globus (Globus) (2017) Research data management simplified. <https://www.globus.org/>. Accessed: 25/03/2017.
- Graziano, F., Humar, B., and Guilford, P. (2003) The role of the e-cadherin gene (*cdh1*) in diffuse gastric cancer susceptibility: from the laboratory to clinical practice. *Annals of Oncology*, **14**(12): 1705–1713.
- Green, R.E., Briggs, A.W., Krause, J., Prufer, K., Burbano, H.A., Siebauer, M., Lachmann, M., and Pääbo, S. (2009) The Neandertal genome and ancient DNA authenticity. *EMBO J*, **28**(17): 2494–2502.
- Güell, O., Sagus, F., and Serrano, M. (2014) Essential plasticity and redundancy of metabolism unveiled by synthetic lethality analysis. *PLoS Comput Biol*, **10**(5): e1003637.

- Guilford, P. (1999) E-cadherin downregulation in cancer: fuel on the fire? *Molecular Medicine Today*, **5**(4): 172 – 177.
- Guilford, P., Hopkins, J., Harraway, J., McLeod, M., McLeod, N., Harawira, P., Taite, H., Scoular, R., Miller, A., and Reeve, A.E. (1998) E-cadherin germline mutations in familial gastric cancer. *Nature*, **392**(6674): 402–5.
- Guilford, P., Humar, B., and Blair, V. (2010) Hereditary diffuse gastric cancer: translation of *cdh1* germline mutations into clinical practice. *Gastric Cancer*, **13**(1): 1–10.
- Guilford, P.J., Hopkins, J.B., Grady, W.M., Markowitz, S.D., Willis, J., Lynch, H., Rajput, A., Wiesner, G.L., Lindor, N.M., Burgart, L.J., *et al.* (1999) E-cadherin germline mutations define an inherited cancer syndrome dominated by diffuse gastric cancer. *Hum Mutat*, **14**(3): 249–55.
- Guo, J., Liu, H., and Zheng, J. (2016) SynLethDB: synthetic lethality database toward discovery of selective and sensitive anticancer drug targets. *Nucleic Acids Res*, **44**(D1): D1011–1017.
- Hajian-Tilaki, K. (2013) Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian J Intern Med*, **4**(2): 627–635.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I.H. (2009) The weka data mining software: an update. *SIGKDD Explor Newsl*, **11**(1): 10–18.
- Hammerman, P.S., Lawrence, M.S., Voet, D., Jing, R., Cibulskis, K., Sivachenko, A., Stojanov, P., McKenna, A., Lander, E.S., Gabriel, S., *et al.* (2012) Comprehensive genomic characterization of squamous cell lung cancers. *Nature*, **489**(7417): 519–525.
- Han, J.D.J., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J.M., Cusick, M.E., Roth, F.P., *et al.* (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*, **430**(6995): 88–93.
- Hanahan, D. and Weinberg, R.A. (2000) The hallmarks of cancer. *Cell*, **100**(1): 57–70.
- Hanahan, D. and Weinberg, R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**(5): 646–674.

- Hanna, S. (2003) Cancer incidence in new zealand (2003-2007). In D. Forman, D. Bray F Brewster, C. Gombe Mbalawa, B. Kohler, M. Piñeros, E. Steliarova-Foucher, R. Swaminathan, and J. Ferlay (editors), *Cancer Incidence in Five Continents*, volume X, 902–907. International Agency for Research on Cancer, Lyon, France. Electronic version <http://ci5.iarc.fr> Accessed 22/03/2017.
- Heiskanen, M., Bian, X., Swan, D., and Basu, A. (2014) caArray microarray database in the cancer biomedical informatics gridTM (caBIGTM). *Cancer Research*, **67**(9 Supplement): 3712–3712.
- Heiskanen, M.A. and Aittokallio, T. (2012) Mining high-throughput screens for cancer drug targets-lessons from yeast chemical-genomic profiling and synthetic lethality. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, **2**(3): 263–272.
- Hell, P. (1976) Graphs with given neighbourhoods i. problèmes combinatoires at theorie des graphes. *Proc Coil Int CNRS, Orsay*, **260**: 219–223.
- Herschkowitz, J.I., Simin, K., Weigman, V.J., Mikaelian, I., Usary, J., Hu, Z., Rasmussen, K.E., Jones, L.P., Assefnia, S., Chandrasekharan, S., *et al.* (2007) Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol*, **8**(5): R76.
- Hillenmeyer, M.E. (2008) The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science*, **320**: 362–365.
- Hoadley, K.A., Yau, C., Wolf, D.M., Cherniack, A.D., Tamborero, D., Ng, S., Leiserson, M.D., Niu, B., McLellan, M.D., Uzunangelov, V., *et al.* (2014) Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*, **158**(4): 929–944.
- Hoehndorf, R., Hardy, N.W., Osumi-Sutherland, D., Tweedie, S., Schofield, P.N., and Gkoutos, G.V. (2013) Systematic analysis of experimental phenotype data reveals gene functions. *PLoS ONE*, **8**(4): e60847.
- Holm, S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**(2): 65–70.
- Holme, P. and Kim, B.J. (2002) Growing scale-free networks with tunable clustering. *Physical Review E*, **65**(2): 026107.

- Hopkins, A.L. (2008) Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol*, **4**(11): 682–690.
- Hu, Z., Fan, C., Oh, D.S., Marron, J.S., He, X., Qaqish, B.F., Livasy, C., Carey, L.A., Reynolds, E., Dressler, L., *et al.* (2006) The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics*, **7**: 96.
- Huang, E., Cheng, S., Dressman, H., Pittman, J., Tsou, M., Horng, C., Bild, A., Iversen, E., Liao, M., Chen, C., *et al.* (2003) Gene expression predictors of breast cancer outcomes. *Lancet*, **361**: 1590–1596.
- Illumina, Inc (Illumina) (2017) Sequencing and array-based solutions for genetic research. <https://www.illumina.com/>. Accessed: 26/03/2017.
- International HapMap 3 Consortium (HapMap) (2003) The International HapMap Project. *Nature*, **426**(6968): 789–796.
- International Human Genome Sequencing Consortium (IHGSC) (2004) Finishing the euchromatic sequence of the human genome. *Nature*, **431**(7011): 931–945.
- Jerby-Arnon, L., Pfetzer, N., Waldman, Y., McGarry, L., James, D., Shanks, E., Seashore-Ludlow, B., Weinstock, A., Geiger, T., Clemons, P., *et al.* (2014) Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. *Cell*, **158**(5): 1199–1209.
- Joachims, T. (1999) Making large-scale support vector machine learning practical. In S. Bernhard, I. Kopr, 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.

- Kandoth, C., Schultz, N., Cherniack, A.D., Akbani, R., Liu, Y., Shen, H., Robertson, A.G., Pashtan, I., Shen, R., Benz, C.C., *et al.* (2013) Integrated genomic characterization of endometrial carcinoma. *Nature*, **497**(7447): 67–73.
- Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., *et al.* (2001) Functional annotation of a full-length mouse cDNA collection. *Nature*, **409**(6821): 685–690.
- Kelley, R. and Ideker, T. (2005) Systematic interpretation of genetic interactions using protein networks. *Nat Biotech*, **23**(5): 561–566.
- Kelly, S., Chen, A., Guilford, P., and Black, M. (2017a) Synthetic lethal interaction prediction of target pathways in E-cadherin deficient breast cancers. Submitted to *BMC Genomics*.
- Kelly, S.T. (2013) *Statistical Predictions of Synthetic Lethal Interactions in Cancer*. Dissertation, University of Otago.
- Kelly, S.T., Single, A.B., Telford, B.J., Beetham, H.G., Godwin, T.D., Chen, A., Black, M.A., and Guilford, P.J. (2017b) Towards HDGC chemoprevention: vulnerabilities in e-cadherin-negative cells identified by genome-wide interrogation of isogenic cell lines and whole tumors. Submitted to *Cancer Prev Res*.
- Kozlov, K.N., Gursky, V.V., Kulakovskiy, I.V., and Samsonova, M.G. (2015) Sequence-based model of gap gene regulation network. *BMC Genomics*, **15**(Suppl 12): S6.
- Kranthi, S., Rao, S., and Manimaran, P. (2013) Identification of synthetic lethal pairs in biological systems through network information centrality. *Mol BioSyst*, **9**(8): 2163–2167.
- Lander, E.S. (2011) Initial impact of the sequencing of the human genome. *Nature*, **470**(7333): 187–197.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**(6822): 860–921.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, **10**(3): R25.

- Latora, V. and Marchiori, M. (2001) Efficient behavior of small-world networks. *Phys Rev Lett*, **87**: 198701.
- Laufer, C., Fischer, B., Billmann, M., Huber, W., and Boutros, M. (2013) Mapping genetic interactions in human cancer cells with rnai and multiparametric phenotyping. *Nat Methods*, **10**(5): 427–31.
- Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014) voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*, **15**(2): R29.
- Lawrence, M.S., Sougnez, C., Lichtenstein, L., Cibulskis, K., Lander, E., Gabriel, S.B., Getz, G., Ally, A., Balasundaram, M., Birol, I., *et al.* (2015) Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*, **517**(7536): 576–582.
- Le Meur, N. and Gentleman, R. (2008) Modeling synthetic lethality. *Genome Biol*, **9**(9): R135.
- Le Meur, N., Jiang, Z., Liu, T., Mar, J., and Gentleman, R.C. (2014) Slgi: Synthetic lethal genetic interaction. r package version 1.26.0.
- Lee, A.Y., Perreault, R., Harel, S., Boulier, E.L., Suderman, M., Hallett, M., and Jenna, S. (2010a) Searching for signaling balance through the identification of genetic interactors of the rab guanine-nucleotide dissociation inhibitor gdi-1. *PLoS ONE*, **5**(5): e10624.
- Lee, I., Lehner, B., Vavouri, T., Shin, J., Fraser, A.G., and Marcotte, E.M. (2010b) Predicting genetic modifier loci using functional gene networks. *Genome Research*, **20**(8): 1143–1153.
- Lee, I. and Marcotte, E.M. (2009) Effects of functional bias on supervised learning of a gene network model. *Methods Mol Biol*, **541**: 463–75.
- Lee, M.J., Ye, A.S., Gardino, A.K., Heijink, A.M., Sorger, P.K., MacBeath, G., and Yaffe, M.B. (2012) Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell*, **149**(4): 780–94.
- Lehner, B., Crombie, C., Tischler, J., Fortunato, A., and Fraser, A.G. (2006) Systematic mapping of genetic interactions in caenorhabditis elegans identifies common modifiers of diverse signaling pathways. *Nat Genet*, **38**(8): 896–903.

- Li, X.J., Mishra, S.K., Wu, M., Zhang, F., and Zheng, J. (2014) Syn-lethality: An integrative knowledge base of synthetic lethality towards discovery of selective anticancer therapies. *Biomed Res Int*, **2014**: 196034.
- Linehan, W.M., Spellman, P.T., Ricketts, C.J., Creighton, C.J., Fei, S.S., Davis, C., Wheeler, D.A., Murray, B.A., Schmidt, L., Vocke, C.D., *et al.* (2016) Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. *N Engl J Med*, **374**(2): 135–145.
- Lokody, I. (2014) Computational modelling: A computational crystal ball. *Nature Reviews Cancer*, **14**(10): 649–649.
- Lord, C.J., Tutt, A.N., and Ashworth, A. (2015) Synthetic lethality and cancer therapy: lessons learned from the development of PARP inhibitors. *Annu Rev Med*, **66**: 455–470.
- Lu, X., Kensche, P.R., Huynen, M.A., and Notebaart, R.A. (2013) Genome evolution predicts genetic interactions in protein complexes and reveals cancer drug targets. *Nat Commun*, **4**: 2124.
- Lu, X., Megchelenbrink, W., Notebaart, R.A., and Huynen, M.A. (2015) Predicting human genetic interactions from cancer genome evolution. *PLoS One*, **10**(5): e0125795.
- Lum, P.Y., Armour, C.D., Stepaniants, S.B., Cavet, G., Wolf, M.K., Butler, J.S., Hinshaw, J.C., Garnier, P., Prestwich, G.D., Leonardson, A., *et al.* (2004) Discovering modes of action for therapeutic compounds using a genome-wide screen of yeast heterozygotes. *Cell*, **116**(1): 121–137.
- Luo, J., Solimini, N.L., and Elledge, S.J. (2009) Principles of Cancer Therapy: Oncogene and Non-oncogene Addiction. *Cell*, **136**(5): 823–837.
- Machado, J., Olivera, C., Carvalh, R., Soares, P., Berx, G., Caldas, C., Sercuca, R., Carneiro, F., and Sorbrinho-Simoes, M. (2001) E-cadherin gene (*cdh1*) promoter methylation as the second hit in sporadic diffuse gastric carcinoma. *Oncogene*, **20**: 1525–1528.
- Masciari, S., Larsson, N., Senz, J., Boyd, N., Kaurah, P., Kandel, M.J., Harris, L.N., Pinheiro, H.C., Troussard, A., Miron, P., *et al.* (2007) Germline e-cadherin mutations in familial lobular breast cancer. *J Med Genet*, **44**(11): 726–31.

- Mattison, J., van der Weyden, L., Hubbard, T., and Adams, D.J. (2009) Cancer gene discovery in mouse and man. *Biochim Biophys Acta*, **1796**(2): 140–161.
- Maxam, A.M. and Gilbert, W. (1977) A new method for sequencing DNA. *Proceedings of the National Academy of Science*, **74**(2): 560–564.
- McCourt, C.M., McArt, D.G., Mills, K., Catherwood, M.A., Maxwell, P., Waugh, D.J., Hamilton, P., O’Sullivan, J.M., and Salto-Tellez, M. (2013) Validation of next generation sequencing technologies in comparison to current diagnostic gold standards for BRAF, EGFR and KRAS mutational analysis. *PLoS ONE*, **8**(7): e69604.
- McLachlan, J., George, A., and Banerjee, S. (2016) The current status of parp inhibitors in ovarian cancer. *Tumori*, **102**(5): 433–440.
- McLendon, R., Friedman, A., Bigner, D., Van Meir, E.G., Brat, D.J., Mastrogianakis, G.M., Olson, J.J., Mikkelsen, T., Lehman, N., Aldape, K., *et al.* (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, **455**(7216): 1061–1068.
- Merlos-Suarez, A., Barriga, F.M., Jung, P., Iglesias, M., Cespedes, M.V., Rossell, D., Sevillano, M., Hernando-Momblona, X., da Silva-Diz, V., Munoz, P., *et al.* (2011) The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. *Cell Stem Cell*, **8**(5): 511–524.
- Miles, D.W. (2001) Update on HER-2 as a target for cancer therapy: herceptin in the clinical setting. *Breast Cancer Res*, **3**(6): 380–384.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*, **5**(7): 621–628.
- Muzny, D.M., Bainbridge, M.N., Chang, K., Dinh, H.H., Drummond, J.A., Fowler, G., Kovar, C.L., Lewis, L.R., Morgan, M.B., Newsham, I.F., *et al.* (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, **487**(7407): 330–337.
- Neeley, E.S., Kornblau, S.M., Coombes, K.R., and Baggerly, K.A. (2009) Variable slope normalization of reverse phase protein arrays. *Bioinformatics*, **25**(11): 1384.

- Noonan, J.P., Coop, G., Kudaravalli, S., Smith, D., Krause, J., Alessi, J., Chen, F., Platt, D., Pääbo, S., Pritchard, J.K., *et al.* (2006) Sequencing and analysis of Neanderthal genomic DNA. *Science*, **314**(5802): 1113–1118.
- Novomestky, F. (2012) *matrixcalc: Collection of functions for matrix calculations*. R package version 1.0-3.
- Oliveira, C., Senz, J., Kaurah, P., Pinheiro, H., Sanges, R., Haegert, A., Corso, G., Schouten, J., Fitzgerald, R., Vogelsang, H., *et al.* (2009) Germline *cdh1* deletions in hereditary diffuse gastric cancer families. *Human Molecular Genetics*, **18**(9): 1545–1555.
- Oliveira, C., Seruca, R., Hoogerbrugge, N., Ligtenberg, M., and Carneiro, F. (2013) Clinical utility gene card for: Hereditary diffuse gastric cancer (HDGC). *Eur J Hum Genet*, **21**(8).
- Oxford Nanopore Technologies (Nanopore) (2017) Oxford Nanopore Technologies. <https://nanoporetech.com/>. Accessed: 27/03/2017.
- PacBio (PacBio) (2017) Pacific Biosciences. <http://www.pacb.com/>. Accessed: 27/03/2017.
- Pandey, G., Zhang, B., Chang, A.N., Myers, C.L., Zhu, J., Kumar, V., and Schadt, E.E. (2010) An integrative multi-network and multi-classifier approach to predict genetic interactions. *PLoS Comput Biol*, **6**(9).
- Parker, J., Mullins, M., Cheung, M., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., *et al.* (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology*, **27**(8): 1160–1167.
- Peltonen, L. and McKusick, V.A. (2001) Genomics and medicine. Dissecting human disease in the postgenomic era. *Science*, **291**(5507): 1224–1229.
- Pereira, B., Chin, S.F., Rueda, O.M., Vollan, H.K., Provenzano, E., Bardwell, H.A., Pugh, M., Jones, L., Russell, R., Sammut, S.J., *et al.* (2016) Erratum: The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. *Nat Commun*, **7**: 11908.
- Perou, C.M., Sørlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., *et al.* (2000) Molecular portraits of human breast tumours. *Nature*, **406**(6797): 747–752.

- Pleasance, E.D., Cheetham, R.K., Stephens, P.J., McBride, D.J., Humphray, S.J., Greenman, C.D., Varela, I., Lin, M.L., Ordóñez, G.R., Bignell, G.R., *et al.* (2010) A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature*, **463**(7278): 191–196.
- Polyak, K. and Weinberg, R.A. (2009) Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nat Rev Cancer*, **9**(4): 265–73.
- Prahalad, A., Sun, C., Huang, S., Di Nicolantonio, F., Salazar, R., Zecchin, D., Beijersbergen, R.L., Bardelli, A., and Bernards, R. (2012) Unresponsiveness of colon cancer to braf(v600e) inhibition through feedback activation of egfr. *Nature*, **483**(7387): 100–3.
- Quantum Biosystems Inc. (Quantum Biosystems) (2017) Quantum Biosystems, . <http://www.quantumbiosystems.com/>. Accessed: 27/03/2017.
- R Core Team (2016) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. R version 3.3.2.
- Ravnan, M.C. and Matalaka, M.S. (2012) Vemurafenib in patients with braf v600e mutation-positive advanced melanoma. *Clin Ther*, **34**(7): 1474–86.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, **43**(7): e47.
- Robin, J.D., Ludlow, A.T., LaRanger, R., Wright, W.E., and Shay, J.W. (2016) Comparison of DNA Quantification Methods for Next Generation Sequencing. *Sci Rep*, **6**: 24067.
- Robinson, M.D. and Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*, **11**(3): R25.
- Roguev, A., Bandyopadhyay, S., Zofall, M., Zhang, K., Fischer, T., Collins, S.R., Qu, H., Shales, M., Park, H.O., Hayles, J., *et al.* (2008) Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science*, **322**(5900): 405–10.
- Rothberg, J.M. and Leamon, J.H. (2008) The development and impact of 454 sequencing. *Nat Biotechnol*, **26**(10): 1117–1124.

- Rung, J. and Brazma, A. (2013) Reuse of public genome-wide gene expression data. *Nat Rev Genet*, **14**(2): 89–99.
- Rustici, G., Kolesnikov, N., Brandizi, M., Burdett, T., Dylag, M., Emam, I., Farne, A., Hastings, E., Ison, J., Keays, M., *et al.* (2013) ArrayExpress update—trends in database growth and links to data analysis tools. *Nucleic Acids Res*, **41**(Database issue): D987–990.
- Ryan, C., Lord, C., and Ashworth, A. (2014) Daisy: Picking synthetic lethals from cancer genomes. *Cancer Cell*, **26**(3): 306–308.
- Sander, J.D. and Joung, J.K. (2014) Crispr-cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*, **32**(4): 347–55.
- Sanger, F. and Coulson, A. (1975) A rapid method for determining sequences in dna by primed synthesis with dna polymerase. *Journal of Molecular Biology*, **94**(3): 441 – 448.
- Scheuer, L., Kauff, N., Robson, M., Kelly, B., Barakat, R., Satagopan, J., Ellis, N., Hensley, M., Boyd, J., Borgen, P., *et al.* (2002) Outcome of preventive surgery and screening for breast and ovarian cancer in BRCA mutation carriers. *J Clin Oncol*, **20**(5): 1260–1268.
- Semb, H. and Christofori, G. (1998) The tumor-suppressor function of e-cadherin. *Am J Hum Genet*, **63**(6): 1588–93.
- Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T. (2005) Rocr: visualizing classifier performance in r. *Bioinformatics*, **21**(20): 7881.
- Slurm development team (Slurm) (2017) Slurm workload manager. <https://slurm.schedmd.com/>. Accessed: 25/03/2017.
- Sørli, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., *et al.* (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA*, **98**(19): 10869–10874.
- Stajich, J.E. and Lapp, H. (2006) Open source tools and toolkits for bioinformatics: significance, and where are we? *Brief Bioinformatics*, **7**(3): 287–296.

- Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009) The cancer genome. *Nature*, **458**(7239): 719–724.
- Ström, C. and Helleday, T. (2012) Strategies for the use of poly(adenosine diphosphate ribose) polymerase (parp) inhibitors in cancer therapy. *Biomolecules*, **2**(4): 635–649.
- Sun, C., Wang, L., Huang, S., Heynen, G.J.J.E., Prahallad, A., Robert, C., Haanen, J., Blank, C., Wesseling, J., Willems, S.M., *et al.* (2014) Reversible and adaptive resistance to braf(v600e) inhibition in melanoma. *Nature*, **508**(7494): 118–122.
- Taylor, I.W., Linding, R., Warde-Farley, D., Liu, Y., Pesquita, C., Faria, D., Bull, S., Pawson, T., Morris, Q., and Wrana, J.L. (2009) Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nat Biotechnol*, **27**(2): 199–204.
- Telford, B.J., Chen, A., Beetham, H., Frick, J., Brew, T.P., Gould, C.M., Single, A., Godwin, T., Simpson, K.J., and Guilford, P. (2015) Synthetic lethal screens identify vulnerabilities in gpcr signalling and cytoskeletal organization in e-cadherin-deficient cells. *Mol Cancer Ther*, **14**(5): 1213–1223.
- The 1000 Genomes Project Consortium (1000 Genomes) (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**(7319): 1061–1073.
- The Cancer Genome Atlas Research Network (TCGA) (2012) Comprehensive molecular portraits of human breast tumours. *Nature*, **490**(7418): 61–70.
- The Cancer Genome Atlas Research Network (TCGA) (2017) The Cancer Genome Atlas Project. <https://cancergenome.nih.gov/>. Accessed: 26/03/2017.
- The Cancer Society of New Zealand (Cancer Society of NZ) (2017) What is cancer? <https://otago-southland.cancernz.org.nz/en/cancer-information/other-links/what-is-cancer-3/>. Accessed: 22/03/2017.
- The Catalogue Of Somatic Mutations In Cancer (COSMIC) (2016) Cosmic: The catalogue of somatic mutations in cancer. <http://cancer.sanger.ac.uk/cosmic>. Release 79 (23/08/2016), Accessed: 05/02/2017.
- The Comprehensive R Archive Network (CRAN) (2017) Cran. <https://cran.r-project.org/>. Accessed: 24/03/2017.

- The ENCODE Project Consortium (ENCODE) (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*, **306**(5696): 636–640.
- The International Cancer Genome Consortium (ICGC) (2017) ICGC Data Portal. <https://dcc.icgc.org/>. Accessed: 26/03/2017.
- Thermo Fisher Scientific (ThermoFisher) (2017a) Ion Proton System for Next Generation Sequencing. <https://www.thermofisher.com>. Accessed: 27/03/2017.
- Thermo Fisher Scientific (ThermoFisher) (2017b) SOLiD Next Generation Sequencing. <https://www.thermofisher.com>. Accessed: 27/03/2017.
- The National Cancer Institute (NCI) (2015) The genetics of cancer. <https://www.cancer.gov/about-cancer/causes-prevention/genetics>. Published: 22/04/2015, Accessed: 22/03/2017.
- The New Zealand eScience Infrastructure (NeSI) (2017) NeSI. <https://www.nesi.org.nz/>. Accessed: 25/03/2017.
- The Pharmaceutical Management Agency (PHARMAC) (2016) Approval of multi-product funding proposal with roche.
- Tierney, L., Rossini, A.J., Li, N., and Sevcikova, H. (2015) *snow: Simple Network of Workstations*. R package version 0.4-2.
- Tiong, K.L., Chang, K.C., Yeh, K.T., Liu, T.Y., Wu, J.H., Hsieh, P.H., Lin, S.H., Lai, W.Y., Hsu, Y.C., Chen, J.Y., *et al.* (2014) Csnk1e/ctnnb1 are synthetic lethal to tp53 in colorectal cancer and are markers for prognosis. *Neoplasia*, **16**(5): 441–50.
- Tischler, J., Lehner, B., and Fraser, A.G. (2008) Evolutionary plasticity of genetic interaction networks. *Nat Genet*, **40**(4): 390–391.
- Tomasetti, C. and Vogelstein, B. (2015) Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science*, **347**(6217): 78–81.
- Tong, A.H., Evangelista, M., Parsons, A.B., Xu, H., Bader, G.D., Page, N., Robinson, M., Raghibizadeh, S., Hogue, C.W., Bussey, H., *et al.* (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science*, **294**(5550): 2364–8.

- Tong, A.H., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G.F., Brost, R.L., Chang, M., *et al.* (2004) Global mapping of the yeast genetic interaction network. *Science*, **303**(5659): 808–13.
- Travers, J. and Milgram, S. (1969) An experimental study of the small world problem. *Sociometry*, **32**(4): 425–443.
- Tsai, H.C., Li, H., Van Neste, L., Cai, Y., Robert, C., Rassool, F.V., Shin, J.J., Harbom, K.M., Beaty, R., Pappou, E., *et al.* (2012) Transient low doses of dna-demethylating agents exert durable antitumor effects on hematological and epithelial tumor cells. *Cancer Cell*, **21**(3): 430–46.
- Tutt, A., Robson, M., Garber, J.E., Domchek, S.M., Audeh, M.W., Weitzel, J.N., Friedlander, M., Arun, B., Loman, N., Schmutzler, R.K., *et al.* (2010) Oral poly(adp-ribose) polymerase inhibitor olaparib in patients with brca1 or brca2 mutations and advanced breast cancer: a proof-of-concept trial. *Lancet*, **376**(9737): 235–44.
- van der Meer, R., Song, H.Y., Park, S.H., Abdulkadir, S.A., and Roh, M. (2014) RNAi screen identifies a synthetic lethal interaction between PIM1 overexpression and PLK1 inhibition. *Clinical Cancer Research*, **20**(12): 3211–3221.
- van Steen, K. (2012) Travelling the world of genegene interactions. *Briefings in Bioinformatics*, **13**(1): 1–19.
- van Steen, M. (2010) *Graph Theory and Complex Networks: An Introduction*. Maarten van Steen, VU Amsterdam.
- Vapnik, V.N. (1995) *The nature of statistical learning theory*. Springer-Verlag New York, Inc.
- Vargas, J.J., Gusella, G., Najfeld, V., Klotman, M., and Cara, A. (2004) Novel integrase-defective lentiviral episomal vectors for gene transfer. *Hum Gene Ther*, **15**: 361–372.
- Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., *et al.* (2001) The sequence of the human genome. *Science*, **291**(5507): 1304–1351.
- Vizeacoumar, F.J., Arnold, R., Vizeacoumar, F.S., Chandrashekhar, M., Buzina, A., Young, J.T., Kwan, J.H., Sayad, A., Mero, P., Lawo, S., *et al.* (2013) A negative

- genetic interaction map in isogenic cancer cell lines reveals cancer cell vulnerabilities. *Mol Syst Biol*, **9**: 696.
- Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., and Kinzler, K.W. (2013) Cancer genome landscapes. *Science*, **339**(6127): 1546–1558.
- Vos, C.B., Cleton-Jansen, A.M., Berx, G., de Leeuw, W.J., ter Haar, N.T., van Roy, F., Cornelisse, C.J., Peterse, J.L., and van de Vijver, M.J. (1997) E-cadherin inactivation in lobular carcinoma in situ of the breast: an early event in tumorigenesis. *Br J Cancer*, **76**(9): 1131–3.
- Wadman, M. and Watson, J. (2008) James Watson’s genome sequenced at high speed. *Nature*, **452**(7189): 788.
- 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.
- Warburg, O. (1956) On the origin of cancer cells. *Science*, **123**(3191): 390–314.
- 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.
- Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.J., Makhijani, V., Roth, G.T., *et al.* (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature*, **452**(7189): 872–876.
- Wickham, H. and Chang, W. (2016) *devtools: Tools to Make Developing R Packages Easier*. R package version 1.12.0.
- Wickham, H., Danenberg, P., and Eugster, M. (2017) *roxygen2: In-Line Documentation for R*. R package version 6.0.1.
- Wong, S.L., Zhang, L.V., Tong, A.H.Y., Li, Z., Goldberg, D.S., King, O.D., Lesage, G., Vidal, M., Andrews, B., Bussey, H., *et al.* (2004) Combining biological networks to predict genetic interactions. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(44): 15682–15687.
- World Health Organization (WHO) (2017) Fact sheet: Cancer. <http://www.who.int/mediacentre/factsheets/fs297/en/>. Updated February 2017, Accessed: 22/03/2017.
- Wu, M., Li, X., Zhang, F., Li, X., Kwoh, C.K., and Zheng, J. (2014) In silico prediction of synthetic lethality by meta-analysis of genetic interactions, functions, and pathways in yeast and human cancer. *Cancer Inform*, **13**(Suppl 3): 71–80.
- Yu, H. (2002) Rmpi: Parallel statistical computing in r. *R News*, **2**(2): 10–14.
- Zhang, F., Wu, M., Li, X.J., Li, X.L., Kwoh, C.K., and Zheng, J. (2015) Predicting essential genes and synthetic lethality via influence propagation in signaling pathways of cancer cell fates. *J Bioinform Comput Biol*, **13**(3): 1541002.
- Zhang, J., Baran, J., Cros, A., Guberman, J.M., Haider, S., Hsu, J., Liang, Y., Rivkin, E., Wang, J., Whitty, B., *et al.* (2011) International cancer genome consortium data

portala one-stop shop for cancer genomics data. *Database: The Journal of Biological Databases and Curation*, **2011**: bar026.

Zhong, W. and Sternberg, P.W. (2006) Genome-wide prediction of c. elegans genetic interactions. *Science*, **311**(5766): 1481–1484.

Zweig, M.H. and Campbell, G. (1993) Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, **39**(4): 561–577.