

# Contents

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

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

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

4.2.1.4	Comparison of screen at pathway level . . . . .	129
4.2.1.4.1	Resampling of genes for pathway enrichment . .	131
4.3	Metagene Analysis . . . . .	137
4.3.1	Pathway expression . . . . .	137
4.3.2	Somatic mutation . . . . .	140
4.3.3	Mutation locus . . . . .	141
4.3.4	Synthetic lethal metagenes . . . . .	143
4.4	Replication in stomach cancer . . . . .	145
4.4.1	Synthetic Lethal Genes and Pathways . . . . .	145
4.4.2	Synthetic Lethal Expression Profiles . . . . .	147
4.4.3	Comparison to Primary Screen . . . . .	149
4.4.3.1	Resampling Analysis . . . . .	150
4.4.4	Metagene Analysis . . . . .	150
4.5	Global Synthetic Lethality . . . . .	151
4.5.1	Hub Genes . . . . .	152
4.5.2	Hub Pathways . . . . .	154
4.6	Replication in cell line encyclopaedia . . . . .	155
4.7	Discussion . . . . .	157
4.7.1	Strengths of the SLIPT Methodology . . . . .	157
4.7.2	Syntheic Lethal Pathways for E-cadherin . . . . .	158
4.7.3	Replication and Validation . . . . .	160
4.7.3.1	Integration with siRNA Screening . . . . .	160
4.7.3.2	Replication across Tissues and Cell lines . . . . .	161
4.8	Summary . . . . .	162
<b>5</b>	<b>Synthetic Lethal Pathway Structure</b>	<b>165</b>
5.1	Synthetic Lethal Genes in Reactome Pathways . . . . .	167
5.1.1	The PI3K/AKT Pathway . . . . .	167
5.1.2	The Extracellular Matrix . . . . .	169
5.1.3	G Protein Coupled Receptors . . . . .	172
5.1.4	Gene Regulation and Translation . . . . .	172
5.2	Network Analysis of Synthetic Lethal Genes . . . . .	173
5.2.1	Gene Connectivity and Vertex Degree . . . . .	173
5.2.2	Gene Importance and Centrality . . . . .	175
5.2.2.1	Information Centrality . . . . .	175
5.2.2.2	PageRank Centrality . . . . .	178
5.3	Testing Pathway Structure of Synthetic Lethal Genes . . . . .	179
5.3.1	Hierarchical Pathway Structure . . . . .	179
5.3.1.1	Contextual Hierarchy of PI3K . . . . .	179
5.3.1.2	Testing Contextual Hierarchy of Synthetic Lethal Genes	179
5.3.2	Upstream or Downstream Synthetic Lethality . . . . .	183
5.3.2.1	Measuring Structure of Candidates within PI3K . . .	183
5.3.2.2	Testing Synthetic Lethal Pathway Structure by Resam- pling . . . . .	185
5.4	Discussion . . . . .	186
5.5	Conclusion . . . . .	186

<b>6 Simulation and Modeling of Synthetic Lethal Pathways</b>	<b>186</b>
6.1 Simulations and Modelling Synthetic Lethality in Expression Data . . . . .	189
6.2 Simulations over simple graph structures . . . . .	190
6.2.1 Performance . . . . .	190
6.2.2 Synthetic lethality across graph stuctures . . . . .	190
6.2.3 Performance with inhibition links . . . . .	190
6.2.4 Performance with 20,000 genes . . . . .	190
6.3 Simulations over pathway-based graphs . . . . .	190
6.4 Comparing methods . . . . .	190
6.4.1 SLIPT and Chi-Squared . . . . .	190
6.4.1.1 Correlated query genes . . . . .	190
6.4.2 Correlation . . . . .	190
6.4.3 Bimodality with BiSEp . . . . .	190
<b>7 Discussion</b>	<b>191</b>
7.1 Significance . . . . .	193
7.2 Future Directions . . . . .	194
7.3 Conclusion . . . . .	195
<b>8 Conclusion</b>	<b>197</b>
<b>References</b>	<b>198</b>
<b>A Sample Quality</b>	<b>223</b>
A.1 Sample Correlation . . . . .	223
A.2 Replicate Samples in TCGA Breast . . . . .	226
<b>B Software Used for Thesis</b>	<b>230</b>
<b>C Secondary Screen Data</b>	<b>239</b>
<b>D Mutation Analysis in Breast Cancer</b>	<b>241</b>
D.1 Synthetic Lethal Genes and Pathways . . . . .	241
D.2 Synthetic Lethal Expression Profiles . . . . .	244
D.3 Comparison to Primary Screen . . . . .	247
D.3.1 Resampling Analysis . . . . .	249
D.4 Compare SLIPT genes . . . . .	251
D.5 Metagene Analysis . . . . .	253
D.6 Mutation Variation . . . . .	254
D.6.1 Mutation Frequency . . . . .	254
D.6.2 PI3K Mutation Expression . . . . .	255
<b>E Metagene Expression Profiles</b>	<b>258</b>

<b>F Stomach Expression Analysis</b>	<b>264</b>
F.1 Synthetic Lethal Genes and Pathways . . . . .	264
F.2 Comparison to Primary Screen . . . . .	267
F.2.1 Resampling Analysis . . . . .	269
F.3 Metagene Analysis . . . . .	271
<b>G Stomach Mutation Analysis</b>	<b>272</b>
G.1 Synthetic Lethal Genes and Pathways . . . . .	272
G.2 Synthetic Lethal Expression Profiles . . . . .	275
G.3 Comparison to Primary Screen . . . . .	278
G.3.1 Resampling Analysis . . . . .	280
G.4 Metagene Analysis . . . . .	282
<b>H Global Synthetic Lethality in Stomach Cancer</b>	<b>283</b>
H.1 Hub Genes . . . . .	285
H.2 Hub Pathways . . . . .	286
<b>I Replication in cell line encyclopaedia</b>	<b>287</b>
<b>J Synthetic Lethal Genes in Pathways</b>	<b>292</b>
<b>K Pathway Connectivity for Mutation SLIPT</b>	<b>300</b>
<b>L Information Centrality for Gene Essentiality</b>	<b>304</b>
<b>M Pathway Structure for Mutation SLIPT</b>	<b>307</b>

# List of Figures

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

4.6	Compare SLIPT and siRNA genes with viability . . . . .	126
4.7	Compare SLIPT and siRNA genes with siRNA viability . . . . .	128
4.8	Resampled intersection of SLIPT and siRNA candidates . . . . .	132
4.9	Pathway metagene expression profiles . . . . .	138
4.10	Somatic mutation against PI3K metagene . . . . .	140
4.11	Somatic mutation locus against expression . . . . .	142
4.12	Synthetic lethal expression profiles of stomach samples . . . . .	148
4.13	Synthetic lethal partners across query genes . . . . .	152
5.1	Synthetic Lethality in the PI3K Cascade . . . . .	168
5.2	Synthetic Lethality in the Elastic Fibre Formation Pathway . . . . .	170
5.3	Synthetic Lethality in the Fibrin Clot Formation . . . . .	171
5.4	Synthetic Lethality and Vertex Degree . . . . .	174
5.5	Synthetic Lethality and Centrality . . . . .	176
5.6	Synthetic Lethality and PageRank . . . . .	178
5.7	Structure of PI3K Ranking . . . . .	180
5.8	Synthetic Lethality and Hierarchy Score in PI3K . . . . .	181
5.9	Hierarchy Score in PI3K against Synthetic Lethality in PI3K . . . . .	181
5.10	Structure of Synthetic Lethality in PI3K . . . . .	182
5.11	Structure of Synthetic Lethality Resampling in PI3K . . . . .	184
A.1	Correlation profiles of removed samples . . . . .	224
A.2	Correlation analysis and sample removal . . . . .	225
A.3	Replicate excluded samples . . . . .	226
A.4	Replicate samples with all remaining . . . . .	227
A.5	Replicate samples with some excluded . . . . .	228
A.5	Replicate samples with some excluded . . . . .	229
D.1	Synthetic lethal expression profiles of analysed samples . . . . .	245
D.2	Comparison of mtSLIPT to siRNA . . . . .	247
D.3	Compare mtSLIPT and siRNA genes with correlation . . . . .	251
D.4	Compare mtSLIPT and siRNA genes with correlation . . . . .	251
D.5	Compare mtSLIPT and siRNA genes with siRNA viability . . . . .	252
D.6	Somatic mutation locus . . . . .	254
D.7	Somatic mutation against PIK3CA metagene . . . . .	255
D.8	Somatic mutation against PI3K protein . . . . .	256
D.9	Somatic mutation against AKT protein . . . . .	257
E.1	Pathway metagene expression profiles . . . . .	259
E.2	Expression profiles for constituent genes of PI3K . . . . .	260
E.3	Expression profiles for p53 related genes . . . . .	261
E.4	Expression profiles for estrogen receptor related genes . . . . .	262
E.5	Expression profiles for BRCA related genes . . . . .	263
F.1	Comparison of SLIPT in stomach to siRNA . . . . .	267
G.1	Synthetic lethal expression profiles of stomach samples . . . . .	276

G.2	Comparison of mtSLIPT in stomach to siRNA . . . . .	278
H.1	Synthetic lethal partners across query genes . . . . .	284
J.1	Synthetic Lethality in the PI3K/AKT Pathway . . . . .	292
J.2	Synthetic Lethality in the PI3K/AKT Pathway in Cancer . . . . .	293
J.3	Synthetic Lethality in the Extracellular Matrix . . . . .	294
J.4	Synthetic Lethality in the GPCRs . . . . .	295
J.5	Synthetic Lethality in the GPCR Downstream . . . . .	296
J.6	Synthetic Lethality in the Translation Elongation . . . . .	297
J.7	Synthetic Lethality in the Nonsense-mediated Decay . . . . .	298
J.8	Synthetic Lethality in the 3' UTR . . . . .	299
K.1	Synthetic Lethality and Vertex Degree . . . . .	300
K.2	Synthetic Lethality and Centrality . . . . .	301
K.3	Synthetic Lethality and PageRank . . . . .	302
L.1	Information centrality distribution . . . . .	306
M.1	Synthetic Lethality and Heirarchy Score in PI3K . . . . .	307
M.2	Heirarchy Score in PI3K against Synthetic Lethality in PI3K . . . . .	308
M.3	Structure of Synthetic Lethality in PI3K . . . . .	308
M.4	Structure of Synthetic Lethality Resampling . . . . .	309

# List of Tables

1.1	Methods for Predicting Genetic Interactions . . . . .	37
1.2	Methods for Predicting Synthetic Lethality in Cancer . . . . .	38
1.3	Methods used by Wu <i>et al.</i> (2014) . . . . .	39
2.1	Excluded Samples by Batch and Clinical Characteristics . . . . .	62
2.2	Computers used during Thesis . . . . .	71
2.3	Linux Utilities and Applications used during Thesis . . . . .	71
2.4	R Installations used during Thesis . . . . .	72
2.5	R Packages used during Thesis . . . . .	72
2.6	R Packages Developed during Thesis . . . . .	74
4.1	Candidate synthetic lethal gene partners of <i>CDH1</i> from SLIPT . . . . .	115
4.2	Pathways for <i>CDH1</i> partners from SLIPT . . . . .	117
4.3	Pathway composition for clusters of <i>CDH1</i> partners from SLIPT . . . . .	121
4.4	Pathway composition for <i>CDH1</i> partners from SLIPT and siRNA screening . . . . .	130
4.5	Pathways for <i>CDH1</i> partners from SLIPT . . . . .	134
4.6	Pathways for <i>CDH1</i> partners from SLIPT and siRNA primary screen . . . . .	135
4.7	Candidate synthetic lethal metagenes against <i>CDH1</i> from SLIPT . . . . .	144
4.8	Pathways for <i>CDH1</i> partners from SLIPT in stomach cancer . . . . .	146
4.9	Query synthetic lethal genes with the most SLIPT partners . . . . .	153
4.10	Pathways for genes with the most SLIPT partners . . . . .	154
4.11	Pathways for <i>CDH1</i> partners from SLIPT in CCLE . . . . .	155
4.12	Pathways for <i>CDH1</i> partners from SLIPT in breast CCLE . . . . .	157
5.1	analysis of variance (ANOVA) for Synthetic Lethality and Vertex Degree	175
5.2	ANOVA for Synthetic Lethality and Information Centrality . . . . .	177
5.3	ANOVA for Synthetic Lethality and PageRank Centrality . . . . .	179
5.4	ANOVA for Synthetic Lethality and PI3K Hierarchy . . . . .	182
5.5	Resampling for pathway structure of synthetic lethal detection methods	186
B.1	R Packages used during Thesis . . . . .	230
C.1	Comparing SLIPT genes against Secondary siRNA Screen in breast cancer	239
C.2	Comparing mtSLIPT genes against Secondary siRNA Screen in breast cancer . . . . .	240
C.3	Comparing SLIPT genes against Secondary siRNA Screen in stomach cancer . . . . .	240

D.1	Candidate synthetic lethal gene partners of <i>CDH1</i> from mtSLIPT . . . . .	242
D.2	Pathways for <i>CDH1</i> partners from mtSLIPT . . . . .	243
D.3	Pathway composition for clusters of <i>CDH1</i> partners from mtSLIPT . . . . .	246
D.4	Pathway composition for <i>CDH1</i> partners from mtSLIPT and siRNA . . . . .	248
D.5	Pathways for <i>CDH1</i> partners from mtSLIPT . . . . .	249
D.6	Pathways for <i>CDH1</i> partners from mtSLIPT and siRNA primary screen	250
D.7	Candidate synthetic lethal metagenes against <i>CDH1</i> from mtSLIPT . . . . .	253
F.1	Synthetic lethal gene partners of <i>CDH1</i> from SLIPT in stomach cancer	265
F.2	Pathway composition for clusters of <i>CDH1</i> partners in stomach SLIPT	266
F.3	Pathway composition for <i>CDH1</i> partners from SLIPT and siRNA screening . . . . .	268
F.4	Pathways for <i>CDH1</i> partners from SLIPT in stomach cancer . . . . .	269
F.5	Pathways for <i>CDH1</i> partners from SLIPT in stomach and siRNA screen	270
F.6	Candidate synthetic lethal metagenes against <i>CDH1</i> from SLIPT in stomach cancer . . . . .	271
G.1	Synthetic lethal gene partners of <i>CDH1</i> from mtSLIPT in stomach cancer	273
G.2	Pathways for <i>CDH1</i> partners from mtSLIPT in stomach cancer . . . . .	274
G.3	Pathway composition for clusters of <i>CDH1</i> partners in stomach mtSLIPT	277
G.4	Pathway composition for <i>CDH1</i> partners from mtSLIPT and siRNA . . . . .	279
G.5	Pathways for <i>CDH1</i> partners from mtSLIPT in stomach cancer . . . . .	280
G.6	Pathways for <i>CDH1</i> partners from mtSLIPT in stomach and siRNA screen	281
G.7	Candidate synthetic lethal metagenes against <i>CDH1</i> from mtSLIPT in stomach cancer . . . . .	282
H.1	Query synthetic lethal genes with the most SLIPT partners . . . . .	285
H.2	Pathways for genes with the most SLIPT partners . . . . .	286
I.1	Candidate synthetic lethal gene partners of <i>CDH1</i> from SLIPT in CCLE	288
I.2	Candidate synthetic lethal gene partners of <i>CDH1</i> from SLIPT in breast CCLE . . . . .	289
I.3	Candidate synthetic lethal gene partners of <i>CDH1</i> from SLIPT in stomach CCLE . . . . .	290
I.4	Pathways for <i>CDH1</i> partners from SLIPT in stomach CCLE . . . . .	291
I.5	Pathways for <i>CDH1</i> partners from SLIPT in breast and stomach CCLE	291
K.1	ANOVA for Synthetic Lethality and Vertex Degree . . . . .	303
K.2	ANOVA for Synthetic Lethality and Information Centrality . . . . .	303
K.3	ANOVA for Synthetic Lethality and PageRank Centrality . . . . .	303
L.1	Information centrality for genes and molecules in the Reactome network	305
M.1	ANOVA for Synthetic Lethality and PI3K Hierarchy . . . . .	307
M.2	Resampling for pathway structure of synthetic lethal detection methods	309

# Glossary

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

# **Acronyms**

ANOVA Analysis of Variance.

siRNA Short interfering ribonucleic acid.

# References

- Aarts, M., Bajrami, I., Herrera-Abreu, M.T., Elliott, R., Brough, R., Ashworth, A., Lord, C.J., and Turner, N.C. (2015) Functional genetic screen identifies increased sensitivity to wee1 inhibition in cells with defects in fanconi anemia and hr pathways. *Mol Cancer Ther*, **14**(4): 865–76.
- Abeshouse, A., Ahn, J., Akbani, R., Ally, A., Amin, S., Andry, C.D., Annala, M., Aprikian, A., Armenia, J., Arora, A., *et al.* (2015) The Molecular Taxonomy of Primary Prostate Cancer. *Cell*, **163**(4): 1011–1025.
- 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 Padiatrica*, **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 polyadribose polymerase. *Nature*, **434**(7035): 913–7.
- Burk, R.D., Chen, Z., Saller, C., Tarvin, K., Carvalho, A.L., Scapulatempo-Neto, C., Silveira, H.C., Fregnani, J.H., Creighton, C.J., Anderson, M.L., *et al.* (2017) Integrated genomic and molecular characterization of cervical cancer. *Nature*, **543**(7645): 378–384.
- Bussey, H., Andrews, B., and Boone, C. (2006) From worm genetic networks to complex human diseases. *Nat Genet*, **38**(8): 862–3.
- Butland, G., Babu, M., Diaz-Mejia, J.J., Bohdana, F., Phanse, S., Gold, B., Yang, W., Li, J., Gagarinova, A.G., Pogoutse, O., *et al.* (2008) esga: *E. coli* synthetic genetic array analysis. *Nat Methods*, **5**(9): 789–95.
- Cancer Research UK (2017) Family history and cancer genes. <http://www.cancerresearchuk.org/about-cancer/causes-of-cancer/inherited-cancer-genes-and-increased-cancer-risk/family-history-and-inherited-cancer-genes>. Accessed: 22/03/2017.
- Cancer Cell Line Encyclopedia (CCLE) (2014) Broad-Novartis Cancer Cell Line Encyclopedia. <http://www.broadinstitute.org/ccle>. Accessed: 07/11/2014.
- cBioPortal for Cancer Genomics (cBioPortal) (2017) cBioPortal for Cancer Genomics. <http://www.cbioportal.org/>. Accessed: 26/03/2017.
- Cerami, E.G., Gross, B.E., Demir, E., Rodchenkov, I., Babur, O., Anwar, N., Schultz, N., Bader, G.D., and Sander, C. (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res*, **39**(Database issue): D685–690.
- Chen, A., Beetham, H., Black, M.A., Priya, R., Telford, B.J., Guest, J., Wiggins, G.A.R., Godwin, T.D., Yap, A.S., and Guilford, P.J. (2014) E-cadherin loss alters cytoskeletal organization and adhesion in non-malignant breast cells but is insufficient to induce an epithelial-mesenchymal transition. *BMC Cancer*, **14**(1): 552.
- Chen, K., Yang, D., Li, X., Sun, B., Song, F., Cao, W., Brat, D.J., Gao, Z., Li, H., Liang, H., *et al.* (2015) Mutational landscape of gastric adenocarcinoma in Chinese: implications for prognosis and therapy. *Proc Natl Acad Sci USA*, **112**(4): 1107–1112.

- Chen, S. and Parmigiani, G. (2007) Meta-analysis of BRCA1 and BRCA2 penetrance. *J Clin Oncol*, **25**(11): 1329–1333.
- Chen, X. and Tompa, M. (2010) Comparative assessment of methods for aligning multiple genome sequences. *Nat Biotechnol*, **28**(6): 567–572.
- Cherniack, A.D., Shen, H., Walter, V., Stewart, C., Murray, B.A., Bowlby, R., Hu, X., Ling, S., Soslow, R.A., Broaddus, R.R., *et al.* (2017) Integrated Molecular Characterization of Uterine Carcinosarcoma. *Cancer Cell*, **31**(3): 411–423.
- Chipman, K. and Singh, A. (2009) Predicting genetic interactions with random walks on biological networks. *BMC Bioinformatics*, **10**(1): 17.
- Christofori, G. and Semb, H. (1999) The role of the cell-adhesion molecule E-cadherin as a tumour-suppressor gene. *Trends in Biochemical Sciences*, **24**(2): 73 – 76.
- Ciriello, G., Gatza, M.L., Beck, A.H., Wilkerson, M.D., Rhie, S.K., Pastore, A., Zhang, H., McLellan, M., Yau, C., Kandoth, C., *et al.* (2015) Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell*, **163**(2): 506–519.
- Clark, M.J. (2004) Endogenous Regulator of G Protein Signaling Proteins Suppress G<sub>o</sub>-Dependent  $\mu$ -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 knowledgebase. *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., Pirotowski, J., Shin, S.H., Yoshida, M., Costanzo, M., *et al.* (2013) A comparative genomic approach for identifying synthetic lethal interactions in human cancer. *Cancer Res*, **73**(20): 6128–36.
- Dickson, D. (1999) Wellcome funds cancer database. *Nature*, **401**(6755): 729.
- Dienstmann, R. and Tabernero, J. (2011) *BRAF* as a target for cancer therapy. *Anti-cancer Agents Med Chem*, **11**(3): 285–95.
- Dijkstra, E.W. (1959) A note on two problems in connexion with graphs. *Numerische Mathematik*, **1**(1): 269–271.
- Dixon, S.J., Andrews, B.J., and Boone, C. (2009) Exploring the conservation of synthetic lethal genetic interaction networks. *Commun Integr Biol*, **2**(2): 78–81.
- Dixon, S.J., 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(adp-ribose) polymerase in tumors from BRCA mutation carriers. *N Engl J Med*, **361**(2): 123–34.
- Fong, P.C., Yap, T.A., Boss, D.S., Carden, C.P., Mergui-Roelvink, M., Gourley, C., De Greve, J., Lubinski, J., Shanley, S., Messiou, C., *et al.* (2010) Poly(adp)-ribose polymerase inhibition: frequent durable responses in BRCA carrier ovarian cancer correlating with platinum-free interval. *J Clin Oncol*, **28**(15): 2512–9.
- Forbes, S.A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S., *et al.* (2015) COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res*, **43**(Database issue): D805–811.

- Fraser, A. (2004) Towards full employment: using RNAi to find roles for the redundant. *Oncogene*, **23**(51): 8346–52.
- Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004) A census of human cancer genes. *Nat Rev Cancer*, **4**(3): 177–183.
- Futreal, P.A., Kasprzyk, A., Birney, E., Mullikin, J.C., Wooster, R., and Stratton, M.R. (2001) Cancer and genomics. *Nature*, **409**(6822): 850–852.
- Gao, B. and Roux, P.P. (2015) Translational control by oncogenic signaling pathways. *Biochimica et Biophysica Acta*, **1849**(7): 753–65.
- Gatza, M.L., Kung, H.N., Blackwell, K.L., Dewhirst, M.W., Marks, J.R., and Chi, J.T. (2011) Analysis of tumor environmental response and oncogenic pathway activation identifies distinct basal and luminal features in HER2-related breast tumor subtypes. *Breast Cancer Res*, **13**(3): R62.
- Gatza, M.L., Silva, G.O., Parker, J.S., Fan, C., and Perou, C.M. (2014) An integrated genomics approach identifies drivers of proliferation in luminal-subtype human breast cancer. *Nat Genet*, **46**(10): 1051–1059.
- Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*, **5**(10): R80.
- Genz, A. and Bretz, F. (2009) Computation of multivariate normal and t probabilities. In *Lecture Notes in Statistics*, volume 195. Springer-Verlag, Heidelberg.
- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2016) *mvtnorm: Multivariate Normal and t Distributions*. R package version 1.0-5. URL.
- Gilbert, W. and Maxam, A. (1973) The nucleotide sequence of the lac operator. *Proceedings of the National Academy of Sciences*, **70**(12): 3581–3584.
- Git, A., Dvinge, H., Salmon-Divon, M., Osborne, M., Kutter, C., Hadfield, J., Bertone, P., and Caldas, C. (2010) Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. *RNA*, **16**(5): 991–1006.

Globus (Globus) (2017) Research data management simplified. <https://www.globus.org/>. Accessed: 25/03/2017.

Graziano, F., Humar, B., and Guilford, P. (2003) The role of the E-cadherin gene (*CDH1*) in diffuse gastric cancer susceptibility: from the laboratory to clinical practice. *Annals of Oncology*, **14**(12): 1705–1713.

Güell, O., Sagus, F., and Serrano, M. (2014) Essential plasticity and redundancy of metabolism unveiled by synthetic lethality analysis. *PLoS Comput Biol*, **10**(5): e1003637.

Guilford, P. (1999) E-cadherin downregulation in cancer: fuel on the fire? *Molecular Medicine Today*, **5**(4): 172 – 177.

Guilford, P., Hopkins, J., Harraway, J., McLeod, M., McLeod, N., Harawira, P., Taite, H., Scouller, 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 grid™ (caBIG™). *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 combinatorics at theorie des graphes. *Proc Coll 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.
- Hoechndorf, 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.
- Internationl 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, lkopf, J.C.B. Christopher, and J.S. Alexander (editors), *Advances in kernel methods*, 169–184. MIT Press.
- Ju, Z., Liu, W., Roebuck, P.L., Siwak, D.R., Zhang, N., Lu, Y., Davies, M.A., Akbani, R., Weinstein, J.N., Mills, G.B., *et al.* (2015) Development of a robust classifier for quality control of reverse-phase protein arrays. *Bioinformatics*, **31**(6): 912.
- Kaelin, Jr, W. (2005) The concept of synthetic lethality in the context of anticancer therapy. *Nat Rev Cancer*, **5**(9): 689–98.
- Kaelin, Jr, W. (2009) Synthetic lethality: a framework for the development of wiser cancer therapeutics. *Genome Med*, **1**: 99.
- Kakiuchi, M., Nishizawa, T., Ueda, H., Gotoh, K., Tanaka, A., Hayashi, A., Yamamoto, S., Tatsuno, K., Katoh, H., Watanabe, Y., *et al.* (2014) Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. *Nat Genet*, **46**(6): 583–587.
- Kamada, T. and Kawai, S. (1989) An algorithm for drawing general undirected graphs. *Information Processing Letters*, **31**(1): 7–15.
- Kandoth, C., Schultz, N., Cherniack, A.D., Akbani, R., Liu, Y., Shen, H., Robertson, A.G., Pashtan, I., Shen, R., Benz, C.C., *et al.* (2013) Integrated genomic characterization of endometrial carcinoma. *Nature*, **497**(7447): 67–73.
- Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., *et al.* (2001) Functional annotation of a full-length mouse cDNA collection. *Nature*, **409**(6821): 685–690.
- Kelley, R. and Ideker, T. (2005) Systematic interpretation of genetic interactions using protein networks. *Nat Biotech*, **23**(5): 561–566.
- Kelly, S., Chen, A., Guilford, P., and Black, M. (2017a) Synthetic lethal interaction prediction of target pathways in E-cadherin deficient breast cancers. Submitted to *BMC Genomics*.
- Kelly, S.T. (2013) *Statistical Predictions of Synthetic Lethal Interactions in Cancer*. Dissertation, University of Otago.
- Kelly, S.T., Single, A.B., Telford, B.J., Beetham, H.G., Godwin, T.D., Chen, A., Black, M.A., and Guilford, P.J. (2017b) Towards HDGC chemoprevention: vulnerabilities

in E-cadherin-negative cells identified by genome-wide interrogation of isogenic cell lines and whole tumors. Submitted to *Cancer Prev Res.*

Kozlov, K.N., Gursky, V.V., Kulakovskiy, I.V., and Samsonova, M.G. (2015) Sequence-based model of gap gene regulation network. *BMC Genomics*, **15**(Suppl 12): S6.

Kranthi, S., Rao, S., and Manimaran, P. (2013) Identification of synthetic lethal pairs in biological systems through network information centrality. *Mol BioSyst*, **9**(8): 2163–2167.

Lander, E.S. (2011) Initial impact of the sequencing of the human genome. *Nature*, **470**(7333): 187–197.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**(6822): 860–921.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, **10**(3): R25.

Latora, V. and Marchiori, M. (2001) Efficient behavior of small-world networks. *Phys Rev Lett*, **87**: 198701.

Laufer, C., Fischer, B., Billmann, M., Huber, W., and Boutros, M. (2013) Mapping genetic interactions in human cancer cells with RNAi and multiparametric phenotyping. *Nat Methods*, **10**(5): 427–31.

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014) voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*, **15**(2): R29.

Lawrence, M.S., Sougnez, C., Lichtenstein, L., Cibulskis, K., Lander, E., Gabriel, S.B., Getz, G., Ally, A., Balasundaram, M., Birol, I., et al. (2015) Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*, **517**(7536): 576–582.

Le Meur, N. and Gentleman, R. (2008) Modeling synthetic lethality. *Genome Biol*, **9**(9): R135.

Le Meur, N., Jiang, Z., Liu, T., Mar, J., and Gentleman, R.C. (2014) Slgi: Synthetic lethal genetic interaction. r package version 1.26.0.

- Lee, A.Y., Perreault, R., Harel, S., Boulier, E.L., Suderman, M., Hallett, M., and Jenna, S. (2010a) Searching for signaling balance through the identification of genetic interactors of the rab guanine-nucleotide dissociation inhibitor gdi-1. *PLoS ONE*, **5**(5): e10624.
- Lee, I., Lehner, B., Vavouri, T., Shin, J., Fraser, A.G., and Marcotte, E.M. (2010b) Predicting genetic modifier loci using functional gene networks. *Genome Research*, **20**(8): 1143–1153.
- Lee, I. and Marcotte, E.M. (2009) Effects of functional bias on supervised learning of a gene network model. *Methods Mol Biol*, **541**: 463–75.
- Lee, M.J., Ye, A.S., Gardino, A.K., Heijink, A.M., Sorger, P.K., MacBeath, G., and Yaffe, M.B. (2012) Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell*, **149**(4): 780–94.
- Lehner, B., Crombie, C., Tischler, J., Fortunato, A., and Fraser, A.G. (2006) Systematic mapping of genetic interactions in *caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. *Nat Genet*, **38**(8): 896–903.
- Li, X.J., Mishra, S.K., Wu, M., Zhang, F., and Zheng, J. (2014) Syn-lethality: An integrative knowledge base of synthetic lethality towards discovery of selective anticancer therapies. *Biomed Res Int*, **2014**: 196034.
- Linehan, W.M., Spellman, P.T., Ricketts, C.J., Creighton, C.J., Fei, S.S., Davis, C., Wheeler, D.A., Murray, B.A., Schmidt, L., Vocke, C.D., et al. (2016) Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. *N Engl J Med*, **374**(2): 135–145.
- Lokody, I. (2014) Computational modelling: A computational crystal ball. *Nature Reviews Cancer*, **14**(10): 649–649.
- Lord, C.J., Tutt, A.N., and Ashworth, A. (2015) Synthetic lethality and cancer therapy: lessons learned from the development of PARP inhibitors. *Annu Rev Med*, **66**: 455–470.
- Lu, X., Kensche, P.R., Huynen, M.A., and Notebaart, R.A. (2013) Genome evolution predicts genetic interactions in protein complexes and reveals cancer drug targets. *Nat Commun*, **4**: 2124.

- Lu, X., Megchelenbrink, W., Notebaart, R.A., and Huynen, M.A. (2015) Predicting human genetic interactions from cancer genome evolution. *PLoS One*, **10**(5): e0125795.
- Lum, P.Y., Armour, C.D., Stepaniants, S.B., Cavet, G., Wolf, M.K., Butler, J.S., Hinchshaw, J.C., Garnier, P., Prestwich, G.D., Leonardson, A., *et al.* (2004) Discovering modes of action for therapeutic compounds using a genome-wide screen of yeast heterozygotes. *Cell*, **116**(1): 121–137.
- Luo, J., Solimini, N.L., and Elledge, S.J. (2009) Principles of Cancer Therapy: Oncogene and Non-oncogene Addiction. *Cell*, **136**(5): 823–837.
- Machado, J., Olivera, C., Carvalh, R., Soares, P., Berx, G., Caldas, C., Sercuca, R., Carneiro, F., and Sorbrinho-Simoes, M. (2001) E-cadherin gene (*CDH1*) promoter methylation as the second hit in sporadic diffuse gastric carcinoma. *Oncogene*, **20**: 1525–1528.
- Masciari, S., Larsson, N., Senz, J., Boyd, N., Kaurah, P., Kandel, M.J., Harris, L.N., Pinheiro, H.C., Troussard, A., Miron, P., *et al.* (2007) Germline E-cadherin mutations in familial lobular breast cancer. *J Med Genet*, **44**(11): 726–31.
- Mattison, J., van der Weyden, L., Hubbard, T., and Adams, D.J. (2009) Cancer gene discovery in mouse and man. *Biochim Biophys Acta*, **1796**(2): 140–161.
- Maxam, A.M. and Gilbert, W. (1977) A new method for sequencing DNA. *Proceedings of the National Academy of Science*, **74**(2): 560–564.
- McCourt, C.M., McArt, D.G., Mills, K., Catherwood, M.A., Maxwell, P., Waugh, D.J., Hamilton, P., O'Sullivan, J.M., and Salto-Tellez, M. (2013) Validation of next generation sequencing technologies in comparison to current diagnostic gold standards for BRAF, EGFR and KRAS mutational analysis. *PLoS ONE*, **8**(7): e69604.
- McLachlan, J., George, A., and Banerjee, S. (2016) The current status of parp inhibitors in ovarian cancer. *Tumori*, **102**(5): 433–440.
- McLendon, R., Friedman, A., Bigner, D., Van Meir, E.G., Brat, D.J., Mastrogianakis, G.M., Olson, J.J., Mikkelsen, T., Lehman, N., Aldape, K., *et al.* (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, **455**(7216): 1061–1068.

- Miles, D.W. (2001) Update on HER-2 as a target for cancer therapy: herceptin in the clinical setting. *Breast Cancer Res*, **3**(6): 380–384.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*, **5**(7): 621–628.
- Muzny, D.M., Bainbridge, M.N., Chang, K., Dinh, H.H., Drummond, J.A., Fowler, G., Kovar, C.L., Lewis, L.R., Morgan, M.B., Newsham, I.F., et al. (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, **487**(7407): 330–337.
- Neeley, E.S., Kornblau, S.M., Coombes, K.R., and Baggerly, K.A. (2009) Variable slope normalization of reverse phase protein arrays. *Bioinformatics*, **25**(11): 1384.
- Novomestky, F. (2012) *matrixcalc: Collection of functions for matrix calculations*. R package version 1.0-3.
- Oliveira, C., Senz, J., Kaurah, P., Pinheiro, H., Sanges, R., Haegert, A., Corso, G., Schouten, J., Fitzgerald, R., Vogelsang, H., et al. (2009) Germline *CDH1* deletions in hereditary diffuse gastric cancer families. *Human Molecular Genetics*, **18**(9): 1545–1555.
- Oliveira, C., Seruca, R., Hoogerbrugge, N., Ligtenberg, M., and Carneiro, F. (2013) Clinical utility gene card for: Hereditary diffuse gastric cancer (HDGC). *Eur J Hum Genet*, **21**(8).
- Pandey, G., Zhang, B., Chang, A.N., Myers, C.L., Zhu, J., Kumar, V., and Schadt, E.E. (2010) An integrative multi-network and multi-classifier approach to predict genetic interactions. *PLoS Comput Biol*, **6**(9).
- Parker, J., Mullins, M., Cheung, M., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., et al. (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology*, **27**(8): 1160–1167.
- Peltonen, L. and McKusick, V.A. (2001) Genomics and medicine. Dissecting human disease in the postgenomic era. *Science*, **291**(5507): 1224–1229.
- Pereira, B., Chin, S.F., Rueda, O.M., Vollan, H.K., Provenzano, E., Bardwell, H.A., Pugh, M., Jones, L., Russell, R., Sammut, S.J., et al. (2016) Erratum: The somatic

mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. *Nat Commun*, **7**: 11908.

Perou, C.M., Sørlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., *et al.* (2000) Molecular portraits of human breast tumours. *Nature*, **406**(6797): 747–752.

Pleasance, E.D., Cheetham, R.K., Stephens, P.J., McBride, D.J., Humphray, S.J., Greenman, C.D., Varela, I., Lin, M.L., Ordonez, 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.

Prahallaad, A., Sun, C., Huang, S., Di Nicolantonio, F., Salazar, R., Zecchin, D., Beijersbergen, R.L., Bardelli, A., and Bernards, R. (2012) Unresponsiveness of colon cancer to *BRAF*(v600e) inhibition through feedback activation of egfr. *Nature*, **483**(7387): 100–3.

R Core Team (2016) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. R version 3.3.2.

Ravnan, M.C. and Matalka, M.S. (2012) Vemurafenib in patients with *BRAF* v600e mutation-positive advanced melanoma. *Clin Ther*, **34**(7): 1474–86.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, **43**(7): e47.

Robin, J.D., Ludlow, A.T., LaRanger, R., Wright, W.E., and Shay, J.W. (2016) Comparison of DNA Quantification Methods for Next Generation Sequencing. *Sci Rep*, **6**: 24067.

Robinson, M.D. and Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*, **11**(3): R25.

Roguev, A., Bandyopadhyay, S., Zofall, M., Zhang, K., Fischer, T., Collins, S.R., Qu, H., Shales, M., Park, H.O., Hayles, J., *et al.* (2008) Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science*, **322**(5900): 405–10.

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

- Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009) The cancer genome. *Nature*, **458**(7239): 719–724.
- Ström, C. and Helleday, T. (2012) Strategies for the use of poly(adenosine diphosphate ribose) polymerase (parp) inhibitors in cancer therapy. *Biomolecules*, **2**(4): 635–649.
- Sun, C., Wang, L., Huang, S., Heynen, G.J.J.E., Prahallad, A., Robert, C., Haanen, J., Blank, C., Wesseling, J., Willems, S.M., *et al.* (2014) Reversible and adaptive resistance to *BRAF*(v600e) inhibition in melanoma. *Nature*, **508**(7494): 118–122.
- Taylor, I.W., Linding, R., Warde-Farley, D., Liu, Y., Pesquita, C., Faria, D., Bull, S., Pawson, T., Morris, Q., and Wrana, J.L. (2009) Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nat Biotechnol*, **27**(2): 199–204.
- Telford, B.J., Chen, A., Beetham, H., Frick, J., Brew, T.P., Gould, C.M., Single, A., Godwin, T., Simpson, K.J., and Guilford, P. (2015) Synthetic lethal screens identify vulnerabilities in gpcr signalling and cytoskeletal organization in E-cadherin-deficient cells. *Mol Cancer Ther*, **14**(5): 1213–1223.
- The 1000 Genomes Project Consortium (1000 Genomes) (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**(7319): 1061–1073.
- The Cancer Genome Atlas Research Network (TCGA) (2012) Comprehensive molecular portraits of human breast tumours. *Nature*, **490**(7418): 61–70.
- The Cancer Genome Atlas Research Network (TCGA) (2017a) The Cancer Genome Atlas Project. <https://cancergenome.nih.gov/>. Accessed: 26/03/2017.
- The Cancer Genome Atlas Research Network (TCGA) (2017b) The Cancer Genome Atlas Project Data Portal. <https://tcga-data.nci.nih.gov/>. Accessed: 06/02/2017 (via cBioPortal).
- The Cancer Society of New Zealand (Cancer Society of NZ) (2017) What is cancer? <https://otago-southland.cancernz.org.nz/en/cancer-information/other-links/what-is-cancer-3/>. Accessed: 22/03/2017.
- The Catalogue Of Somatic Mutations In Cancer (COSMIC) (2016) Cosmic: The catalogue of somatic mutations in cancer. <http://cancer.sanger.ac.uk/cosmic>. Release 79 (23/08/2016), Accessed: 05/02/2017.

The Comprehensive R Archive Network (CRAN) (2017) Cran. <https://cran.r-project.org/>. Accessed: 24/03/2017.

The ENCODE Project Consortium (ENCODE) (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*, **306**(5696): 636–640.

The Internation Cancer Genome Consortium (ICGC) (2017) ICGC Data Portal. <https://dcc.icgc.org/>. Accessed: 06/02/2017.

The National Cancer Institute (NCI) (2015) The genetics of cancer. <https://www.cancer.gov/about-cancer/causes-prevention/genetics>. Published: 22/04/2015, Accessed: 22/03/2017.

The New Zealand eScience Infrastructure (NeSI) (2017) NeSI. <https://www.nesi.org.nz/>. Accessed: 25/03/2017.

The Pharmaceutical Management Agency (PHARMAC) (2016) Approval of multi-product funding proposal with roche.

Tierney, L., Rossini, A.J., Li, N., and Sevcikova, H. (2015) *snow: Simple Network of Workstations*. R package version 0.4-2.

Tiong, K.L., Chang, K.C., Yeh, K.T., Liu, T.Y., Wu, J.H., Hsieh, P.H., Lin, S.H., Lai, W.Y., Hsu, Y.C., Chen, J.Y., et al. (2014) Csnk1e/ctnnb1 are synthetic lethal to tp53 in colorectal cancer and are markers for prognosis. *Neoplasia*, **16**(5): 441–50.

Tischler, J., Lehner, B., and Fraser, A.G. (2008) Evolutionary plasticity of genetic interaction networks. *Nat Genet*, **40**(4): 390–391.

Tomasetti, C. and Vogelstein, B. (2015) Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science*, **347**(6217): 78–81.

Tong, A.H., Evangelista, M., Parsons, A.B., Xu, H., Bader, G.D., Page, N., Robinson, M., Raghibizadeh, S., Hogue, C.W., Bussey, H., et al. (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science*, **294**(5550): 2364–8.

Tong, A.H., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berziz, 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(adribose) polymerase inhibitor olaparib in patients with *BRCA1* or *BRCA2* mutations and advanced breast cancer: a proof-of-concept trial. *Lancet*, **376**(9737): 235–44.
- van der Meer, R., Song, H.Y., Park, S.H., Abdulkadir, S.A., and Roh, M. (2014) RNAi screen identifies a synthetic lethal interaction between PIM1 overexpression and PLK1 inhibition. *Clinical Cancer Research*, **20**(12): 3211–3221.
- van Steen, K. (2012) Travelling the world of genegene interactions. *Briefings in Bioinformatics*, **13**(1): 1–19.
- van Steen, M. (2010) *Graph Theory and Complex Networks: An Introduction*. Maarten van Steen, VU Amsterdam.
- Vapnik, V.N. (1995) *The nature of statistical learning theory*. Springer-Verlag New York, Inc.
- Vargas, J.J., Gusella, G., Najfeld, V., Klotman, M., and Cara, A. (2004) Novel integrase-defective lentiviral episomal vectors for gene transfer. *Hum Gene Ther*, **15**: 361–372.
- Vizeacoumar, F.J., Arnold, R., Vizeacoumar, F.S., Chandrashekhar, M., Buzina, A., Young, J.T., Kwan, J.H., Sayad, A., Mero, P., Lawo, S., et al. (2013) A negative genetic interaction map in isogenic cancer cell lines reveals cancer cell vulnerabilities. *Mol Syst Biol*, **9**: 696.
- Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., and Kinzler, K.W. (2013) Cancer genome landscapes. *Science*, **339**(6127): 1546–1558.
- Vos, C.B., Cleton-Jansen, A.M., Berx, G., de Leeuw, W.J., ter Haar, N.T., van Roy, F., Cornelisse, C.J., Peterse, J.L., and van de Vijver, M.J. (1997) E-cadherin inactivation

in lobular carcinoma in situ of the breast: an early event in tumorigenesis. *Br J Cancer*, **76**(9): 1131–3.

Wang, K., Singh, D., Zeng, Z., Coleman, S.J., Huang, Y., Savich, G.L., He, X., Mieczkowski, P., Grimm, S.A., Perou, C.M., *et al.* (2010) MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res*, **38**(18): e178.

Wang, K., Yuen, S.T., Xu, J., Lee, S.P., Yan, H.H., Shi, S.T., Siu, H.C., Deng, S., Chu, K.M., Law, S., *et al.* (2014) Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat Genet*, **46**(6): 573–582.

Wang, X. and Simon, R. (2013) Identification of potential synthetic lethal genes to p53 using a computational biology approach. *BMC Medical Genomics*, **6**(1): 30.

Wappett, M. (2014) Bisep: Toolkit to identify candidate synthetic lethality. r package version 2.0.

Wappett, M., Dulak, A., Yang, Z.R., Al-Watban, A., Bradford, J.R., and Dry, J.R. (2016) Multi-omic measurement of mutually exclusive loss-of-function enriches for candidate synthetic lethal gene pairs. *BMC Genomics*, **17**: 65.

Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W.H.A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., *et al.* (2015) *gplots: Various R Programming Tools for Plotting Data*. R package version 2.17.0.

Watts, D.J. and Strogatz, S.H. (1998) Collective dynamics of 'small-world' networks. *Nature*, **393**(6684): 440–2.

Weinstein, I.B. (2000) Disorders in cell circuitry during multistage carcinogenesis: the role of homeostasis. *Carcinogenesis*, **21**(5): 857–864.

Weinstein, J.N., Akbani, R., Broom, B.M., Wang, W., Verhaak, R.G., McConkey, D., Lerner, S., Morgan, M., Creighton, C.J., Smith, C., *et al.* (2014) Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature*, **507**(7492): 315–322.

Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J.M., Chang, K., *et al.* (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*, **45**(10): 1113–1120.

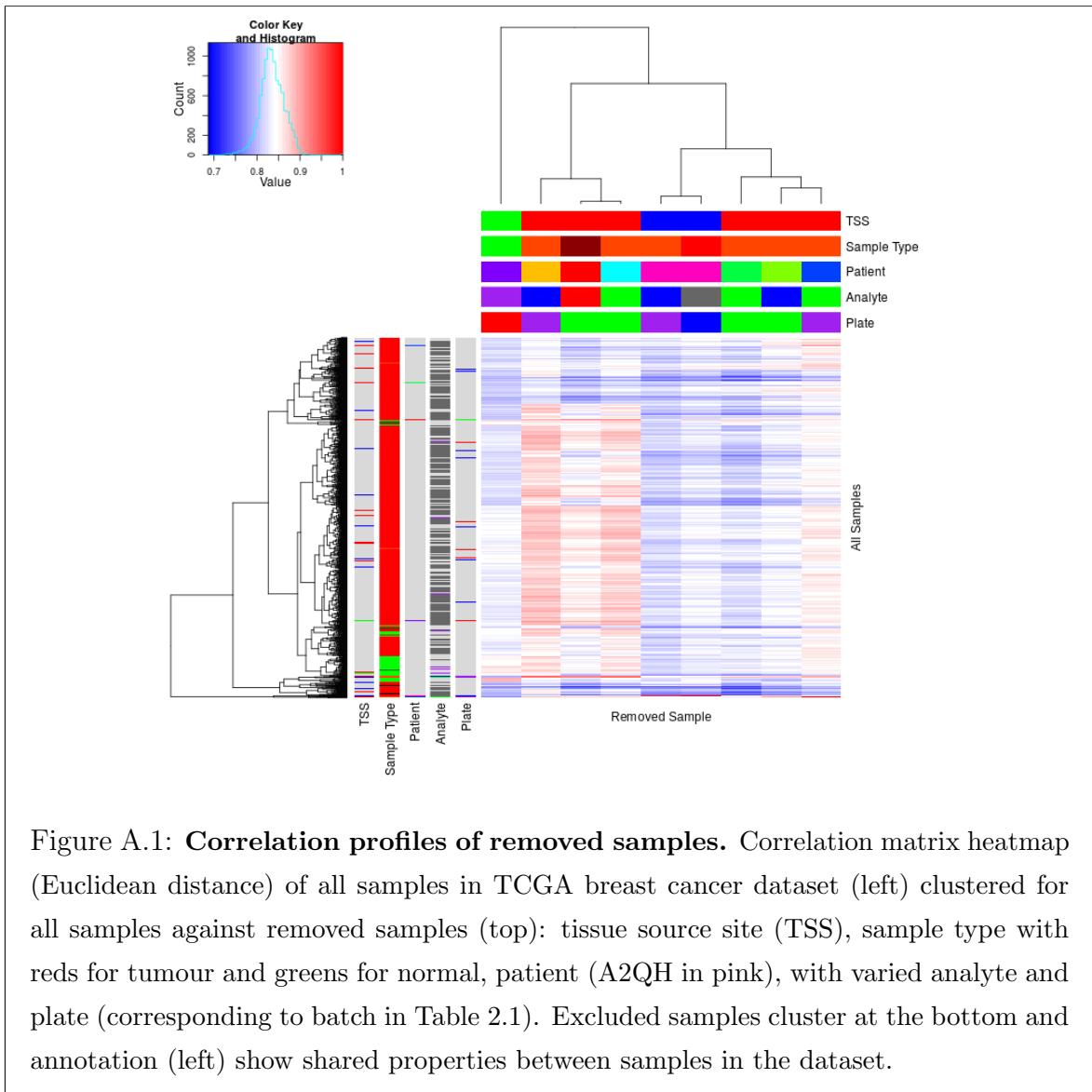
- Wickham, H. and Chang, W. (2016) *devtools: Tools to Make Developing R Packages Easier*. R package version 1.12.0.
- Wickham, H., Danenberg, P., and Eugster, M. (2017) *roxygen2: In-Line Documentation for R*. R package version 6.0.1.
- Wong, S.L., Zhang, L.V., Tong, A.H.Y., Li, Z., Goldberg, D.S., King, O.D., Lesage, G., Vidal, M., Andrews, B., Bussey, H., *et al.* (2004) Combining biological networks to predict genetic interactions. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(44): 15682–15687.
- World Health Organization (WHO) (2017) Fact sheet: Cancer. <http://www.who.int/mediacentre/factsheets/fs297/en/>. Updated February 2017, Accessed: 22/03/2017.
- Wu, M., Li, X., Zhang, F., Li, X., Kwoh, C.K., and Zheng, J. (2014) In silico prediction of synthetic lethality by meta-analysis of genetic interactions, functions, and pathways in yeast and human cancer. *Cancer Inform*, **13**(Suppl 3): 71–80.
- Yu, H. (2002) Rmpi: Parallel statistical computing in r. *R News*, **2**(2): 10–14.
- Zhang, F., Wu, M., Li, X.J., Li, X.L., Kwoh, C.K., and Zheng, J. (2015) Predicting essential genes and synthetic lethality via influence propagation in signaling pathways of cancer cell fates. *J Bioinform Comput Biol*, **13**(3): 1541002.
- Zhang, J., Baran, J., Cros, A., Guberman, J.M., Haider, S., Hsu, J., Liang, Y., Rivkin, E., Wang, J., Whitty, B., *et al.* (2011) International cancer genome consortium data portal a one-stop shop for cancer genomics data. *Database: The Journal of Biological Databases and Curation*, **2011**: bar026.
- Zhong, W. and Sternberg, P.W. (2006) Genome-wide prediction of c. elegans genetic interactions. *Science*, **311**(5766): 1481–1484.
- Zweig, M.H. and Campbell, G. (1993) Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, **39**(4): 561–577.

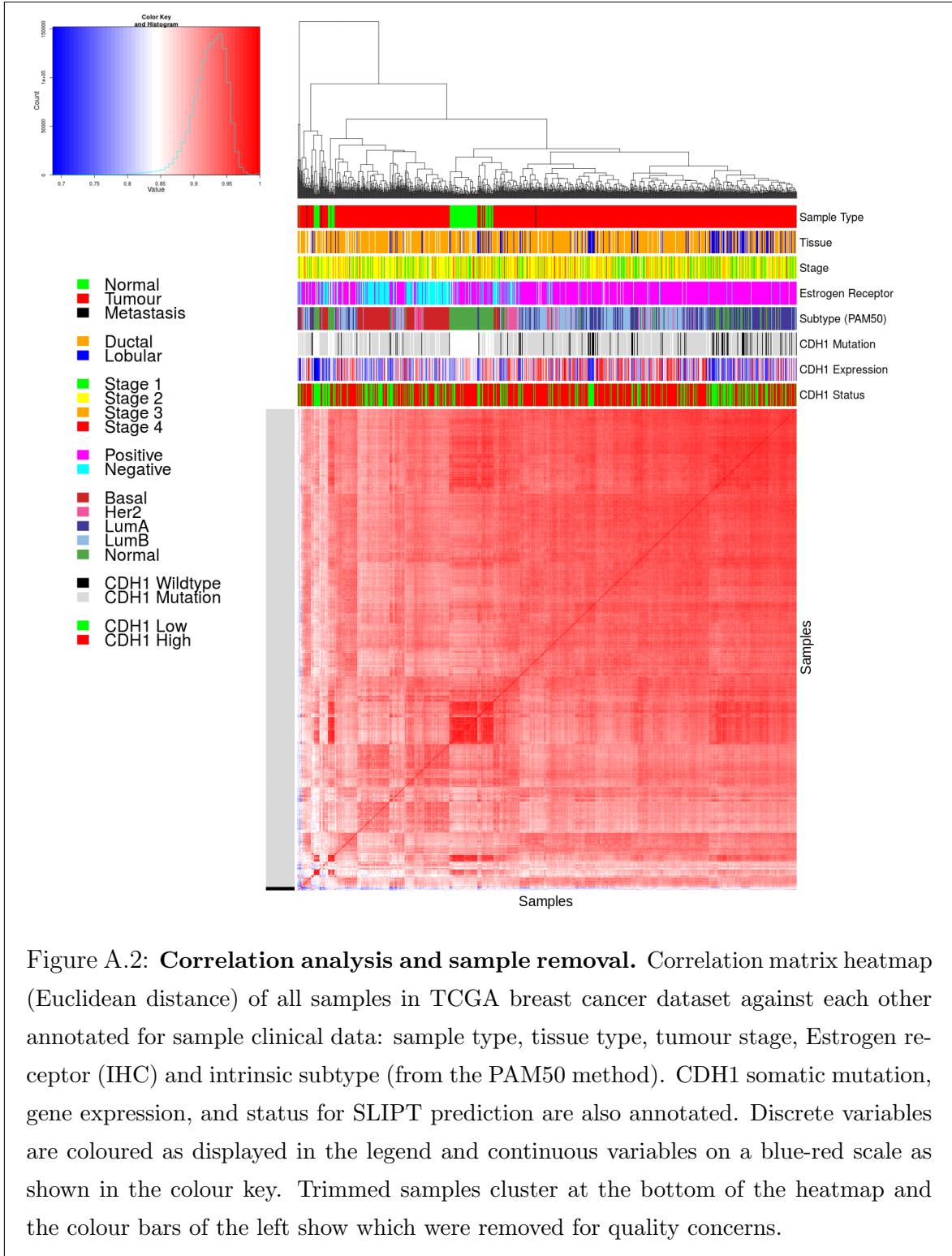
# **Appendix A**

## **Sample Quality**

### **A.1 Sample Correlation**

Samples were excluded from expression analysis based on sample correlations and the clustering analysis presented below, as described in Section 2.2.2.





## A.2 Replicate Samples in TCGA Breast

Replicate samples were picked where possible from the TCGA breast cancer gene expression data to examine for sample quality. Independent samples of the same tumour are expected to have very high Pearson's correlation between their expression profiles unless there were issues with sample collection or preparation and are thus an indicator of sample quality. The log-transformed raw read counts for replicate samples were examined in Figures A.3–A.5. These were examined before normalisation which would be expected to increase sample concordance.

Another consideration are the samples which were removed for quality concerns (in Section 2.2.2). While these were selected by unbiased hierarchical clustering (See Figure A.2), it is notable that many of the excluded (tumour) samples were performed in replicate despite relatively few replicate samples in the overall dataset. These samples correlate poorly with the rest of the dataset, in addition to with replicate samples.

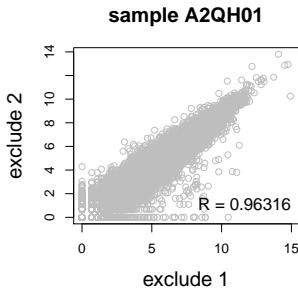
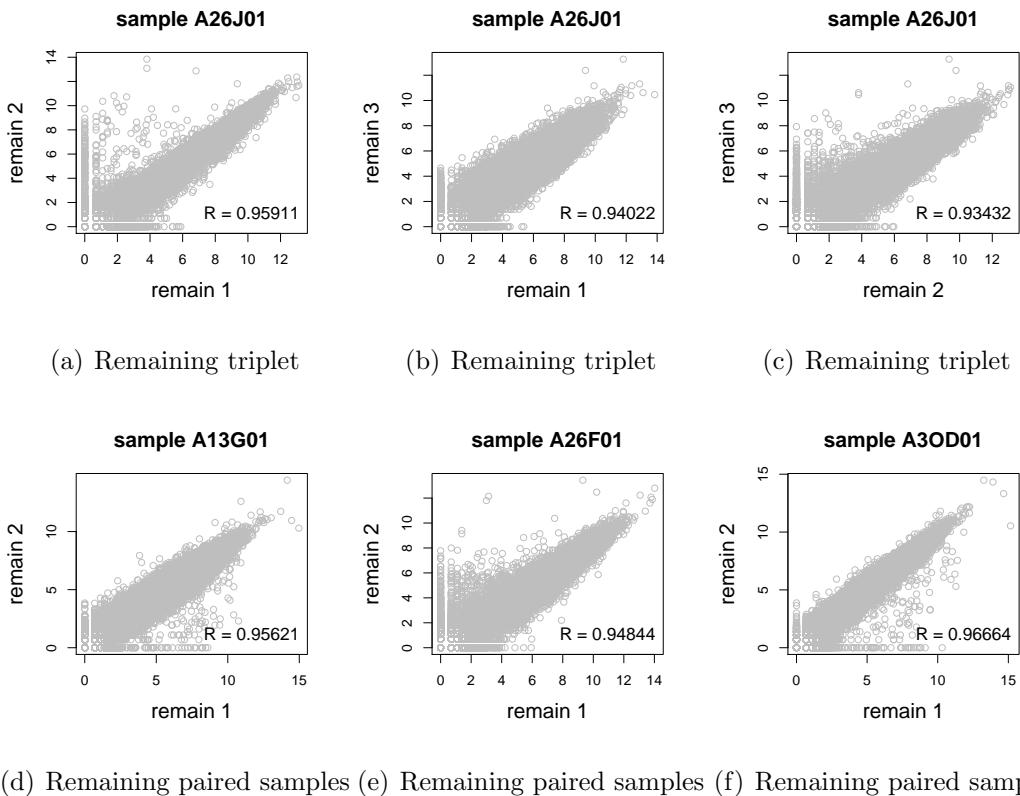
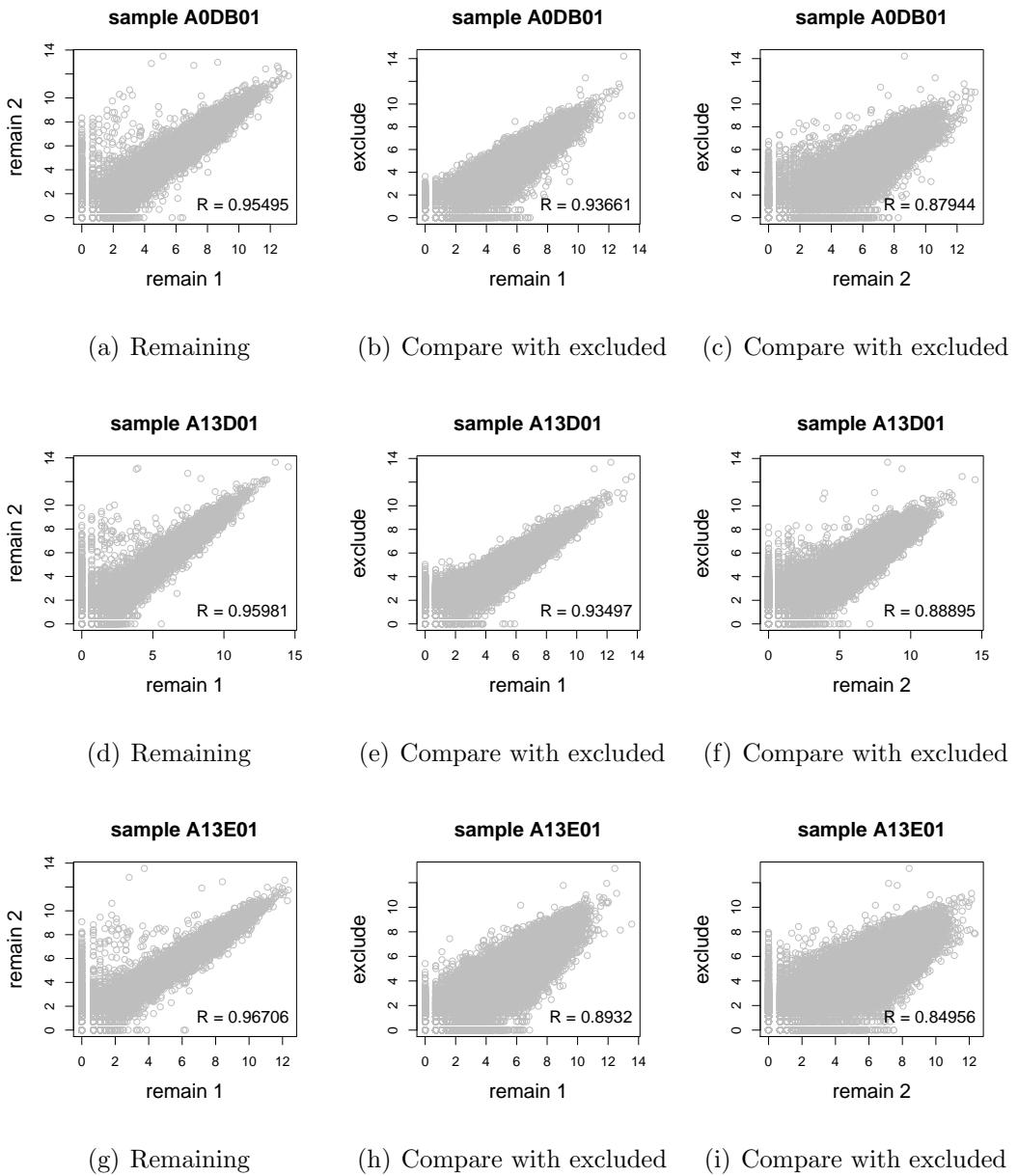


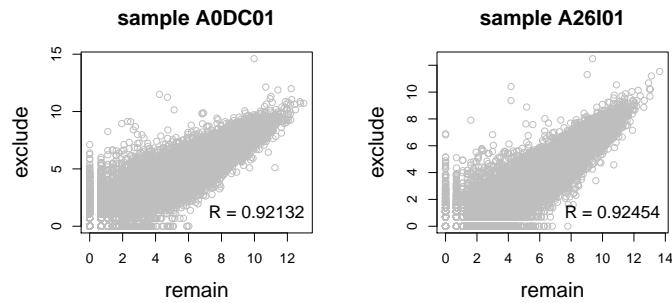
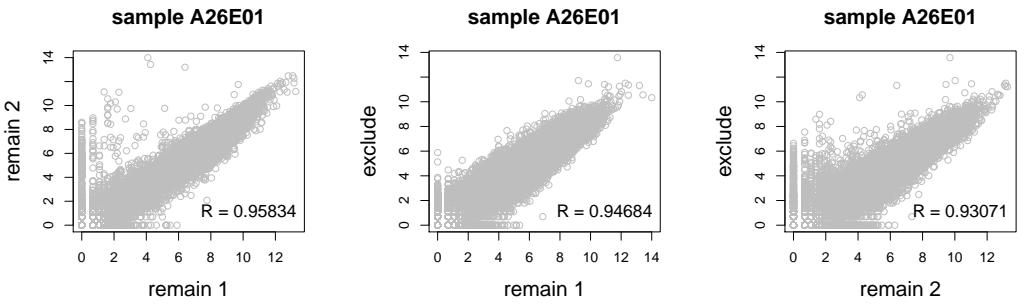
Figure A.3: **Replicate excluded samples.** Both tumour samples of patient A2QH were excluded as they were poorly correlated with other samples, although they are highly similar to each other as shown by Pearson's correlation of log-raw counts.



**Figure A.4: Replicate samples with all remaining.** Patient A26J was sampled 3 times and compared pairwise. Pairs of samples were also compared for other patients with replicate samples. In all cases, replicate samples remaining in the dataset were highly concordant as shown by Pearson's correlation of log-raw counts.



**Figure A.5: Replicate samples with some excluded.** Patients A0DB, A13D, A13E, and A26E were each sampled 3 times and compared pairwise. Pairs of samples were also compared for other patients with replicate samples. In all cases, the replicate samples remaining in the dataset more were highly concordant (as shown by Pearson's correlation of log-raw counts) than those excluded from the analysis.



**Figure A.5: Replicate samples with some excluded.** Patients A0DB, A13D, A13E, and A26E were each sampled 3 times and compared pairwise. Pairs of samples were also compared for other patients with replicate samples. In all cases, the replicate samples remaining in the dataset more were highly concordant (as shown by Pearson's correlation of log-raw counts) than those excluded from the analysis.

# Appendix B

## Software Used for Thesis

Table B.1: R Packages used during Thesis

Package	Repository	Laptop	Lab	Server	NeSI
base	base	3.3.2	3.3.2	3.3.1	3.3.0
abind	CRAN		1.4-5		1.4-3
acepack	CRAN		1.4.1		1.3-3.3
ade4	CRAN		1.7-5		
annaffy	Bioconductor		1.46.0		
AnnotationDbi	Bioconductor		1.36.0	1.36.0	1.34.4
apComplex	CRAN		2.40.0		
ape	CRAN		4		3.4
arm	CRAN		1.9-3		
assertthat	CRAN	0.1	0.1	0.1	0.1
backports	CRAN	1.0.5	1.0.4	1.0.5	1.0.2
base64	CRAN			2	2
base64enc	CRAN		0.1-3		0.1-3
beanplot	CRAN		1.2	1.2	1.2
BH	CRAN	1.60.0-2	1.62.0-1	1.62.0-1	1.60.0-2
Biobase	Bioconductor		2.34.0	2.34.0	2.32.0
BiocGenerics	Bioconductor		0.20.0	0.20.0	0.18.0
BiocInstaller	Bioconductor		1.24.0	1.20.3	1.22.3
BiocParallel	Bioconductor		1.8.1	1.8.1	
Biostrings	Bioconductor		2.42.1	2.42.0	
BiSEp	Bioconductor		2.0.1	2.0.1	2.0.1

bitops	CRAN	1.0-6	1.0-6	1.0-6	1.0-6
boot	base	1.3-18	1.3-18	1.3-18	1.3-18
brew	CRAN	1.0-6	1.0-6	1.0-6	1.0-6
broom	CRAN	0.4.1			
caTools	CRAN	1.17.1	1.17.1	1.17.1	1.17.1
cgdssr	CRAN		1.2.5		
checkmate	CRAN		1.8.2		1.7.4
chron	CRAN	2.3-47	2.3-48	2.3-50	2.3-47
class	base	7.3-14	7.3-14	7.3-14	7.3-14
cluster	base	2.0.5	2.0.5	2.0.5	2.0.4
coda	CRAN		0.19-1		0.18-1
codetools	base	0.2-15	0.2-15	0.2-15	0.2-14
colorRamps	CRAN		2.3		
colorspace	CRAN	1.2-6	1.3-2	1.3-2	1.2-6
commonmark	CRAN	1.1		1.2	
compiler	base	3.3.2	3.3.2	3.3.1	3.3.0
corpcor	CRAN		1.6.8	1.6.8	1.6.8
Cprob	CRAN		1.2.4		
crayon	CRAN	1.3.2	1.3.2	1.3.2	1.3.2
crop	CRAN		0.0-2	0.0-2	
curl	CRAN	1.2	2.3	2.3	0.9.7
d3Network	CRAN		0.5.2.1		
data.table	CRAN	1.9.6	1.10.0	1.10.1	1.9.6
data.tree	CRAN		0.7.0	0.7.0	
datasets	base	3.3.2	3.3.2	3.3.1	3.3.0
DBI	CRAN	0.5-1	0.5-1	0.5-1	0.5-1
dendextend	CRAN	1.4.0	1.4.0	1.4.0	
DEoptimR	CRAN	1.0-8	1.0-8	1.0-8	1.0-4
desc	CRAN	1.1.0		1.1.0	
devtools	CRAN	1.12.0	1.12.0	1.12.0	1.12.0
DiagrammeR	CRAN		0.9.0	0.9.0	
dichromat	CRAN	2.0-0	2.0-0	2.0-0	2.0-0
digest	CRAN	0.6.10	0.6.11	0.6.12	0.6.9
diptest	CRAN	0.75-7	0.75-7	0.75-7	
doParallel	CRAN	1.0.10	1.0.10	1.0.10	1.0.10

dplyr	CRAN	0.5.0	0.5.0	0.5.0	0.5.0
ellipse	CRAN		0.3-8	0.3-8	0.3-8
evaluate	CRAN		0.1	0.1	0.9
fdrtool	CRAN		1.2.15		
fields	CRAN		8.1		
flexmix	CRAN	2.3-13	2.3-13	2.3-13	
forcats	CRAN	0.2.0			
foreach	CRAN	1.4.3	1.4.3	1.4.3	1.4.3
foreign	base	0.8-67	0.8-67	0.8-67	0.8-66
formatR	CRAN		1.4	1.4	1.4
Formula	CRAN		1.2-1		1.2-1
fpc	CRAN	2.1-10	2.1-10	2.1-10	
futile.logger	CRAN		1.4.3	1.4.3	1.4.1
futile.options	CRAN		1.0.0	1.0.0	1.0.0
gdata	CRAN	2.17.0	2.17.0	2.17.0	2.17.0
geepack	CRAN		1.2-1		
GenomeInfoDb	Bioconductor		1.10.2	1.10.1	
GenomicAlignments	Bioconductor		1.10.0	1.10.0	
GenomicRanges	Bioconductor		1.26.2	1.26.1	
ggm	CRAN		2.3		
ggplot2	CRAN	2.1.0	2.2.1	2.2.1	2.1.0
git2r	CRAN	0.15.0	0.18.0	0.16.0	0.15.0
glasso	CRAN		1.8		
GO.db	Bioconductor		3.4.0	3.2.2	3.3.0
GOSemSim	Bioconductor		2.0.3	1.28.2	1.30.3
gplots	CRAN	3.0.1	3.0.1	3.0.1	3.0.1
graph	Bioconductor		1.52.0		
graphics	base	3.3.2	3.3.2	3.3.1	3.3.0
graphsim	GitHub TomKellyGenetics	0.1.0	0.1.0	0.1.0	0.1.0
grDevices	base	3.3.2	3.3.2	3.3.1	3.3.0
grid	base	3.3.2	3.3.2	3.3.1	3.3.0
gridBase	CRAN	0.4-7	0.4-7	0.4-7	0.4-7
gridExtra	CRAN	2.2.1	2.2.1	2.2.1	2.2.1
gridGraphics	CRAN		0.1-5		

gtable	CRAN	0.2.0	0.2.0	0.2.0	0.2.0
gtools	CRAN	3.5.0	3.5.0	3.5.0	3.5.0
haven	CRAN	1.0.0			
heatmap.2x	GitHub		0.0.0.9000	0.0.0.9000	0.0.0.9000
	TomKellyGenetics				
hgu133plus2.db	Bioconductor		3.2.3		
highr	CRAN		0.6	0.6	0.6
Hmisc	CRAN		4.0-2	4.0-2	3.17-4
hms	CRAN	0.2	0.3		
htmlTable	CRAN		1.8	1.9	
htmltools	CRAN	0.3.5	0.3.5	0.3.5	0.3.5
htmlwidgets	CRAN		0.8	0.8	
httpuv	CRAN	1.3.3		1.3.3	
httr	CRAN	1.2.1	1.2.1	1.2.1	1.1.0
huge	CRAN		1.2.7		
hunspell	CRAN		2.3		2
hypergraph	CRAN		1.46.0		
igraph	CRAN	1.0.1	1.0.1	1.0.1	1.0.1
igraph.extensions	GitHub				
	TomKellyGenetics	0.1.0.9001	0.1.0.9001	0.1.0.9001	0.1.0.9001
influenceR	CRAN		0.1.0	0.1.0	
info.centrality	GitHub				
	TomKellyGenetics	0.1.0	0.1.0	0.1.0	0.1.0
IRanges	Bioconductor		2.8.1	2.8.1	2.6.1
irlba	CRAN	2.1.1	2.1.2	2.1.2	2.0.0
iterators	CRAN	1.0.8	1.0.8	1.0.8	1.0.8
jpeg	CRAN		0.1-8		
jsonlite	CRAN	1.1	1.2	1.3	0.9.20
KEGG.db	Bioconductor		3.2.3		
kernlab	CRAN	0.9-25	0.9-25	0.9-25	
KernSmooth	base	2.23-15	2.23-15	2.23-15	2.23-15
knitr	CRAN		1.15.1	1.15.1	1.14
labeling	CRAN	0.3	0.3	0.3	0.3
lambda.r	CRAN		1.1.9	1.1.9	1.1.7
lattice	base	0.20-34	0.20-34	0.20-34	0.20-33

latticeExtra	CRAN		0.6-28		0.6-28
lava	CRAN		1.4.6		
lavaan	CRAN		0.5-22		
lazyeval	CRAN	0.2.0	0.2.0	0.2.0	0.2.0
les	CRAN		1.24.0		
lgtdl	CRAN		1.1.3		
limma	Bioconductor		3.30.7	3.30.3	
lme4	CRAN		1.1-12		1.1-12
lubridate	CRAN	1.6.0			
magrittr	CRAN	1.5	1.5	1.5	1.5
maps	CRAN		3.1.1		
markdown	CRAN		0.7.7	0.7.7	0.7.7
MASS	base	7.3-45	7.3-45	7.3-45	7.3-45
Matrix	base	1.2-7.1	1.2-7.1	1.2-8	1.2-6
matrixcalc	CRAN	1.0-3	1.0-3	1.0-3	1.0-3
mclust	CRAN	5.2	5.2.1	5.2.2	5.2
memoise	CRAN	1.0.0	1.0.0	1.0.0	1.0.0
methods	base	3.3.2	3.3.2	3.3.1	3.3.0
mgcv	base	1.8-16	1.8-16	1.8-17	1.8-12
mi	CRAN		1		
mime	CRAN	0.5	0.5	0.5	0.4
minqa	CRAN		1.2.4		1.2.4
mnormt	CRAN	1.5-5	1.5-5		1.5-4
modelr	CRAN	0.1.0			
modeltools	CRAN	0.2-21	0.2-21	0.2-21	
multtest	Bioconductor		2.30.0	2.30.0	
munsell	CRAN	0.4.3	0.4.3	0.4.3	0.4.3
mvtnorm	CRAN	1.0-5	1.0-5	1.0-6	1.0-5
network	CRAN		1.13.0		
nlme	base	3.1-128	3.1-128	3.1-131	3.1-128
nloptr	CRAN		1.0.4		1.0.4
NMF	CRAN	0.20.6	0.20.6	0.20.6	0.20.6
nnet	base	7.3-12	7.3-12	7.3-12	7.3-12
numDeriv	CRAN		2016.8-1		2014.2-1
openssl	CRAN	0.9.4	0.9.6	0.9.6	0.9.4

org.Hs.eg.db	Bioconductor		3.1.2		3.3.0
org.Sc.sgd.db	Bioconductor		3.4.0		
parallel	base	3.3.2	3.3.2	3.3.1	3.3.0
pathway.structure	GitHub	0.1.0	0.1.0	0.1.0	0.1.0
.permutation	TomKellyGenetics				
pbivnorm	CRAN		0.6.0		
PGSEA	Bioconductor		1.48.0		
pkgmaker	CRAN	0.22	0.22	0.22	0.22
PKI	CRAN		0.1-3		
plogr	CRAN		0.1-1	0.1-1	
plot.igraph	GitHub	0.0.0.9001	0.0.0.9001	0.0.0.9001	0.0.0.9001
	TomKellyGenetics				
plotrix	CRAN		3.6-4		
plyr	CRAN	1.8.4	1.8.4	1.8.4	1.8.3
png	CRAN		0.1-7		0.1-7
prabclus	CRAN	2.2-6	2.2-6	2.2-6	
praise	CRAN	1.0.0	1.0.0		1.0.0
pROC	CRAN		1.8	1.9.1	
prodlim	CRAN		1.5.7		
prof.tree	CRAN		0.1.0		
protools	CRAN		0.99-2		
progress	CRAN			1.1.2	
psych	CRAN	1.6.12	1.6.12		
purrr	CRAN	0.2.2	0.2.2	0.2.2	0.2.2
qgraph	CRAN		1.4.1		
quadprog	CRAN		1.5-5	1.5-5	1.5-5
R.methodsS3	CRAN		1.7.1		1.7.1
R.oo	CRAN		1.21.0		1.20.0
R.utils	CRAN		2.5.0		
R6	CRAN	2.1.3	2.2.0	2.2.0	2.1.3
RBGL	CRAN		1.50.0		
RColorBrewer	CRAN	1.1-2	1.1-2	1.1-2	1.1-2
Rcpp	CRAN	0.12.7	0.12.9	0.12.9	0.12.7
RcppArmadillo	CRAN			0.7.700.0.0	0.6.700.6.0
RcppEigen	CRAN		0.3.2.9.0		0.3.2.8.1

RCurl	CRAN		1.95-4.8	1.95-4.8	1.95-4.8
reactome.db	Bioconductor		1.52.1	1.52.1	
reactometree	GitHub		0.1		
	TomKellyGenetics				
readr	CRAN	1.0.0	1.0.0		
readxl	CRAN	0.1.1			
registry	CRAN	0.3	0.3	0.3	0.3
reshape2	CRAN	1.4.1	1.4.2	1.4.2	1.4.1
rgeff	CRAN		0.15.3	0.15.3	
rgl	CRAN			0.97.0	0.95.1441
Rgraphviz	CRAN		2.18.0		
rjson	CRAN		0.2.15		
RJSONIO	CRAN		1.3-0		
rmarkdown	CRAN		1.3	1.3	1
Rmpi	CRAN		0.6-6		0.6-5
rngtools	CRAN	1.2.4	1.2.4	1.2.4	1.2.4
robustbase	CRAN	0.92-7	0.92-7	0.92-7	0.92-5
ROCR	CRAN	1.0-7	1.0-7	1.0-7	1.0-7
Rook	CRAN		1.1-1	1.1-1	
roxygen2	CRAN	6.0.1	5.0.1	6.0.1	5.0.1
rpart	base	4.1-10	4.1-10	4.1-10	4.1-10
rprojroot	CRAN	1.2	1.1	1.2	
Rsamtools	Bioconductor		1.26.1	1.26.1	
rsconnect	CRAN		0.7		
RSQLite	CRAN		1.1-2	1.1-2	1.0.0
rstudioapi	CRAN	0.6	0.6	0.6	0.6
rvest	CRAN	0.3.2			
S4Vectors	Bioconductor		0.12.1	0.12.0	0.10.3
safe	Bioconductor		3.14.0	3.10.0	
scales	CRAN	0.4.0	0.4.1	0.4.1	0.4.0
selectr	CRAN	0.3-1			
sem	CRAN		3.1-8		
shiny	CRAN	0.14		1.0.0	
slipt	GitHub		0.1.0	0.1.0	0.1.0
	TomKellyGenetics				

sm	CRAN	2.2-5.4	2.2-5.4		
sna	CRAN		2.4		
snow	CRAN	0.4-1	0.4-2	0.4-2	0.3-13
sourcetools	CRAN	0.1.5		0.1.5	
SparseM	CRAN		1.74		1.7
spatial	base	7.3-11	7.3-11	7.3-11	7.3-11
splines	base	3.3.2	3.3.2	3.3.1	3.3.0
statnet.common	CRAN		3.3.0		
stats	base	3.3.2	3.3.2	3.3.1	3.3.0
stats4	base	3.3.2	3.3.2	3.3.1	3.3.0
stringi	CRAN	1.1.1	1.1.2	1.1.2	1.0-1
stringr	CRAN	1.1.0	1.1.0	1.2.0	1.0.0
SummarizedExperiment	Bioconductor		1.4.0	1.4.0	
survival	base	2.39-4	2.40-1	2.40-1	2.39-4
tcltk	base	3.3.2	3.3.2	3.3.1	3.3.0
testthat	CRAN	1.0.2	1.0.2		1.0.2
tibble	CRAN	1.2	1.2	1.2	1.2
tidyverse	GitHub hadley	1.1.1			
timeline	CRAN		0.9		
tools	base	3.3.2	3.3.2	3.3.1	3.3.0
tpr	CRAN		0.3-1		
trimcluster	CRAN	0.1-2	0.1-2	0.1-2	
Unicode	CRAN	9.0.0-1	9.0.0-1	9.0.0-1	
utils	base	3.3.2	3.3.2	3.3.1	3.3.0
vioplot	CRAN		0.2		
vioplotx	GitHub TomKellyGenetics	0.0.0.9000	0.0.0.9000		
viridis	CRAN	0.3.4	0.3.4	0.3.4	
visNetwork	CRAN		1.0.3	1.0.3	
whisker	CRAN	0.3-2	0.3-2	0.3-2	0.3-2
withr	CRAN	1.0.2	1.0.2	1.0.2	1.0.2
XML	base	3.98-1.3	3.98-1.1	3.98-1.5	3.98-1.4

xml2	CRAN	1.1.1	1.1.1	1.0.0
xtable	CRAN	1.8-2	1.8-2	1.8-2
XVector	Bioconductor		0.14.0	0.14.0
yaml	CRAN		2.1.14	2.1.14
zlibbioc	CRAN		1.20.0	1.20.0
zoo	CRAN	1.7-13	1.7-14	1.7-13

# Appendix C

## Secondary Screen Data

A series of experimental genome-wide siRNA screens have been performed on synthetic lethal partners of *CDH1* (Telford *et al.*, 2015). The strongest candidates from a primary screen were subject to a further secondary screen for validation by independent replication with 4 gene knockdowns with different targeting siRNA. As shown in Table C.1, there is significant ( $p = 7.49 \times 10^{-3}$  by Fisher’s exact test) association between SLIPT candidates and stronger validations of siRNA candidates. Since there were more SLIPT– genes among those not validated and more SLIPT+ genes among those validated with several siRNAs, this supports the use of SLIPT as a synthetic lethal discovery procedure which may augment such screening experiments.

Table C.1: Comparing SLIPT genes against Secondary siRNA Screen in breast cancer

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

Similar analysis with mtSLIPT, comparing SLIPT against *CDH1* somatic mutation with siRNA validation results was not significant ( $p = 7.02 \times 10^{-1}$  by Fisher’s exact test). However, as shown in Table C.2, the observed and expected values were in a direction consistent with that observed above for SLIPT against low *CDH1* expression.

It is not unexpected that this result does not have comparable statistical support due to the lower sample size for mutation data.

Table C.2: Comparing mtSLIPT genes against Secondary siRNA Screen in breast cancer

		Secondary Screen					<b>Total</b>
		0/4	1/4	2/4	3/4	4/4	
<b>mtSLIPT+</b>	Observed	54	35	17	4	6	111
	Expected	60	31	14	4	1	
<b>mtSLIPT-</b>	Observed	206	101	45	14	5	371
	Expected	200	105	48	14	4	
<b>Total</b>		269	143	63	19	6	482

This analysis was replicated on a (smaller) stomach cancer dataset but it was less conclusive ( $p = 2.36 \times 10^{-1}$  by Fisher's exact test). As shown in Table C.3, fewer SLIPT candidates were validated than expected statistically. However, these results in stomach cancer may not be directly comparable to experiments in a breast cell line. Genes validated by 0 or 1 siRNA behave consistently with the results above.

Table C.3: Comparing SLIPT genes against Secondary siRNA Screen in stomach cancer

		Secondary Screen					<b>Total</b>
		0/4	1/4	2/4	3/4	4/4	
<b>SLIPT+</b>	Observed	67	47	13	4	1	132
	Expected	71	37	17	5	2	
<b>SLIPT-</b>	Observed	195	90	50	14	5	354
	Expected	190	100	46	13	4	
<b>Total</b>		262	137	63	19	6	486

# **Appendix D**

## **Mutation Analysis in Breast Cancer**

### **D.1 Synthetic Lethal Genes and Pathways**

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

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

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

Strongest candidate SL partners for *CDH1* by mtSLIPT with observed and expected mutant samples with low expression of partner genes

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

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

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

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

## D.2 Synthetic Lethal Expression Profiles

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

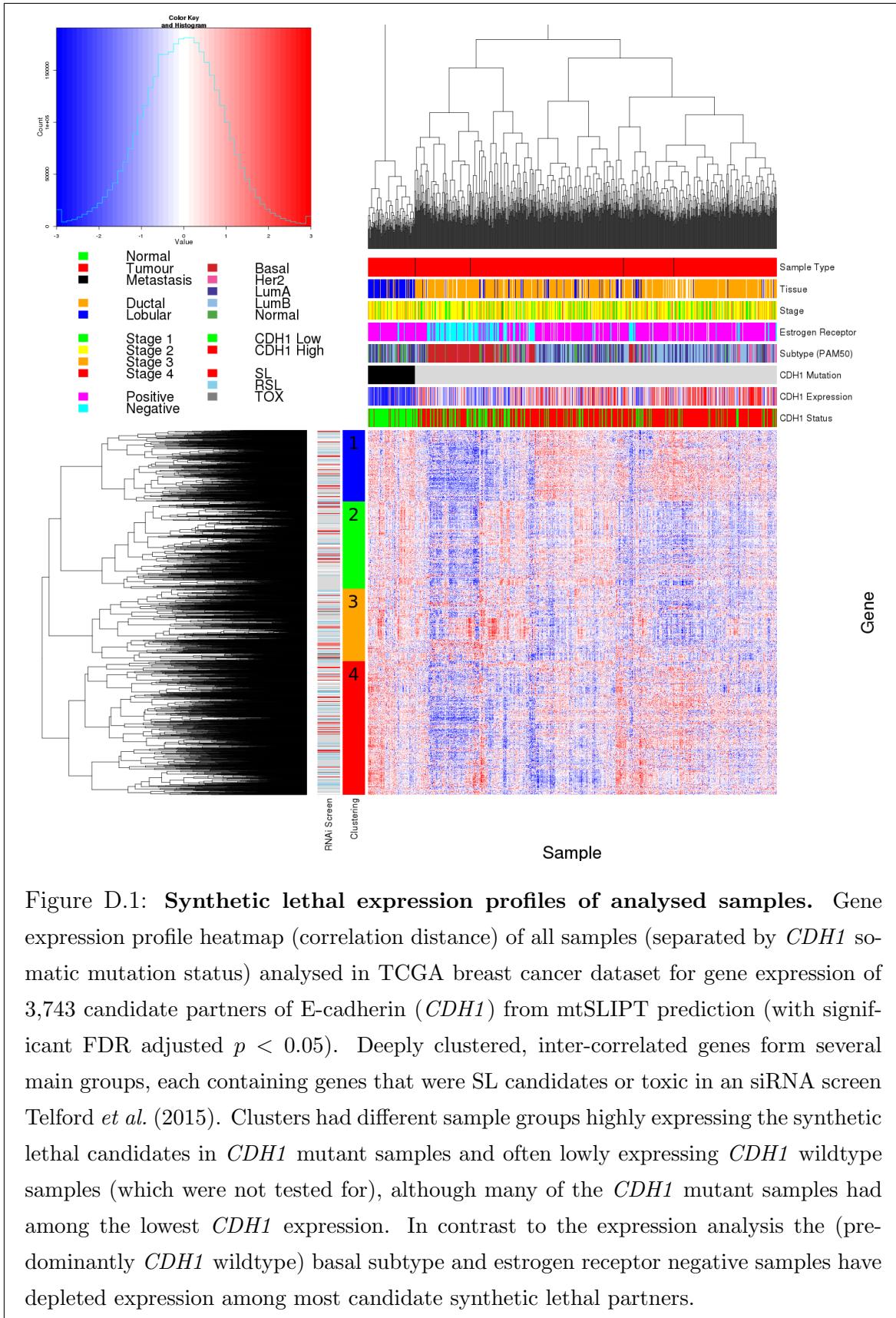


Table D.3: Pathway composition for clusters of *CDH1* partners from mtSLIPT

Pathways Over-represented in Cluster 1	Pathway Size	Cluster Genes	p-value (FDR)
Olfactory Signaling Pathway	57	8	$7.1 \times 10^{-9}$
Assembly of the primary cilium	149	14	$8.0 \times 10^{-9}$
Sphingolipid metabolism	62	8	$9.6 \times 10^{-9}$
Signaling by ERBB4	133	12	$5.1 \times 10^{-8}$
PI3K Cascade	65	7	$4.9 \times 10^{-7}$
Circadian Clock	33	5	$4.9 \times 10^{-7}$
Nuclear signaling by ERBB4	34	5	$4.9 \times 10^{-7}$
Intraflagellar transport	35	5	$4.9 \times 10^{-7}$
PI3K events in ERBB4 signaling	87	8	$4.9 \times 10^{-7}$
PIP3 activates AKT signaling	87	8	$4.9 \times 10^{-7}$
PI3K events in ERBB2 signaling	87	8	$4.9 \times 10^{-7}$
PI-3K cascade:FGFR1	87	8	$4.9 \times 10^{-7}$
PI-3K cascade:FGFR2	87	8	$4.9 \times 10^{-7}$
PI-3K cascade:FGFR3	87	8	$4.9 \times 10^{-7}$
PI-3K cascade:FGFR4	87	8	$4.9 \times 10^{-7}$
Deadenylation of mRNA	22	4	$5.6 \times 10^{-7}$
PI3K/AKT activation	90	8	$5.6 \times 10^{-7}$
Cargo trafficking to the periciliary membrane	38	5	$5.6 \times 10^{-7}$
Signaling by Hedgehog	108	9	$5.6 \times 10^{-7}$
Downstream signal transduction	143	11	$5.6 \times 10^{-7}$

Pathways Over-represented in Cluster 2	Pathway Size	Cluster Genes	p-value (FDR)
G <sub>αs</sub> signalling events	83	19	$5.1 \times 10^{-25}$
Extracellular matrix organization	238	30	$1.4 \times 10^{-18}$
Hemostasis	422	46	$2.7 \times 10^{-16}$
Aquaporin-mediated transport	32	9	$2.7 \times 10^{-16}$
Transcriptional regulation of white adipocyte differentiation	56	11	$1.7 \times 10^{-15}$
Degradation of the extracellular matrix	102	15	$1.7 \times 10^{-15}$
Integration of energy metabolism	84	13	$8.8 \times 10^{-15}$
GPCR downstream signaling	472	48	$2.8 \times 10^{-14}$
G <sub>αs</sub> signalling events	15	6	$5.0 \times 10^{-14}$
Molecules associated with elastic fibres	33	8	$5.4 \times 10^{-14}$
Phase 1 - Functionalization of compounds	67	11	$5.6 \times 10^{-14}$
Platelet activation, signaling and aggregation	179	20	$5.6 \times 10^{-14}$
Vasopressin regulates renal water homeostasis via Aquaporins	24	7	$6.1 \times 10^{-14}$
Elastic fibre formation	37	8	$.03 \times 10^{-13}$
Calmodulin induced events	27	7	$3.3 \times 10^{-13}$
CaM pathway	27	7	$3.3 \times 10^{-13}$
cGMP effects	18	6	$3.6 \times 10^{-13}$
G <sub>αs</sub> signalling events	167	18	$6.3 \times 10^{-13}$
Ca-dependent events	29	7	$8.2 \times 10^{-13}$
Binding and Uptake of Ligands by Scavenger Receptors	40	8	$8.2 \times 10^{-13}$

Pathways Over-represented in Cluster 3	Pathway Size	Cluster Genes	p-value (FDR)
Eukaryotic Translation Elongation	86	55	$1.1 \times 10^{-112}$
Peptide chain elongation	83	54	$1.3 \times 10^{-112}$
Viral mRNA Translation	81	53	$1.6 \times 10^{-111}$
Eukaryotic Translation Termination	83	53	$7.1 \times 10^{-110}$
Nonsense Mediated Decay independent of the Exon Junction Complex	88	54	$1.0 \times 10^{-108}$
Formation of a pool of free 40S subunits	93	53	$4.1 \times 10^{-102}$
Nonsense-Mediated Decay	103	54	$3.9 \times 10^{-98}$
Nonsense-Mediated Decay enhanced by the Exon Junction Complex	103	54	$3.9 \times 10^{-98}$
L13a-mediated translational silencing of Ceruloplasmin expression	103	53	$1.2 \times 10^{-95}$
3' -UTR-mediated translational regulation	103	53	$1.2 \times 10^{-95}$
SRP-dependent cotranslational protein targeting to membrane	104	53	$4.3 \times 10^{-95}$
GTP hydrolysis and joining of the 60S ribosomal subunit	104	53	$4.3 \times 10^{-95}$
Influenza Viral RNA Transcription and Replication	108	53	$9.6 \times 10^{-93}$
Eukaryotic Translation Initiation	111	53	$4.2 \times 10^{-91}$
Cap-dependent Translation Initiation	111	53	$4.2 \times 10^{-91}$
Influenza Life Cycle	112	53	$1.4 \times 10^{-90}$
Influenza Infection	117	53	$6.2 \times 10^{-88}$
Translation	141	55	$3 \times 10^{-81}$
Formation of the ternary complex, and subsequently, the 43S complex	47	23	$2.3 \times 10^{-48}$
Translation initiation complex formation	54	23	$9.1 \times 10^{-45}$

Pathways Over-represented in Cluster 4	Pathway Size	Cluster Genes	p-value (FDR)
ECM proteoglycans	66	10	$2.9 \times 10^{-11}$
deactivation of the beta-catenin transactivating complex	38	7	$5.1 \times 10^{-10}$
Arachidonic acid metabolism	41	7	$1.1 \times 10^{-9}$
Gαq signalling events	149	14	$4.0 \times 10^{-9}$
HS-GAG degradation	21	5	$4.5 \times 10^{-9}$
Uptake and actions of bacterial toxins	22	5	$6.1 \times 10^{-9}$
Gastrin-CREB signalling pathway via PKC and MAPK	170	15	$6.1 \times 10^{-9}$
RNA Polymerase I, RNA Polymerase III, and Mitochondrial Transcription	64	8	$6.1 \times 10^{-9}$
Non-integrin membrane-ECM interactions	53	7	$1.5 \times 10^{-8}$
Syndecan interactions	25	5	$1.5 \times 10^{-8}$
NOTCH1 Intracellular Domain Regulates Transcription	40	6	$2.3 \times 10^{-8}$
Synthesis of Leukotrienes and Exoxins	15	4	$3.2 \times 10^{-8}$
Signaling by NOTCH1	59	7	$5.3 \times 10^{-8}$
Regulation of insulin secretion	44	6	$6.0 \times 10^{-8}$
Metabolism of lipids and lipoproteins	471	37	$8.2 \times 10^{-8}$
Signaling by NOTCH	80	8	$1.2 \times 10^{-7}$
Platelet activation, signaling and aggregation	179	14	$1.2 \times 10^{-7}$
Recruitment of mitotic centrosome proteins and complexes	64	7	$1.2 \times 10^{-7}$
Centrosome maturation	64	7	$1.2 \times 10^{-7}$
Biological oxidations	133	11	$1.5 \times 10^{-7}$

### D.3 Comparison to Primary Screen

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

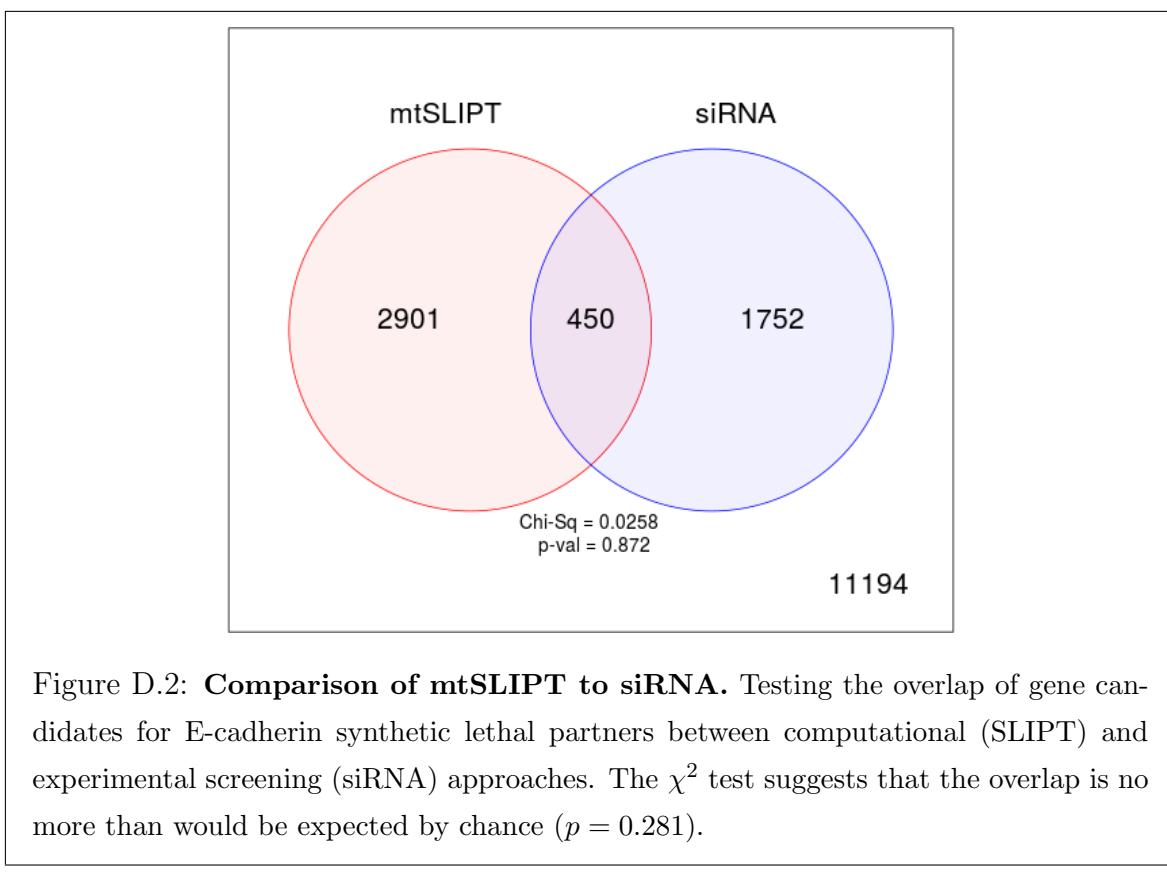


Figure D.2: **Comparison of mtSLIPT to siRNA.** Testing the overlap of gene candidates for E-cadherin synthetic lethal partners between computational (SLIPT) and experimental screening (siRNA) approaches. The  $\chi^2$  test suggests that the overlap is no more than would be expected by chance ( $p =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

Despite a lower sample size (and low number of predicted partners) for mutation analysis, the pathway composition (Tables D.2 and D.4) is similar to expression analysis, as described in Section 4.2.1.4. In particular, the resampling analysis (Section D.3.1) supported many of the results of expression analysis (Section 4.2.1.4.1) with Tables D.5 and D.6 detecting many of the same or functionally-related pathways.

Table D.4: Pathway composition for *CDH1* partners from mtSLIPT and siRNA

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

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

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

### D.3.1 Resampling Analysis

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

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

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).

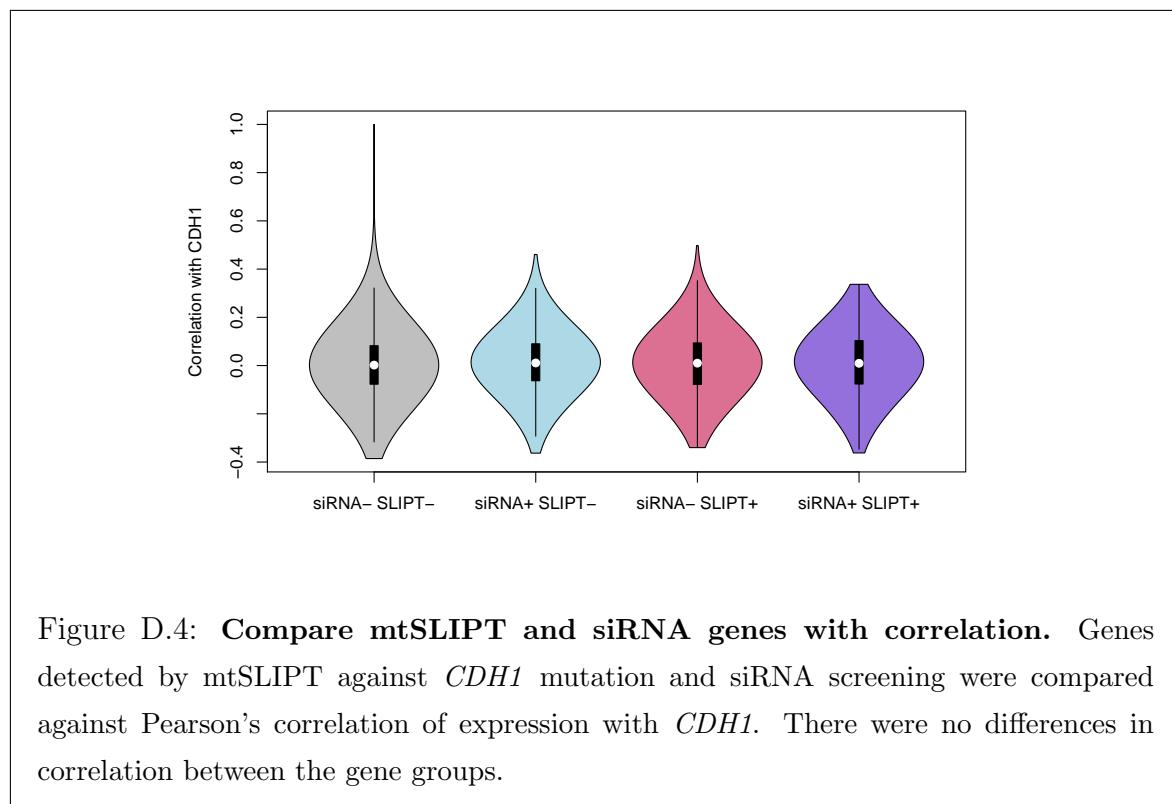
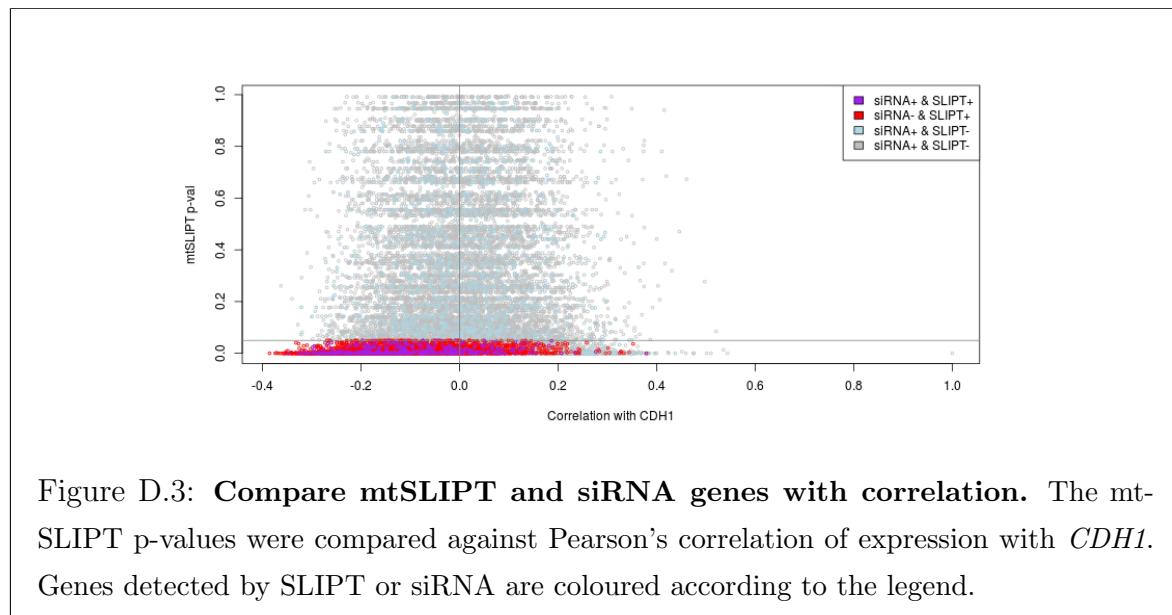
Table D.6: Pathways for *CDH1* partners from mtSLIPT and siRNA primary screen

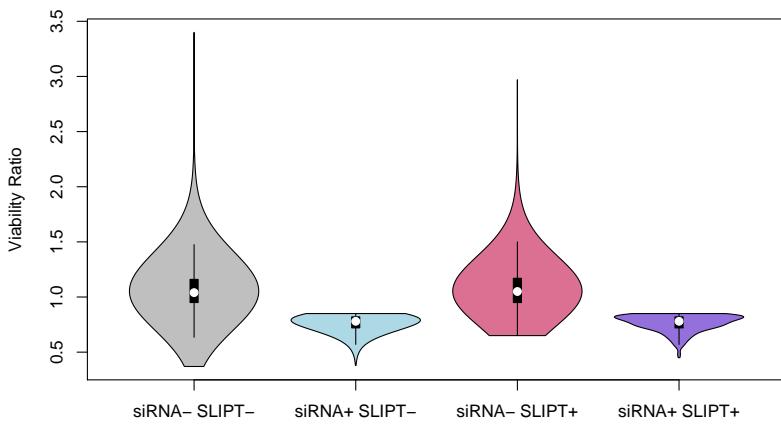
Reactome Pathway	Over-representation	Permutation
Visual phototransduction	$1.2 \times 10^{-9}$	0.86279
<b>G<sub>αs</sub> signalling events</b>	$2.9 \times 10^{-7}$	0.023066
Retinoid metabolism and transport	$2.9 \times 10^{-7}$	0.299
Acylic chain remodelling of PS	$1.1 \times 10^{-5}$	0.42584
Transcriptional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

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

## D.4 Compare SLIPT genes

The mutation synthetic lethal partners with *CDH1* were also compared to siRNA primary screen data (Telford *et al.*, 2015), by correlation and siRNA viability as described in sections 4.2.1.1 and 4.2.1.2.





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

## D.5 Metagene Analysis

Metagene analysis was also performed for synthetic lethal candidates for *CDH1* mutation. These are described and compared to expression analysis in Section 4.3.4.

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

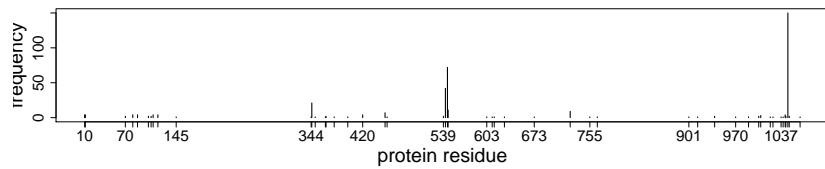
Pathway	ID	Observed	Expected	$\chi^2$ value	p-value	p-value (FDR)
Neurotoxicity of clostridium toxins	168799	8	36.7	79.4	$5.71 \times 10^{-18}$	$3.14 \times 10^{-15}$
Aquaporin-mediated transport	445717	8	36.7	76.3	$2.73 \times 10^{-17}$	$9.01 \times 10^{-15}$
Toxicity of botulinum toxin type G (BoNT/G)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

Strongest candidate SL partners for *CDH1* by mtSLIPT with observed and expected mutant samples with low expression of partner metagenes

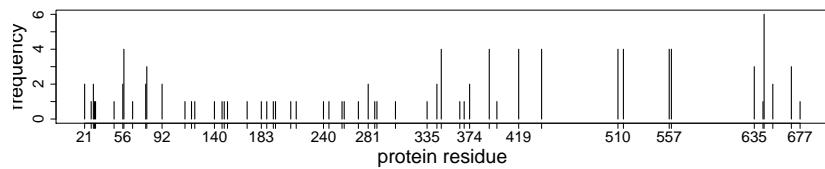
## D.6 Mutation Variation

Mutations have different effects as shown by the following examples in cancer genes.

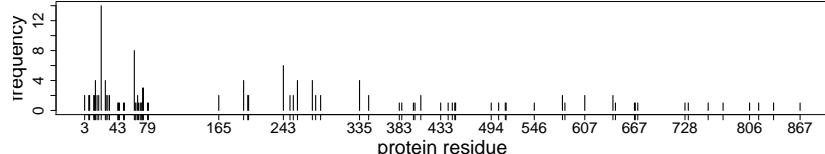
### D.6.1 Mutation Frequency



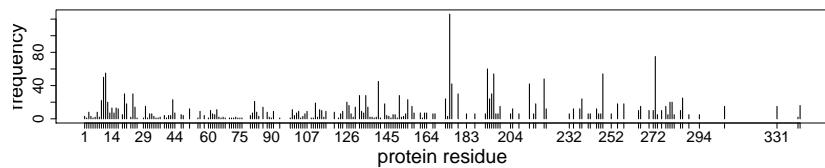
(a) *PI3KCA*



(b) *PI3KR1*



(c) *CDH1*



(d) *TP53*

Figure D.6: **Somatic mutation locus.** Mutation frequency at each locus in TCGA breast cancer. *PIK3CA* shows clear recurrent E545K and H1047R oncogene mutations consistent with it being an oncogene. *PIK3R1* and *CDH1* are tumour suppressors with inactivating mutations distributed throughout the gene, whereas *TP53* exhibits both of these properties and a very high mutation frequency compared to other genes.

## D.6.2 PI3K Mutation Expression

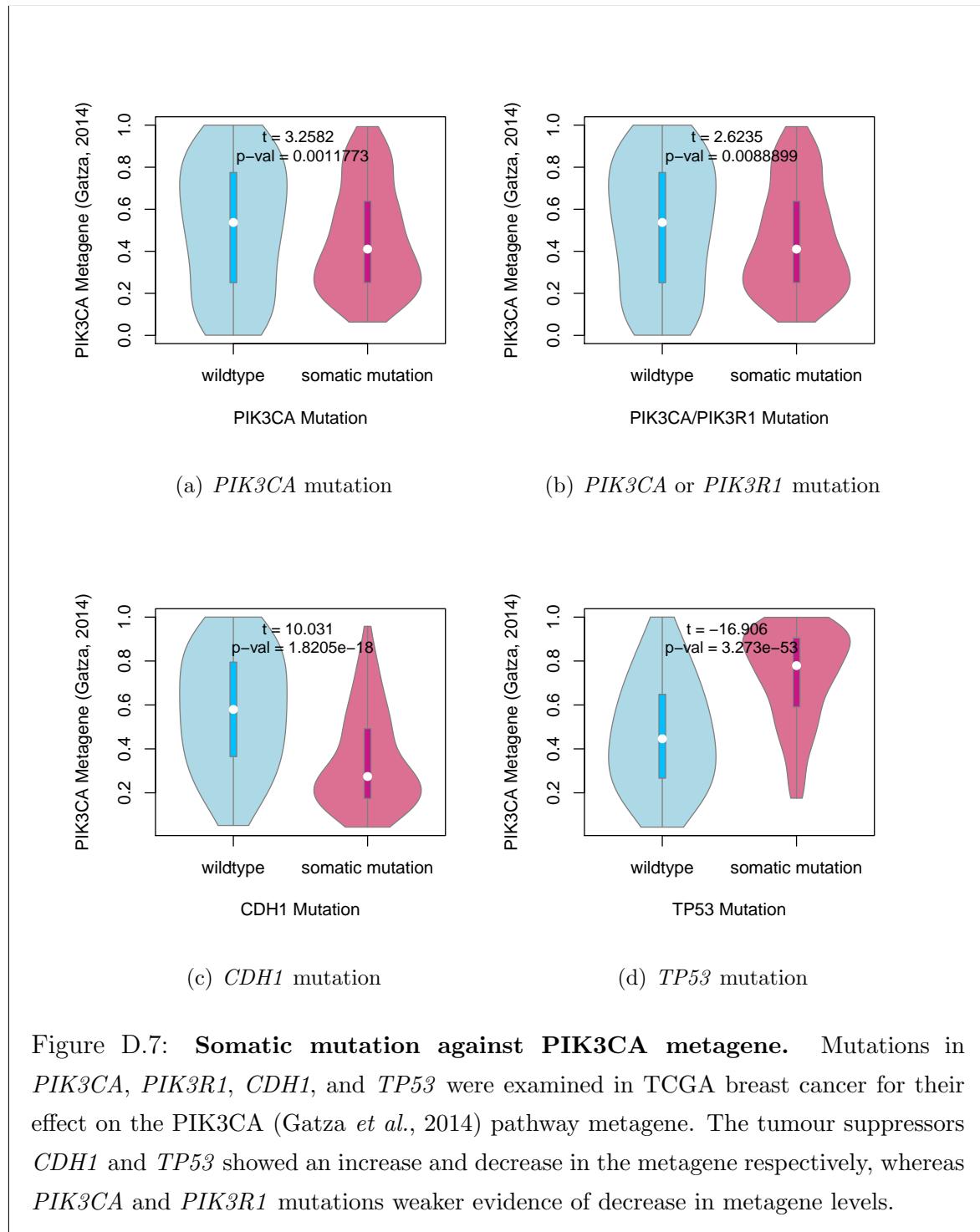
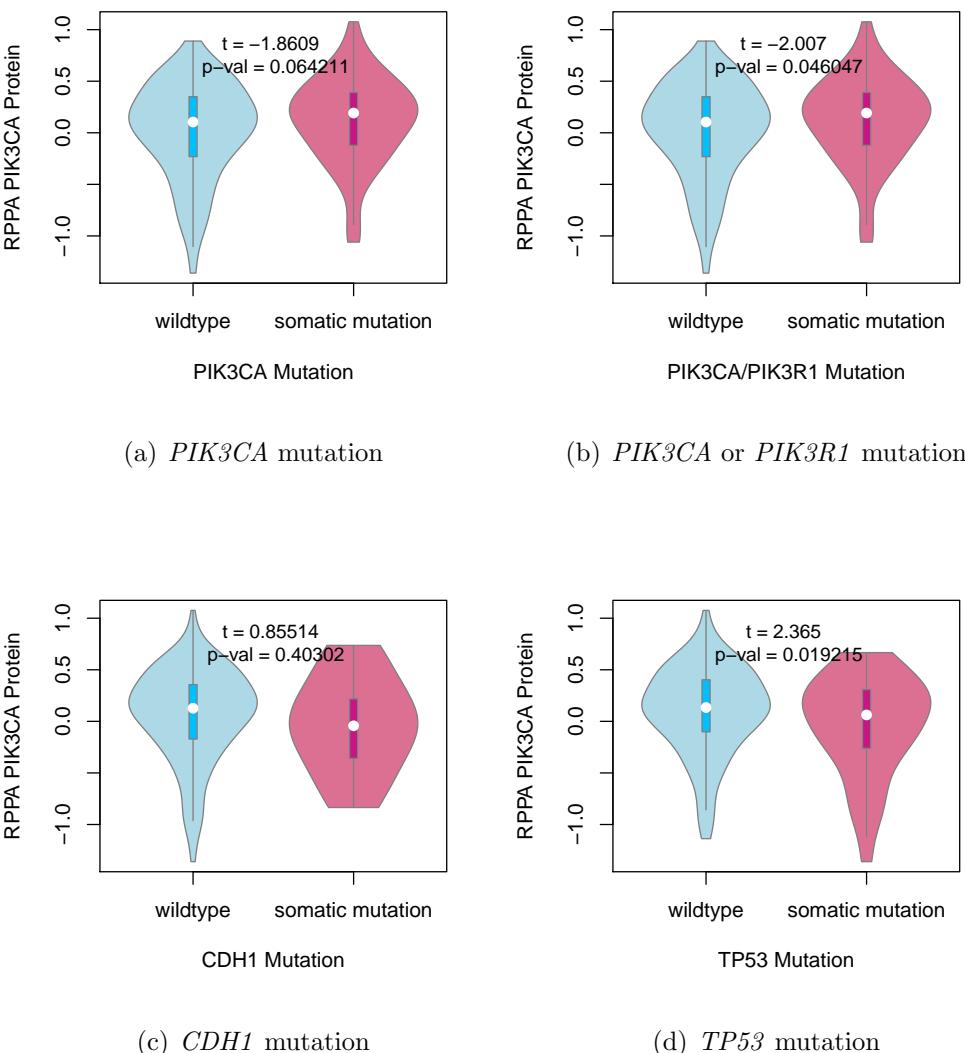
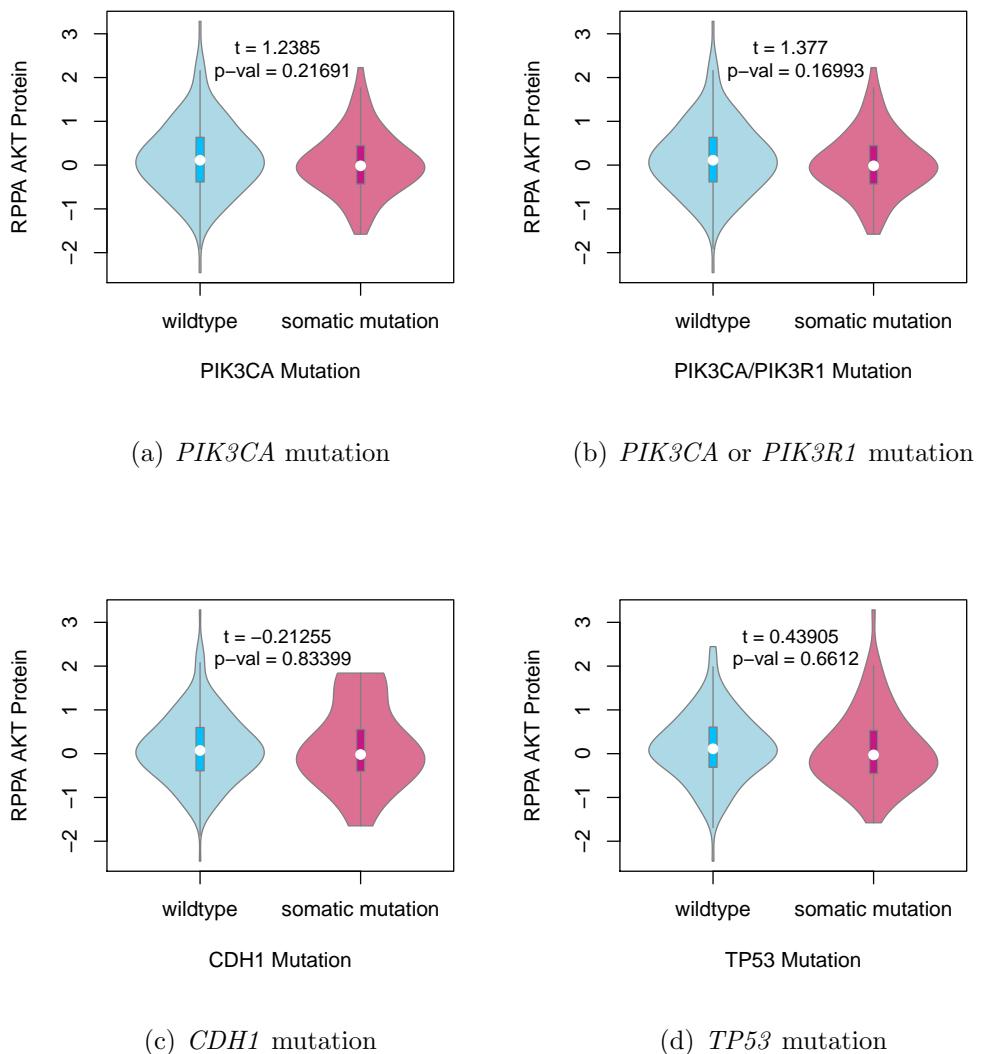


Figure D.7: **Somatic mutation against PIK3CA metagene.** Mutations in *PIK3CA*, *PIK3R1*, *CDH1*, and *TP53* were examined in TCGA breast cancer for their effect on the PIK3CA (Gatza *et al.*, 2014) pathway metagene. The tumour suppressors *CDH1* and *TP53* showed an increase and decrease in the metagene respectively, whereas *PIK3CA* and *PIK3R1* mutations weaker evidence of decrease in metagene levels.



**Figure D.8: Somatic mutation against PI3K protein.** Mutations in *PIK3CA*, *PIK3R1*, *CDH1*, and *TP53* were examined in TCGA breast cancer for their effect on the expression of the p110 $\alpha$  protein (encoded by *PIK3CA*). Protein levels were significantly elevated in samples with *PIK3CA* or *PIK3R1* mutations and lower in samples with *TP53* mutations.



**Figure D.9: Somatic mutation against AKT protein.** Mutations in *PIK3CA*, *PIK3R1*, *CDH1*, and *TP53* were examined in TCGA breast cancer for their effect on the expression of the AKT protein (a downstream target of *PIK3CA*). Protein levels were not significantly different in samples mutations in any of these cancer genes.

## **Appendix E**

### **Metagene Expression Profiles**

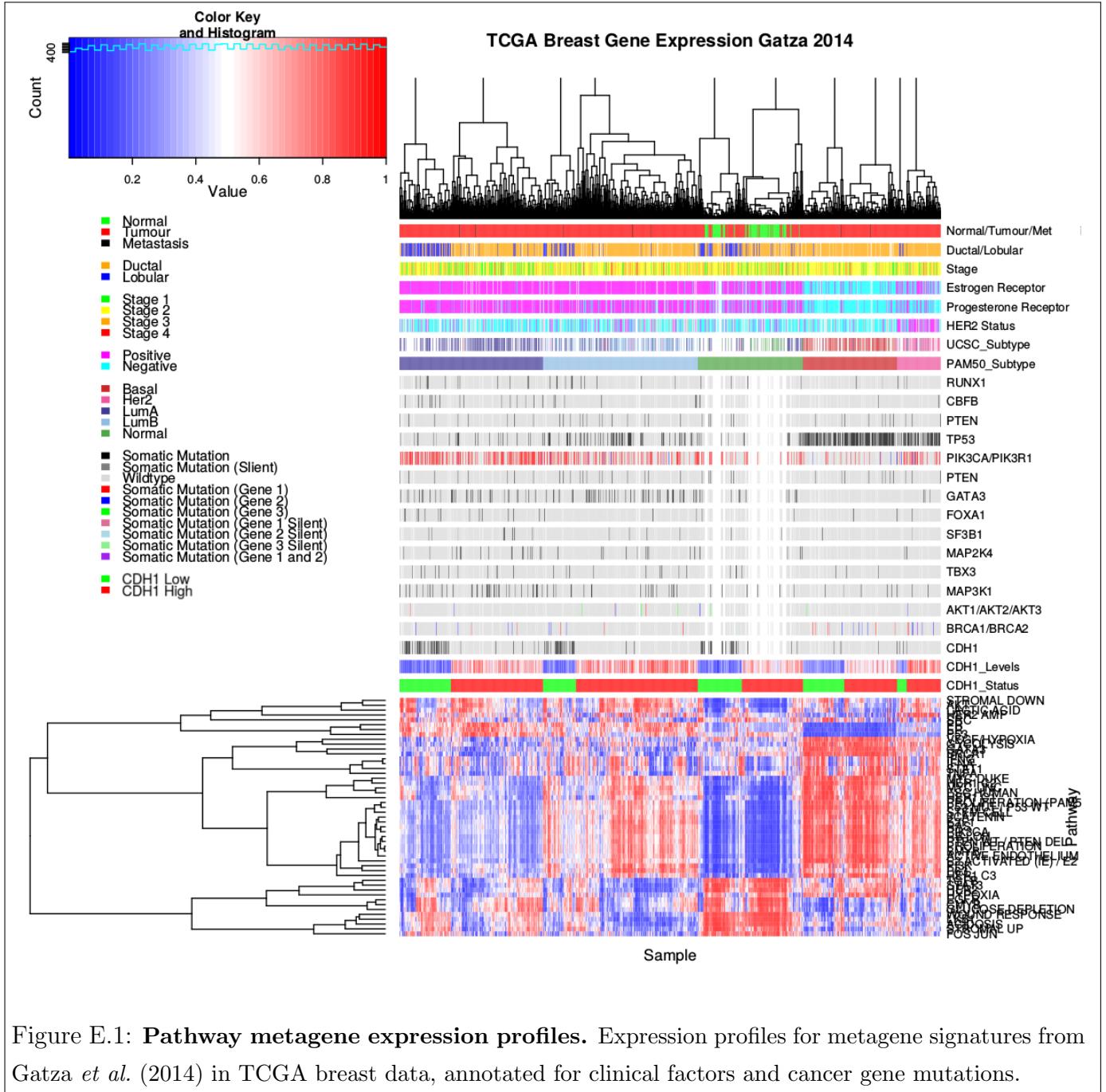
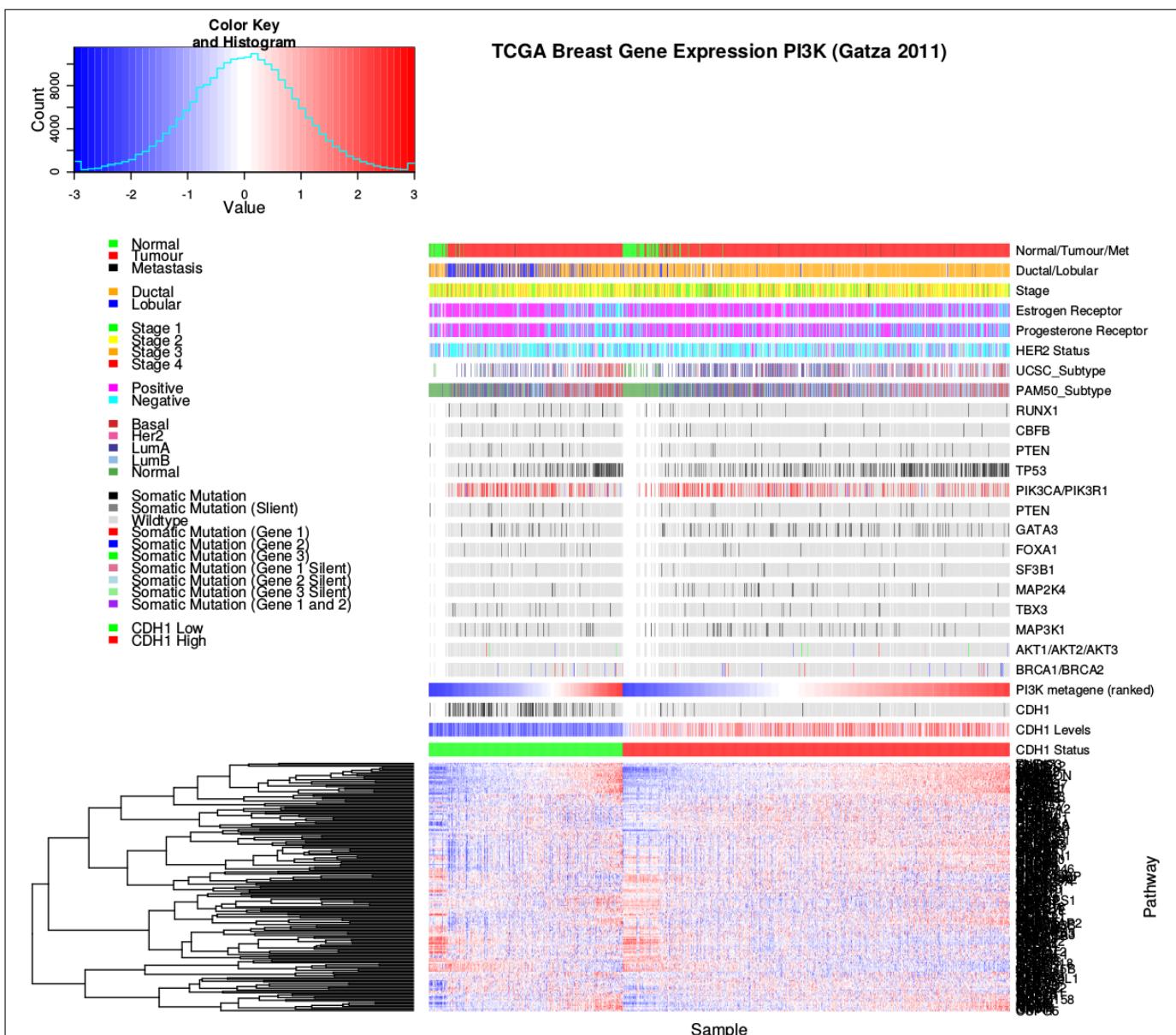
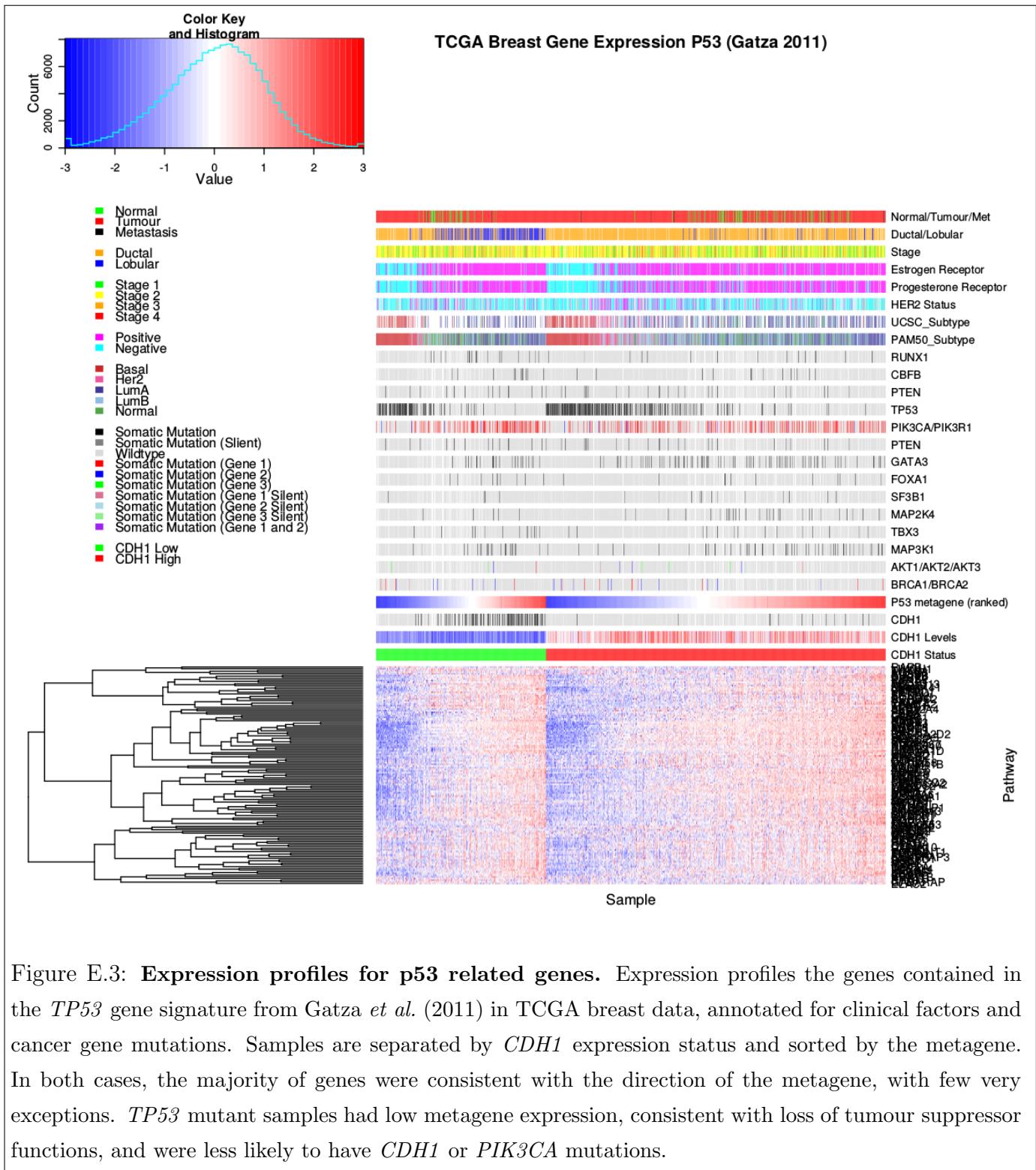
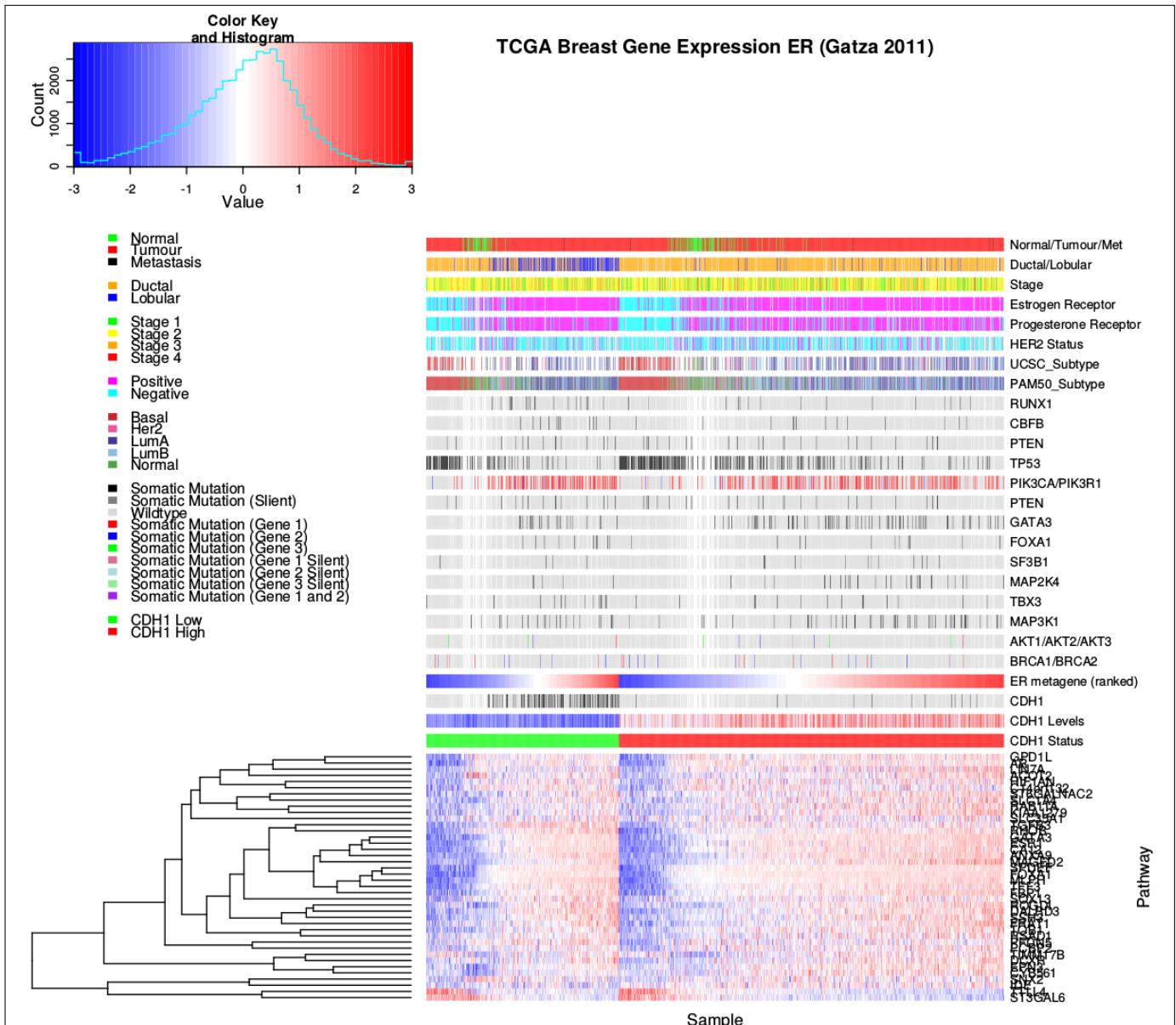


Figure E.1: **Pathway metagene expression profiles.** Expression profiles for metagene signatures from Gatz *et al.* (2014) in TCGA breast data, annotated for clinical factors and cancer gene mutations.

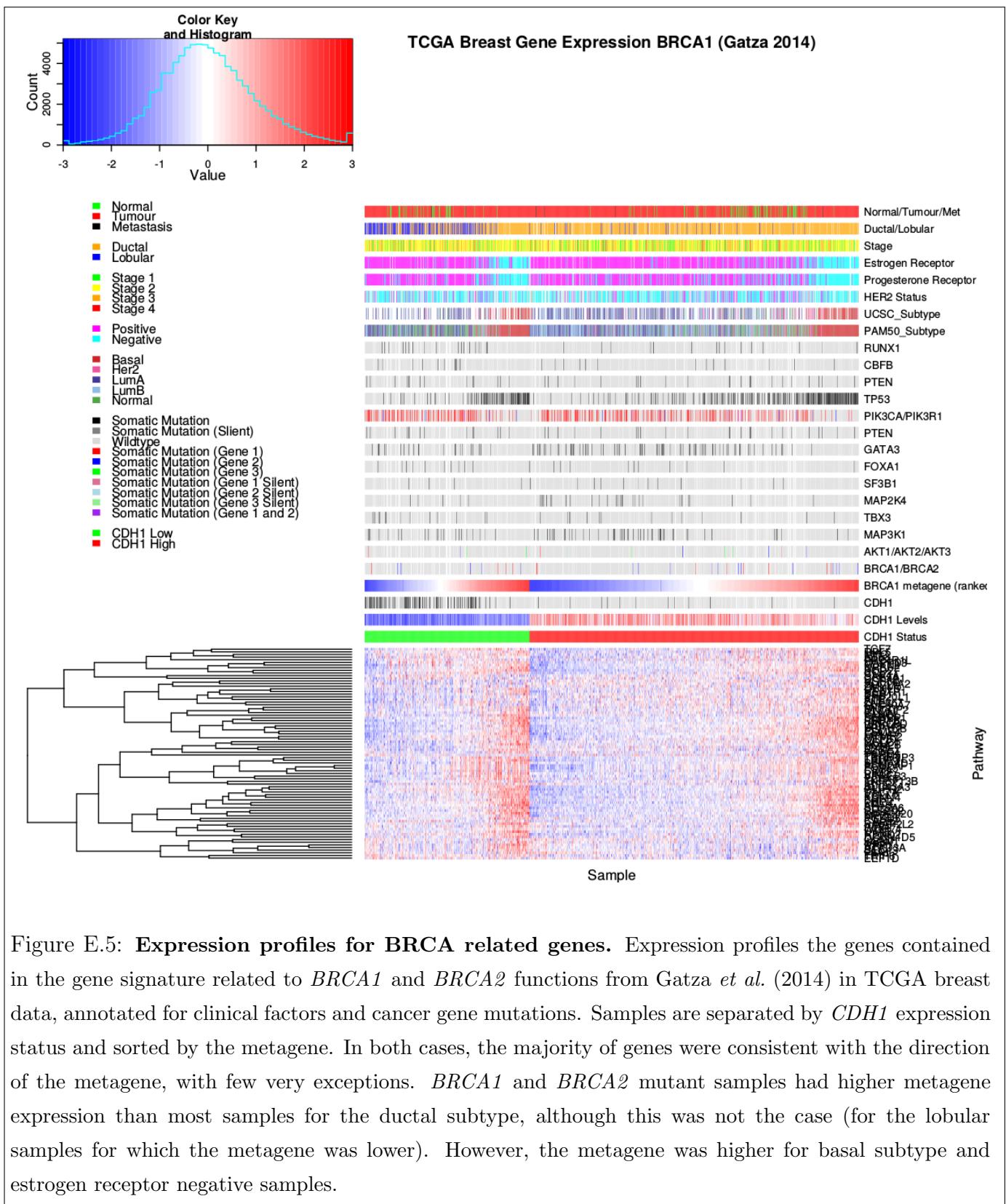


**Figure E.2: Expression profiles for constituent genes of PI3K.** Expression profiles the genes contained in the PI3K gene signature from Gatzka *et al.* (2011) in TCGA breast data, annotated for clinical factors and cancer gene mutations. Samples are separated by *CDH1* expression status and sorted by the metagene. In both cases, the majority of genes were consistent with the direction of the PI3K metagene, although considerable proportion were inversely correlated with the metagene. Normal samples had low PI3K metagene expression and *TP53* mutant samples had high PI3K expression. Although, oncogenic *PIK3CA* and tumour suppressor *PIK3R1* mutations across samples including those with low metagene response.





**Figure E.4: Expression profiles for estrogen receptor related genes.** Expression profiles for the genes contained in the estrogen receptor (ER) gene signature from Gatza *et al.* (2011) in TCGA breast data, annotated for clinical factors and cancer gene mutations. Samples are separated by *CDH1* expression status and sorted by the metagene. In both cases, the majority of genes were consistent with the direction of the metagene, with very few exceptions being inversely correlated. Estrogen receptor (by antibody staining) negative samples had low metagene expression, as expected. These were more likely to be ductal and basal subtypes, lacking *CDH1* or *PIK3CA* mutations.



# **Appendix F**

## **Stomach Expression Analysis**

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

### **F.1 Synthetic Lethal Genes and Pathways**

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

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

SLIPT partners of *CDH1* with observed and expected mutant samples of both genes

Table F.2: 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	$1.3 \times 10^{-97}$
Formation of a pool of free 40S subunits	94	51	$1.3 \times 10^{-97}$
Eukaryotic Translation Elongation	87	49	$4.8 \times 10^{-97}$
Peptide chain elongation	84	48	$1.4 \times 10^{-96}$
Eukaryotic Translation Termination	84	48	$1.4 \times 10^{-96}$
GTP hydrolysis and joining of the 60S ribosomal subunit	105	52	$7.9 \times 10^{-94}$
Nonsense Mediated Decay independent of the Exon Junction Complex	89	48	$3.1 \times 10^{-93}$
Li3a-mediated translational silencing of Ceruloplasmin expression	104	51	$5.1 \times 10^{-92}$
3' UTR-mediated translational regulation	104	51	$5.1 \times 10^{-92}$
SRP-dependent cotranslational protein targeting to membrane	105	51	$1.7 \times 10^{-91}$
Eukaryotic Translation Initiation	112	52	$3.3 \times 10^{-90}$
Cap-dependent Translation Initiation	112	52	$3.3 \times 10^{-90}$
Translation	142	56	$3.6 \times 10^{-85}$
Nonsense-Mediated Decay	104	48	$1.2 \times 10^{-84}$
Nonsense Mediated Decay enhanced by the Exon Junction Complex	104	48	$1.2 \times 10^{-84}$
Influenza Viral RNA Transcription and Replication	109	48	$4.1 \times 10^{-82}$
Influenza Life Cycle	113	48	$3.4 \times 10^{-80}$
Influenza Infection	118	48	$6.4 \times 10^{-78}$
Infections diseases	349	68	$1.8 \times 10^{-50}$
Formation of the ternary complex, and subsequently, the 43S complex	48	21	$3.7 \times 10^{-43}$

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

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

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

## F.2 Comparison to Primary Screen

The synthetic lethal partners with *CDH1* expression in stomach cancers were also compared to siRNA primary screen data (Telford *et al.*, 2015), as performed in Section 4.2.1. These are expected to be more concordant with the experimental results performed on a null mutant, however this is not the case at the gene level: less genes overlapped with experimental candidates in Figure F.1. This may be affected by lower sample size for mutations in TCGA data or lower frequency (expected value) of *CDH1* mutations compared to low expression.

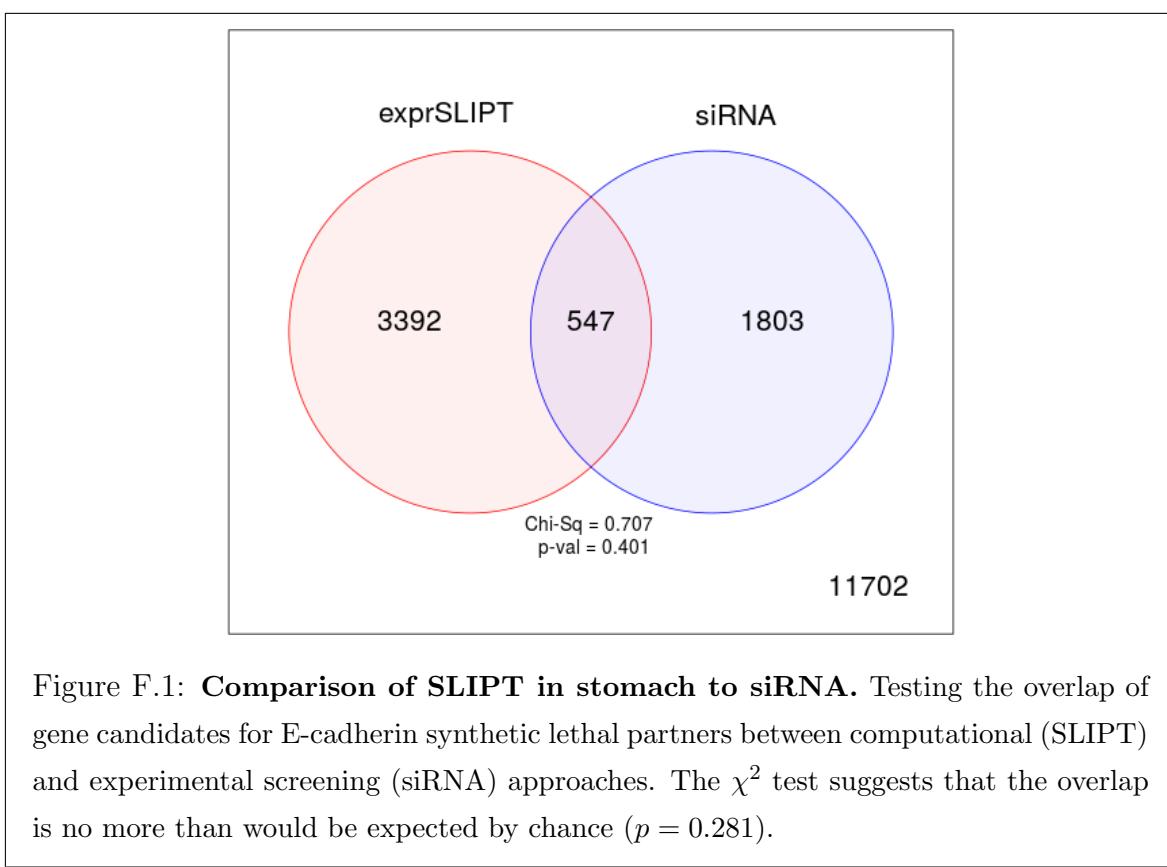


Figure F.1: **Comparison of SLIPT in stomach to siRNA.** Testing the overlap of gene candidates for E-cadherin synthetic lethal partners between computational (SLIPT) and experimental screening (siRNA) approaches. The  $\chi^2$  test suggests that the overlap is no more than would be expected by chance ( $p = 0.281$ ).

Table F.3: Pathway composition for *CDH1* partners from SLIPT and siRNA screening

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

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

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

## F.2.1 Resampling Analysis

Table F.4: Pathways for *CDH1* partners from SLIPT in stomach cancer

Reactome Pathway	Over-representation	Permutation
<i>Extracellular matrix organization</i>	$7.5 \times 10^{-140}$	0.070215
Hemostasis	$1.8 \times 10^{-121}$	0.25804
Developmental Biology	$9.2 \times 10^{-107}$	0.53032
Axon guidance	$1.5 \times 10^{-102}$	0.6704
<b>Eukaryotic Translation Termination</b>	$1.9 \times 10^{-99}$	$> 1.031 \times 10^{-5}$
GPCR ligand binding	$3.8 \times 10^{-99}$	0.54914
<b>Viral mRNA Translation</b>	$3.3 \times 10^{-98}$	$> 1.031 \times 10^{-5}$
Formation of a pool of free 40S subunits	$3.3 \times 10^{-98}$	$> 1.031 \times 10^{-5}$
<b>Eukaryotic Translation Elongation</b>	$1.6 \times 10^{-97}$	$> 1.031 \times 10^{-5}$
Peptide chain elongation	$7.2 \times 10^{-97}$	$> 1.031 \times 10^{-5}$
Class A/1 (Rhodopsin-like receptors)	$2.7 \times 10^{-96}$	0.58174
<b>Nonsense Mediated Decay independent of the Exon Junction Complex</b>	$3 \times 10^{-96}$	$> 1.031 \times 10^{-5}$
Infectious disease	$2.6 \times 10^{-94}$	0.25484
GTP hydrolysis and joining of the 60S ribosomal subunit	$3.4 \times 10^{-94}$	$> 1.031 \times 10^{-5}$
L13a-mediated translational silencing of Ceruloplasmin expression	$2.8 \times 10^{-92}$	$> 1.031 \times 10^{-5}$
<b>3' -UTR-mediated translational regulation</b>	$2.8 \times 10^{-92}$	$> 1.031 \times 10^{-5}$
Neuronal System	$8.4 \times 10^{-92}$	0.53433
SRP-dependent cotranslational protein targeting to membrane	$9.5 \times 10^{-92}$	$> 1.031 \times 10^{-5}$
<b>Eukaryotic Translation Initiation</b>	$2.0 \times 10^{-90}$	$> 1.031 \times 10^{-5}$
Cap-dependent Translation Initiation	$2.0 \times 10^{-90}$	$> 1.031 \times 10^{-5}$
<b>Nonsense-Mediated Decay</b>	$7.4 \times 10^{-90}$	$> 1.031 \times 10^{-5}$
Nonsense Mediated Decay enhanced by the Exon Junction Complex	$7.4 \times 10^{-90}$	$> 1.031 \times 10^{-5}$
Adaptive Immune System	$8.1 \times 10^{-88}$	0.14116
<b>Translation</b>	$1.3 \times 10^{-87}$	$> 1.031 \times 10^{-5}$
Platelet activation, signaling and aggregation	$1.3 \times 10^{-86}$	0.28959
<b>Influenza Infection</b>	$1 \times 10^{-82}$	$> 1.031 \times 10^{-5}$
<b>Influenza Viral RNA Transcription and Replication</b>	$2.4 \times 10^{-82}$	$> 1.031 \times 10^{-5}$
<b>Influenza Life Cycle</b>	$2 \times 10^{-80}$	$> 1.031 \times 10^{-5}$
Response to elevated platelet cytosolic Ca <sup>2+</sup>	$4.9 \times 10^{-78}$	0.50817
Signalling by NGF	$1.6 \times 10^{-75}$	0.38518
Rho GTPase cycle	$5.1 \times 10^{-75}$	0.14864
Signaling by PDGF	$7.4 \times 10^{-74}$	0.40493
<i>Signaling by Rho GTPases</i>	$5.1 \times 10^{-73}$	0.077217
Glycosaminoglycan metabolism	$1.4 \times 10^{-68}$	0.52984
Gαi signalling events	$1.8 \times 10^{-66}$	0.9254
Metabolism of carbohydrates	$1.1 \times 10^{-65}$	0.39501
<b>Gαs signalling events</b>	$2.7 \times 10^{-65}$	0.0050293
Potassium Channels	$2.7 \times 10^{-65}$	0.53359
Transmission across Chemical Synapses	$1.8 \times 10^{-64}$	0.81833
ECM proteoglycans	$3.4 \times 10^{-64}$	0.083482
Peptide ligand-binding receptors	$4.8 \times 10^{-64}$	0.62817
Degradation of the extracellular matrix	$1.1 \times 10^{-63}$	0.80879
Platelet homeostasis	$5.3 \times 10^{-63}$	0.53134
NGF signalling via TRKA from the plasma membrane	$6.1 \times 10^{-63}$	0.57117
Integration of energy metabolism	$4.5 \times 10^{-61}$	0.10889
Collagen formation	$5.4 \times 10^{-61}$	0.29896
Integrin cell surface interactions	$7 \times 10^{-59}$	0.18167
Collagen biosynthesis and modifying enzymes	$7 \times 10^{-59}$	0.30208
Neurotransmitter Receptor Binding And Downstream Transmission	$8.7 \times 10^{-57}$	0.82522
In The Postsynaptic Cell		
Signaling by Wnt	$8.7 \times 10^{-57}$	0.25468

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 F.5: Pathways for *CDH1* partners from SLIPT in stomach and siRNA screen

Reactome Pathway	Over-representation	Permutation
Platelet activation, signaling and aggregation	$3.9 \times 10^{-9}$	0.49557
Class A/1 (Rhodopsin-like receptors)	$3.9 \times 10^{-9}$	0.98432
Response to elevated platelet cytosolic Ca <sup>2+</sup>	$5.5 \times 10^{-8}$	0.54349
Platelet homeostasis	$5.7 \times 10^{-8}$	0.45017
Nucleotide-like (purinergic) receptors	$1.8 \times 10^{-7}$	0.36966
Peptide ligand-binding receptors	$3.8 \times 10^{-7}$	0.91294
<b>Molecules associated with elastic fibres</b>	$7.1 \times 10^{-7}$	0.0025868
Amine ligand-binding receptors	$8.6 \times 10^{-7}$	0.43303
<i>Gαi</i> signalling events	$9.8 \times 10^{-7}$	0.99626
GPCR ligand binding	$1.1 \times 10^{-6}$	0.97733
<b>Elastic fibre formation</b>	$1.5 \times 10^{-6}$	0.0025868
<i>Gαq</i> signalling events	$1.9 \times 10^{-6}$	0.86089
P2Y receptors	$3.8 \times 10^{-6}$	0.18795
Serotonin receptors	$3.8 \times 10^{-6}$	0.37853
Signal amplification	$2.3 \times 10^{-5}$	0.47856
Gastrin-CREB signalling pathway via PKC and MAPK	$2.3 \times 10^{-5}$	0.98567
<b>Complement cascade</b>	$2.4 \times 10^{-5}$	$> 3.4628 \times 10^{-6}$
Glycosaminoglycan metabolism	$2.5 \times 10^{-5}$	0.38953
Glycogen breakdown (glycogenolysis)	$2.7 \times 10^{-5}$	0.83772
Defective B4GALT7 causes EDS, progeroid type	$4.9 \times 10^{-5}$	0.10792
Defective B3GAT3 causes JDSSDHD	$4.9 \times 10^{-5}$	0.10792
Role of LAT2/NTAL/LAB on calcium mobilization	$5.6 \times 10^{-5}$	0.35373
Cell surface interactions at the vascular wall	$5.6 \times 10^{-5}$	0.47642
<i>Gαs</i> signalling events	$6 \times 10^{-5}$	0.019858
Signaling by NOTCH	$6 \times 10^{-5}$	0.19008
A tetrasaccharide linker sequence is required for GAG synthesis	0.00017	0.47642
<b>Extracellular matrix organization</b>	0.00018	0.0047308
Collagen formation	0.00018	0.19245
Effects of PIP2 hydrolysis	0.0002	0.37779
Syndecan interactions	0.0002	0.37779
<b>Diseases associated with glycosaminoglycan metabolism</b>	0.00023	0.01028
<b>Diseases of glycosylation</b>	0.00023	0.01028
<i>Chondroitin sulfate/dermatan sulfate metabolism</i>	0.00023	0.085541
Integrin alphaIIb beta3 signaling	0.00028	0.76936
Keratan sulfate biosynthesis	0.00034	0.68744
Rho GTPase cycle	0.00034	0.15675
Creation of C4 and C2 activators	0.00035	0.12275
Abacavir transport and metabolism	0.00035	0.12443
Amine compound SLC transporters	0.00037	0.69773
FCER1 mediated NF-κB activation	0.00037	0.69846
Fc epsilon receptor (FCER1) signaling	0.00056	0.43303
Defective EXT2 causes exostoses 2	0.00067	0.16053
Defective EXT1 causes exostoses 1, TRPS2 and CHDS	0.00067	0.16053
<i>Collagen biosynthesis and modifying enzymes</i>	0.00071	0.052911
Keratan sulfate/keratin metabolism	0.00073	0.46533
G alpha (12/13) signalling events	0.00078	0.59164
<b>SEMA3A-Plexin repulsion signaling by inhibiting Integrin adhesion</b>	0.00084	0.038504
Signal attenuation	0.00084	0.37779
Eicosanoid ligand-binding receptors	0.0011	0.11117
SOS-mediated signalling	0.0011	0.25387

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

### F.3 Metagene Analysis

Metagene analysis was also performed for synthetic lethal candidates for *CDH1* expression in stomach cancer. These are described and compared to mutation analysis in Section G.4.

Table F.6: Candidate synthetic lethal metagenes against *CDH1* from SLIPT in stomach cancer

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

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

# **Appendix G**

## **Stomach Mutation Analysis**

The following results are a replication of the TCGA results (in Appendix D) with stomach cancer data, using synthetic lethality (mtSLIPT) against *CDH1* mutation.

### **G.1 Synthetic Lethal Genes and Pathways**

Table G.1: Synthetic lethal gene partners of *CDH1* from mtSLIPT in stomach cancer

Gene	Observed	Expected	$\chi^2$ value	p-value	p-value (FDR)
<i>OLFML1</i>	5	10.1	29.2	$4.53 \times 10^{-7}$	0.0031
<i>NRIP2</i>	6	10.1	25.4	$3.11 \times 10^{-6}$	0.00706
<i>VIM</i>	3	10.1	24.7	$4.29 \times 10^{-6}$	0.00706
<i>TCF4</i>	5	10.1	24.7	$4.33 \times 10^{-6}$	0.00706
<i>ZEB2</i>	5	10.1	24.7	$4.33 \times 10^{-6}$	0.00706
<i>BCL2</i>	2	10.1	22	$1.66 \times 10^{-5}$	0.0155
<i>SMARCA2</i>	2	10.1	22	$1.66 \times 10^{-5}$	0.0155
<i>CCND2</i>	3	10.1	21.1	$2.61 \times 10^{-5}$	0.0155
<i>MMP19</i>	3	10.1	21.1	$2.61 \times 10^{-5}$	0.0155
<i>NEURL1B</i>	3	10.1	21.1	$2.61 \times 10^{-5}$	0.0155
<i>IGFBP6</i>	6	10.1	21.1	$2.65 \times 10^{-5}$	0.0155
<i>OGN</i>	6	10.1	21.1	$2.65 \times 10^{-5}$	0.0155
<i>THY1</i>	6	10.2	21	$2.7 \times 10^{-5}$	0.0155
<i>DZIP1</i>	4	10.1	20.6	$3.29 \times 10^{-5}$	0.0155
<i>LOC650368</i>	4	10.1	20.6	$3.29 \times 10^{-5}$	0.0155
<i>PCOLCE</i>	4	10.1	20.6	$3.29 \times 10^{-5}$	0.0155
<i>PTGFR</i>	4	10.1	20.6	$3.29 \times 10^{-5}$	0.0155
<i>RUNX1T1</i>	4	10.1	20.6	$3.29 \times 10^{-5}$	0.0155
<i>CLEC2B</i>	5	10.1	20.6	$3.3 \times 10^{-5}$	0.0155
<i>MSC</i>	5	10.1	20.6	$3.3 \times 10^{-5}$	0.0155
<i>NISCH</i>	5	10.1	20.6	$3.3 \times 10^{-5}$	0.0155
<i>TSPAN11</i>	5	10.1	20.6	$3.3 \times 10^{-5}$	0.0155
<i>KCTD12</i>	2	10.1	19.1	$7.19 \times 10^{-5}$	0.0246
<i>LRRK55</i>	2	10.1	19.1	$7.19 \times 10^{-5}$	0.0246
<i>PCBP3</i>	2	10.1	19.1	$7.19 \times 10^{-5}$	0.0246

mtSLIPT partners with observed and expected *CDH1* mutant samples with low expression

Table G.2: Pathways for *CDH1* partners from mtSLIPT in stomach cancer

Pathways Over-represented	Pathway Size	SL Genes	p-value (FDR)
Extracellular matrix organization	241	20	$9.6 \times 10^{-9}$
Elastic fibre formation	38	6	$3.7 \times 10^{-8}$
Diseases associated with glycosaminoglycan metabolism	26	5	$3.7 \times 10^{-8}$
Diseases of glycosylation	26	5	$3.7 \times 10^{-8}$
Nitric oxide stimulates guanylate cyclase	24	4	$3.1 \times 10^{-6}$
Molecules associated with elastic fibres	34	4	$3.7 \times 10^{-5}$
Platelet homeostasis	54	5	$3.7 \times 10^{-5}$
Initial triggering of complement	17	3	$3.7 \times 10^{-5}$
Regulation of IGF transport and uptake by IGFBPs	17	3	$3.7 \times 10^{-5}$
Collagen degradation	58	5	$5.6 \times 10^{-5}$
Defective B4GALT7 causes EDS, progeroid type	19	3	$5.6 \times 10^{-5}$
Defective B3GAT3 causes JDSSDHD	19	3	$5.6 \times 10^{-5}$
Degradation of the extracellular matrix	104	7	$8.0 \times 10^{-5}$
ECM proteoglycans	66	5	0.00017
A tetrasaccharide linker sequence is required for GAG synthesis	25	3	0.00025
RHO GTPases Activate WASPs and WAVEs	29	3	0.00059
Non-integrin membrane-ECM interactions	53	4	0.00065
Creation of C4 and C2 activators	11	2	0.00079
Dermatan sulfate biosynthesis	11	2	0.00079
Integrin cell surface interactions	82	5	0.00098

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

## G.2 Synthetic Lethal Expression Profiles

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

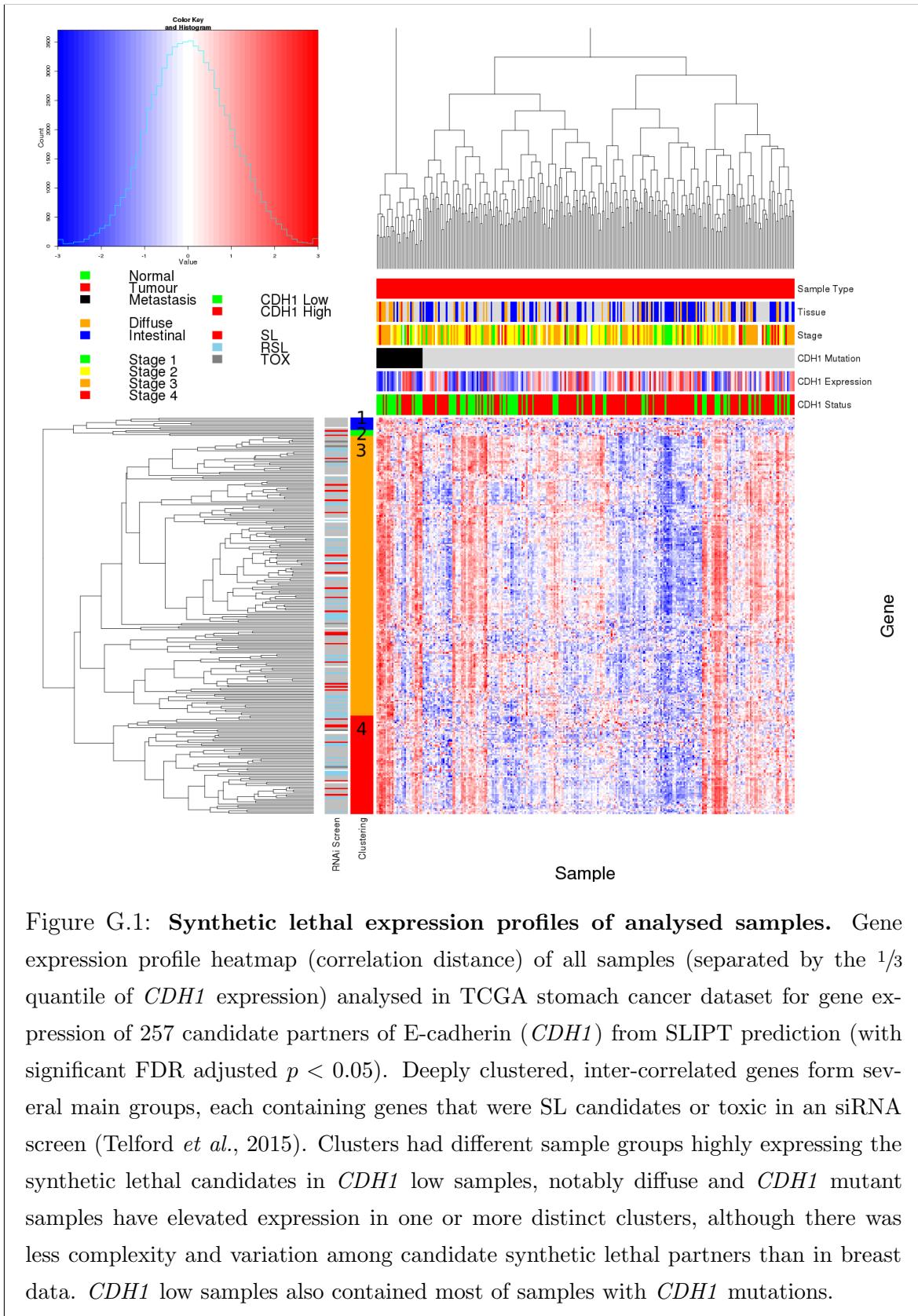


Table G.3: Pathway composition for clusters of *CDH1* partners in stomach mtSLIPT

Pathways Over-represented in Cluster 1	Pathway Size	Cluster Genes	p-value (FDR)
CD28 dependent PI3K/Akt signaling	15	1	1
Hormone-sensitive lipase (HSL)-mediated triacylglycerol hydrolysis	19	1	1
CD28 co-stimulation	26	1	1
Lipid digestion, mobilization, and transport	48	1	1
Costimulation by the CD28 family	51	1	1
Dectin-1 mediated noncanonical NF- $\kappa$ B signaling	58	1	1
CLEC7A (Dectin-1) signaling	99	1	1
C-type lectin receptors (CLRs)	123	1	1
Adaptive Immune System	418	1	1
Metabolism of lipids and lipoproteins	494	1	1
Interleukin-6 signaling	10	0	1
Apoptosis	150	0	1
Hemostasis	445	0	1
Intrinsic Pathway for Apoptosis	36	0	1
Cleavage of Growing Transcript in the Termination Region	33	0	1
PKB-mediated events	28	0	1
PI3K Cascade	68	0	1
RAF/MAP kinase cascade	10	0	1
Global Genomic NER (GG-NER)	35	0	1
Repair synthesis for gap-filling by DNA polymerase in TC-NER	15	0	1

Pathways Over-represented in Cluster 2	Pathway Size	Cluster Genes	p-value (FDR)
Kinesins	22	1	1
O-linked glycosylation of mucins	49	1	1
O-linked glycosylation	59	1	1
MHC class II antigen presentation	85	1	1
Factors involved in megakaryocyte development and platelet production	120	1	1
Post-translational protein modification	303	1	1
Adaptive Immune System	418	1	1
Hemostasis	445	1	1
Interleukin-6 signaling	10	0	1
Apoptosis	150	0	1
Intrinsic Pathway for Apoptosis	36	0	1
Cleavage of Growing Transcript in the Termination Region	33	0	1
PKB-mediated events	28	0	1
PI3K Cascade	68	0	1
RAF/MAP kinase cascade	10	0	1
Global Genomic NER (GG-NER)	35	0	1
Repair synthesis for gap-filling by DNA polymerase in TC-NER	15	0	1
Gap-filling DNA repair synthesis and ligation in TC-NER	17	0	1
Formation of transcription-coupled NER (TC-NER) repair complex	29	0	1
Dual incision reaction in TC-NER	29	0	1

Pathways Over-represented in Cluster 3	Pathway Size	Cluster Genes	p-value (FDR)
Extracellular matrix organization	241	20	$9.6 \times 10^{-9}$
Elastic fibre formation	38	6	$3.7 \times 10^{-8}$
Diseases associated with glycosaminoglycan metabolism	26	5	$3.7 \times 10^{-8}$
Diseases of glycosylation	26	5	$3.7 \times 10^{-8}$
Molecules associated with elastic fibres	34	4	$4.8 \times 10^{-5}$
Initial triggering of complement	17	3	$4.8 \times 10^{-5}$
Regulation of IGF transport and uptake by IGFBPs	17	3	$4.8 \times 10^{-5}$
Collagen degradation	58	5	$6.7 \times 10^{-5}$
Defective B4GALT7 causes EDS, progeroid type	19	3	$6.7 \times 10^{-5}$
Defective B3GAT3 causes JDSSDH	19	3	$6.7 \times 10^{-5}$
Degradation of the extracellular matrix	104	7	$9.5 \times 10^{-5}$
ECM proteoglycans	66	5	0.0002
A tetrasaccharide linker sequence is required for GAG synthesis	25	5	0.00029
Non-integrin membrane-ECM interactions	53	4	0.00079
Creation of C4 and C2 activators	11	2	0.00093
Dermatan sulfate biosynthesis	11	2	0.00093
Integrin cell surface interactions	82	5	0.0012
Keratan sulfate degradation	12	2	0.0012
Complement cascade	34	3	0.0013
CS/DS degradation	13	2	0.0015

Pathways Over-represented in Cluster 4	Pathway Size	Cluster Genes	p-value (FDR)
cGMP effects	18	2	0.11
Nitric oxide stimulates guanylate cyclase	24	2	0.19
Neurotoxicity of clostridium toxins	10	1	1
Platelet homeostasis	54	2	1
Eicosanoid ligand-binding receptors	14	1	1
Prolactin receptor signaling	15	1	1
Acyl chain remodelling of PI	15	1	1
Signaling by FGFR1 fusion mutants	15	1	1
PKA activation	16	1	1
PKA-mediated phosphorylation of CREB	17	1	1
Synthesis of glycosylphosphatidylinositol (GPI)	17	1	1
PKA activation in glucagon signalling	17	1	1
Butyrate Response Factor 1 (BRF1) destabilizes mRNA	17	1	1
Other semaphorin interactions	19	1	1
Acyl chain remodelling of PE	21	1	1
Signaling by Leptin	21	1	1
DARPP-32 events	22	1	1
Glucagon-like Peptide-1 (GLP1) regulates insulin secretion	22	1	1
Uptake and actions of bacterial toxins	22	1	1
Acyl chain remodelling of PC	23	1	1

### G.3 Comparison to Primary Screen

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

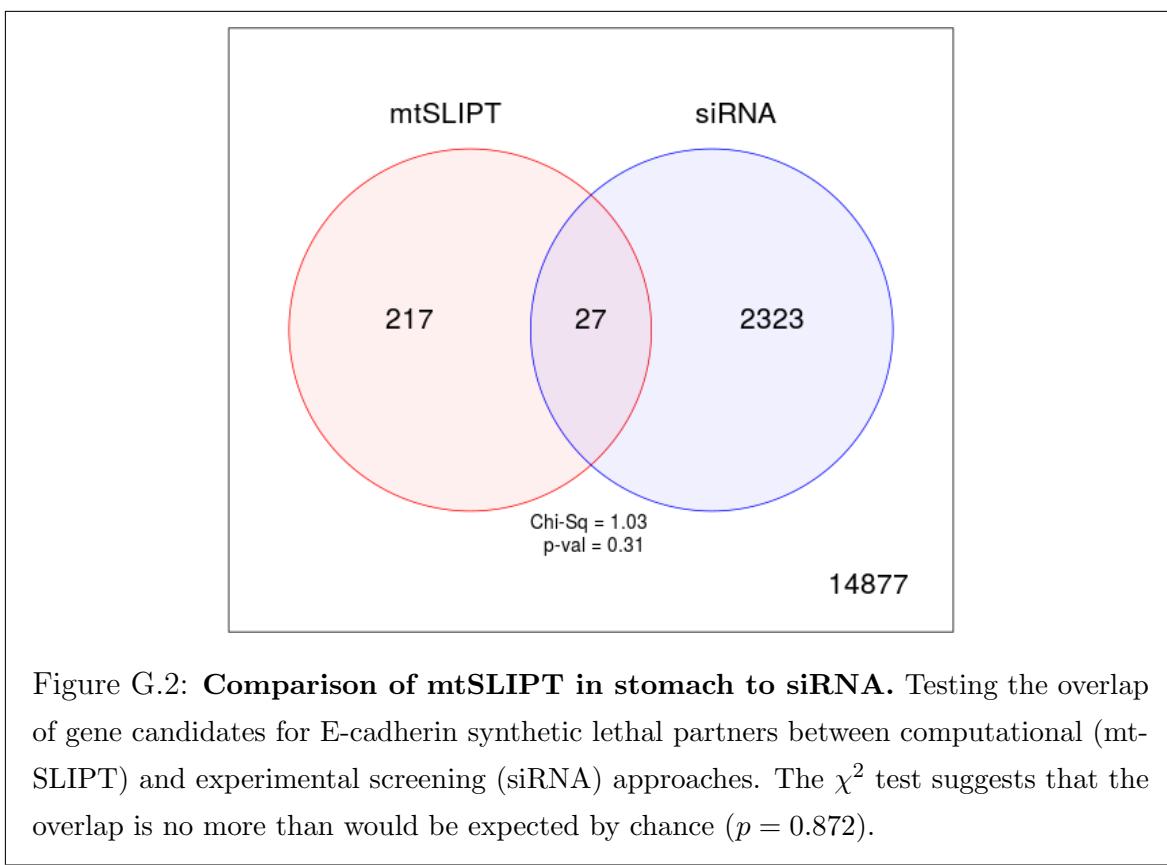


Table G.4: Pathway composition for *CDH1* partners from mtSLIPT and siRNA

Predicted only by SLIPT (217 genes)	Pathway Size	Genes Identified	p-value (FDR)
Diseases associated with glycosaminoglycan metabolism	26	5	$1.6 \times 10^{-7}$
Diseases of glycosylation	26	5	$1.6 \times 10^{-7}$
Extracellular matrix organization	238	18	$1.7 \times 10^{-6}$
Elastic fibre formation	38	5	$4.6 \times 10^{-6}$
Initial triggering of complement	16	3	$7.3 \times 10^{-5}$
Regulation of IGF transport and uptake by IGFBPs	17	3	$8.9 \times 10^{-5}$
Defective B4GALT7 causes EDS, progeroid type	19	3	0.00013
Defective B3GAT3 causes JDSSDHD	19	3	0.00013
Collagen degradation	57	5	0.00013
ECM proteoglycans	65	5	0.00039
A tetrasaccharide linker sequence is required for GAG synthesis	24	3	0.00039
Nitric oxide stimulates guanylate cyclase	24	3	0.00039
RHO GTPases Activate WASPs and WAVEs	28	3	0.00094
Creation of C4 and C2 activators	10	2	0.00098
Non-integrin membrane-ECM interactions	52	4	0.0012
Dermatan sulfate biosynthesis	11	2	0.0013
Degradation of the extracellular matrix	101	6	0.0016
Keratan sulfate degradation	12	2	0.0016
Complement cascade	33	3	0.0018
Molecules associated with elastic fibres	34	3	0.002

Detected only by siRNA screen (2323 genes)	Pathway Size	Genes Identified	p-value (FDR)
Class A/1 (Rhodopsin-like receptors)	282	86	$6.5 \times 10^{-85}$
GPCR ligand binding	363	97	$9.2 \times 10^{-79}$
Peptide ligand-binding receptors	175	52	$4.5 \times 10^{-61}$
G <sub>αi</sub> signalling events	184	49	$1.6 \times 10^{-53}$
G <sub>αq</sub> signalling events	159	43	$5.2 \times 10^{-50}$
Gastrin-CREB signalling pathway via PKC and MAPK	180	46	$9.4 \times 10^{-50}$
DAP12 interactions	159	35	$8.3 \times 10^{-37}$
Platelet activation, signaling and aggregation	182	37	$2.3 \times 10^{-35}$
Hemostasis	438	71	$3.3 \times 10^{-35}$
Downstream signal transduction	146	32	$7.7 \times 10^{-35}$
Signaling by PDGF	172	35	$2.1 \times 10^{-34}$
DAP12 signaling	149	32	$2.7 \times 10^{-34}$
Signaling by ERBB2	146	31	$2.5 \times 10^{-33}$
Signalling by NGF	266	44	$5.3 \times 10^{-31}$
Downstream signaling of activated FGFR1	134	28	$5.3 \times 10^{-31}$
Downstream signaling of activated FGFR2	134	28	$5.3 \times 10^{-31}$
Downstream signaling of activated FGFR3	134	28	$5.3 \times 10^{-31}$
Downstream signaling of activated FGFR4	134	28	$5.3 \times 10^{-31}$
Signaling by FGFR	146	29	$2.0 \times 10^{-30}$
Signaling by FGFR1	146	29	$2.0 \times 10^{-30}$

Intersection of SLIPT and siRNA screen (23 genes)	Pathway Size	Genes Identified	p-value (FDR)
ADP signalling through P2Y purinoceptor 1	10	1	1
G-protein beta:gamma signalling	11	1	1
G-protein activation	12	1	1
Eicosanoid ligand-binding receptors	14	1	1
Platelet homeostasis	53	2	1
G <sub>αz</sub> signalling events	15	1	1
Signal amplification	16	1	1
Activation of Kainate Receptors upon glutamate binding	17	1	1
Thrombin signalling through protease activated receptors (PARs)	17	1	1
Nitric oxide stimulates guanylate cyclase	24	1	1
Activation of G protein gated Potassium channels	25	1	1
G protein gated Potassium channels	25	1	1
Inhibition of voltage gated Ca <sup>2+</sup> channels via Gbeta/gamma subunits	25	1	1
Laminin interactions	29	1	1
Inwardly rectifying K <sup>+</sup> channels	31	1	1
Glucagon signaling in metabolic regulation	33	1	1
Molecules associated with elastic fibres	34	1	1
Ca <sup>2+</sup> pathway	36	1	1
Elastic fibre formation	38	1	1
GABA B receptor activation	38	1	1

### G.3.1 Resampling Analysis

Table G.5: Pathways for *CDH1* partners from mtSLIPT in stomach cancer

Reactome Pathway	Over-representation	Permutation
<i>Extracellular matrix organization</i>	$9.6 \times 10^{-9}$	0.057678
<b>Elastic fibre formation</b>	$3.7 \times 10^{-8}$	0.033817
<i>Diseases associated with glycosaminoglycan metabolism</i>	$3.7 \times 10^{-8}$	0.049336
<i>Diseases of glycosylation</i>	$3.7 \times 10^{-8}$	0.049336
<i>Nitric oxide stimulates guanylate cyclase</i>	$3.1 \times 10^{-6}$	0.037904
<b>Initial triggering of complement</b>	$3.7 \times 10^{-5}$	0.020828
<b>Molecules associated with elastic fibres</b>	$3.7 \times 10^{-5}$	0.027865
<i>Regulation of IGF transport and uptake by IGFBPs</i>	$3.7 \times 10^{-5}$	0.069102
<i>Platelet homeostasis</i>	$3.7 \times 10^{-5}$	0.097294
<i>Defective B4GALT7 causes EDS, progeroid type</i>	$5.6 \times 10^{-5}$	0.081505
<i>Defective B3GAT3 causes JDSSDHD</i>	$5.6 \times 10^{-5}$	0.081505
<b>Collagen degradation</b>	$5.6 \times 10^{-5}$	0.1104
<i>Degradation of the extracellular matrix</i>	$8 \times 10^{-5}$	0.43477
<i>ECM proteoglycans</i>	0.00017	0.06469
<i>A tetrasaccharide linker sequence is required for GAG synthesis</i>	0.00025	0.10536
<i>RHO GTPases Activate WASPs and WAVES</i>	0.00059	0.053929
<i>Non-integrin membrane-ECM interactions</i>	0.00065	0.10424
<i>Creation of C4 and C2 activators</i>	0.00079	0.05461
<i>Dermatan sulfate biosynthesis</i>	0.00079	0.21163
<i>Integrin cell surface interactions</i>	0.00098	0.092405
<i>Glucagon signaling in metabolic regulation</i>	0.00098	0.13425
<i>Keratan sulfate degradation</i>	0.00098	0.22137
<b>Complement cascade</b>	0.0011	0.01552
<i>CS/DS degradation</i>	0.0012	0.065012
<i>Eicosanoid ligand-binding receptors</i>	0.0016	0.066128
<i>Nuclear signaling by ERBB4</i>	0.0016	0.15511
<i>Collagen formation</i>	0.0026	0.13447
<b>cGMP effects</b>	0.0041	0.020195
<i>Voltage gated Potassium channels</i>	0.0041	0.068923
<b>Chondroitin sulfate biosynthesis</b>	0.0059	$> 1.5862 \times 10^{-5}$
<i>Chondroitin sulfate/dermatan sulfate metabolism</i>	0.0065	0.087745
<i>Heparan sulfate/heparin (HS-GAG) metabolism</i>	0.0071	0.085622
<i>Synthesis of substrates in N-glycan biosynthesis</i>	0.0085	0.09456
<i>Regulation of actin dynamics for phagocytic cup formation</i>	0.0085	0.096227
<i>CDO in myogenesis</i>	0.01	0.32599
<i>Myogenesis</i>	0.01	0.32599
<i>Syndecan interactions</i>	0.012	0.10975
<i>Activation of Matrix Metalloproteinases</i>	0.012	0.33499
<i>Glycosaminoglycan metabolism</i>	0.012	0.29716
<i>Collagen biosynthesis and modifying enzymes</i>	0.013	0.10774
<i>Keratan sulfate biosynthesis</i>	0.016	0.12644
<i>O-linked glycosylation</i>	0.016	0.65101
<i>Laminin interactions</i>	0.021	0.12766
<i>Biosynthesis of the N-glycan precursor (dolichol lipid-linked oligosaccharide) and transfer to a nascent protein</i>	0.027	0.065782
<i>Sialic acid metabolism</i>	0.027	0.13413
<i>Keratan sulfate/keratin metabolism</i>	0.029	0.15708
<i>Potassium Channels</i>	0.032	0.43477
<i>Fcgamma receptor (FCGR) dependent phagocytosis</i>	0.042	0.15851
<i>Ion transport by P-type ATPases</i>	0.048	0.66686
<i>Retinoid metabolism and transport</i>	0.051	0.058715

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 G.6: Pathways for *CDH1* partners from mtSLIPT in stomach and siRNA screen

Reactome Pathway	Over-representation	Permutation
SLBP independent Processing of Histone Pre-mRNAs	1	$> 1.2349 \times 10^{-5}$
Mitochondrial protein import	1	$> 1.2349 \times 10^{-5}$
Voltage gated Potassium channels	1	$> 1.2349 \times 10^{-5}$
Tandem pore domain potassium channels	1	$> 1.2349 \times 10^{-5}$
L13a-mediated translational silencing of Ceruloplasmin expression	1	$> 1.2349 \times 10^{-5}$
Eukaryotic Translation Elongation	1	$> 1.2349 \times 10^{-5}$
Peptide chain elongation	1	$> 1.2349 \times 10^{-5}$
3' -UTR-mediated translational regulation	1	$> 1.2349 \times 10^{-5}$
Activation of Matrix Metalloproteinases	1	$> 1.2349 \times 10^{-5}$
HIV Infection	1	$> 1.2349 \times 10^{-5}$
Cell Cycle	1	$> 1.2349 \times 10^{-5}$
Influenza Infection	1	$> 1.2349 \times 10^{-5}$
Influenza Life Cycle	1	$> 1.2349 \times 10^{-5}$
Influenza Viral RNA Transcription and Replication	1	$> 1.2349 \times 10^{-5}$
Neurotoxicity of clostridium toxins	1	$> 1.2349 \times 10^{-5}$
p38MAPK events	1	$> 1.2349 \times 10^{-5}$
SCF-beta-TrCP mediated degradation of Emi1	1	$> 1.2349 \times 10^{-5}$
SRP-dependent cotranslational protein targeting to membrane	1	$> 1.2349 \times 10^{-5}$
Vpu mediated degradation of CD4	1	$> 1.2349 \times 10^{-5}$
Serotonin Neurotransmitter Release Cycle	1	$> 1.2349 \times 10^{-5}$
Acetylcholine Binding And Downstream Events	1	$> 1.2349 \times 10^{-5}$
Viral mRNA Translation	1	$> 1.2349 \times 10^{-5}$
Cobalamin (Cbl, vitamin B12) transport and metabolism	1	$> 1.2349 \times 10^{-5}$
ERK/MAPK targets	1	$> 1.2349 \times 10^{-5}$
Vitamin B5 (pantothenate) metabolism	1	$> 1.2349 \times 10^{-5}$
Signaling by BMP	1	$> 1.2349 \times 10^{-5}$
Synthesis of Leukotrienes (LT) and Eoxins (EX)	1	$> 1.2349 \times 10^{-5}$
Separation of Sister Chromatids	1	$> 1.2349 \times 10^{-5}$
Mitotic Metaphase and Anaphase	1	$> 1.2349 \times 10^{-5}$
TRP channels	1	$> 1.2349 \times 10^{-5}$
Defects in cobalamin (B12) metabolism	1	$> 1.2349 \times 10^{-5}$
Regulation by c-FLIP	1	$> 1.2349 \times 10^{-5}$
Attenuation phase	1	$> 1.2349 \times 10^{-5}$
Autodegradation of the E3 ubiquitin ligase COP1	1	$> 1.2349 \times 10^{-5}$
Apoptotic cleavage of cell adhesion proteins	1	$> 1.2349 \times 10^{-5}$
Negative regulation of TCF-dependent signaling by WNT ligand antagonists	1	$> 1.2349 \times 10^{-5}$
PERK regulates gene expression	1	$> 1.2349 \times 10^{-5}$
Regulation of the Fanconi anemia pathway	1	$> 1.2349 \times 10^{-5}$
Passive transport by Aquaporins	1	$> 1.2349 \times 10^{-5}$
Lysosome Vesicle Biogenesis	1	$> 1.2349 \times 10^{-5}$
Zinc transporters	1	$> 1.2349 \times 10^{-5}$
Zinc influx into cells by the SLC39 gene family	1	$> 1.2349 \times 10^{-5}$
Asparagine N-linked glycosylation	1	$> 1.2349 \times 10^{-5}$
AUF1 (hnRNP D0) destabilizes mRNA	1	$> 1.2349 \times 10^{-5}$
Asymmetric localization of PCP proteins	1	$> 1.2349 \times 10^{-5}$
degradation of DVL	1	$> 1.2349 \times 10^{-5}$
CASP8 activity is inhibited	1	$> 1.2349 \times 10^{-5}$
Degradation of GLI1 by the proteasome	1	$> 1.2349 \times 10^{-5}$
BBSome-mediated cargo-targeting to cilium	1	$> 1.2349 \times 10^{-5}$
Regulation of necroptotic cell death	1	$> 1.2349 \times 10^{-5}$

## G.4 Metagene Analysis

Metagene analysis was also performed for synthetic lethal candidates for *CDH1* mutation in stomach cancer. These are described and compared to expression analysis in Section F.3.

Table G.7: Candidate synthetic lethal metagenes against *CDH1* from mtSLIPT in stomach cancer

Pathway	ID	Observed	Expected	$\chi^2$ value	p-value	p-value (FDR)
Prostacyclin signalling through prostacyclin receptor	392851	1	10.1	26.5	$1.73 \times 10^{-6}$	0.00286
Cell surface interactions at the vascular wall	202733	3	10.1	21.1	$2.61 \times 10^{-5}$	0.00642
The NLRP1 inflammasome	844455	3	10.1	21.1	$2.61 \times 10^{-5}$	0.00642
Innate Immune System	168249	6	10.1	21.1	$2.65 \times 10^{-5}$	0.00642
Keratan sulfate/keratin metabolism	1638074	4	10.1	20.6	$3.29 \times 10^{-5}$	0.00642
Keratan sulfate biosynthesis	2022854	4	10.1	20.6	$3.29 \times 10^{-5}$	0.00642
Signaling by SCF-KIT	1433557	5	10.1	20.6	$3.30 \times 10^{-5}$	0.00642
VEGFA-VEGFR2 Pathway	4420097	5	10.1	20.6	$3.30 \times 10^{-5}$	0.00642
p130Cas linkage to MAPK signaling for integrins	372708	2	10.1	19.1	$7.19 \times 10^{-5}$	0.00651
cGMP effects	418457	8	10.1	19	$7.46 \times 10^{-5}$	0.00651
Regulation of cytoskeletal remodeling and cell spreading by IPP complex components	446388	8	10.1	19	$7.46 \times 10^{-5}$	0.00651
Fcgamma receptor (FCGR) dependent phagocytosis	2029480	3	10.1	17.9	0.000127	0.00651
A third proteolytic cleavage releases NICD	157212	7	10.1	17.9	0.00013	0.00651
Signalling by NGF	166520	7	10.1	17.9	0.00013	0.00651
Signaling by VEGF	194138	7	10.1	17.9	0.00013	0.00651
Regulation of thyroid hormone activity	350864	7	10.1	17.9	0.00013	0.00651
Nitric oxide stimulates guanylate cyclase	392154	7	10.1	17.9	0.00013	0.00651
Platelet homeostasis	418346	7	10.1	17.9	0.00013	0.00651
PI3K events in ERBB4 signaling	1250342	4	10.1	17.3	0.000179	0.00651
PIP3 activates AKT signaling	1257604	4	10.1	17.3	0.000179	0.00651
GAB1 signalosome	180292	4	10.1	17.3	0.000179	0.00651
PI3K events in ERBB2 signaling	1963642	4	10.1	17.3	0.000179	0.00651
PI3K/AKT Signaling in Cancer	2219528	4	10.1	17.3	0.000179	0.00651
Rap1 signalling	392517	4	10.1	17.3	0.000179	0.00651
Lysosphingolipid and LPA receptors	419408	4	10.1	17.3	0.000179	0.00651

Strongest candidate SL partners for *CDH1* by mtSLIPT with observed and expected mutant samples with low expression of partner metagenes

## **Appendix H**

### **Global Synthetic Lethality in Stomach Cancer**

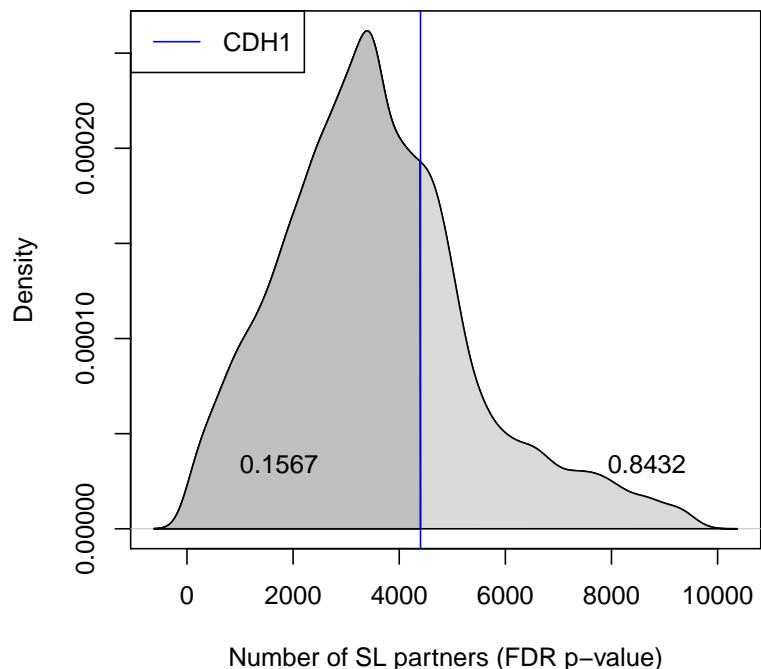


Figure H.1: **Synthetic lethal partners across query genes.** Global synthetic lethal pairs were examined across the genome in TCGA stomach expression data by applying SLIPT across query genes. The high number of predicted partners for *CDH1* was typical for a human gene and lower than many other genes.

## H.1 Hub Genes

Table H.1: Query synthetic lethal genes with the most SLIPT partners

Gene	Direction	raw p-value	p-value (FDR)	SLIPT raw p-value	SLIPT (FDR)
<i>HEG1</i>	10719	16956	16724	9616	9532
<i>SYNE1</i>	10755	17210	16984	9749	9676
<i>A2M</i>	10743	16650	16378	9529	9433
<i>ANK2</i>	11008	16616	16355	9764	9653
<i>TTC28</i>	10757	16523	16248	9530	9429
<i>FAT4</i>	10451	16286	15978	9225	9115
<i>MRVI1</i>	10904	16967	16718	9775	9686
<i>PAPLN</i>	10483	16405	16104	9305	9193
<i>NFASC</i>	10773	16575	16307	9578	9475
<i>MACF1</i>	9697	16378	16058	8620	8540
<i>HMCN1</i>	10475	16101	15733	9156	9008
<i>MPDZ</i>	10878	16550	16299	9599	9491
<i>FLRT2</i>	10776	16760	16473	9590	9464
<i>SETBP1</i>	10869	16632	16349	9615	9489
<i>LAMA4</i>	10463	16447	16121	9273	9151
<i>IL1R1</i>	10611	16185	15803	9299	9174
<i>ABCA6</i>	10499	16573	16318	9260	9158
<i>LAMC1</i>	10238	15777	15392	8837	8691
<i>TNS1</i>	10920	17038	16806	9836	9751
<i>AMOTL1</i>	10612	16458	16178	9367	9250

Genes with the most candidate SL partners SLIPT in TCGA stomach expression data with the number of partner genes predicted by direction criteria and  $\chi^2$  testing separately and combined as a SLIPT analysis. Where specified, the p-values for the  $\chi^2$  test were adjusted for multiple tests (FDR).

## H.2 Hub Pathways

Table H.2: Pathways for genes with the most SLIPT partners

Pathways Over-represented	Pathway Size	SL Genes	p-value	p-value (FDR)
Molecules associated with elastic fibres	34	10	$4.6 \times 10^{-21}$	$2.7 \times 10^{-18}$
Extracellular matrix organization	241	29	$5.3 \times 10^{-21}$	$2.7 \times 10^{-18}$
Smooth Muscle Contraction	29	9	$5.6 \times 10^{-20}$	$1.6 \times 10^{-17}$
Elastic fibre formation	38	10	$6 \times 10^{-20}$	$1.6 \times 10^{-17}$
Nitric oxide stimulates guanylate cyclase	24	8	$6.9 \times 10^{-19}$	$1.4 \times 10^{-16}$
Muscle contraction	64	12	$8.3 \times 10^{-19}$	$1.4 \times 10^{-16}$
Platelet homeostasis	54	11	$1.3 \times 10^{-18}$	$1.9 \times 10^{-16}$
cGMP effects	18	6	$3.3 \times 10^{-15}$	$4.3 \times 10^{-13}$
Laminin interactions	30	7	$1.3 \times 10^{-14}$	$1.6 \times 10^{-12}$
Axon guidance	289	25	$5 \times 10^{-13}$	$5.2 \times 10^{-11}$
Signaling by BMP	23	5	$3.7 \times 10^{-11}$	$3.2 \times 10^{-9}$
RHO GTPases activate PAKs	23	5	$3.7 \times 10^{-11}$	$3.2 \times 10^{-9}$
Non-integrin membrane-ECM interactions	53	7	$7.2 \times 10^{-11}$	$5.8 \times 10^{-9}$
Rho GTPase cycle	120	11	$1.2 \times 10^{-10}$	$8.7 \times 10^{-9}$
Degradation of the extracellular matrix	104	10	$1.3 \times 10^{-10}$	$8.8 \times 10^{-9}$
Netrin-1 signaling	42	6	$2.5 \times 10^{-10}$	$1.6 \times 10^{-8}$
Developmental Biology	432	32	$8.3 \times 10^{-10}$	$5 \times 10^{-8}$
L1CAM interactions	80	8	$8.7 \times 10^{-10}$	$5 \times 10^{-8}$
Semaphorin interactions	64	7	$1.1 \times 10^{-9}$	$6.1 \times 10^{-8}$
Cell-extracellular matrix interactions	18	4	$1.3 \times 10^{-9}$	$6.6 \times 10^{-8}$

Gene set over-representation analysis (hypergeometric test) for Reactome pathways in the top 500 “hub” genes with the most candidate synthetic lethal partners by SLIPT analysis of TCGA stomach expression data

## **Appendix I**

### **Replication in cell line encyclopaedia**

Table I.1: Candidate synthetic lethal gene partners of *CDH1* from SLIPT in CCLE

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

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

Table I.2: Candidate synthetic lethal gene partners of *CDH1* from SLIPT in breast CCLE

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

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

Table I.3: Candidate synthetic lethal gene partners of *CDH1* from SLIPT in stomach CCLE

Gene	Observed	Expected	$\chi^2$ value	p-value	p-value (FDR)
<i>ZEB1</i>	1	4.45	36	$2.84 \times 10^{-7}$	0.00175
<i>WDR47</i>	0	4.45	26.7	$2.3 \times 10^{-5}$	0.013
<i>KANK2</i>	1	4.45	25.1	$4.81 \times 10^{-5}$	0.0222
<i>LEPRE1</i>	0	4.45	24.5	$6.26 \times 10^{-5}$	0.0228
<i>KATNAL1</i>	0	4.45	24.3	$6.88 \times 10^{-5}$	0.0231
<i>TET1</i>	0	4.45	23.9	$8.23 \times 10^{-5}$	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 I.4: 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*

Table I.5: Pathways for *CDH1* partners from SLIPT in breast and stomach CCLE

Pathways Over-represented	Pathway Size	SL Genes	p-value (FDR)
Collagen formation	66	8	$1.1 \times 10^{-7}$
Glycosaminoglycan metabolism	111	11	$1.1 \times 10^{-7}$
Extracellular matrix organization	236	20	$1.1 \times 10^{-7}$
Collagen biosynthesis and modifying enzymes	55	7	$1.7 \times 10^{-7}$
Keratan sulfate biosynthesis	28	5	$2.2 \times 10^{-7}$
Keratan sulfate/keratin metabolism	32	5	$7.5 \times 10^{-7}$
ECM proteoglycans	65	7	$1.1 \times 10^{-6}$
Non-integrin membrane-ECM interactions	52	6	$2.0 \times 10^{-6}$
Cell junction organization	71	7	$3.0 \times 10^{-6}$
Assembly of collagen fibrils and other multimeric structures	39	5	$3.6 \times 10^{-6}$
Post-chaperonin tubulin folding pathway	14	3	$1.7 \times 10^{-5}$
Adherens junctions interactions	29	4	$1.7 \times 10^{-5}$
Cell-Cell communication	118	9	$1.7 \times 10^{-5}$
Sialic acid metabolism	31	4	$2.5 \times 10^{-5}$
Synthesis and interconversion of nucleotide di- and triphosphates	16	3	$3.1 \times 10^{-5}$
Transport to the Golgi and subsequent modification	34	4	$4.8 \times 10^{-5}$
Asparagine N-linked glycosylation	113	8	$7.8 \times 10^{-5}$
Elastic fibre formation	37	4	$8.5 \times 10^{-5}$
L1CAM interactions	77	6	$9.5 \times 10^{-5}$
Signal transduction by L1	20	3	$9.5 \times 10^{-5}$

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

## Appendix J

# Synthetic Lethal Genes in Pathways

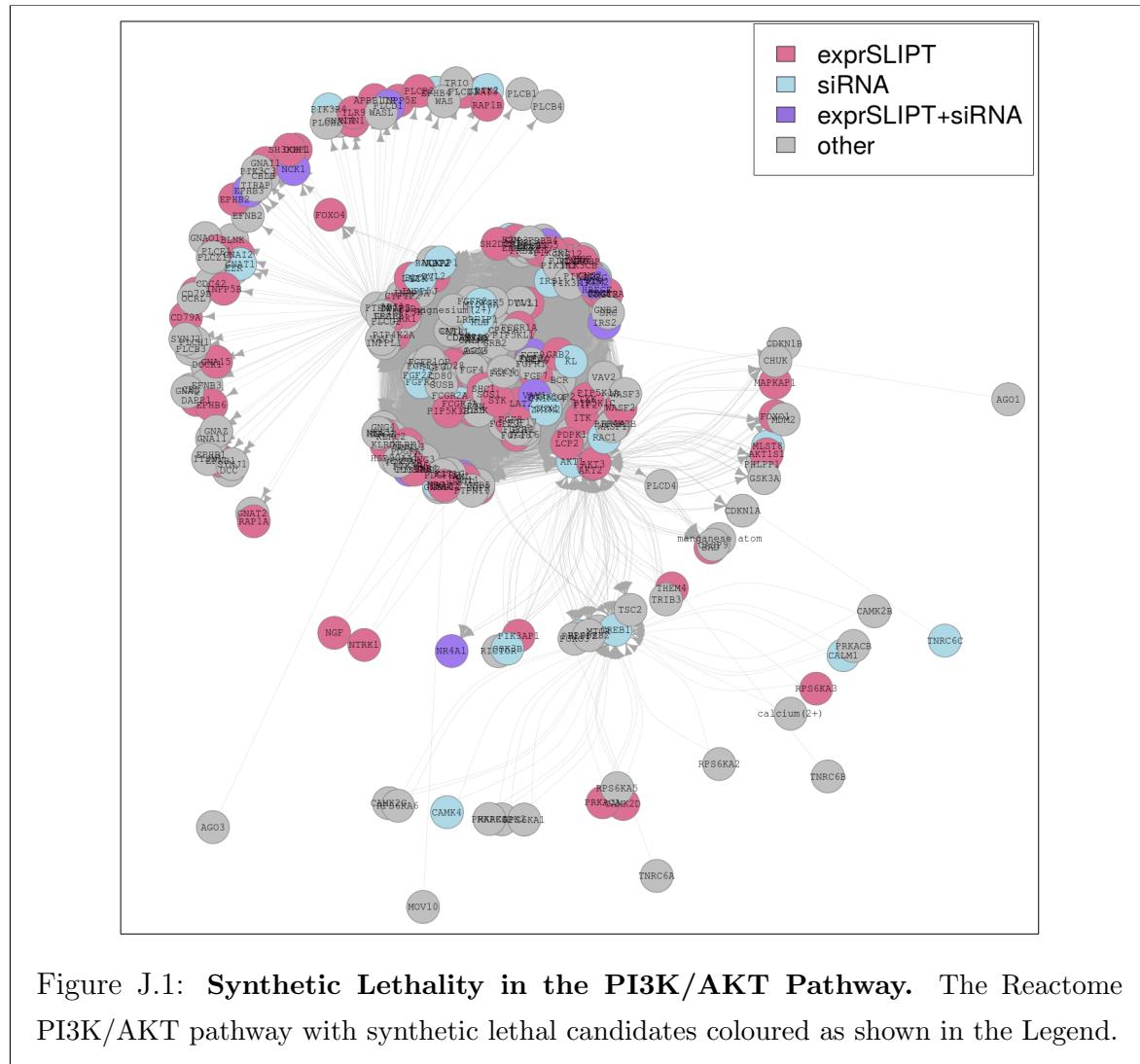
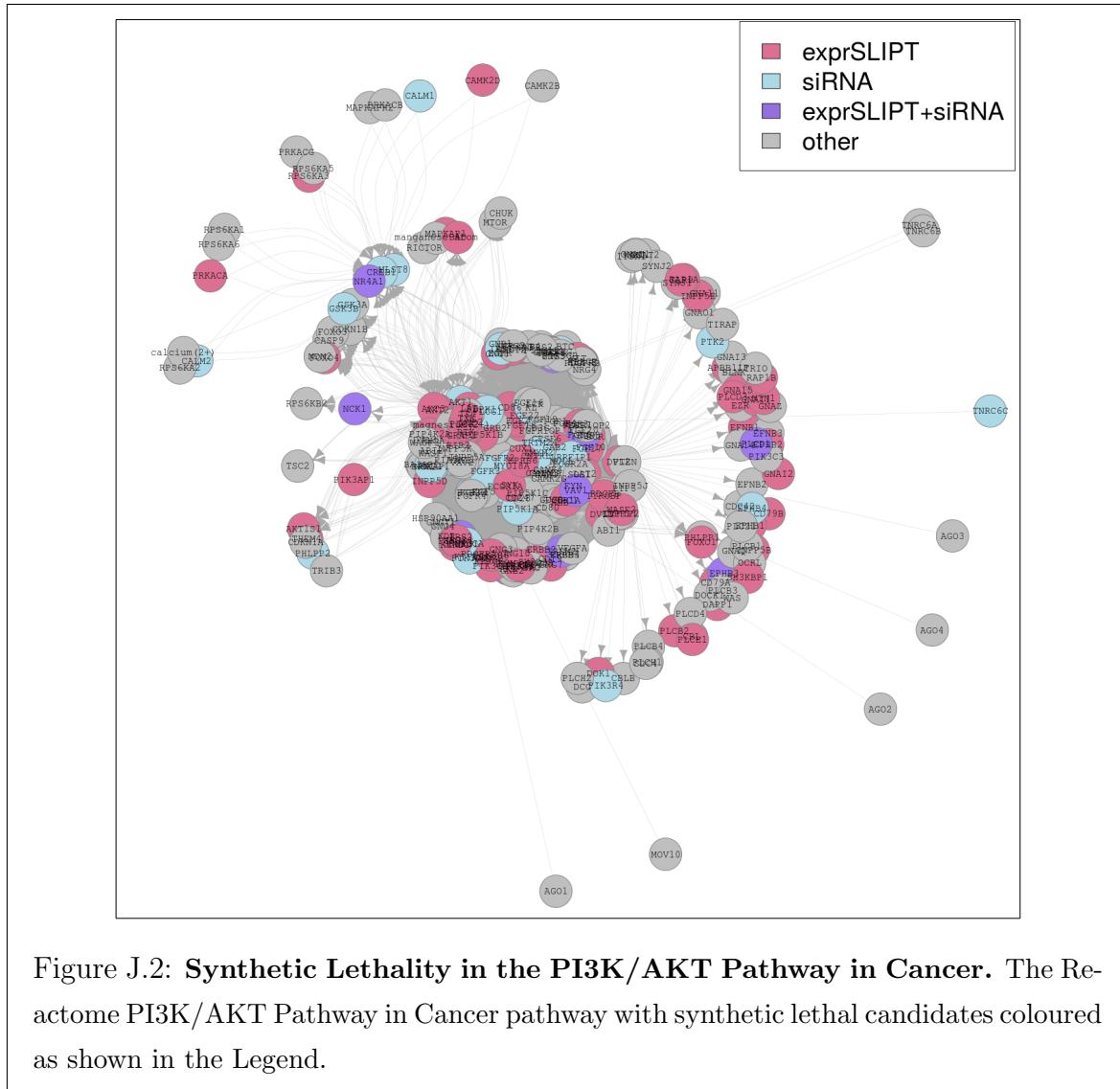
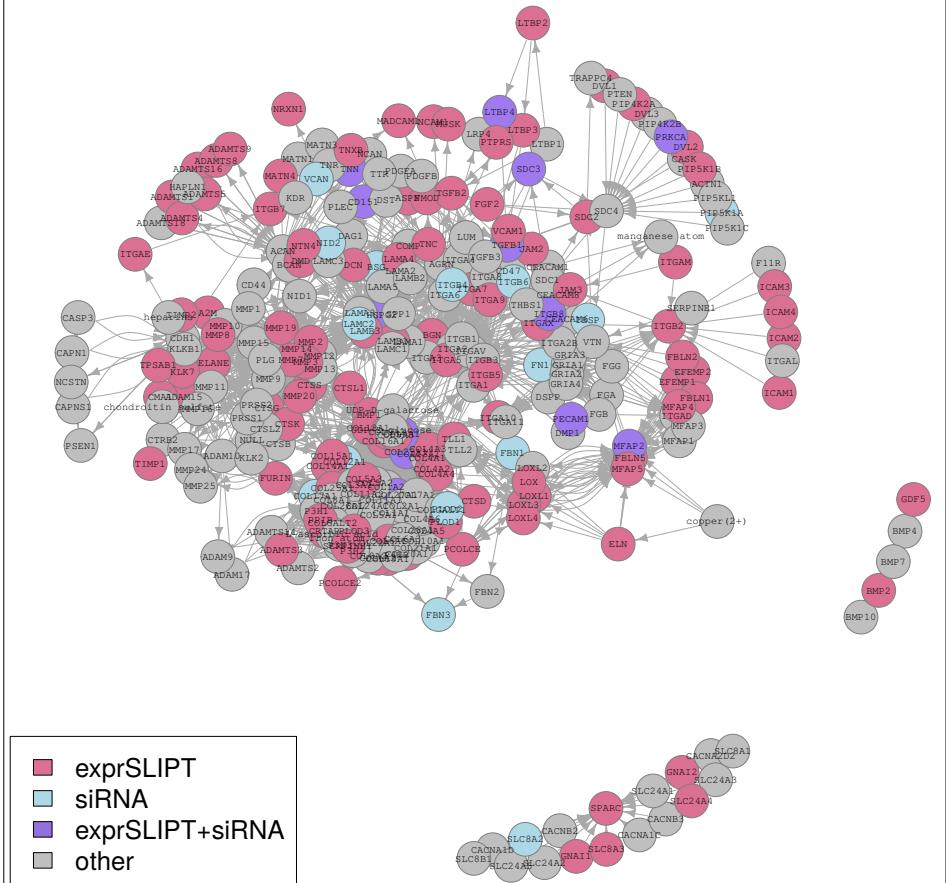


Figure J.1: Synthetic Lethality in the PI3K/AKT Pathway. The Reactome PI3K/AKT pathway with synthetic lethal candidates coloured as shown in the Legend.





**Figure J.3: Synthetic Lethality in the Extracellular Matrix.** The Reactome Extracellular Matrix pathway with synthetic lethal candidates coloured as shown in the Legend.

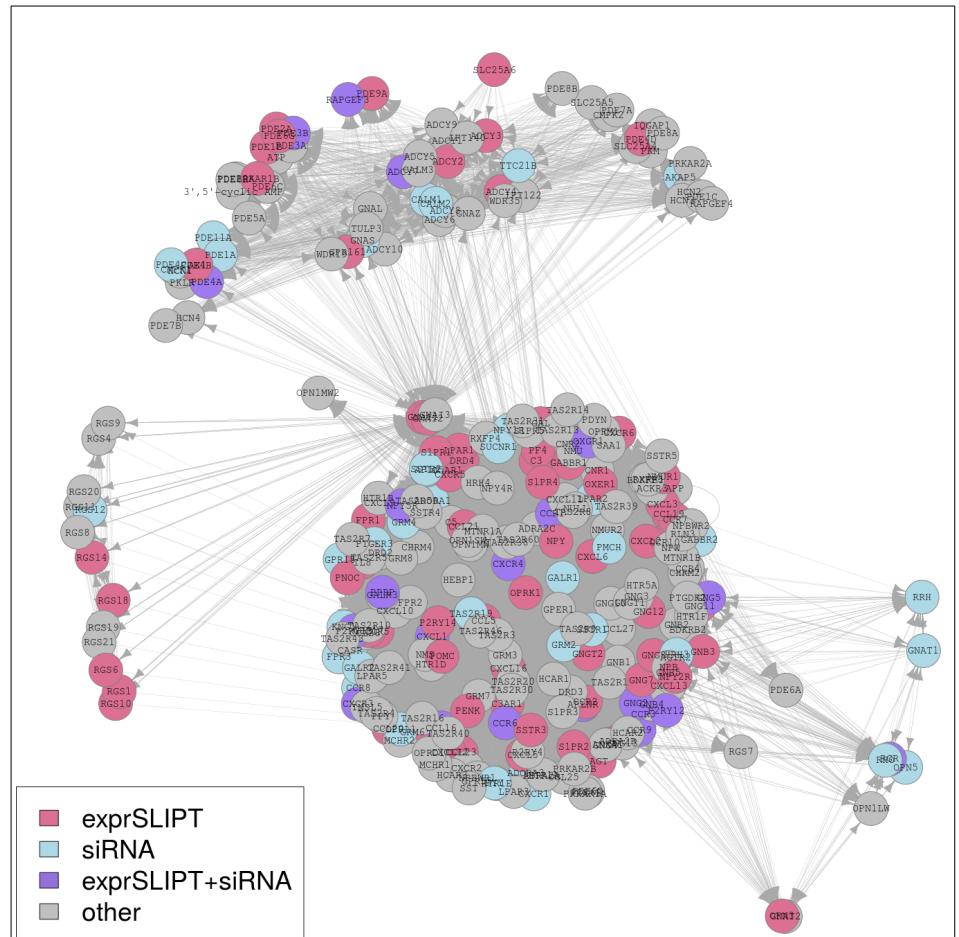


Figure J.4: **Synthetic Lethality in the GPCRs.** The Reactome  $G_{\alpha i}$  pathway with synthetic lethal candidates coloured as shown in the Legend.

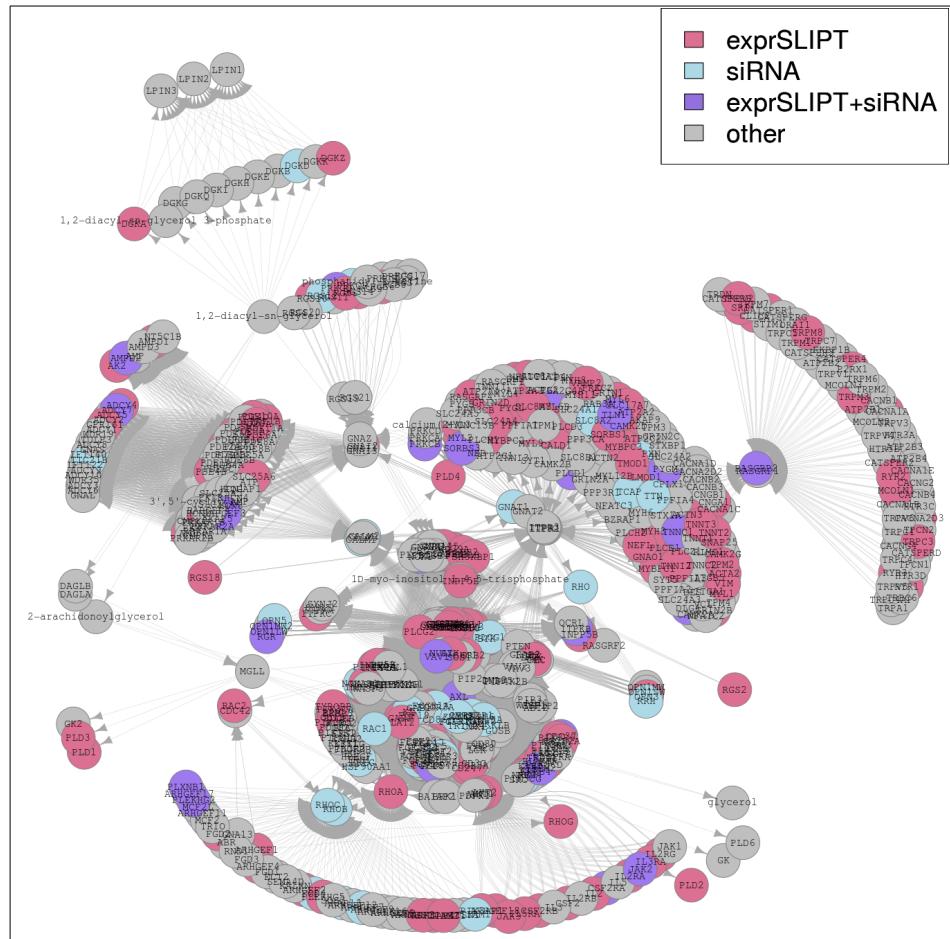
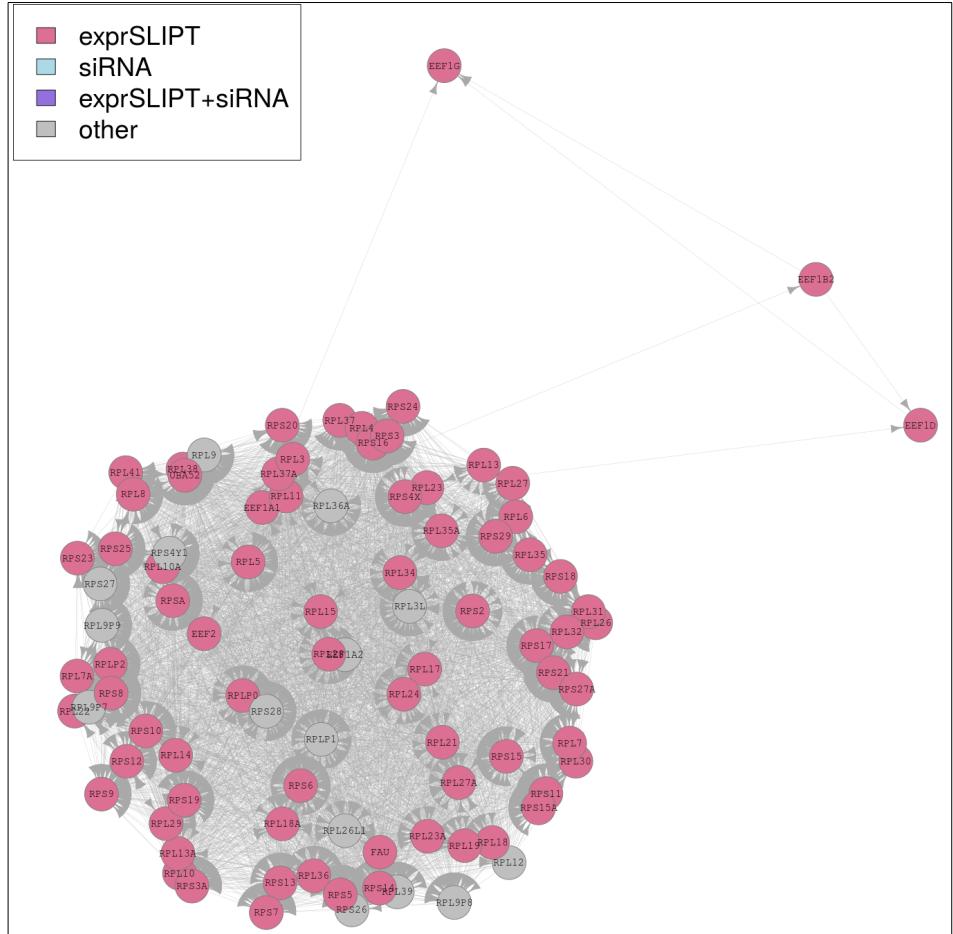
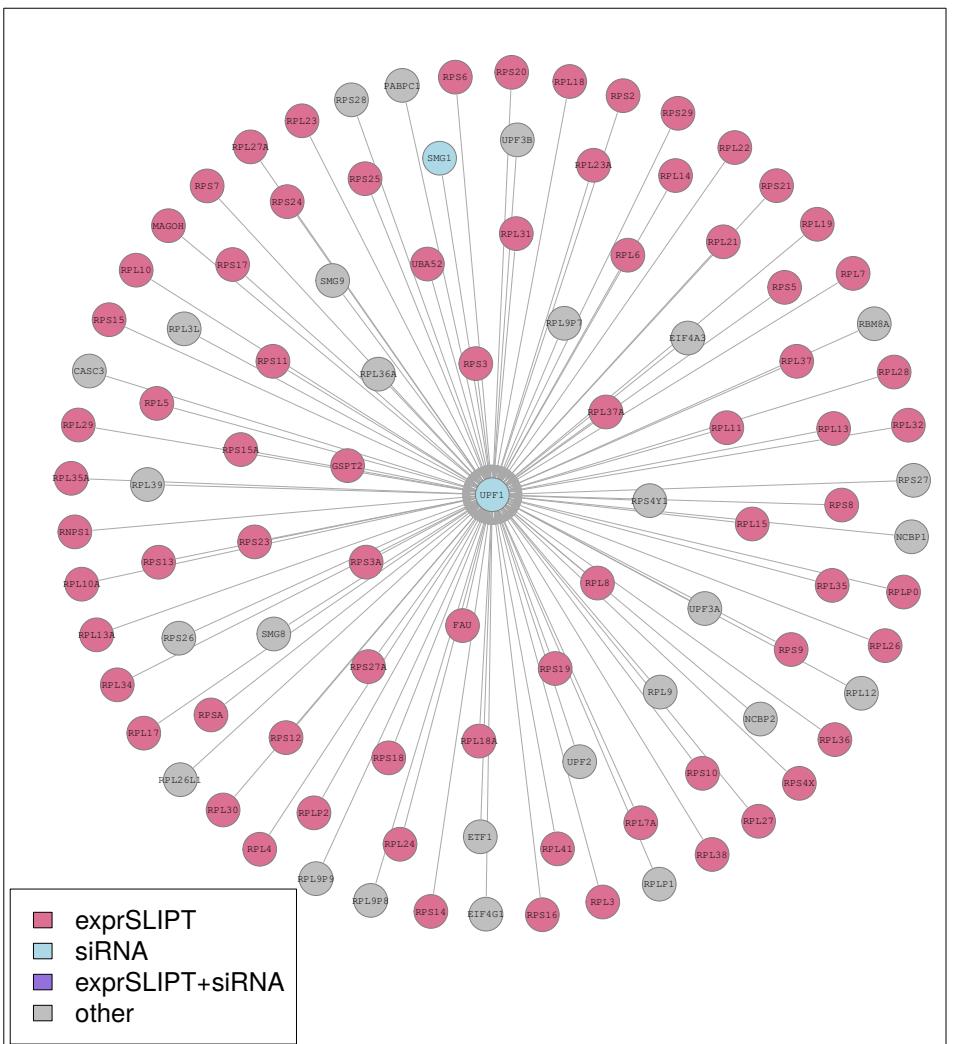


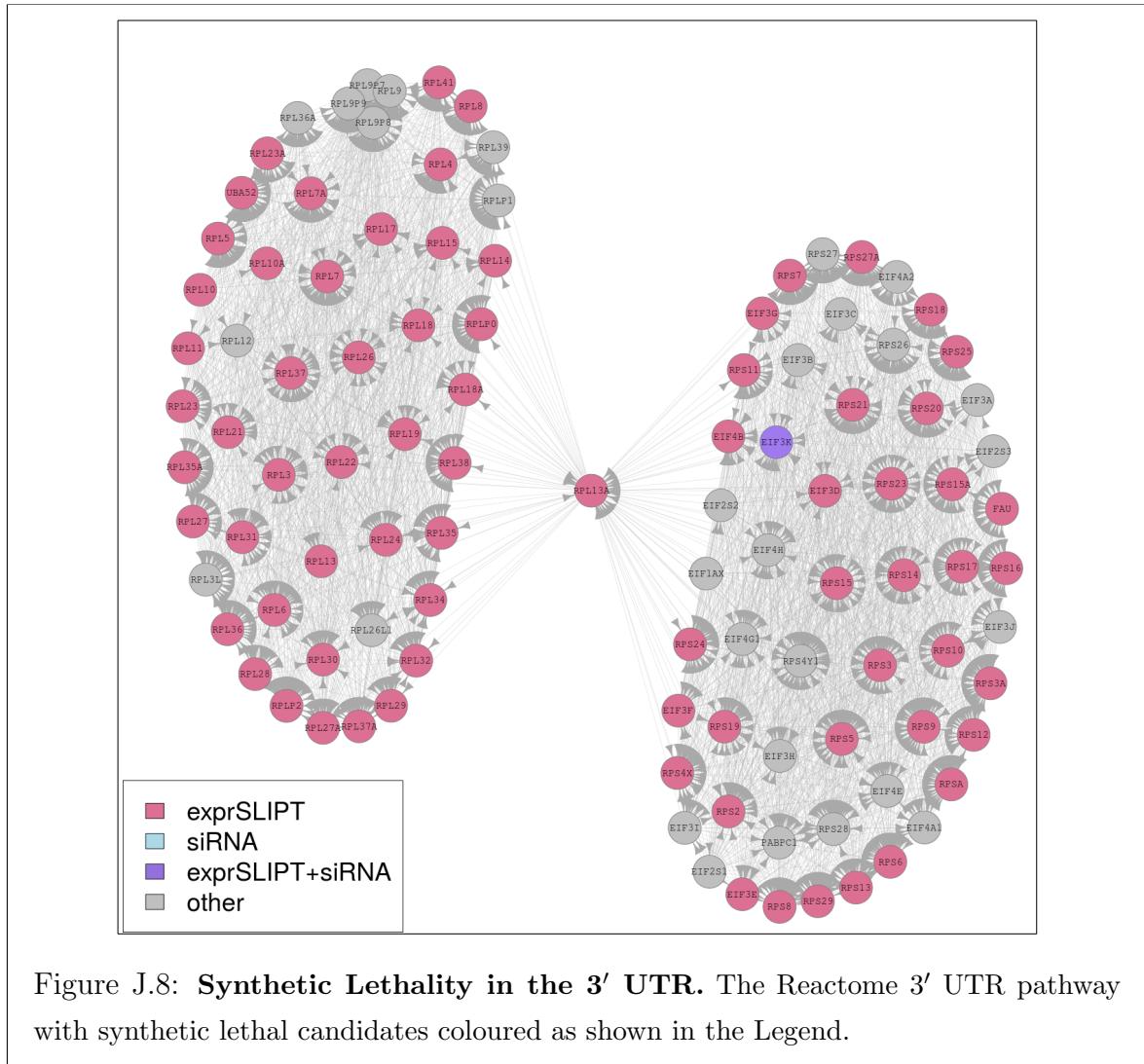
Figure J.5: **Synthetic Lethality in the GPCR Downstream.** The Reactome GPCR Downstream pathway with synthetic lethal candidates coloured as shown in the Legend.



**Figure J.6: Synthetic Lethality in the Translation Elongation.** The Reactome Translation Elongation pathway with synthetic lethal candidates coloured as shown in the Legend.



**Figure J.7: Synthetic Lethality in the Nonsense-mediated Decay.** The Reactome Nonsense-mediated Decay pathway with synthetic lethal candidates coloured as shown in the Legend.



## Appendix K

# Pathway Connectivity for Mutation SLIPT

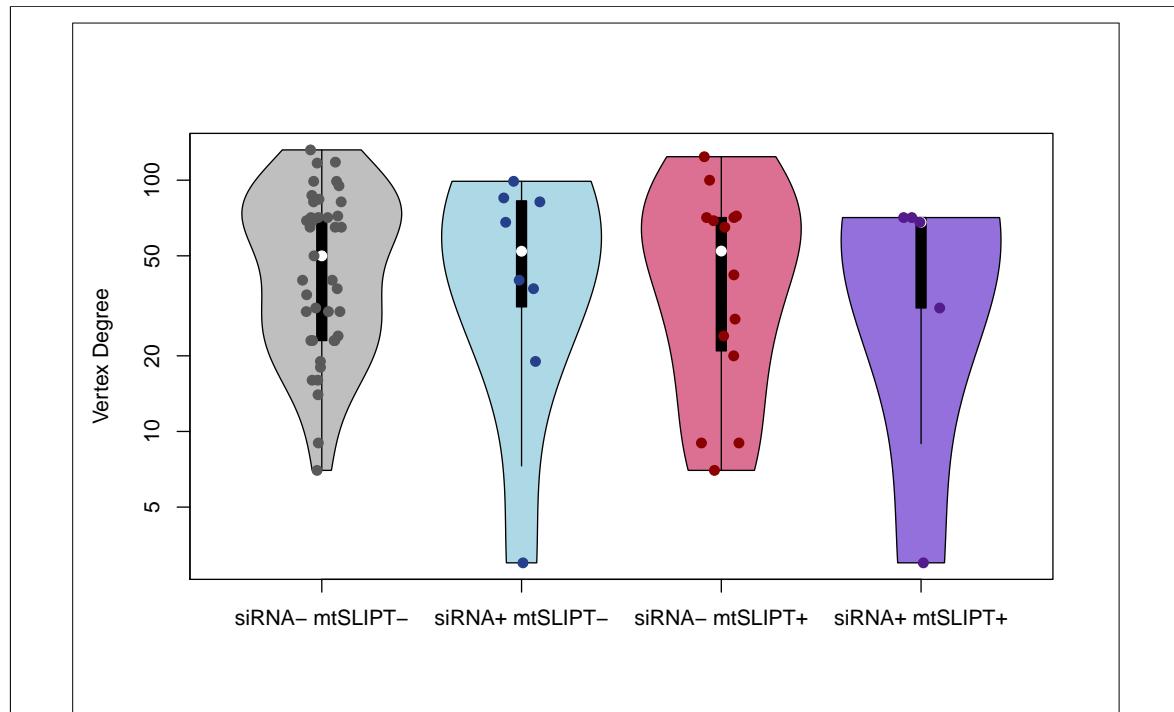


Figure K.1: **Synthetic Lethality and Vertex Degree.** The number of connected genes (vertex degree) was compared (on a log-scale across genes detected by mtSLIPT and short interfering ribonucleic acid (siRNA) screening in the Reactome PI3K cascade pathway. There were very few differences in vertex degree between the groups, although genes detected by siRNA included those with the fewest connections.

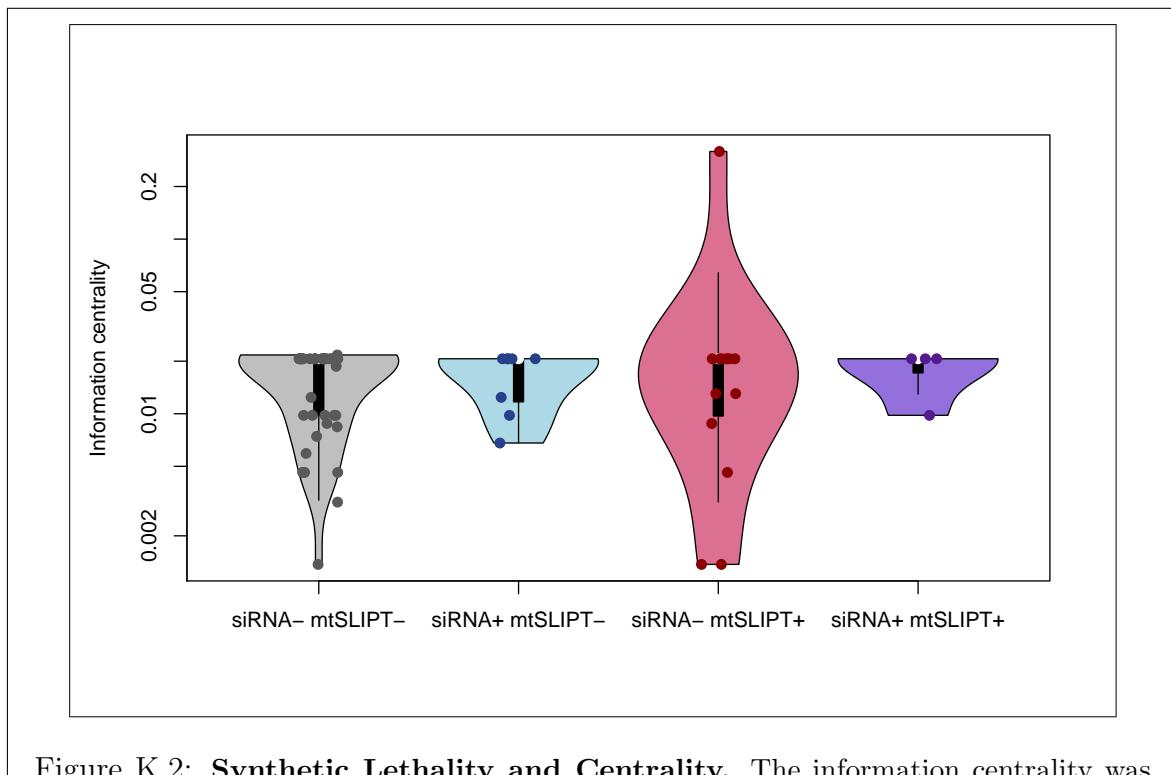


Figure K.2: **Synthetic Lethality and Centrality.** The information centrality was compared (on a log-scale across genes detected by mtSLIPT and siRNA screening in the Reactome PI3K cascade pathway. Genes detected by siRNA had higher connectivity than many genes not detected by either approach. The gene with the highest centrality was detected by mtSLIPT.

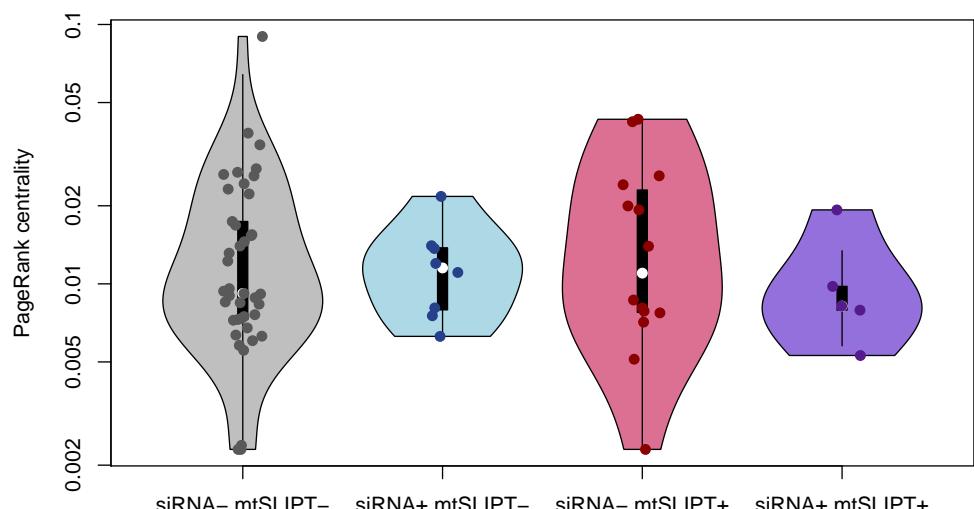


Figure K.3: **Synthetic Lethality and PageRank.** The PageRank centrality was compared (on a log-scale across genes detected by mtSLIPT and siRNA screening in the Reactome PI3K cascade pathway. Genes detected by siRNA had a more restricted range of centrality values than other genes not detected by either approach, although these groups also had fewer genes.

Table K.1: ANOVA for Synthetic Lethality and Vertex Degree

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

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

Table K.2: ANOVA for Synthetic Lethality and Information Centrality

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

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

Table K.3: ANOVA for Synthetic Lethality and PageRank Centrality

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

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

## Appendix L

### Information Centrality for Gene Essentiality

Network structure is another useful strategy to analyse gene function and this has been used to investigate network properties of a network constructed from of Reactome pathways imported via Pathway Commons with Paxtools (Cerami *et al.*, 2011; Demir *et al.*, 2013). Most notably, information centrality which has been proposed as a measure of gene essentiality was calculated as performed by Kranthi *et al.* (2013) using the efficiency and shortest path between each pair of nodes in the network before and after a node of interest is removed to test the importance of a node to network connectivity. Reactome contains substrates and cofactors in addition to genes or proteins. In support of centrality as a measure of essentiality, a number nodes with the highest centrality (shown in Table L.1) were essential nutrients including  $Mg^{2+}$ ,  $Ca^{2+}$ ,  $Zn^{2+}$ , and Fe. In addition, there were genes important in development of epithelial tissues and breast cancer such as *IL8*, *GATA3*, and *CTNNB1* detected with relatively high information centrality.

Table L.1: Information centrality for genes and molecules in the Reactome network

<b>Node</b>	<b>Centrality</b>
<i>ZNF473</i>	0.0510
magnesium(2+)	0.0082
<i>XBP1</i>	0.0053
calcium(2+)	0.0050
zinc(2+)	0.0048
iron atom	0.0041
<i>FMN</i>	0.0040
<i>AGT</i>	0.0037
<i>HSP90AA1</i>	0.0029
phosphatidyl-L-serine	0.0029
<i>P2RX7</i>	0.0026
<i>PANX1</i>	0.0024
<i>NCAM1</i>	0.0022
<i>NUDT1</i>	0.0021
<i>PLAUR</i>	0.0020
<i>IL8</i>	0.0020
<i>HSPA8</i>	0.0019
<i>TYROBP</i>	0.0019
<i>CASP3</i>	0.0017
<i>GNAL</i>	0.0015
<i>CBLB</i>	0.0015
<i>HBB</i>	0.0014
<i>GATA4</i>	0.0013
<i>TGS1</i>	0.0013
<i>CTNNB1</i>	0.0012

Highest information centrality for genes (proteins), cofactors, and minerals in the Reactome network

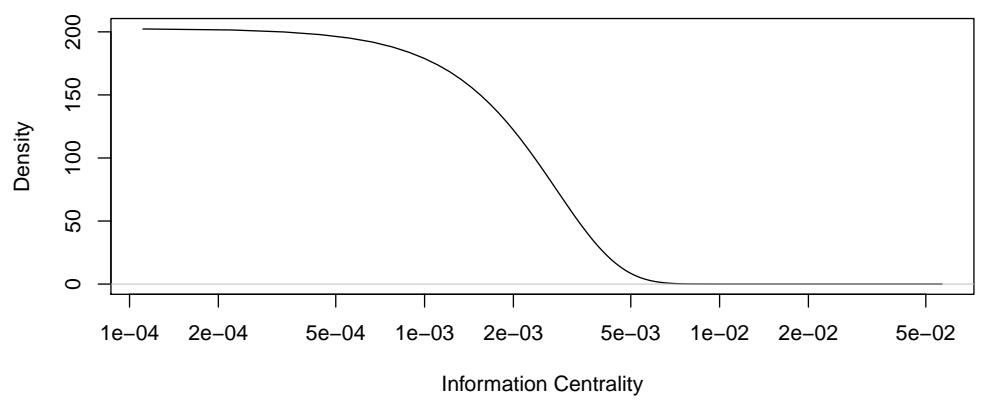


Figure L.1: **Information centrality distribution.** Information centrality in the Reactome network for nodes, including genes/proteins and other biomolecules.

# Appendix M

## Pathway Structure for Mutation SLIPT

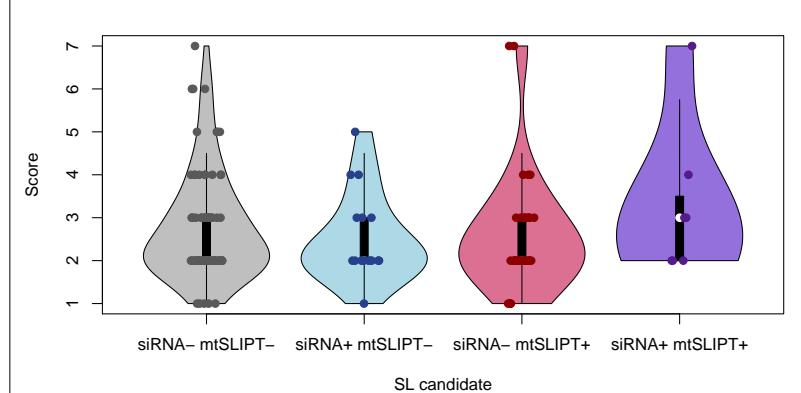
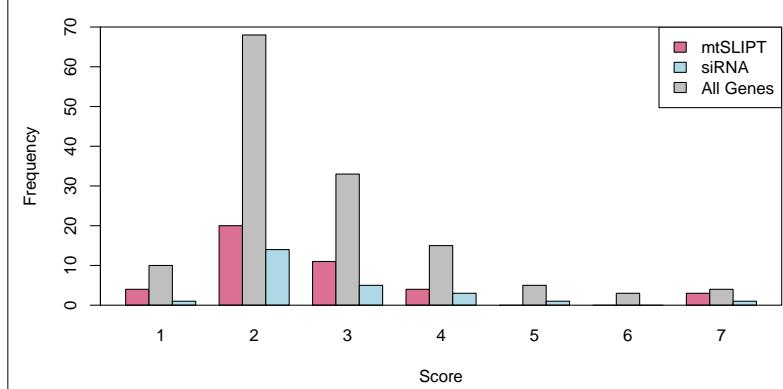


Figure M.1: **Synthetic Lethality and Heirarchy Score in PI3K.** The hierarchical distance scores were similarly distributed across mtSLIPT and siRNA genes. Genes detected by both methods had a higher (downstream) median than either group.

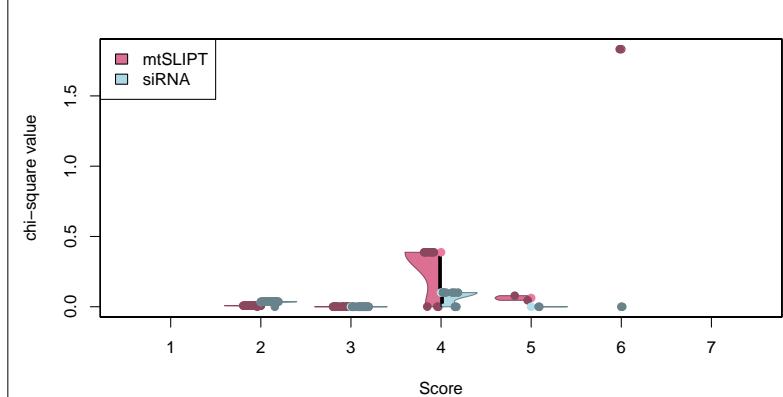
Table M.1: ANOVA for Synthetic Lethality and PI3K Hierarchy

	DF	Sum Squares	Mean Squares	F-value	p-value
siRNA	1	0.001	0.00070	0.0004	0.9841
mtSLIPT	1	0.007	0.0066	0.0040	0.9496
siRNA×mtSLIPT	1	3.906	3.9056	2.3829	0.1250

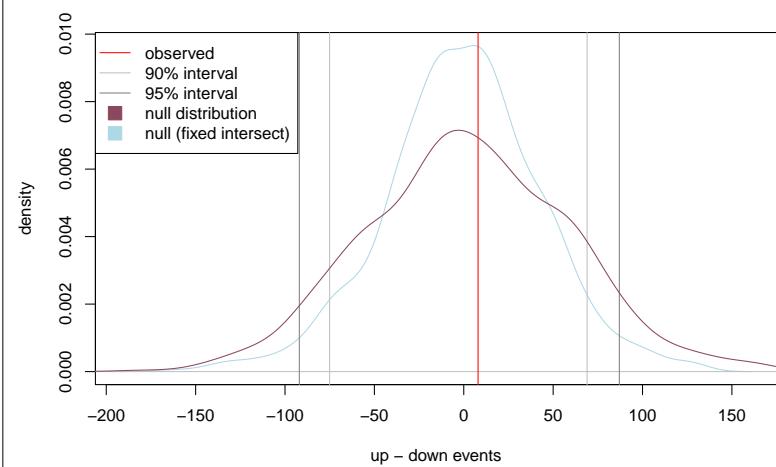
Analysis of variance for PI3K hierarchy score against synthetic lethal detection approaches (with an interaction term)



**Figure M.2: Heirarchy Score in PI3K against Synthetic Lethality in PI3K.** The number of mtSLIPT and siRNA genes against the hierarchical distance scores showing no significant tendency for either method to either of the pathway upstream or downstream extremities.



**Figure M.3: Structure of Synthetic Lethality in PI3K.** The number of mtSLIPT and siRNA genes upstream or downstream of each gene in the Reactome PI3K pathway were tested (by the  $\chi^2$ -test). These are plotted as a split violin plot against the hierarchical distance scores showing no significant tendency for either method to either of the pathway upstream or downstream extremities.



**Figure M.4: Structure of Synthetic Lethality Resampling.** A null distribution (10,000 iterations) of the siRNA genes upstream or downstream of mtSLIPT genes (shown by the difference) in the PI3K pathway. The observed events (red) were compared to the the distribution (violet) and were not significant. Genes detected by both methods were fixed for the distribution (blue). The genes detected by both approaches were used.

Table M.2: Resampling for pathway structure of synthetic lethal detection methods

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

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