

Contents

4 Synthetic Lethal Analysis of Gene Expression Data	6
4.1 Synthetic lethal genes in breast cancer	8
4.1.1 Synthetic lethal pathways in breast cancer	10
4.1.2 Expression profiles of synthetic lethal partners	10
4.1.2.1 Subgroup pathway analysis	11
4.2 Comparison of synthetic lethal gene candidates	13
4.2.1 Comparison with differential expression	13
4.2.2 Comparison with correlation	13
4.2.3 Comparison with primary siRNA screen candidates	13
4.2.3.1 Comparison of screen at pathway level	16
4.2.3.1.1 Resampling of genes for pathway enrichment . .	16
4.2.4 Comparison with secondary screen siRNA screen candidates . .	20
4.2.4.1 Comparison of candidate SL Pathways	20
4.3 Mutation, Copy Number, and Methylation	20
4.3.1 Synthetic lethality by DNA copy number	22
4.3.2 Synthetic lethality by somatic mutation	22
4.3.2.1 Mutation analysis	22
4.3.3 ANOVA of Expression Predictors	22
4.4 Global Synthetic Lethality	23
4.4.1 Hub Genes	23
4.5 Metagene Analysis	23
4.5.1 Pathway expression	23
4.5.2 Somatic mutation	23
4.5.3 Synthetic lethal metagenes	23
4.6 Replication in stomach cancer	23
4.7 Replication in cell line encyclopaedia	24
4.8 Summary	26
References	40
A Sample Correlation	42
B Software Used for Thesis	44
C Secondary Screen Data	53
D Replicate Samples in TCGA Breast	55
E Mutation Analysis in Breast Cancer	59
F Expression Analysis in Stomach Cancer	67

List of Figures

4.1	Synthetic lethal expression profiles of analysed samples	12
4.2	Comparison of SLIPT to siRNA	15
4.3	Resampled intersection of SLIPT and siRNA candidates	18
4.4	Synthetic lethal expression profiles of stomach samples	31
4.5	Comparison of SLIPT in stomach to siRNA	32
A.1	Correlation profiles of removed samples	42
A.2	Correlation analysis and sample removal	43
D.1	Replicate excluded samples	55
D.2	Replicate samples with all remaining	56
D.3	Replicate samples with some excluded	57
D.3	Replicate samples with some excluded	58
E.1	Synthetic lethal expression profiles of analysed samples	65
E.2	Comparison of mtSLIPT to siRNA	66
G.1	Synthetic lethal expression profiles of stomach samples	71
G.2	Comparison of mtSLIPT in stomach to siRNA	73

List of Tables

4.1	Candidate synthetic lethal genes against E-cadherin from SLIPT	9
4.2	Pathways for <i>CDH1</i> partners from SLIPT	10
4.3	Pathway composition for clusters of <i>CDH1</i> partners from SLIPT	14
4.4	Pathway composition for <i>CDH1</i> partners from SLIPT and siRNA screening	17
4.5	Pathways for <i>CDH1</i> partners from SLIPT	19
4.6	Pathways for <i>CDH1</i> partners from SLIPT and siRNA primary screen .	21
4.7	Candidate synthetic lethal metagenes against <i>CDH1</i> from SLIPT	24
4.8	Candidate synthetic lethal genes against E-cadherin from SLIPT in stomach cancer	25
4.9	Pathways for <i>CDH1</i> partners from SLIPT in stomach cancer	26
4.10	Pathway composition for clusters of <i>CDH1</i> partners in stomach SLIPT	27
4.11	Pathway composition for <i>CDH1</i> partners from SLIPT and siRNA screening	28
4.12	Pathways for <i>CDH1</i> partners from SLIPT in stomach cancer	29
4.13	Pathways for <i>CDH1</i> partners from SLIPT in stomach and siRNA screen .	30
4.14	Candidate synthetic lethal metagenes against <i>CDH1</i> from SLIPT in stomach cancer	33
4.15	Candidate synthetic lethal genes against E-cadherin from SLIPT in CCLE .	34
4.16	Pathways for <i>CDH1</i> partners from SLIPT in CCLE	35
4.17	Candidate synthetic lethal genes against E-cadherin from SLIPT in breast CCLE	36
4.18	Pathways for <i>CDH1</i> partners from SLIPT in breast CCLE	37
4.19	Candidate synthetic lethal genes against E-cadherin from SLIPT in stomach CCLE	38
4.20	Pathways for <i>CDH1</i> partners from SLIPT in stomach CCLE	39
B.1	R Packages used during Thesis	44
C.1	Comparing SLIPT genes against Secondary siRNA Screen in breast cancer	53
C.2	Comparing mtSLIPT genes against Secondary siRNA Screen in breast cancer	54
C.3	Comparing SLIPT genes against Secondary siRNA Screen in stomach cancer	54
E.1	Candidate synthetic lethal genes against E-cadherin from mtSLIPT . .	59
E.2	Pathways for <i>CDH1</i> partners from mtSLIPT	60
E.3	Pathway composition for clusters of <i>CDH1</i> partners from mtSLIPT . .	61
E.4	Pathway composition for <i>CDH1</i> partners from mtSLIPT and siRNA . .	62
E.5	Pathways for <i>CDH1</i> partners from mtSLIPT	63
E.6	Pathways for <i>CDH1</i> partners from mtSLIPT and siRNA primary screen	64

E.7 Candidate synthetic lethal metagenes against <i>CDH1</i> from mtSLIPT	66
G.1 Candidate synthetic lethal genes against E-cadherin from mtSLIPT in stomach cancer	69
G.2 Pathways for <i>CDH1</i> partners from mtSLIPT in stomach cancer	70
G.3 Pathway composition for clusters of <i>CDH1</i> partners in stomach mtSLIPT	72
G.4 Pathway composition for <i>CDH1</i> partners from mtSLIPT and siRNA	74
G.5 Pathways for <i>CDH1</i> partners from mtSLIPT in stomach cancer	75
G.6 Pathways for <i>CDH1</i> partners from mtSLIPT in stomach and siRNA screen	76
G.7 Candidate synthetic lethal metagenes against <i>CDH1</i> from mtSLIPT in stomach cancer	77

Chapter 4

Synthetic Lethal Analysis of Gene Expression Data

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

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

directly on metagenes to detect synthetic lethal pathways.

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

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

4.1 Synthetic lethal genes in breast cancer

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

While these are among the strongest synthetic lethal candidates, translational genes are crucial to the viability of healthy cells and dosing for a selective synthetic lethal effect against these may be difficult compared to other biological functions which may also be supported among the SLIPT candidate genes. Furthermore, few known bio-

logical functions of *CDH1* were among the strongest SL candidates so the remaining candidate genes may also be informative since they are likely to contain these expected functions in addition to novel relationships for *CDH1*. Thus further pathway level analyses were also conducted to examine biological functions over-represented among synthetic candidate genes and identify synthetic lethal pathways.

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

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

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

The modified mtSLIPT methodology (as described in section ??) was also applied to the normalised TCGA breast cancer gene expression dataset, against somatic loss of function mutations in *CDH1*. As shown in Table E.1, the most significant genes also had strong evidence of expression associated with *CDH1* mutations (high χ^2 values) with fewer samples exhibiting both low expression and mutations of each gene than

expected statistically. Although, these were not a strongly supported as the expression analysis (in Table 4.1) nor were as many genes detected. This is unsurprising due to the lower sample size with matching somatic mutation data and the lower frequency of *CDH1* mutations compared to low expression by $1/3$ quantiles.

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

4.1.1 Synthetic lethal pathways in breast cancer

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

Immune pathways, including the adaptive immune system and responses to infectious diseases were also strongly implicated as synthetic lethal with loss of E-cadherin. This is consistent with the alterations of immune response being a hallmark of cancer ?, since evading the immune system is necessary for cancer survival. Either of these systems are potential means to target *CDH1* deficient cells, although these were not detected in an isolated cell line experimental screen (Telford *et al.*, 2015) and the differences between to findings in patient data will be described in more detail in section 4.2.1.4.

It is also notable that the pathways over-represented in SLIPT candidate genes have strongly significant over-representation of Reactome pathways from the hypergeometric test (as described in section ??). Even after adjusting stringently for multiple tests, biologically related pathways give consensus support to these pathways. These pathways are further supported by testing for synthetic lethality against *CDH1* mutations (mtSLIPT) with many of these pathways also among the most strongly supported in this analysis (shown in Table E.2). This analysis more closely represents the null *CDH1* mutations in HDGC (?) and the experimental MCF10A cell model (Chen *et al.*, 2014).

Table 4.2: Pathways for *CDH1* partners from SLIPT

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

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

Although it still supports translational and immune pathways not detected in the isolated experimental system, G-protein-coupled receptors (GPCRs) were also among the most strongly supported pathways, supporting the experimental findings of Telford *et al.* (2015) for these intracellular signalling pathways already being targeted for other diseases.

4.1.2 Expression profiles of synthetic lethal partners

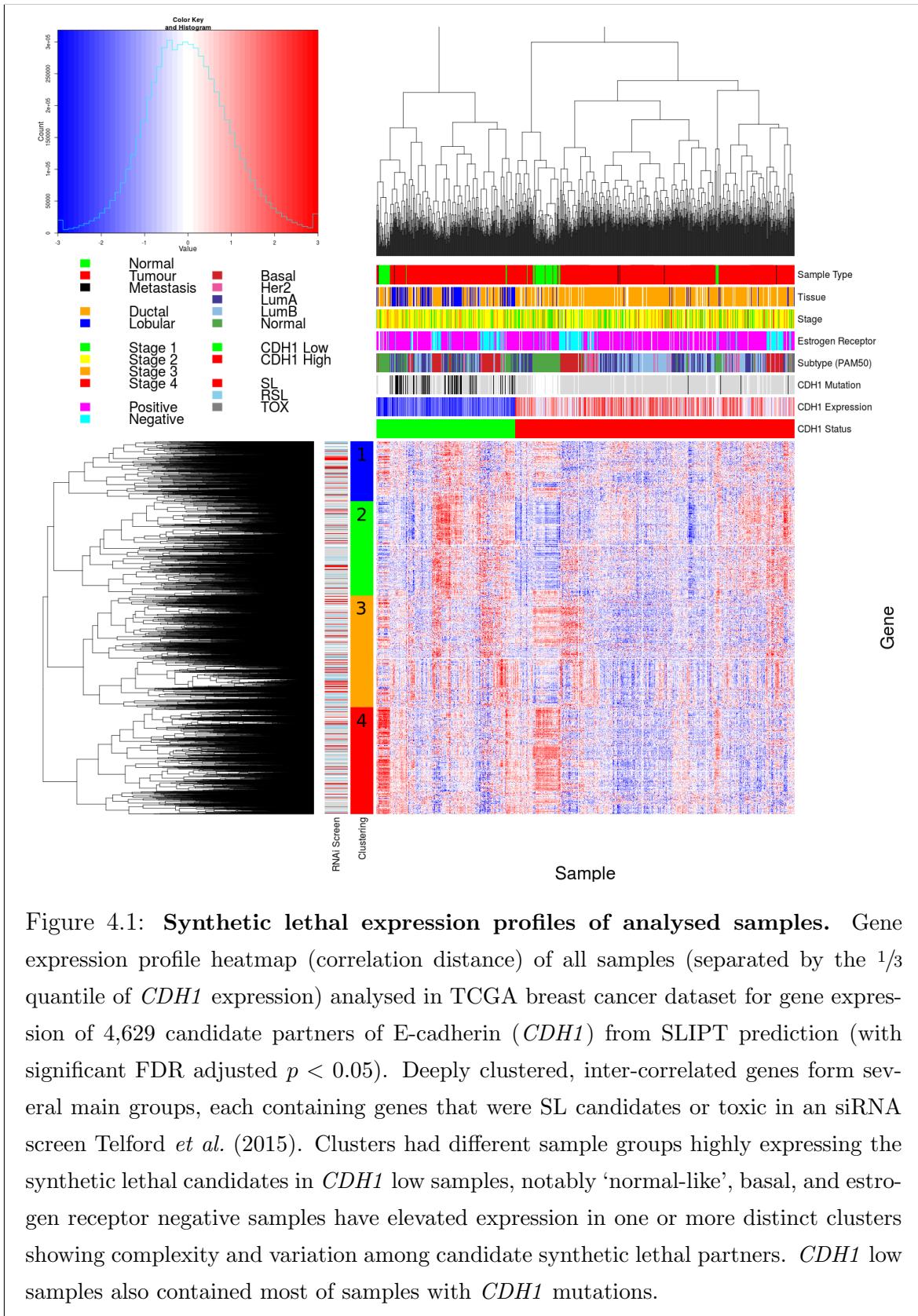
Due to the sheer number of gene candidates and to examine their expression patterns, investigations proceeded into correlation structure and pathway over-representation. This serves to explore the functional similarity of the synthetic lethal partners of *CDH1*, with the eventual aim to assess their utility as drug targets. As shown in Figure 4.1 (which clusters *CDH1* lowly expressing samples separately), there were several large clusters of genes among the expression profiles of the *CDH1* synthetic lethal candidate partners. The clustering suggests co-regulation of genes or pathway correlation between partner gene candidates. A number of candidates from an experimental RNAi screen study performed by Telford *et al.* (2015) were also identified by this approach. In addition, we identified novel gene candidates, which had little effect on viability in

isogenic cell line experiments.

In these expression profiles, a gene with a moderate or high signal across samples exhibiting low *CDH1* expression would represent a potential drug target. However, it appears that several molecular subtypes of cancer have elevation of different clusters of synthetic lethal candidates in samples with low *CDH1*. This clustering suggests that different targets or combinations could be effective in different patients suggesting potential utility for stratification. In particular, estrogen receptor negative, basal subtype, and “normal-like” samples Dai *et al.* (2015); Eroles *et al.* (2012); Parker *et al.* (2009) have elevation of genes specific to particular clusters which is indicative of some synthetic lethal interactions being specific to a particular molecular subtype or genetic background. Thus synthetic lethal drug therapy against these subtypes may be ineffective if it were designed against genes in another cluster.

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

The *CDH1* mutant samples (in Figure 4.1) were predominantly among the *CDH1* lowly expressing samples and distributed throughout *CDH1* samples with clustering analysis. Thus the molecular profiles of *CDH1* low samples are indistinguishable from *CDH1* mutant samples with the exception of normal samples (that do not have somatic mutation data as it is inferred from comparison to them to tumour-specific genotypes).



Conversely, many of the *CDH1* mutant samples (in Figure E.1) had among the lowest *CDH1* expression and some of the synthetic lethal partners were also highly expressed in lowly expressing *CDH1* wildtype samples, despite these not being considered as “inactivated” by mtSLIPT analysis.

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

4.1.2.1 Subgroup pathway analysis

Synthetic lethal gene candidates for *CDH1* from SLIPT performed on RNA-Seq expression data were also used for pathway over-representation analyses (as described in section ??). The correlation structure in the expression of candidates synthetic lethal genes in *CDH1* low tumours (lowest 1/3rd quantile of expression) was examined for distinct biological pathways in subgroups of genes elevated in different clusters of samples. These gene were highly expressed in different samples with their clinical factors including estrogen receptor status and intrinsic (PAM50) subtype (Parker *et al.*, 2009) shown in Figure 4.1.

As shown by the most over-represented pathways in Table 4.3, each correlated cluster of candidate synthetic lethal partners of *CDH1* contains functionally different genes.

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

Pathways Over-represented in Cluster 1	Pathway Size	Cluster Genes	p-value (FDR)
Collagen formation	67	10	4.0×10^{-11}
Extracellular matrix organisation	238	21	1.8×10^{-9}
Collagen biosynthesis and modifying enzymes	56	8	1.8×10^{-9}
Uptake and actions of bacterial toxins	22	5	9.5×10^{-9}
Elastic fibre formation	37	6	1.9×10^{-8}
Muscle contraction	62	7	2.4×10^{-7}
Fatty acid, triacylglycerol, and ketone body metabolism	117	10	4.9×10^{-7}
XBP1(S) activates chaperone genes	51	6	6.6×10^{-7}
IRE1alpha activates chaperones	54	6	1.2×10^{-6}
Neurotoxicity of clostridium toxins	10	3	1.3×10^{-6}
Retrograde neurotrophin signalling	10	3	1.3×10^{-6}
Assembly of collagen fibrils and other multimeric structures	40	5	1.9×10^{-6}
Collagen degradation	58	6	2.0×10^{-6}
Arachidonic acid metabolism	41	5	2.1×10^{-6}
Synthesis of PA	26	4	3.0×10^{-6}
Signaling by NOTCH	80	7	3.3×10^{-6}
Signalling to RAS	27	4	3.7×10^{-6}
Integrin cell surface interactions	82	7	4.2×10^{-6}
Smooth Muscle Contraction	28	4	4.4×10^{-6}
ECM proteoglycans	66	6	6.3×10^{-6}
Pathways Over-represented in Cluster 2	Pathway Size	Cluster Genes	p-value (FDR)
Eukaryotic Translation Elongation	86	75	1.1×10^{-181}
Viral mRNA Translation	81	72	9.8×10^{-179}
Peptide chain elongation	83	72	1.9×10^{-175}
Eukaryotic Translation Termination	83	72	1.9×10^{-175}
Formation of a pool of free 40S subunits	93	75	1.9×10^{-171}
Nonsense Mediated Decay independent of the Exon Junction Complex	88	72	9.9×10^{-168}
L13a-mediated translational silencing of Ceruloplasmin expression	103	75	3.0×10^{-159}
3' -UTR-mediated translational regulation	103	75	3.0×10^{-159}
Nonsense-Mediated Decay	103	75	3.0×10^{-159}
Nonsense-Mediated Decay enhanced by the Exon Junction Complex	103	75	3.0×10^{-159}
SRP-dependent cotranslational protein targeting to membrane	104	75	3.2×10^{-158}
GTP hydrolysis and joining of the 60S ribosomal subunit	104	75	3.2×10^{-158}
Eukaryotic Translation Initiation	111	75	4.5×10^{-151}
Cap-dependent Translation Initiation	111	75	4.5×10^{-151}
Influenza Infection	117	75	1.4×10^{-145}
Influenza Viral RNA Transcription and Replication	108	72	5.7×10^{-145}
Translation	141	81	8.0×10^{-143}
Influenza Life Cycle	112	72	2.3×10^{-141}
Infectious disease	347	103	2.2×10^{-95}
Formation of the ternary complex, and subsequently, the 43S complex	47	33	6.8×10^{-80}
Pathways Over-represented in Cluster 3	Pathway Size	Cluster Genes	p-value (FDR)
Adaptive Immune System	412	90	6.1×10^{-61}
Chemokine receptors bind chemokines	52	27	6.7×10^{-56}
Generation of second messenger molecules	29	21	6.5×10^{-55}
Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	64	29	6.5×10^{-55}
TCR signalling	62	27	8.9×10^{-51}
Peptide ligand-binding receptors	161	40	1.5×10^{-45}
Translocation of ZAP-70 to Immunological synapse	16	14	3.1×10^{-43}
Costimulation by the CD28 family	51	22	4.0×10^{-43}
PD-1 signalling	21	15	4.0×10^{-41}
Class A/I (Rhodopsin-like receptors)	258	50	6.7×10^{-41}
Phosphorylation of CD3 and TCR zeta chains	18	14	1.3×10^{-40}
Interferon gamma signalling	74	24	5.0×10^{-39}
GPCR ligand binding	326	57	1.8×10^{-38}
Cytokine Signaling in Immune system	268	48	8.9×10^{-37}
Downstream TCR signalling	45	18	1.8×10^{-35}
G _{αi} signalling events	167	33	2.2×10^{-33}
Cell surface interactions at the vascular wall	99	21	1.3×10^{-26}
Interferon Signalling	164	28	1.7×10^{-26}
Extracellular matrix organisation	238	35	2.7×10^{-25}
Antigen activates B Cell Receptor leading to generation of second messengers	32	12	7.2×10^{-25}
Pathways Over-represented in Cluster 4	Pathway Size	Cluster Genes	p-value (FDR)
Extracellular matrix organisation	238	48	8.0×10^{-41}
Class A/I (Rhodopsin-like receptors)	258	47	2.8×10^{-36}
GPCR ligand binding	326	54	2.1×10^{-34}
G _{αi} signalling events	83	22	1.4×10^{-31}
GPCR downstream signalling	472	68	1.1×10^{-29}
Haemostasis	423	61	3.3×10^{-29}
Platelet activation, signalling and aggregation	180	31	7.1×10^{-28}
Binding and Uptake of Ligands by Scavenger Receptors	40	14	9.9×10^{-27}
RA biosynthesis pathway	22	11	2.5×10^{-26}
Response to elevated platelet cytosolic Ca ²⁺	82	19	3.0×10^{-26}
Developmental Biology	420	57	3.5×10^{-26}
G _{αi} signalling events	167	28	7.3×10^{-26}
Platelet degranulation	77	18	1.6×10^{-25}
Gastrin-CREB signalling pathway via PKC and MAPK	171	28	2.5×10^{-25}
Muscle contraction	62	16	4.7×10^{-25}
G _{αq} signalling events	150	25	3.2×10^{-24}
Retinoid metabolism and transport	34	12	5.0×10^{-24}
Phase 1 - Functionalisation of compounds	67	16	6.5×10^{-24}
Signalling by Retinoic Acid	42	13	6.7×10^{-24}
Degradation of the extracellular matrix	102	19	1.4×10^{-22}

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

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

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

4.2 Comparison of synthetic lethal gene candidates

4.2.1 Comparison with siRNA screen candidates

Gene candidates were compared between computational (SLIPT in TCGA breast cancer data) and experimental (the primary siRNA screen performed by Telford *et al.* (2015)) approaches in Figure 4.2. The number of genes detected by both methods did not produce a significant overlap but these may be difficult to compare due to vast differences between the detection methods. There were similar issues comparison of mtSLIPT genes tested against *CDH1* mutations (in Figure G.2), despite excluded genes not tested by both methods in either test. However, these intersecting genes may still be functionally informative or amenable to drug triage as they were replicated across both methods and pathway over-representation differed between the sections of the Venn diagram (see Figure 4.2).

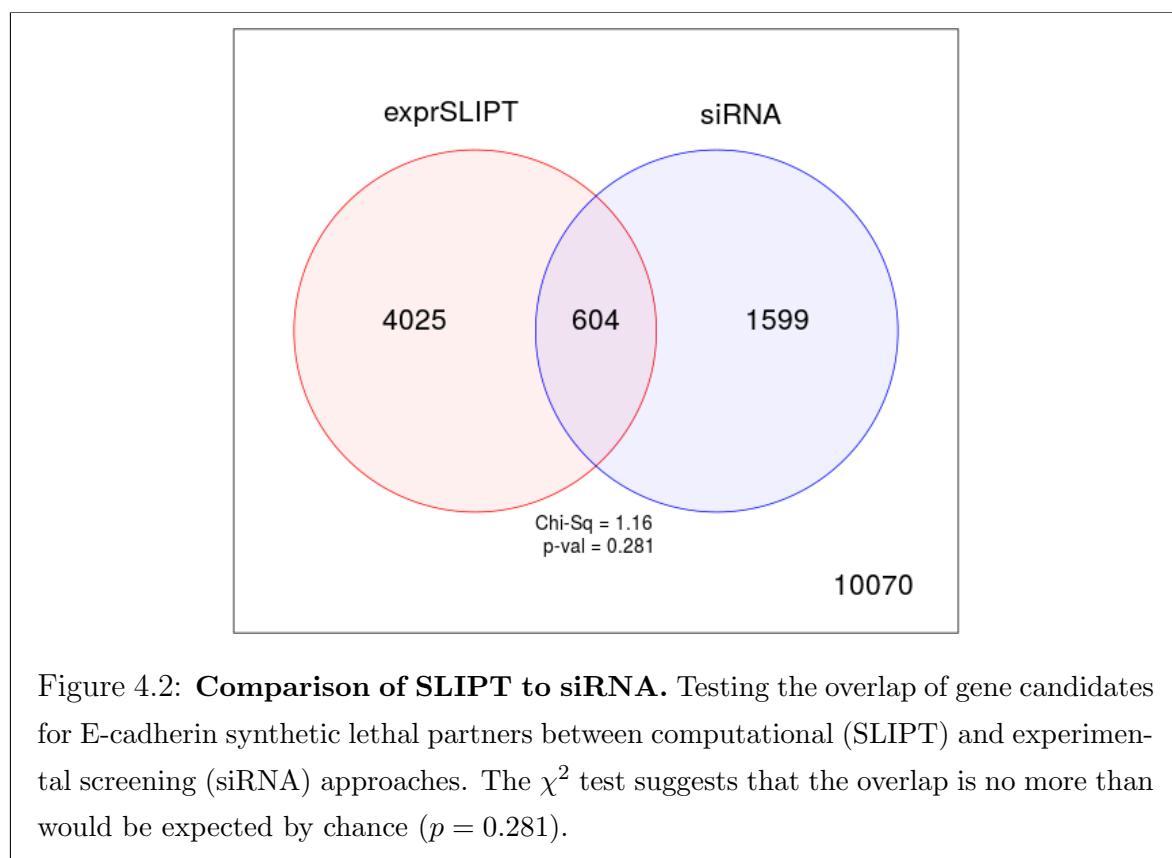


Figure 4.2: **Comparison of SLIPT to siRNA.** Testing the overlap of gene candidates for E-cadherin synthetic lethal partners between computational (SLIPT) and experimental screening (siRNA) approaches. The χ^2 test suggests that the overlap is no more than would be expected by chance ($p = 0.281$).

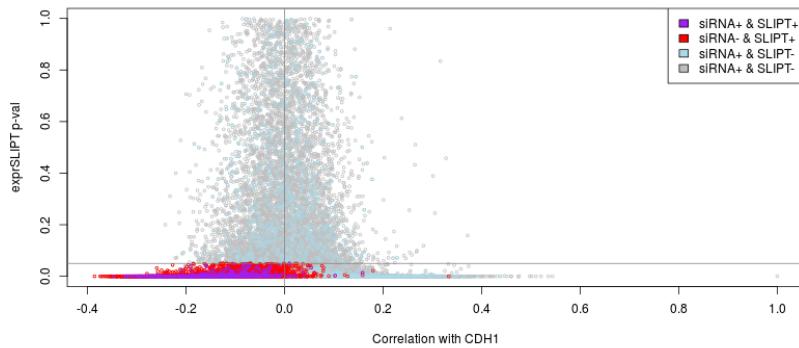


Figure 4.3: Compare SLIPT and siRNA genes with correlation. The χ^2 p-values for genes tested by SLIPT (in TCGA breast cancer) expression analysis were compared against Pearson's correlation of gene expression with *CDH1*. Genes detected by SLIPT or siRNA are coloured according to the legend.

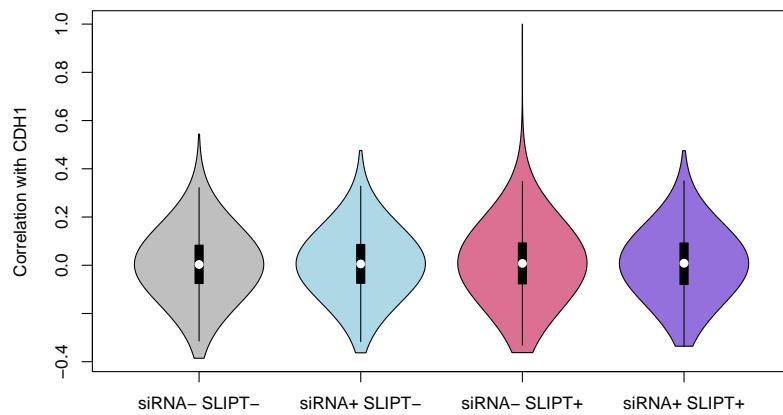


Figure 4.4: Compare SLIPT and siRNA genes with correlation. Genes detected as candidate synthetic lethal partners by SLIPT (in TCGA breast cancer) expression analysis and experimental screening (with siRNA) were compared against Pearson's correlation of gene expression with *CDH1*. There were no differences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

4.2.1.1 Comparison with correlation

Another potential means to triage drug target candidates is correlation of expression profiles with *CDH1*. Correlation with *CDH1* was compared to SLIPT and siRNA results in Figure 4.3. The genes not detected by SLIPT (including siRNA candidates) included genes with high (insignificant) SLIPT p-values. As expected, these genes were distributed around a correlation of zero and genes with higher correlation with *CDH1* (either direction) had more significant SLIPT p-values, although there were exceptions to this trend and larger positive correlations were negative correations.

The majority of SLIPT candidates appeared to have negative correlations and moreso for those genes detected by both approaches, although these were typically weak correlations and are unlikely to be sufficient to detect such genes on their own. This is supported by simulation results in section ??.

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

However, the apparent tendancy for genes detected by SLIPT or siRNA to have negative correlations with *CDH1* expression may be due to the smaller number of genes in these groups. The distribution of *CDH1* correations does not differ across these gene groups (as shown by Figures 4.4 and ??). Therefore further triage of gene candidates by correaltion is not suitable, nor is use of correlation itself to predict synthetic lethal partners in the first place.

4.2.1.2 Comparison with viability

A similar comparison of SLIPT results was made with the viability ratio (of *CDH1* mutant to wildtype) in the primary siRNA screen performed by Telford *et al.* (2015). The significance and viability thresholds used for SLIPT and siRNA detection of synthetic lethal candidate partners of *CDH1* are clear in Figure 4.5. However note that not all of the gene below these thresholds are neccessarily selected to be candidate partners as

additional criteria were used in each case: directional criteria as for SLIPT (see section ??) and minimum wildtype viability for siRNA (Telford *et al.*, 2015).

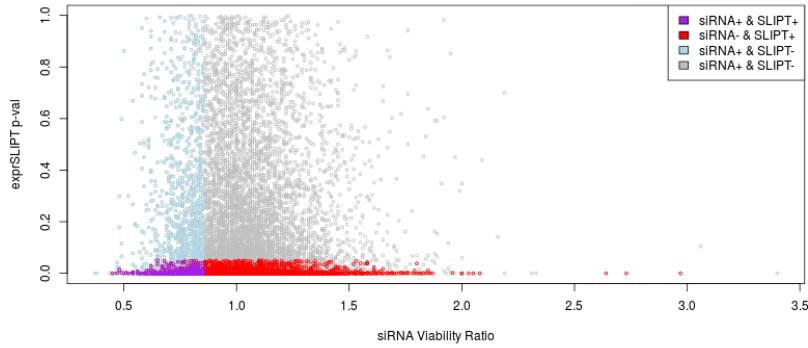


Figure 4.5: Compare SLIPT and siRNA genes with siRNA viability. The χ^2 p-values for genes tested by SLIPT (in TCGA breast cancer) expression analysis were compared against the viability ratio of *CDH1* mutant and wildtype cells in the primary siRNA screen. Genes detected by SLIPT or siRNA are coloured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

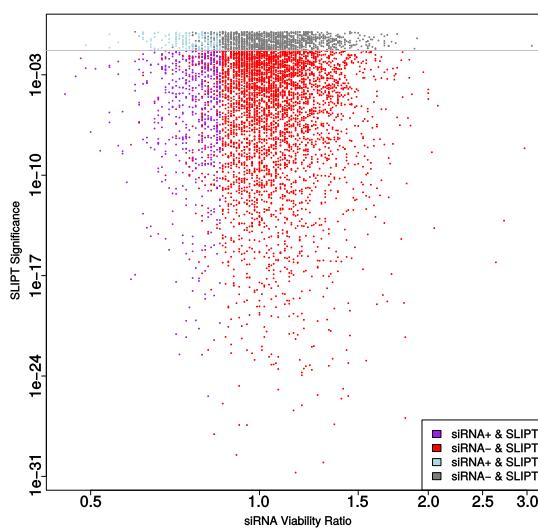


Figure 4.6: Compare SLIPT and siRNA genes with viability. The χ^2 p-values for genes tested by SLIPT (in TCGA breast cancer) expression analysis were compared (on a log-scale) against the viability ratio of *CDH1* mutant and wildtype cells in the primary siRNA screen. Genes detected by SLIPT or siRNA are coloured according to the legend with a grey line for $p = 0.05$.

There does not appear to be a clear relationship between SLIPT and siRNA candi-

dates. Many genes not detected by both approaches were numerous in Figures 4.2 and E.2. These genes detected by either are not necessarily near the thresholds for the other. In this respect the SLIPT approach with patient data and cell line experiments are independent means to identify synthetic lethal candidates. While genes detected by both approaches were not necessarily more strongly supported by either, the genes with a viability closer to 1 (no synthetic lethal effect) in siRNA included those with more significant SLIPT p-values whereas more extreme viability ratios tended to be less significant (as shown by a logarithmic plot in Figure 4.6). Although it should be noted that genes with more moderate viability ratios were more common and SLIPT was capable (despite adjusting for multiple testing) of detecting significant genes with extreme viability ratios, particularly those considerably lower than 1.

However, there was not support for SLIPT candidates or those detected by both approaches having considerably different viability ratios (as shown in Figures 4.7 and ??). The difference between the gene groups stems largely from the viability thresholds used by Telford *et al.* (2015) to detect synthetic lethal candidates in the primary screen, rather than more extreme viability ratios for genes identified by SLIPT.

4.2.1.3 Comparison with secondary screen siRNA screen candidates

However, it should be noted that genes with a lower viability ratio were not necessarily the strongest supported by experimental screening. The primary screen (with 4 pooled siRNAs) has been used for the majority of comparisons in this thesis because the genome-wide panel of target genes screened enables a large number of genes to be compared with SLIPT results from gene expression and somatic mutation analysis. A secondary screen was also performed by Telford *et al.* (2015) on the isogenic MCF10A breast cell lines to individually validate the siRNAs separately, with the strongest candidates being those exhibiting synthetic lethal viability ratios replicated across independently targeting siRNAs. This was performed for the top 500 candidates (with the lowest viability ratio) from the primary screen and the 482 of these genes also tested by SLIPT in breast cancer (and the 486 genes tested by SLIPT in stomach cancer).

The seconday screen results are given in Appendix D which show that SLIPT candidate genes are more significantly ($p = 7.49 \times 10^{-3}$ by Fisher's exact test) more likely to be validated in the secondary screen and are thus informative of more robust partner genes, in addition to providing support that these interactions are consistent with expression profiles from heterogeneous patient samples across genetic backgrounds. While

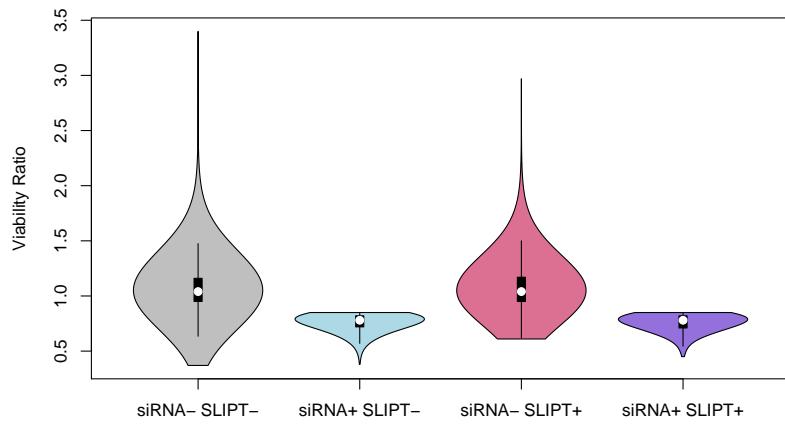


Figure 4.7: Compare SLIPT and siRNA genes with siRNA viability. Genes detected as candidate synthetic lethal partners by SLIPT (in TCGA breast cancer) expression analysis and experimental screening (with siRNA) were compared against the viability ratio of *CDH1* mutant and wildtype cells in the primary siRNA screen. There were clear no differences in viability between genes detected by SLIPT and those not with the differences being primarily due to viability thresholds being used to detect synthetic lethality by Telford *et al.* (2015).

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

4.2.1.4 Comparison of screen at pathway level

These pathway over-representation analyses (performed as described in section ??) correspond to genes separated into SLIPT or siRNA screen candidates unique to either method or detected by both (Table 4.4). The SLIPT-specific gene candidates were involved most strongly with translational and immune regulatory pathways, although extracellular matrix pathways were also supported. These pathways were largely consistent with those identified in Tabel 4.2 and in the clustering analysis (Table 4.3). The genes detected only by the siRNA screen had over-representation of cell signalling pathways, including many containing genes known to be involved in cancer (e.g., MAPK, PDGF, ERBB2, and FGFR), with the detection of Class A GPCRs supporting the independent analyses by Telford *et al.* Telford *et al.* (2015). The intersection of computational and experimental synthetic lethal partners of *CDH1* has stronger evidence for over-representation of GPCR pathways and more specific subclasses, such as visual phototransduction ($p = 6.9 \times 10^{-10}$) and $G_{\alpha s}$ signalling events ($p = 1.7 \times 10^{-7}$), than other signalling pathways.

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

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

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

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

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

identified by SLIPT may still be informative of the role of *CDH1* loss of function in cancers or mechanisms by which further gene loss leads to specific inviability.

4.2.1.4.1 Resampling of genes for pathway enrichment

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

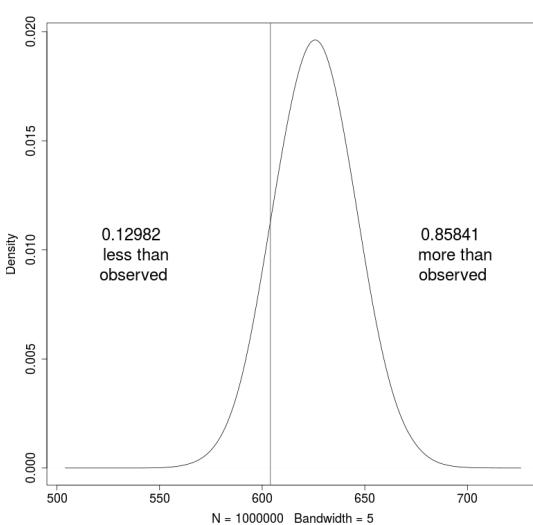


Figure 4.8: **Resampled intersection of SLIPT and siRNA candidates.** Resampling analysis of intersect size from genes detected by SLIPT and siRNA screening approaches over 1 million replicates. The proportion of expected intersection sizes for random samples below or above the observed intersection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

The overlap between synthetic lethal from bioinformatics SLIPT predictions and siRNA screening has raised other questions including whether the pathways overrepresented would be expected by chance. This of particular concern since the siRNA

candidate genes themselves are highly over-represented for particular pathways (such as GPCRs) so selecting any intersect with them would be enriched for these pathways. Another pathway approach is to test whether pathways are over-represented in randomly sampled genes, comparing many “resamplings” or “permutations” of these genes to the enrichment statistics observed for these pathways in the SLIPT candidates and their intersection with the siRNA hits shows whether we detect these pathways more than we expect by chance (as described in section ??).

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

A resampling approach is also applicable to testing whether the number of genes detected by each approach significantly intersected. As shown in Figure 4.8, resampling did not find evidence of significant depletion or over-representation for experimental synthetic lethal candidates in the computationally predicted synthetic lethal partners of *CDH1* and the overlap may be observed by chance.

A permutation analysis was performed to resample the genes tested by both approaches to investigate whether the observed pathway over-representation could have occurred in a randomly selected sample of genes from the experimental candidates, that is, whether the pathway predictions from SLIPT could be expected by chance. While the number of siRNA candidate genes detected by SLIPT was not statistically significant ($p = 0.281$), this may be due to the vastly different limitations of the approaches and the correlation structure of gene expression not being independent (as assumed for multiple testing procedures). The intersection may still be functionally relevant to *CDH1*-deficient cancers, such as the pathway data in Table 4.4. The resampling analysis for pathways was compared to the pathway over-representation for SLIPT predicted synthetic lethal partners in Table 4.5. Similarly, the pathway resampling for intersection between SLIPT predictions and experimental screen candidates was compared to pathway over-representation in Table 4.6 for intersection with siRNA data.

The pathway resampling approach for SLIPT-specific gene candidates (Table 4.5) replicates the gene set over-representation analysis for all SLIPT genes, detecting evidence of synthetic lethal pathways for *CDH1* in translational, immune, and cell sig-

nalling pathways including $G_{\alpha i}$ signalling, GPCR downstream signalling, and chemokine receptor binding. While the immune and signal transduction pathways were not significantly over-represented in the resampling analysis, the results for the two approaches were largely consistent for translation and post-transcriptional gene regulation, supporting gene set over-representation of the SLIPT-specific pathways in Table 4.5. In particular, some of the most significantly over-represented pathways had higher observed χ^2 values than any of the 1 million random permutations.

Table 4.5: Pathways for *CDH1* partners from SLIPT

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

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

The intersection between computational and experimental candidates (in Table 4.6) differed between over-representation and resampling analyses. Namely, many of the over-represented pathways were not significant in the resampling analysis, including visual phototransduction and retinoic acid signalling, although pathways involving defective *EXT1* or *EXT2* genes approach significance after FDR adjustment for multiple tests. Of the highest over-represented pathways in the intersection, only $G_{\alpha s}$ signalling events were supported by both over-representation and resampling analyses. Other pathways supported by both analyses were cytoplasmic elastic fibre formation,

associated protein modification pathways, energy metabolism, and the fibrin clotting cascade.

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

However, we note that several pathways, including some immune functions and neurotransmitters, were supported by the resampling analysis (in Table 4.6) when the initial pathway over-representation test was not significant. These functions appear to have been detected by both approaches more than expected by chance but must be interpreted with caution since they were still not common enough to be detected in pathway over-representation analysis.

4.2.1.5 Comparison of candidate SL Pathways

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

4.3 Metagene Analysis

[include?]

4.3.1 Pathway expression

4.3.2 Somatic mutation

4.3.3 Synthetic lethal metagenes

4.4 Mutation analysis

Data in Appendix ??

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

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

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

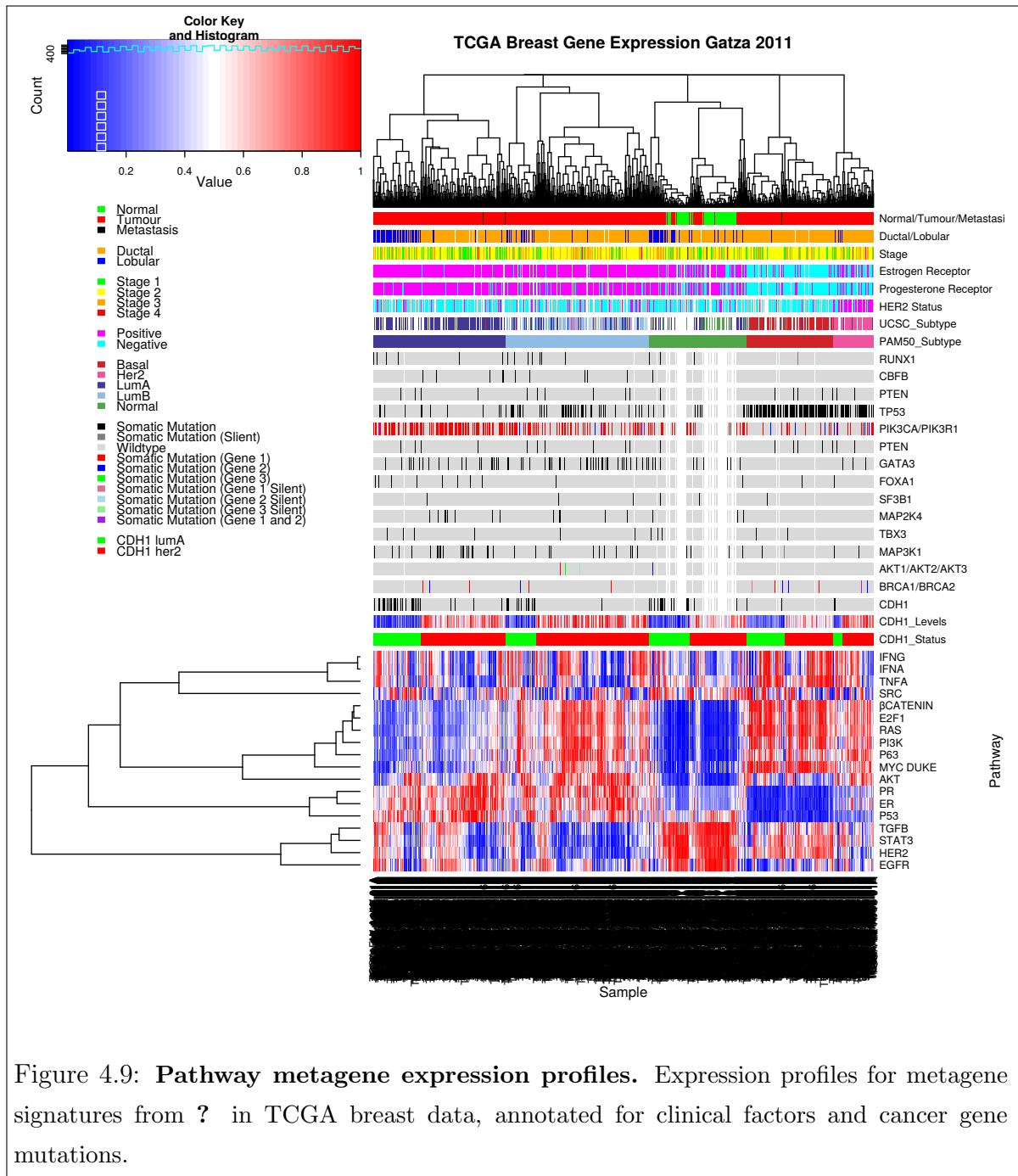


Table 4.7: Candidate synthetic lethal metagenes against *CDH1* from SLIPT

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

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

4.5 Global Synthetic Lethality

[include?]

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

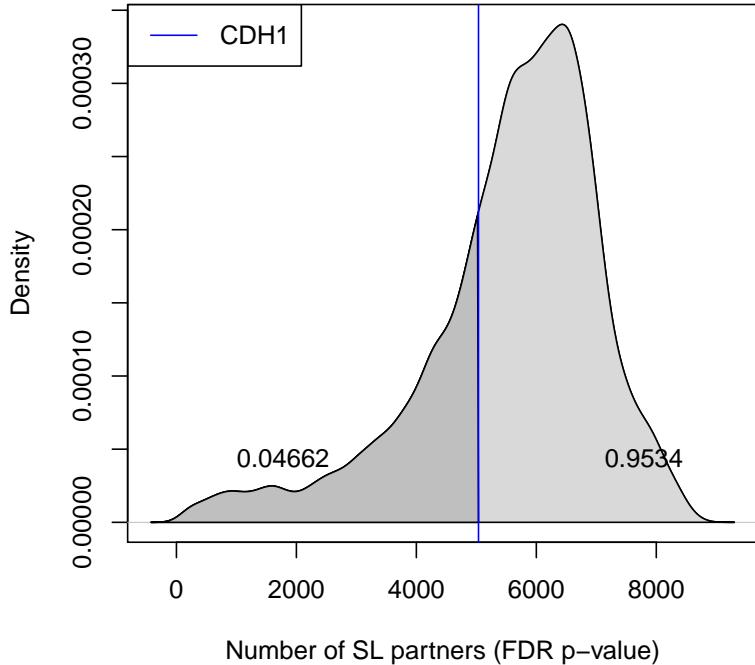


Figure 4.10: **Synthetic lethal partners across query genes.** Global synthetic lethal pairs were examined across the genome in TCGA breast expression data by applying SLIPT across query genes. The high number of predicted partners for *CDH1* was typical for a human gene and lower than many other genes.

4.5.1 Hub Genes

4.5.2 Hub Pathways

4.6 Replication in stomach cancer

4.6.1 Synthetic Lethal Genes and Pathways

4.6.2 Synthetic Lethal Expression Profiles

4.6.3 Comparison to Primary Screen

4.6.3.1 Resampling Analysis

4.6.4 Metagene Analysis

4.7 ³¹ Replication in cell line encyclopaedia

As breast cancer cell lines are the experimental system in which many cancer genetics and drug targets are investigated, these were analysed in addition to patient samples

Table 4.8: Query synthetic lethal genes with the most SLIPT partners

Gene	Direction	raw p-value	p-value (FDR)	SLIPT raw p-value	SLIPT (FDR)
<i>TGFBR2</i>	8134	17982	17973	8007	8006
<i>A2M</i>	8571	17605	17583	8345	8339
<i>TNS1</i>	8019	17949	17934	7874	7873
<i>PROS1</i>	8539	17668	17642	8317	8310
<i>ANXA1</i>	9085	17330	17302	8689	8682
<i>CELF2</i>	8665	17406	17368	8370	8355
<i>BOC</i>	8694	17371	17348	8384	8381
<i>PLAGL1</i>	8792	17361	17327	8448	8436
<i>PDGFRA</i>	8296	17650	17621	8095	8087
<i>FAM171A1</i>	8874	17560	17533	8567	8562
<i>FAM126A</i>	8510	17383	17356	8184	8178
<i>TSHZ2</i>	7942	17983	17976	7787	7786
<i>KCTD12</i>	8366	17651	17621	8115	8108
<i>MAML2</i>	8336	17537	17503	8069	8061
<i>FOXO1</i>	8027	17753	17737	7840	7836
<i>AMOTL1</i>	8425	17388	17347	8147	8139
<i>FAT4</i>	8111	17750	17732	7925	7919
<i>CAV1</i>	8645	17491	17464	8342	8331
<i>SVEP1</i>	7945	17859	17842	7791	7784
<i>EPB41L2</i>	8415	17327	17296	8097	8092

Genes with the most candidate SL partners SLIPT in TCGA breast expression data with the number of partner genes predicted by direction criteria and χ^2 testing separately and combined as a SLIPT analysis.

Where specified, the p-values for the χ^2 test were adjusted for multiple tests (FDR).

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

4.8 Summary

We have developed a simple, interpretable, computational approach to predict synthetic lethal partners from genomics data. Originally developed for microarray gene expression data, it has been expanded to test DNA copy number, or RNA-Seq gene expression data which are both also supported by the TCGA dataset. DNA copy num-

Table 4.9: Pathways for genes with the most SLIPT partners

Pathways Over-represented	Pathway Size	SL Genes	p-value	p-value (FDR)
Constitutive Signaling by Aberrant PI3K in Cancer	56	10	8.4×10^{-16}	8.7×10^{-13}
PI3K/AKT Signaling in Cancer	78	11	2.1×10^{-14}	1.1×10^{-11}
Role of LAT2/NTAL/LAB on calcium mobilization	96	12	7.7×10^{-14}	2.2×10^{-11}
Complement cascade	33	7	1.2×10^{-13}	2.2×10^{-11}
Cell surface interactions at the vascular wall	99	12	1.6×10^{-13}	2.2×10^{-11}
PI3K events in ERBB4 signaling	87	11	2.6×10^{-13}	2.2×10^{-11}
PIP3 activates AKT signaling	87	11	2.6×10^{-13}	2.2×10^{-11}
PI3K events in ERBB2 signaling	87	11	2.6×10^{-13}	2.2×10^{-11}
PI-3K cascade:FGFR1	87	11	2.6×10^{-13}	2.2×10^{-11}
PI-3K cascade:FGFR2	87	11	2.6×10^{-13}	2.2×10^{-11}
PI-3K cascade:FGFR3	87	11	2.6×10^{-13}	2.2×10^{-11}
PI-3K cascade:FGFR4	87	11	2.6×10^{-13}	2.2×10^{-11}
Extracellular matrix organization	238	22	4.7×10^{-13}	3.6×10^{-11}
Muscle contraction	62	9	4.9×10^{-13}	3.6×10^{-11}
PI3K/AKT activation	90	11	5.5×10^{-13}	3.8×10^{-11}
GAB1 signalosome	91	11	7.1×10^{-13}	4.6×10^{-11}
Smooth Muscle Contraction	28	6	2.4×10^{-12}	1.5×10^{-10}
Response to elevated platelet cytosolic Ca^{2+}	82	10	2.6×10^{-12}	1.5×10^{-10}
Signaling by SCF-KIT	126	13	3.0×10^{-12}	1.6×10^{-10}
Signaling by FGFR	143	14	5.0×10^{-12}	2.2×10^{-10}

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 breast expression data

ber was included for comparison with the DAISY tool of Jerby-Arnon *et al.* (2014). Predictions based on microarray data were inconclusive when compared with an RNAi screen for *CDH1* in MCF10A breast cells as performed by Telford *et al.* (2015), few predictions replicated between BC2116, CCLE, or TCGA microarray datasets, results with gene expression and DNA copy number were vastly different, and predictions from TCGA microarray and RNA-Seq datasets for the same samples differed were inconsistent. The Aligent TCGA microarray data in particular is difficult to compare to other datasets and will in the future use Affymetrix microarrays or RNA-Seq platforms for predictions from gene expression data. The analyses focus on gene expression data as it is widely available for applications in other cancers and current attempts to use gene expression data for synthetic lethal discovery vary widely (Jerby-Arnon *et al.*, 2014; Lu *et al.*, 2015; Tiong *et al.*, 2014). There is no consensus for which approach is more appropriate since they lack much a basis on biological experimental data or statistical modelling and often use difficult to interpret machine learning methodology.

Genomics analyses are prone to false-positives and require statistical caution, particularly where working with gene-pairs scale up the number of multiple tests drastically,

Table 4.10: Candidate synthetic lethal genes against E-cadherin from SLIPT in stomach cancer

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

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

at the expense of statistical power. Experimental SGA and RNAi screens for synthetic lethality are also error-prone, especially with false-positives, raising the need for understanding the expected behaviour and number of functional relationships and genetic interactions in the genome, or in discovery of synthetic lethal partners of a particular query gene. A characteristic of gene interaction networks is a scale-free topology leading to highly interacting hub genes, these represent important genes in a functional network. As shown in Tables 1-3, Gene Ontology terms for genes important in cancer proliferation, progression, and drug response were enriched in hub genes, showing that synthetic lethal interactions are among important genes in cancer cells. Gene functions

Table 4.11: Pathways for *CDH1* partners from SLIPT in stomach cancer

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

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

replicated across the breast cancer datasets are highlighted in bold, despite differences in particular hits, gene expression platforms, and only correcting for multiple tests for each gene query separately, there are many gene functions replicated across breast cancer gene expression analyses. TCGA microarray data was less consistent with the other datasets, as expected from lower sample size, lower concordance of particular hits for the example query of *CDH1*, and suspected lower quality of data on the Aligent microarray platform.

As specific genes were difficult to replicate across experiments, gene expression profiles for synthetic lethal partners must be more complex than originally expected to directly compensate for loss of query gene or completely lack (or clearly under-represent) co-loss (Jerby-Arnon *et al.*, 2014; Kelly, 2013; Lu *et al.*, 2015). The predicted synthetic lethal partners of *CDH1* (with FDR correction) were investigated with gene expression profiles and clinical variables to find relationships in gene expression, gene function, and clinical characteristics. The large number of hits indicate that synthetic lethality is error-prone and identifying genes or pathways relevant for clinical application will be difficult.

The expression profiles of the SL partners of *CDH1* predicted from the TCGA

Table 4.12: Pathway composition for clusters of *CDH1* partners in stomach SLIPT

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

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

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

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

Table 4.13: Pathway composition for *CDH1* partners from SLIPT and siRNA screening

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

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

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

Table 4.14: Pathways for *CDH1* partners from SLIPT in stomach cancer

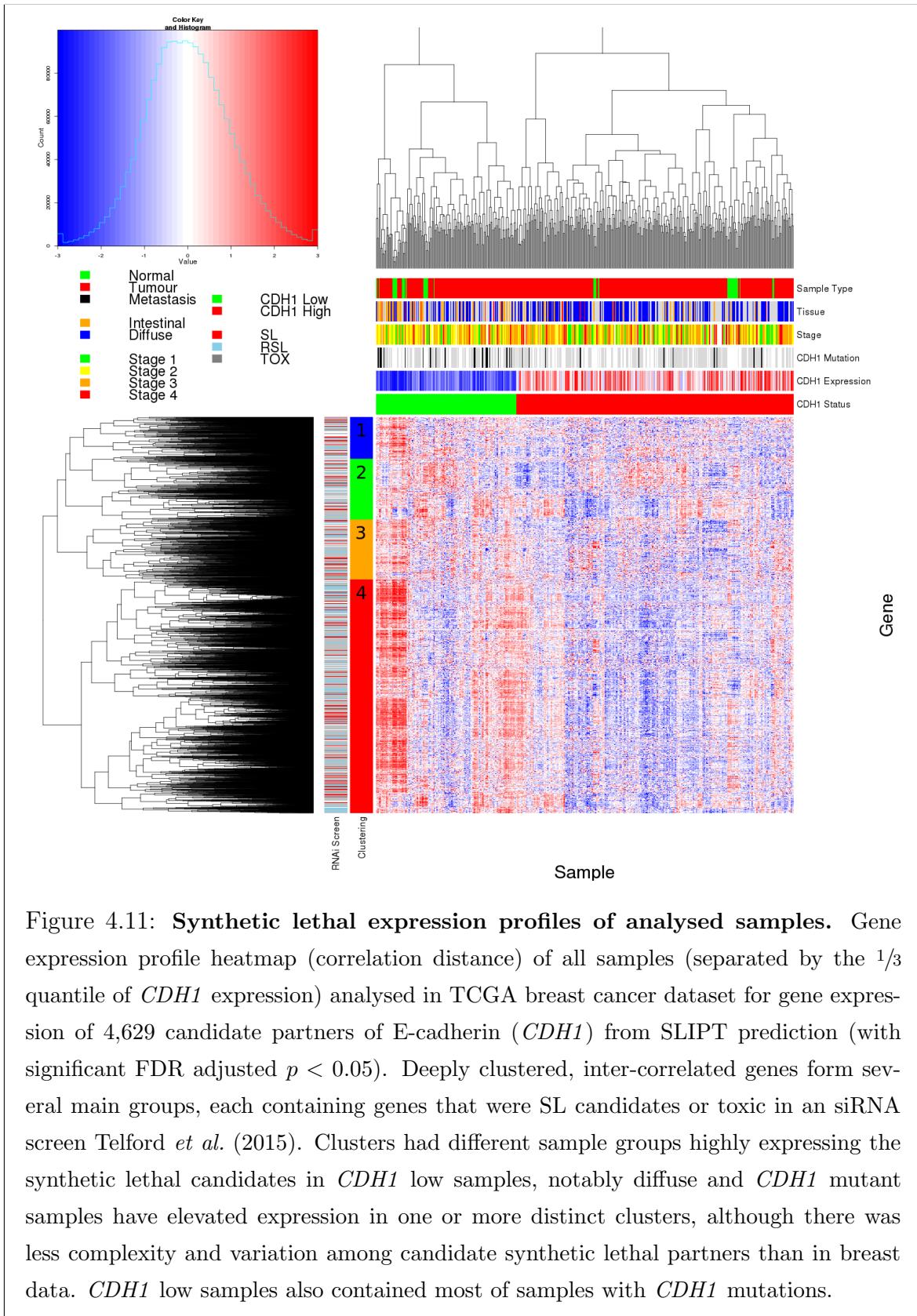
Reactome Pathway	Over-representation	Permutation
Eukaryotic Translation Elongation	1.3×10^{-207}	$< 1.001 \times 10^{-3}$
Peptide chain elongation	5.6×10^{-201}	$< 1.001 \times 10^{-3}$
Viral mRNA Translation	1.2×10^{-196}	$< 1.001 \times 10^{-3}$
Eukaryotic Translation Termination	1.2×10^{-196}	$< 1.001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

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

Table 4.15: Pathways for *CDH1* partners from SLIPT in stomach and siRNA screen

Reactome Pathway	Over-representation	Permutation
Chemokine receptors bind chemokines	5.2×10^{-16}	0.0026524
Class A/1 (Rhodopsin-like receptors)	6.4×10^{-14}	0.05974
Peptide ligand-binding receptors	8.8×10^{-13}	0.10988
Visual phototransduction	1.8×10^{-11}	0.30639
GPCR ligand binding	8.1×10^{-11}	0.17895
Retinoid metabolism and transport	1.3×10^{-10}	0.17481
Gastrin-CREB signalling pathway via PKC and MAPK	1.5×10^{-10}	0.52377
<i>Gαq</i> signalling events	5.6×10^{-10}	0.57601
Platelet activation, signaling and aggregation	1.7×10^{-9}	0.34977
<i>Gαi</i> signalling events	3.5×10^{-8}	0.23131
Response to elevated platelet cytosolic Ca^{2+}	1.8×10^{-7}	0.18637
HS-GAG degradation	4.2×10^{-7}	0.24605
Platelet homeostasis	4.7×10^{-7}	0.18996
VEGFA-VEGFR2 Pathway	5.1×10^{-7}	0.87816
Transcriptional regulation of white adipocyte differentiation	6.4×10^{-7}	0.18505
Signaling by NOTCH4	1.2×10^{-6}	0.36495
Signaling by NOTCH	1.5×10^{-6}	0.76112
Signaling by VEGF	1.5×10^{-6}	0.52553
Defective EXT2 causes exostoses 2	1.7×10^{-6}	0.24605
Defective EXT1 causes exostoses 1, TRPS2 and CHDS	1.7×10^{-6}	0.24605
<i>Gαs</i> signalling events	1.7×10^{-6}	0.31637
Generation of second messenger molecules	3.5×10^{-6}	0.032952
DAP12 interactions	3.5×10^{-6}	0.8492
Mitochondrial Fatty Acid Beta-Oxidation	4×10^{-6}	0.033295
Acyl chain remodelling of PS	6×10^{-6}	0.46799
Phase 1 - Functionalization of compounds	6.5×10^{-6}	0.068729
Costimulation by the CD28 family	6.5×10^{-6}	0.031427
Translocation of ZAP-70 to Immunological synapse	8.1×10^{-6}	$< 2.299 \times 10^{-4}$
Complement cascade	9.8×10^{-6}	$< 2.299 \times 10^{-4}$
Molecules associated with elastic fibres	9.8×10^{-6}	0.025491
Signal amplification	1.1×10^{-5}	0.36204
Phosphorylation of CD3 and TCR zeta chains	1.5×10^{-5}	$< 2.299 \times 10^{-4}$
Cell surface interactions at the vascular wall	1.6×10^{-5}	0.039572
Hemostasis	1.7×10^{-5}	0.22035
FCER1 mediated MAPK activation	1.7×10^{-5}	0.35433
Defective B4GALT7 causes EDS, progeroid type	1.8×10^{-5}	0.36204
Defective B3GAT3 causes JDSSDH	1.8×10^{-5}	0.36204
Elastic fibre formation	1.9×10^{-5}	0.0026524
Signaling by NOTCH1	1.9×10^{-5}	0.52553
Acyl chain remodelling of PE	2.9×10^{-5}	0.46799
TCR signaling	2.9×10^{-5}	0.1269
Signaling by Leptin	2.9×10^{-5}	0.36091
PD-1 signaling	2.9×10^{-5}	$< 2.299 \times 10^{-4}$
Opioid Signalling	3.3×10^{-5}	0.81326
Signaling by SCF-KIT	3.4×10^{-5}	0.79924
Arachidonic acid metabolism	3.4×10^{-5}	0.0033013
DAP12 signaling	3.4×10^{-5}	0.9366
Netrin-1 signaling	3.4×10^{-5}	0.76768
Signaling by Retinoic Acid	3.4×10^{-5}	0.011724
Respiratory electron transport	4×10^{-5}	0.28245

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



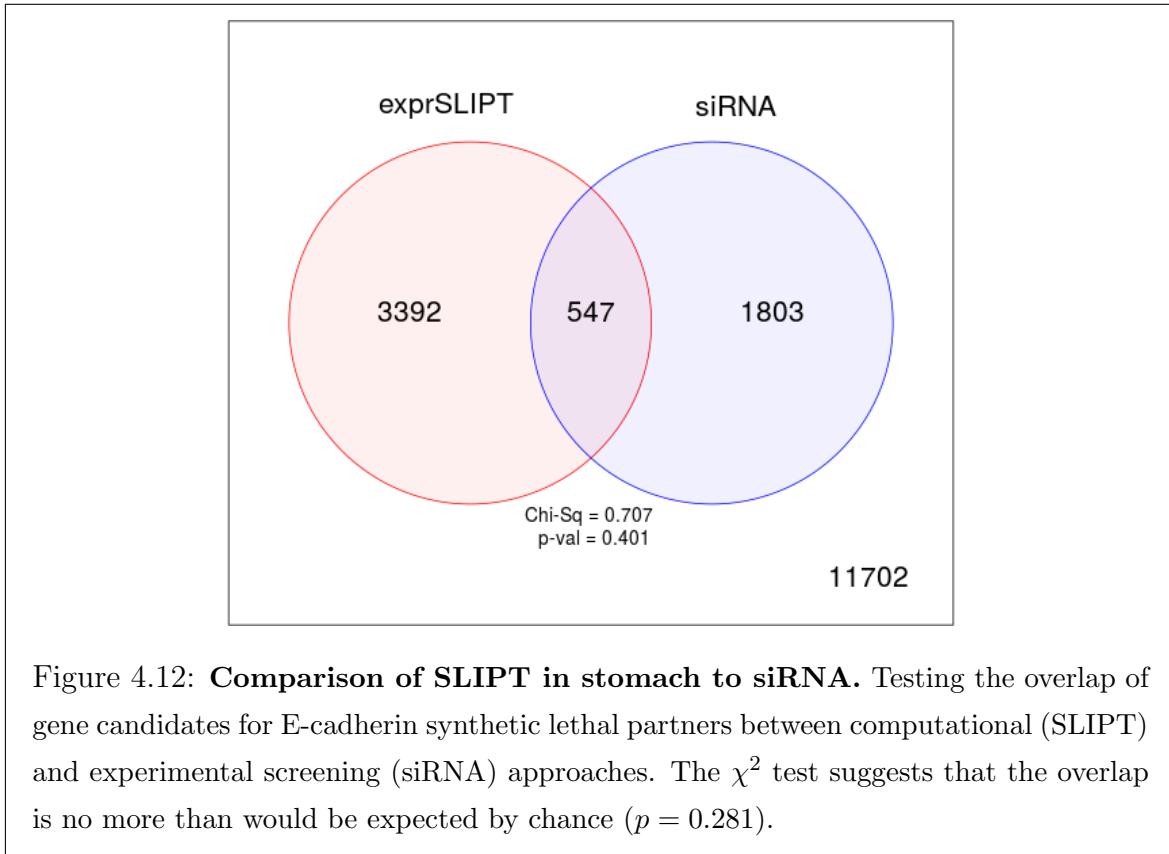


Figure 4.12: **Comparison of SLIPT in stomach to siRNA.** Testing the overlap of gene candidates for E-cadherin synthetic lethal partners between computational (SLIPT) and experimental screening (siRNA) approaches. The χ^2 test suggests that the overlap is no more than would be expected by chance ($p = 0.281$).

breast cancer RNA-Seq data in *CDH1* low tumours (where synthetic lethal partners are expected to have compensating high or stable expression) are shown in Figure 7 and their corresponding functional enrichment is given below in Table 5, computed as WikiPathways in GeneSetDB (Araki *et al.*, 2012). The 3 subgroups of genes are showed functional organisation of expression profiles in *CDH1* low breast tumours. The first group is enriched for G protein coupled receptors, an established drug target and supported in cell line experiments (Telford *et al.*, 2015). The second group contains genes involved in development and metabolism consistent with cancer cells showing stem cell properties and the Warburg hypothesis (Merlos-Suarez *et al.*, 2011; Warburg, 1956). The third group contains cell signalling and focal adhesion functions, including pathways involved in cancer proliferation, metastasis, and consistent with internal synthetic lethality within the pathways containing *CDH1* (Barabási and Oltvai, 2004).

Ductal breast cancers show higher expression of synthetic lethal partners suggesting treatment would be more effective in this tumour subtype. However, there is consistently low expression of SL partners in ER negative tumours, although this is independent of tumour stage and consistent with poor prognosis in these patients and

Table 4.16: Candidate synthetic lethal metagenes against *CDH1* from SLIPT in stomach cancer

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

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

could inform other treatment strategies or prevent ineffective treatment further impacting quality of life in these patients. These results suggest that synthetic lethal partner expression varies between patients; that these different tumour classes would react differently to the same treatment; that treatment of different pathways and combinations in different patients is the most effective approach to target genes compensating for *CDH1* gene loss; and the expression of synthetic partners could be a clinically important biomarker. While these are important clinical implications, the synthetic lethal predictions lack enough confidence for direct translation into pre-clinical models or clinical applications leading to a need for statistical modelling and simulation of synthetic lethality in genomics expression data.

Aims

- Pathway Structure of Candidate Synthetic Lethal Genes for *CDH1* from TCGA breast data

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

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

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

- Comparisons to Experimental siRNA Screen Candidates
- Replication of Pathways across in TCGA Stomach data

Summary

- We have developed a Synthetic Lethal detection method that generates a high number of synthetic lethal candidates
- Pathways in cell signalling, extracellular matrix, and cytoskeletal functions were supported with experimental candidates and the known functions of E-cadherin

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

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

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

- Several candidate pathways were supported by mutation analysis and replicated across breast and stomach cancer
- Translation and immune functions were uniquely detected by the computational approach which may be explained by differences between patient samples and cell line models
- There remains the need to identify actionable genes within these pathways, relationships with experimental candidates, and how these pathways may affect viability when lost

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

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

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

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

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

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

Table 4.21: Candidate synthetic lethal genes against E-cadherin from SLIPT in stomach CCLE

Gene	Observed	Expected	χ^2 value	p-value	p-value (FDR)
<i>ZEB1</i>	1	4.45	36	2.84×10^{-7}	0.00175
<i>WDR47</i>	0	4.45	26.7	2.3×10^{-5}	0.013
<i>KANK2</i>	1	4.45	25.1	4.81×10^{-5}	0.0222
<i>LEPRE1</i>	0	4.45	24.5	6.26×10^{-5}	0.0228
<i>KATNAL1</i>	0	4.45	24.3	6.88×10^{-5}	0.0231
<i>TET1</i>	0	4.45	23.9	8.23×10^{-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 4.22: Pathways for *CDH1* partners from SLIPT in stomach CCLE

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

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

References

- 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.
- Barabási, A.L. and Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet*, **5**(2): 101–13.
- 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.
- 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.
- 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.
- 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.
- 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.*

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.

Merlos-Suarez, A., Barriga, F.M., Jung, P., Iglesias, M., Cespedes, M.V., Rossell, D., Sevillano, M., Hernando-Momblona, X., da Silva-Diz, V., Munoz, P., *et al.* (2011) The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. *Cell Stem Cell*, **8**(5): 511–524.

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.

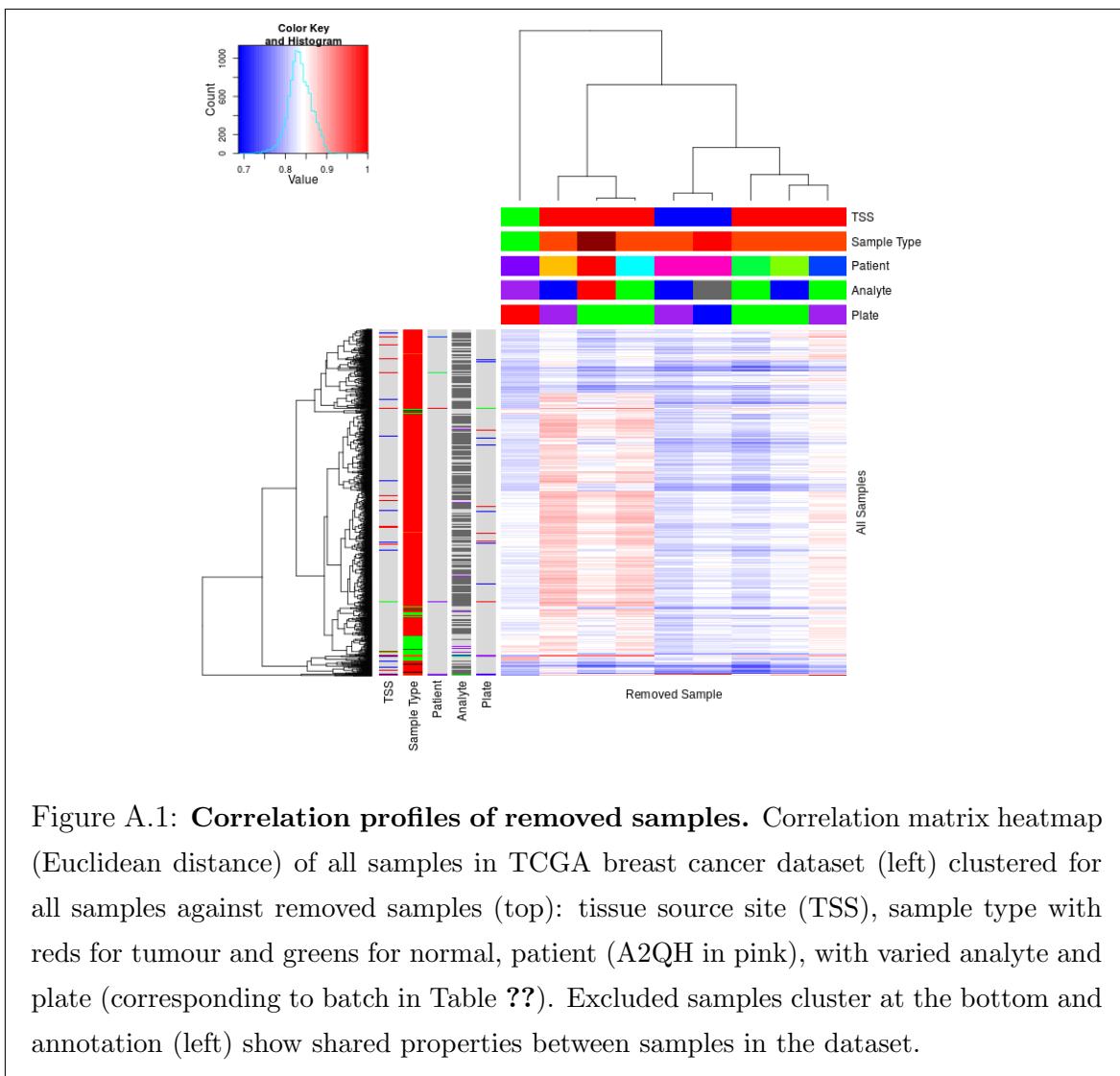
Telford, B.J., Chen, A., Beetham, H., Frick, J., Brew, T.P., Gould, C.M., Single, A., Godwin, T., Simpson, K.J., and Guilford, P. (2015) Synthetic lethal screens identify vulnerabilities in gpcr signalling and cytoskeletal organization in e-cadherin-deficient cells. *Mol Cancer Ther*, **14**(5): 1213–1223.

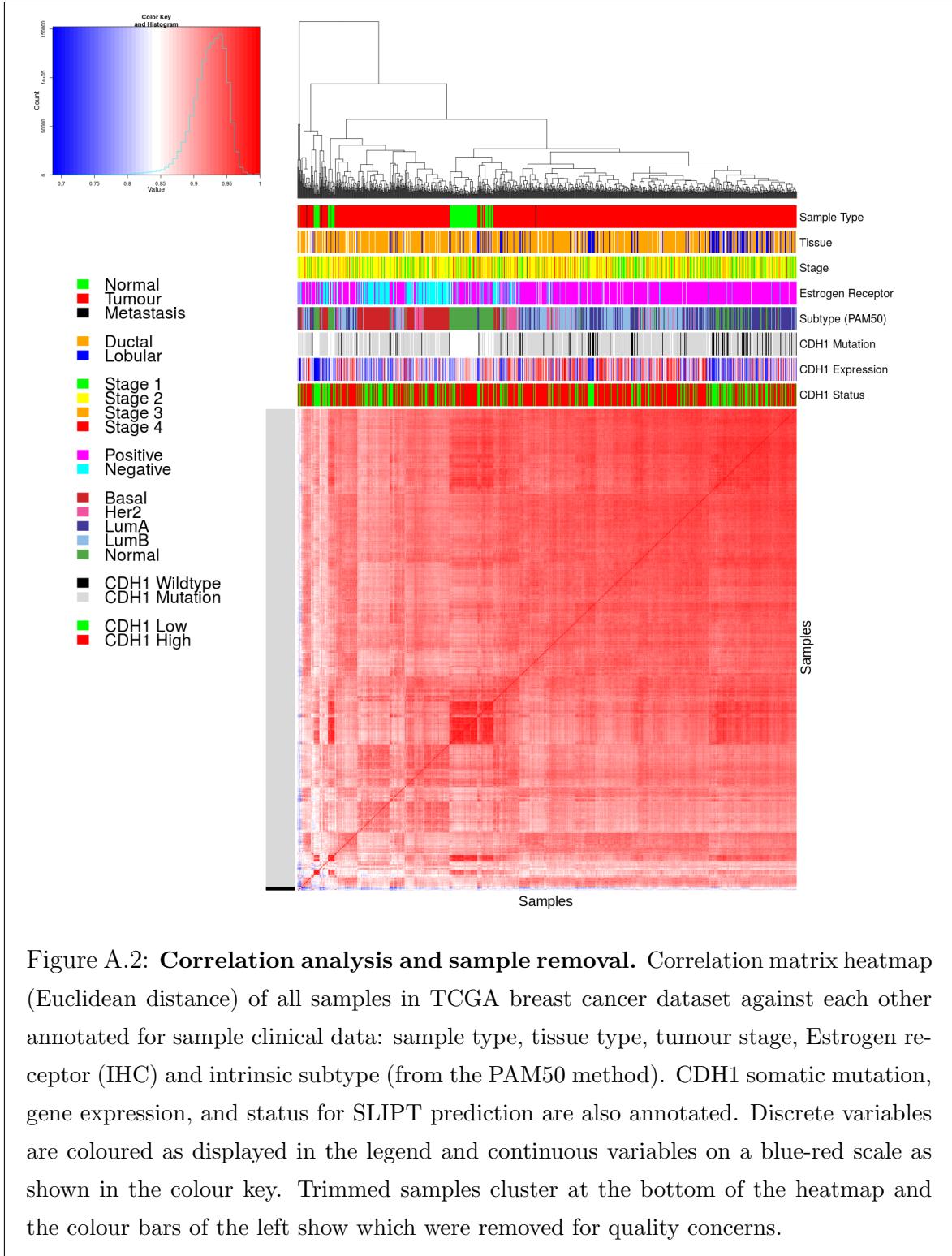
Tiong, K.L., Chang, K.C., Yeh, K.T., Liu, T.Y., Wu, J.H., Hsieh, P.H., Lin, S.H., Lai, W.Y., Hsu, Y.C., Chen, J.Y., *et al.* (2014) Csnk1e/ctnnb1 are synthetic lethal to tp53 in colorectal cancer and are markers for prognosis. *Neoplasia*, **16**(5): 441–50.

Warburg, O. (1956) On the origin of cancer cells. *Science*, **123**(3191): 390–314.

Appendix A

Sample Correlation





Appendix B

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 B.1–B.3. 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 ??). 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.

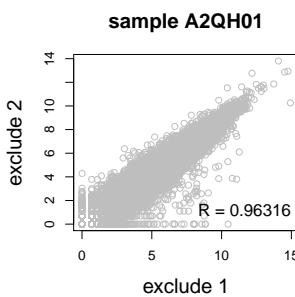


Figure B.1: **Replicate excluded samples.** Both tumour samples of patient A2QH were excluded as they were poorly correlated with other samples, although they are highly similar to each other as shown by Pearson's correlation of log-raw counts.

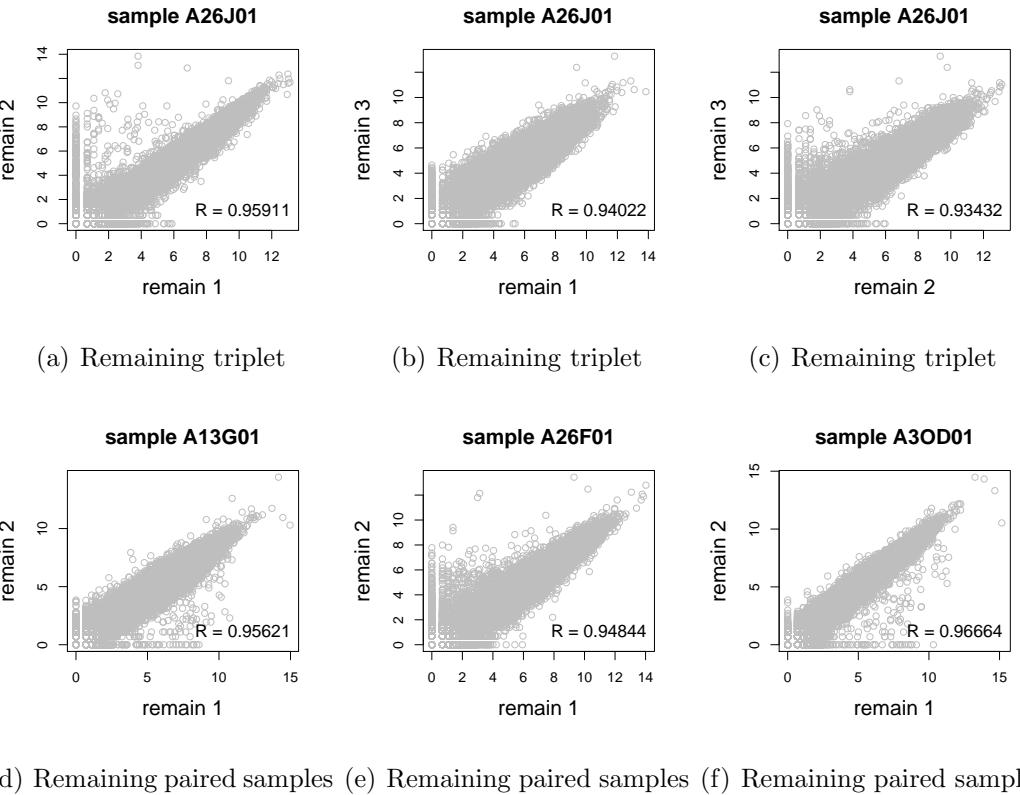


Figure B.2: Replicate samples with all remaining. Patient A26J was sampled 3 times and compared pairwise. Pairs of samples were also compared for other patients with replicate samples. In all cases, replicate samples remaining in the dataset were highly concordant as shown by Pearson's correlation of log-raw counts.

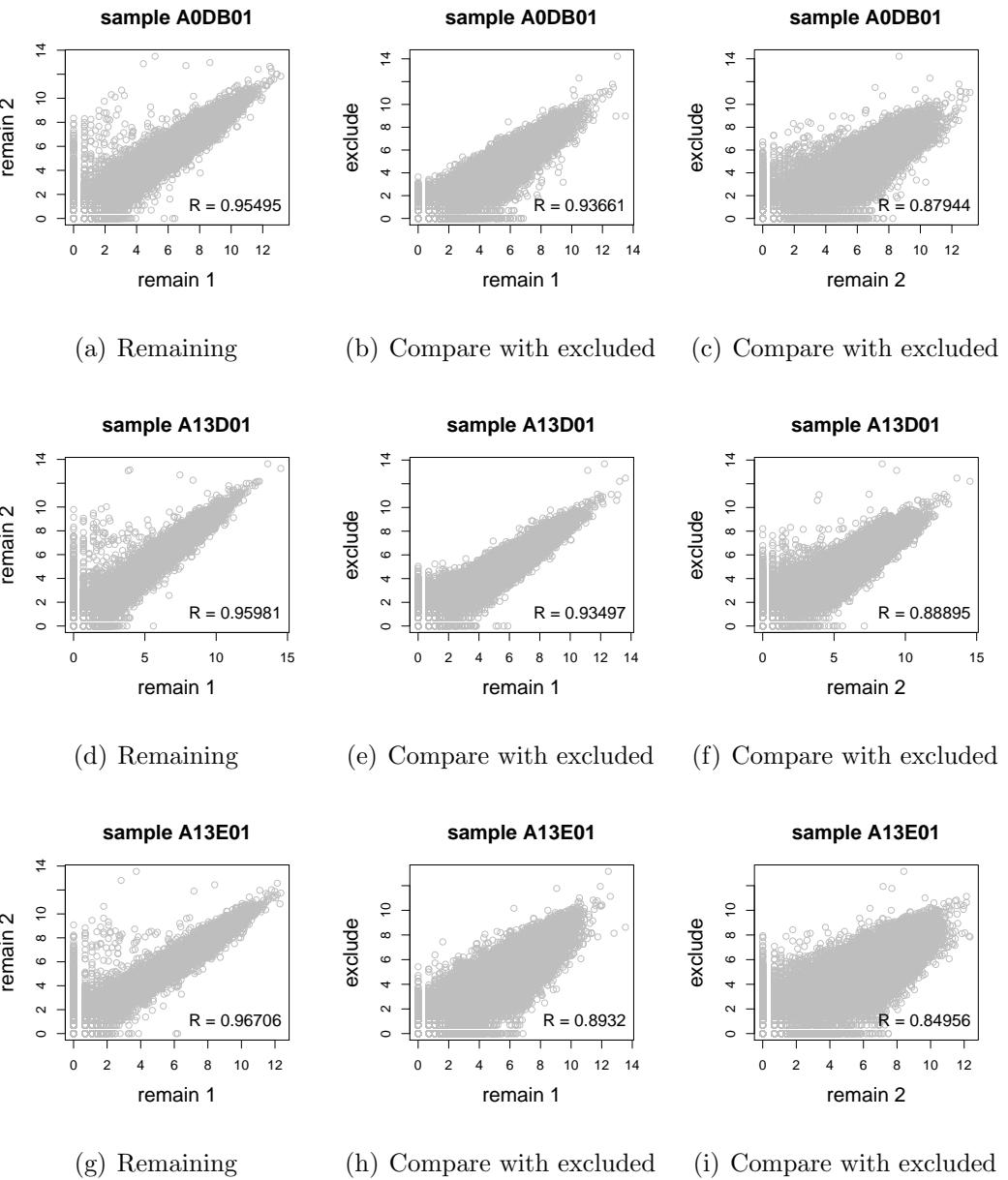


Figure B.3: **Replicate samples with some excluded.** Patients A0DB, A13D, A13E, and A26E were each sampled 3 times and compared pairwise. Pairs of samples were also compared for other patients with replicate samples. In all cases, the replicate samples remaining in the dataset more were highly concordant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

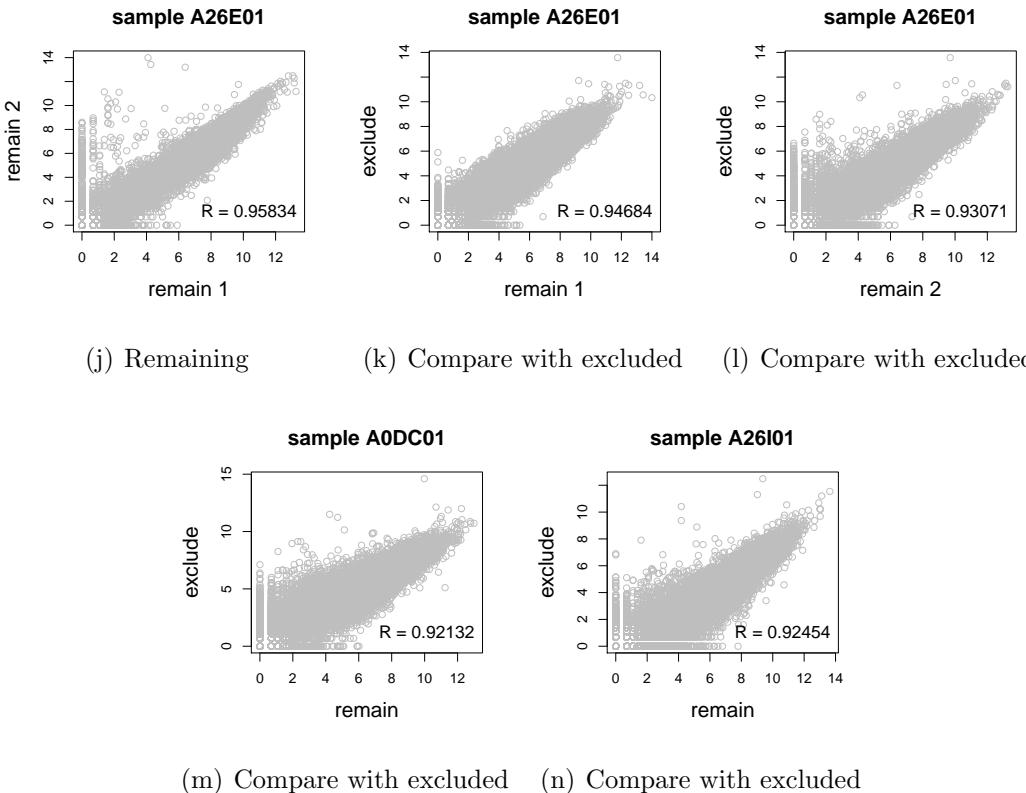


Figure B.3: Replicate samples with some excluded. Patients A0DB, A13D, A13E, and A26E were each sampled 3 times and compared pairwise. Pairs of samples were also compared for other patients with replicate samples. In all cases, the replicate samples remaining in the dataset more were highly concordant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

Appendix C

Software Used for Thesis

Table C.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	
Biostings	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
cgdsr	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 TomKellyGenetics	0.0.0.9000	0.0.0.9000	0.0.0.9000	0.0.0.9000
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				
.permutation	TomKellyGenetics	0.1.0	0.1.0	0.1.0	0.1.0
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				
	TomKellyGenetics	0.0.0.9001	0.0.0.9001	0.0.0.9001	0.0.0.9001
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		
protoools	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	

		GitHub			
		TomKellyGenetics	0.1		
reactometree					
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
rgefx	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	
slpt	GitHub TomKellyGenetics	0.1.0	0.1.0	0.1.0	0.1.0
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
Summarized Experiment	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	1.8-2

XVector	Bioconductor	0.14.0	0.14.0	
yaml	CRAN	2.1.14	2.1.14	2.1.13
zlibbioc	CRAN	1.20.0	1.20.0	
zoo	CRAN	1.7-13	1.7-14	1.7-13

Appendix D

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 D.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 D.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 D.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.

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 D.3, fewer

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

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

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

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

Appendix E

Mutation Analysis in Breast Cancer

Table E.1: Candidate synthetic lethal genes against E-cadherin from mtSLIPT

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

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

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

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

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

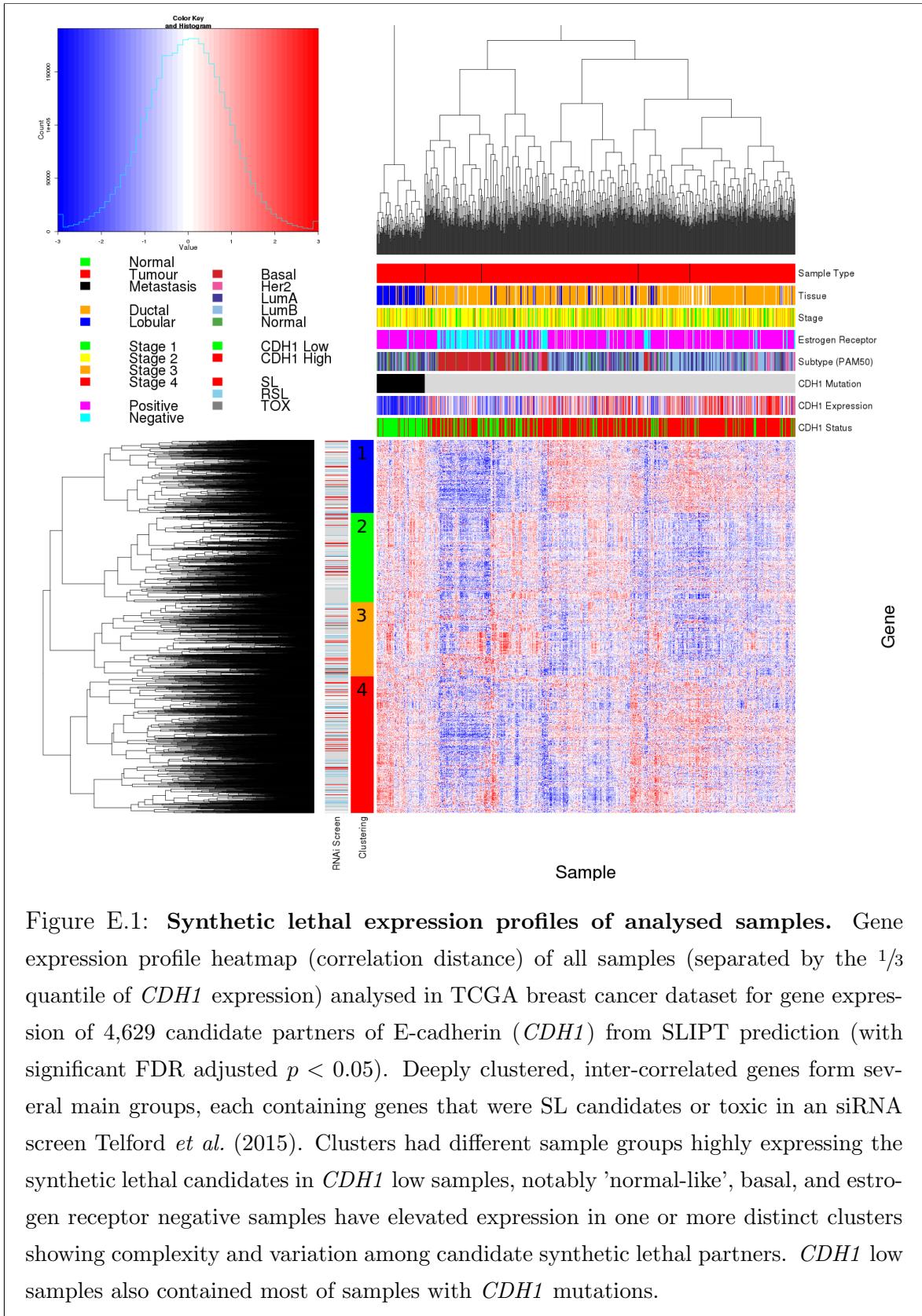


Figure E.1: Synthetic lethal expression profiles of analysed samples. Gene expression profile heatmap (correlation distance) of all samples (separated by the 1/3 quantile of *CDH1* expression) analysed in TCGA breast cancer dataset for gene expression of 4,629 candidate partners of E-cadherin (*CDH1*)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

Table E.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	7.1×10^{-9}
Assembly of the primary cilium	149	14	8×10^{-9}
Sphingolipid metabolism	62	8	9.6×10^{-9}
Signaling by ERBB4	133	12	5.1×10^{-8}
PI3K Cascade	65	7	4.9×10^{-7}
Circadian Clock	33	5	4.9×10^{-7}
Nuclear signaling by ERBB4	34	5	4.9×10^{-7}
Intraflagellar transport	35	5	4.9×10^{-7}
PI3K events in ERBB4 signaling	87	8	4.9×10^{-7}
PIP3 activates AKT signaling	87	8	4.9×10^{-7}
PI3K events in ERBB2 signaling	87	8	4.9×10^{-7}
PI-3K cascade:FGFR1	87	8	4.9×10^{-7}
PI-3K cascade:FGFR2	87	8	4.9×10^{-7}
PI-3K cascade:FGFR3	87	8	4.9×10^{-7}
PI-3K cascade:FGFR4	87	8	4.9×10^{-7}
Deadenylation of mRNA	22	4	5.6×10^{-7}
PI3K/AKT activation	90	8	5.6×10^{-7}
Cargo trafficking to the periciliary membrane	38	5	5.6×10^{-7}
Signaling by Hedgehog	108	9	5.6×10^{-7}
Downstream signal transduction	143	11	5.6×10^{-7}

Pathways Over-represented in Cluster 2	Pathway Size	Cluster Genes	p-value (FDR)
G _{αs} signalling events	83	19	5.1×10^{-25}
Extracellular matrix organization	238	30	1.4×10^{-18}
Hemostasis	422	46	2.7×10^{-16}
Aquaporin-mediated transport	32	9	2.7×10^{-16}
Transcriptional regulation of white adipocyte differentiation	56	11	1.7×10^{-15}
Degradation of the extracellular matrix	102	15	1.7×10^{-15}
Integration of energy metabolism	84	13	8.8×10^{-15}
GPCR downstream signaling	472	48	2.8×10^{-14}
G _{αs} signalling events	15	6	5×10^{-14}
Molecules associated with elastic fibres	33	8	5.4×10^{-14}
Phase 1 - Functionalization of compounds	67	11	5.6×10^{-14}
Platelet activation, signaling and aggregation	179	20	5.6×10^{-14}
Vasopressin regulates renal water homeostasis via Aquaporins	24	7	6.1×10^{-14}
Elastic fibre formation	37	8	3×10^{-13}
Calmodulin induced events	27	7	3.3×10^{-13}
CaM pathway	27	7	3.3×10^{-13}
cGMP effects	18	6	3.6×10^{-13}
G _{αs} signalling events	167	18	6.3×10^{-13}
Ca-dependent events	29	7	8.2×10^{-13}
Binding and Uptake of Ligands by Scavenger Receptors	40	8	8.2×10^{-13}

Pathways Over-represented in Cluster 3	Pathway Size	Cluster Genes	p-value (FDR)
Eukaryotic Translation Elongation	86	55	1.1×10^{-112}
Peptide chain elongation	83	54	1.3×10^{-112}
Viral mRNA Translation	81	53	1.6×10^{-111}
Eukaryotic Translation Termination	83	53	7.1×10^{-110}
Nonsense Mediated Decay independent of the Exon Junction Complex	88	54	1×10^{-108}
Formation of a pool of free 40S subunits	93	53	4.1×10^{-102}
Nonsense-Mediated Decay	103	54	3.9×10^{-98}
Nonsense-Mediated Decay enhanced by the Exon Junction Complex	103	54	3.9×10^{-98}
L13a-mediated translational silencing of Ceruloplasmin expression	103	53	1.2×10^{-95}
3' -UTR-mediated translational regulation	103	53	1.2×10^{-95}
SRP-dependent cotranslational protein targeting to membrane	104	53	4.3×10^{-95}
GTP hydrolysis and joining of the 60S ribosomal subunit	104	53	4.3×10^{-95}
Influenza Viral RNA Transcription and Replication	108	53	9.6×10^{-93}
Eukaryotic Translation Initiation	111	53	4.2×10^{-91}
Cap-dependent Translation Initiation	111	53	4.2×10^{-91}
Influenza Life Cycle	112	53	1.4×10^{-90}
Influenza Infection	117	53	6.2×10^{-88}
Translation	141	55	3×10^{-81}
Formation of the ternary complex, and subsequently, the 43S complex	47	23	2.3×10^{-48}
Translation initiation complex formation	54	23	9.1×10^{-45}

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

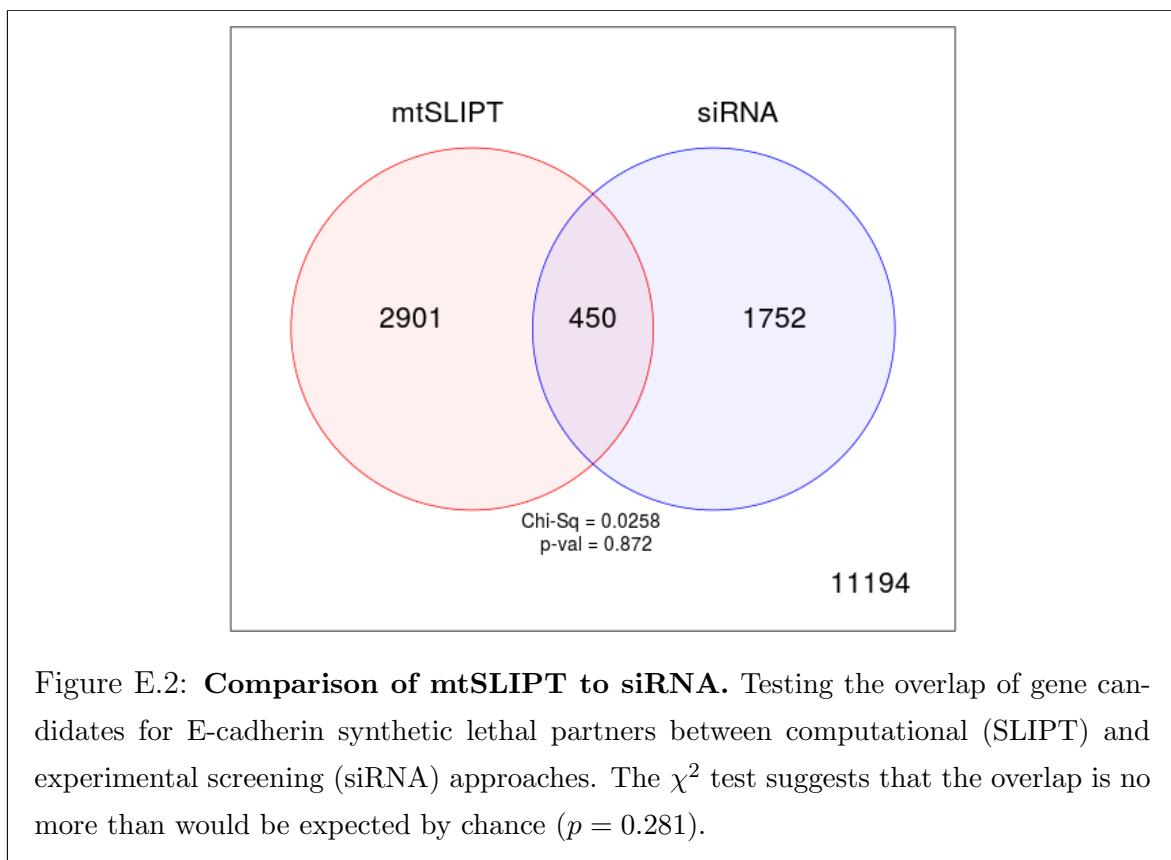


Figure E.2: **Comparison of mtSLIPT to siRNA.** Testing the overlap of gene candidates for E-cadherin synthetic lethal partners between computational (SLIPT) and experimental screening (siRNA) approaches. The χ^2 test suggests that the overlap is no more than would be expected by chance ($p = 0.281$).

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

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

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

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

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

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

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

Reactome Pathway	Over-representation	Permutation
Visual phototransduction	1.2×10^{-9}	0.86279
G_{as} signalling events	2.9×10^{-7}	0.023066
Retinoid metabolism and transport	2.9×10^{-7}	0.299
Acylic chain remodelling of PS	1.1×10^{-5}	0.42584
Transcriptional regulation of white adipocyte differentiation	1.1×10^{-5}	0.53928
Chemokine receptors bind chemokines	1.1×10^{-5}	0.95259
<i>Signaling by NOTCH4</i>	1.2×10^{-5}	0.079229
Defective EXT2 causes exostoses 2	1.2×10^{-5}	0.22292
Defective EXT1 causes exostoses 1, TRPS2 and CHDS	1.2×10^{-5}	0.22292
Platelet activation, signaling and aggregation	1.2×10^{-5}	0.48853
Serotonin receptors	1.4×10^{-5}	0.34596
Nicotinamide salvaging	1.4×10^{-5}	0.70881
Phase 1 - Functionalization of compounds	2×10^{-5}	0.31142
Amine ligand-binding receptors	2.5×10^{-5}	0.34934
Acylic chain remodelling of PE	3.8×10^{-5}	0.42615
Signaling by GPCR	3.8×10^{-5}	0.93888
Molecules associated with elastic fibres	3.9×10^{-5}	0.017982
DAP12 interactions	3.9×10^{-5}	0.71983
Beta defensins	3.9×10^{-5}	0.91458
Cytochrome P ₄₅₀ - arranged by substrate type	4.7×10^{-5}	0.83493
GPCR ligand binding	5.7×10^{-5}	0.95258
Acylic chain remodelling of PC	6.1×10^{-5}	0.42584
Response to elevated platelet cytosolic Ca ²⁺	6.4×10^{-5}	0.54046
Arachidonic acid metabolism	6.7×10^{-5}	0.026696
Defective B4GALT7 causes EDS, progeroid type	7.3×10^{-5}	0.24921
Defective B3GAT3 causes JDSSDHD	7.3×10^{-5}	0.24921
Hydrolysis of LPC	7.3×10^{-5}	0.80663
Elastic fibre formation	7.4×10^{-5}	0.0058768
HS-GAG degradation	9.4×10^{-5}	0.0083179
<i>Bile acid and bile salt metabolism</i>	9.4×10^{-5}	0.079905
Netrin-1 signaling	0.00011	0.92216
Integration of energy metabolism	0.00011	0.011152
Dectin-2 family	0.00012	0.10385
Platelet sensitization by LDL	0.00012	0.34596
DAP12 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).

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

Pathway	ID	Observed	Expected	χ^2 value	p-value	p-value (FDR)
Linoleic acid (LA) metabolism	2046105	79	36.70	87.03	1.2637×10^{-19}	2.0839×10^{-160}
ATF6-alpha activates chaperone genes	381183	78	36.70	80.25	3.7449×10^{-18}	3.0877×10^{-150}
Neurotoxicity of clostridium toxins	168799	8	36.70	79.41	5.7092×10^{-18}	3.1382×10^{-150}
Aquaporin-mediated transport	445717	8	36.70	76.28	2.7327×10^{-17}	9.0124×10^{-150}
Toxicity of botulinum toxin type G (BoNTG)	5250989	8	36.70	76.278	2.7327×10^{-17}	9.0124×10^{-150}
Purine metabolism	73847	75	36.70	75.86	3.3623×10^{-17}	9.2407×10^{-150}
Chk1Chk2(Cds1) mediated inactivation of Cyclin B:Cdk1 complex	75035	74	36.70	71.68	2.7211×10^{-16}	6.41×10^{-140}
Scavenging by Class F Receptors	3000484	75	36.70	69.56	7.8573×10^{-16}	1.4396×10^{-130}
Cytosolic tRNA aminoacylation	379716	75	36.70	69.56	7.8573×10^{-16}	1.4396×10^{-130}
G1S Transition	69206	74	36.70	69.21	9.3593×10^{-16}	1.5433×10^{-130}
ABC-family proteins mediated transport	382556	10	36.70	68.16	1.5826×10^{-15}	1.8641×10^{-130}
MG1 Transition	68874	74	36.70	68.16	1.5826×10^{-15}	1.8641×10^{-130}
DNA Replication Pre-Initiation	69002	74	36.70	68.16	1.5826×10^{-15}	1.8641×10^{-130}
Cell Cycle Checkpoints	69620	74	36.70	68.16	1.5826×10^{-15}	1.8641×10^{-130}
Basigin interactions	210991	74	36.70	67.23	2.5162×10^{-15}	2.7661×10^{-130}
Mitotic G1-G1S phases	453279	72	36.70	64.98	7.7471×10^{-15}	7.9843×10^{-130}
Metabolism of folate and pterines	196757	73	36.70	63.42	1.6932×10^{-14}	1.6424×10^{-120}
Tetrahydrobiopterin (BH4) synthesis, recycling, salvage and regulation	1474151	73	36.70	62.68	2.4547×10^{-14}	2.0427×10^{-120}
DNA Replication	69306	72	36.70	62.51	2.6652×10^{-14}	2.0427×10^{-120}
Separation of Sister Chromatids	2467813	71	36.70	62.47	2.7252×10^{-14}	2.0427×10^{-120}
M Phase	68886	71	36.70	62.47	2.7252×10^{-14}	2.0427×10^{-120}
Cell Cycle, Mitotic	69278	71	36.70	62.47	2.7252×10^{-14}	2.0427×10^{-120}
G0 and Early G1	1538133	70	36.70	61.62	4.1658×10^{-14}	2.8623×10^{-120}
Regulation of PLK1 Activity at G2M Transition	2565942	70	36.70	61.62	4.1658×10^{-14}	2.8623×10^{-120}
alpha-linolenic (omega3) and linoleic (omega6) acid metabolism	2046104	70	36.70	60.07	9.0139×10^{-14}	5.1255×10^{-120}

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

Appendix F

Expression Analysis in Stomach Cancer

Appendix G

Mutation Analysis in Stomach Cancer

Table G.1: Candidate synthetic lethal genes against E-cadherin from mtSLIPT in stomach cancer

Gene	Observed	Expected	χ^2 value	p-value	p-value (FDR)
<i>OLFML1</i>	5	10.1	29.2	4.53×10^{-7}	0.0031
<i>NRIP2</i>	6	10.1	25.4	3.11×10^{-6}	0.00706
<i>VIM</i>	3	10.1	24.7	4.29×10^{-6}	0.00706
<i>TCF4</i>	5	10.1	24.7	4.33×10^{-6}	0.00706
<i>ZEB2</i>	5	10.1	24.7	4.33×10^{-6}	0.00706
<i>BCL2</i>	2	10.1	22	1.66×10^{-5}	0.0155
<i>SMARCA2</i>	2	10.1	22	1.66×10^{-5}	0.0155
<i>CCND2</i>	3	10.1	21.1	2.61×10^{-5}	0.0155
<i>MMP19</i>	3	10.1	21.1	2.61×10^{-5}	0.0155
<i>NEURL1B</i>	3	10.1	21.1	2.61×10^{-5}	0.0155
<i>IGFBP6</i>	6	10.1	21.1	2.65×10^{-5}	0.0155
<i>OGN</i>	6	10.1	21.1	2.65×10^{-5}	0.0155
<i>THY1</i>	6	10.2	21	2.7×10^{-5}	0.0155
<i>DZIP1</i>	4	10.1	20.6	3.29×10^{-5}	0.0155
<i>LOC650368</i>	4	10.1	20.6	3.29×10^{-5}	0.0155
<i>PCOLCE</i>	4	10.1	20.6	3.29×10^{-5}	0.0155
<i>PTGFR</i>	4	10.1	20.6	3.29×10^{-5}	0.0155
<i>RUNX1T1</i>	4	10.1	20.6	3.29×10^{-5}	0.0155
<i>CLEC2B</i>	5	10.1	20.6	3.3×10^{-5}	0.0155
<i>MSC</i>	5	10.1	20.6	3.3×10^{-5}	0.0155
<i>NISCH</i>	5	10.1	20.6	3.3×10^{-5}	0.0155
<i>TSPAN11</i>	5	10.1	20.6	3.3×10^{-5}	0.0155
<i>KCTD12</i>	2	10.1	19.1	7.19×10^{-5}	0.0246
<i>LRRC55</i>	2	10.1	19.1	7.19×10^{-5}	0.0246
<i>PCBP3</i>	2	10.1	19.1	7.19×10^{-5}	0.0246

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

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	9.6×10^{-9}
Elastic fibre formation	38	6	3.7×10^{-8}
Diseases associated with glycosaminoglycan metabolism	26	5	3.7×10^{-8}
Diseases of glycosylation	26	5	3.7×10^{-8}
Nitric oxide stimulates guanylate cyclase	24	4	3.1×10^{-6}
Molecules associated with elastic fibres	34	4	3.7×10^{-5}
Platelet homeostasis	54	5	3.7×10^{-5}
Initial triggering of complement	17	3	3.7×10^{-5}
Regulation of IGF transport and uptake by IGFBPs	17	3	3.7×10^{-5}
Collagen degradation	58	5	5.6×10^{-5}
Defective B4GALT7 causes EDS, progeroid type	19	3	5.6×10^{-5}
Defective B3GAT3 causes JDSSDHD	19	3	5.6×10^{-5}
Degradation of the extracellular matrix	104	7	8.0×10^{-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*

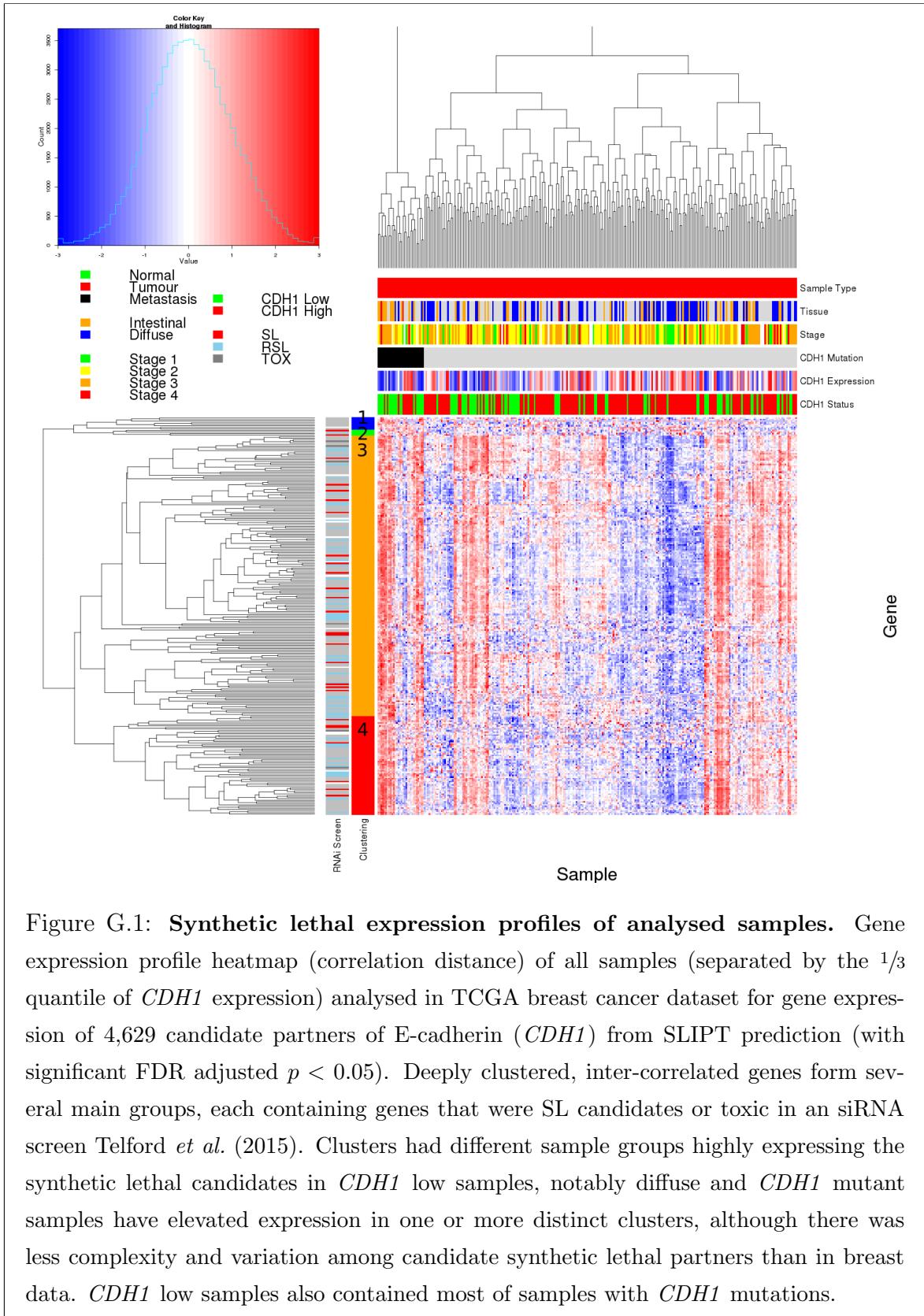


Table G.3: Pathway composition for clusters of *CDH1* partners in stomach mtSLIPT

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- κ 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	9.6×10^{-9}
Elastic fibre formation	38	6	3.7×10^{-8}
Diseases associated with glycosaminoglycan metabolism	26	5	3.7×10^{-8}
Diseases of glycosylation	26	5	3.7×10^{-8}
Molecules associated with elastic fibres	34	4	4.8×10^{-5}
Initial triggering of complement	17	3	4.8×10^{-5}
Regulation of IGF transport and uptake by IGFBPs	17	3	4.8×10^{-5}
Collagen degradation	58	5	6.7×10^{-5}
Defective B4GALT7 causes EDS, progeroid type	19	3	6.7×10^{-5}
Defective B3GAT3 causes JDSSDH	19	3	6.7×10^{-5}
Degradation of the extracellular matrix	104	7	9.5×10^{-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

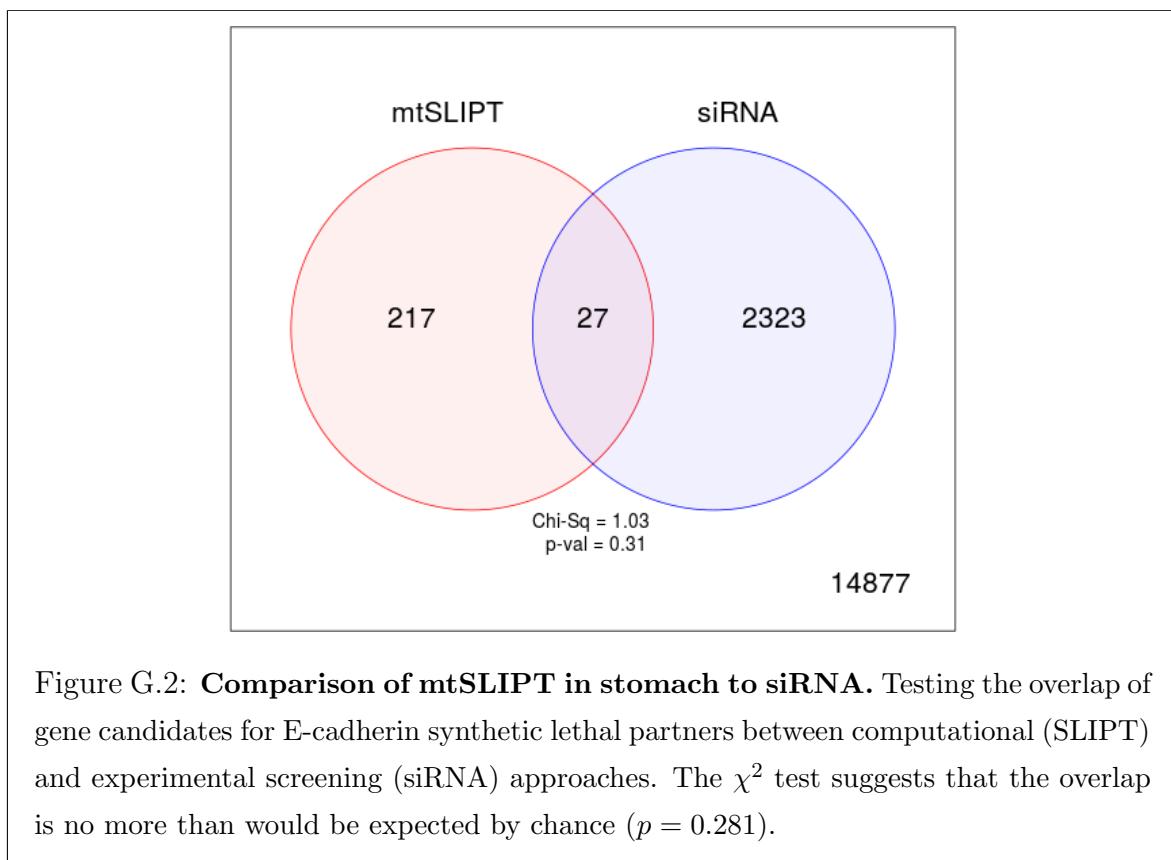


Table G.4: Pathway composition for *CDH1* partners from mtSLIPT and siRNA

Predicted only by SLIPT (217 genes)	Pathway	Size	Genes Identified	p-value (FDR)
Eukaryotic Translation Elongation		87	57	2.8×10^{-120}
Peptide chain elongation		84	56	3.1×10^{-120}
Eukaryotic Translation Termination		84	55	2.8×10^{-117}
Viral mRNA Translation		82	54	4.1×10^{-116}
Nonsense Mediated Decay independent of the Exon Junction Complex		89	55	3.7×10^{-113}
Formation of a pool of free 40S subunits		94	55	2.8×10^{-109}
Nonsense-Mediated Decay		104	57	8.4×10^{-108}
Nonsense Mediated Decay enhanced by the Exon Junction Complex		104	57	8.4×10^{-108}
L13a-mediated translational silencing of Ceruloplasmin expression		104	56	3.4×10^{-105}
3' -UTR-mediated translational regulation		104	56	3.4×10^{-105}
GTP hydrolysis and joining of the 60S ribosomal subunit		105	56	1.4×10^{-104}
Eukaryotic Translation Initiation		112	56	2.8×10^{-100}
Cap-dependent Translation Initiation		112	56	2.8×10^{-100}
SRP-dependent cotranslational protein targeting to membrane		105	54	2.2×10^{-99}
Influenza Viral RNA Transcription and Replication		109	54	5.3×10^{-97}
Influenza Life Cycle		113	54	9.6×10^{-95}
Influenza Infection		118	55	1.7×10^{-94}
Translation		142	60	3.5×10^{-94}
Infectious disease		349	77	5.9×10^{-62}
Extracellular matrix organization		241	54	3×10^{-52}

Detected only by siRNA screen (2323 genes)	Pathway	Size	Genes Identified	p-value (FDR)
Class A/1 (Rhodopsin-like receptors)		282	69	1.9×10^{-59}
GPCR ligand binding		363	78	2.7×10^{-54}
Peptide ligand-binding receptors		175	41	1.5×10^{-42}
G _{αi} signalling events		184	41	1.1×10^{-40}
Gastrin-CREB signalling pathway via PKC and MAPK		180	37	1.5×10^{-35}
G _{αq} signalling events		159	34	3.7×10^{-35}
DAP12 interactions		159	27	1.1×10^{-24}
VEGFA-VEGFR2 Pathway		91	19	1.0×10^{-23}
Downstream signal transduction		146	24	1.9×10^{-22}
Signaling by VEGF		99	19	2.6×10^{-22}
DAP12 signaling		149	24	4.2×10^{-22}
Organelle biogenesis and maintenance		264	34	4.3×10^{-20}
Downstream signaling of activated FGFR1		134	21	4.3×10^{-20}
Downstream signaling of activated FGFR2		134	21	4.3×10^{-20}
Downstream signaling of activated FGFR3		134	21	4.3×10^{-20}
Downstream signaling of activated FGFR4		134	21	4.3×10^{-20}
Signaling by ERBB2		146	22	5.3×10^{-20}
Signaling by FGFR		146	22	5.3×10^{-20}
Signaling by FGFR1		146	22	5.3×10^{-20}
Signaling by FGFR2		146	22	5.3×10^{-20}

Intersection of SLIPT and siRNA screen (23 genes)	Pathway	Size	Genes Identified	p-value (FDR)
HS-GAG degradation		21	4	4.9×10^{-6}
Retinoid metabolism and transport		39	5	4.9×10^{-6}
Platelet activation, signaling and aggregation		186	13	4.9×10^{-6}
Signaling by NOTCH4		11	3	4.9×10^{-6}
G _{αs} signalling events		100	8	5×10^{-6}
Defective EXT2 causes exostoses 2		12	3	5×10^{-6}
Defective EXT1 causes exostoses 1, TRPS2 and CHDS		12	3	5×10^{-6}
Class A/1 (Rhodopsin-like receptors)		289	18	2.2×10^{-5}
Signaling by PDGF		173	11	2.9×10^{-5}
Circadian Clock		34	4	2.9×10^{-5}
Signaling by ERBB4		139	9	4.3×10^{-5}
Role of LAT2/NTAL/LAB on calcium mobilization		99	7	4.4×10^{-5}
Peptide ligand-binding receptors		181	11	4.5×10^{-5}
Defective B4GALT7 causes EDS, progeroid type		19	3	4.5×10^{-5}
Defective B3GAT3 causes JDSSDH		19	3	4.5×10^{-5}
Signaling by NOTCH		80	6	4.5×10^{-5}
G _{αq} signalling events		164	10	5.1×10^{-5}
Response to elevated platelet cytosolic Ca ²⁺		84	6	7.1×10^{-5}
Signaling by ERBB2		148	9	7.1×10^{-5}
Signaling by SCF-KIT		129	8	8.3×10^{-5}

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

Reactome Pathway	Over-representation	Permutation
Eukaryotic Translation Elongation	2×10^{-128}	$< 8.802 \times 10^{-4}$
Peptide chain elongation	2×10^{-128}	$< 8.802 \times 10^{-4}$
Eukaryotic Translation Termination	2.3×10^{-125}	$< 8.802 \times 10^{-4}$
Viral mRNA Translation	2.5×10^{-124}	$< 8.802 \times 10^{-4}$
Nonsense Mediated Decay independent of the Exon Junction Complex	8.6×10^{-124}	$< 8.802 \times 10^{-4}$
Nonsense-Mediated Decay	5.2×10^{-117}	$< 8.802 \times 10^{-4}$
Nonsense Mediated Decay enhanced by the Exon Junction Complex	5.2×10^{-117}	$< 8.802 \times 10^{-4}$
Formation of a pool of free 40S subunits	1.6×10^{-116}	$< 8.802 \times 10^{-4}$
L13a-mediated translational silencing of Ceruloplasmin expression	1.3×10^{-111}	$< 8.802 \times 10^{-4}$
3' -UTR-mediated translational regulation	1.3×10^{-111}	$< 8.802 \times 10^{-4}$
GTP hydrolysis and joining of the 60S ribosomal subunit	6.2×10^{-111}	$< 8.802 \times 10^{-4}$
SRP-dependent cotranslational protein targeting to membrane	2.9×10^{-108}	$< 8.802 \times 10^{-4}$
Eukaryotic Translation Initiation	3×10^{-106}	$< 8.802 \times 10^{-4}$
Cap-dependent Translation Initiation	3×10^{-106}	$< 8.802 \times 10^{-4}$
Influenza Viral RNA Transcription and Replication	5.1×10^{-103}	$< 8.802 \times 10^{-4}$
Influenza Infection	1.5×10^{-102}	$< 8.802 \times 10^{-4}$
Translation	3.7×10^{-101}	$< 8.802 \times 10^{-4}$
Influenza Life Cycle	1.4×10^{-100}	$< 8.802 \times 10^{-4}$
GPCR downstream signaling	1×10^{-80}	0.034498
Hemostasis	1.4×10^{-78}	0.086519
Extracellular matrix organization	1.5×10^{-67}	0.040016
Developmental Biology	1.8×10^{-66}	0.18385
Infectious disease	7.3×10^{-66}	0.068426
Signalling by NGF	8.5×10^{-63}	0.16798
Metabolism of lipids and lipoproteins	4.9×10^{-58}	0.51411
Platelet activation, signaling and aggregation	2.7×10^{-55}	0.081717
GPCR ligand binding	7.3×10^{-55}	0.28898
Signaling by PDGF	8.4×10^{-55}	0.16025
Class A/1 (Rhodopsin-like receptors)	3.2×10^{-54}	0.22801
Fc epsilon receptor (FCER1) signaling	6.2×10^{-53}	0.15229
Adaptive Immune System	5.1×10^{-52}	0.037698
Signaling by ERBB4	5.9×10^{-52}	0.10088
Axon guidance	8.8×10^{-52}	0.40234
Formation of the ternary complex, and subsequently, the 43S complex	1.6×10^{-51}	0.00088017
Ribosomal scanning and start codon recognition	2.2×10^{-50}	0.00088017
Translation initiation complex formation	2.2×10^{-50}	0.0017305
NGF signalling via TRKA from the plasma membrane	6.7×10^{-50}	0.28811
Activation of the mRNA upon binding of the cap-binding complex and eIFs, and subsequent binding to 43S	7.1×10^{-50}	0.0017305
Transmembrane transport of small molecules	1.8×10^{-49}	0.081229
Signaling by ERBB2	5.9×10^{-49}	0.11896
Rho GTPase cycle	3.6×10^{-48}	0.035735
Gαs signalling events	1.1×10^{-47}	0.0088487
Downstream signal transduction	1.7×10^{-47}	0.11909
Signaling by FGFR	1.7×10^{-47}	0.11896
Signaling by FGFR1	1.7×10^{-47}	0.11896
Signaling by FGFR2	1.7×10^{-47}	0.11896
Signaling by FGFR3	1.7×10^{-47}	0.11896
Signaling by FGFR4	1.7×10^{-47}	0.11896
DAP12 interactions	1.9×10^{-47}	0.28811
DAP12 signaling	1×10^{-46}	0.12442

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
Signaling by NOTCH4	4.9×10^{-6}	0.050121
HS-GAG degradation	4.9×10^{-6}	0.013193
Platelet activation, signaling and aggregation	4.9×10^{-6}	0.28053
Retinoid metabolism and transport	4.9×10^{-6}	0.0927
Defective EXT2 causes exostoses 2	5×10^{-6}	0.14898
Defective EXT1 causes exostoses 1, TRPS2 and CHDS	5×10^{-6}	0.14898
<i>Gαs</i> signalling events	5×10^{-6}	0.048426
Class A/1 (Rhodopsin-like receptors)	2.2×10^{-5}	0.60435
Signaling by PDGF	2.9×10^{-5}	0.43907
Circadian Clock	2.9×10^{-5}	0.012519
Signaling by ERBB4	4.3×10^{-5}	0.12835
Role of LAT2/NTAL/LAB on calcium mobilization	4.4×10^{-5}	0.27344
Defective B4GALT7 causes EDS, progeroid type	4.5×10^{-5}	0.23536
Defective B3GAT3 causes JDSSDHD	4.5×10^{-5}	0.23536
Peptide ligand-binding receptors	4.5×10^{-5}	0.41193
Signaling by NOTCH	4.5×10^{-5}	0.10912
<i>Gαq</i> signalling events	5.1×10^{-5}	0.28937
Signaling by ERBB2	7.1×10^{-5}	0.50797
Response to elevated platelet cytosolic Ca^{2+}	7.1×10^{-5}	0.38513
Signaling by SCF-KIT	8.3×10^{-5}	0.55412
PI3K events in ERBB4 signaling	0.0001	0.24486
PIP3 activates AKT signaling	0.0001	0.24486
Collagen formation	0.0001	0.15296
PI3K events in ERBB2 signaling	0.0001	0.24486
PI-3K cascade:FGFR1	0.0001	0.24486
PI-3K cascade:FGFR2	0.0001	0.24486
PI-3K cascade:FGFR3	0.0001	0.24486
PI-3K cascade:FGFR4	0.0001	0.24486
Growth hormone receptor signaling	0.0001	0.057494
PI3K Cascade	0.00011	0.20906
Effects of PIP2 hydrolysis	0.00012	0.14898
A tetrasaccharide linker sequence is required for GAG synthesis	0.00012	0.29766
PI3K/AKT activation	0.00013	0.24486
GAB1 signalosome	0.00013	0.4648
Diseases associated with glycosaminoglycan metabolism	0.00013	0.050121
Diseases of glycosylation	0.00013	0.050121
Heparan sulfate/heparin (HS-GAG) metabolism	0.00016	0.19
HS-GAG biosynthesis	0.00016	0.29681
Integrin alphaIIb beta3 signaling	0.00016	0.63007
Interferon gamma signaling	0.00018	0.43088
Gastrin-CREB signalling pathway via PKC and MAPK	0.00018	0.77958
Chemokine receptors bind chemokines	0.00023	0.62702
Downstream signal transduction	0.00027	0.54921
Platelet homeostasis	0.00029	0.24577
IRS-mediated signalling	0.00029	0.31766
<i>Gαi</i> signalling events	0.00029	$< 2.749 \times 10^{-4}$
Diseases of signal transduction	0.00029	0.65733
Signaling by activated point mutants of FGFR1	0.00029	0.24892
FGFR1c ligand binding and activation	0.00029	0.24892
Signaling by NOTCH3	0.00029	0.017419

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).

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

Pathway	ID	Observed	Expected	χ^2 value	p-value	p-value (FDR)
Prostacyclin signalling through prostacyclin receptor	392851	1	10.07	26.53	1.7307×10^{-6}	0.0028590
Cell surface interactions at the vascular wall	202733	3	10.07	21.11	2.6107×10^{-5}	0.00642330
The NLRP1 inflammasome	844455	3	10.07	21.11	2.6107×10^{-5}	0.00642330
Innate Immune System	168249	6	10.07	21.07	2.6548×10^{-5}	0.00642330
Keratan sulfatekeratin metabolism	1638074	4	10.07	20.65	3.2861×10^{-5}	0.00642330
Keratan sulfate biosynthesis	2022854	4	10.07	20.65	3.2861×10^{-5}	0.00642330
Signaling by SCF-KIT	1433557	5	10.07	20.64	3.3045×10^{-5}	0.00642330
VEGFA-VEGFR2 Pathway	4420097	5	10.07	20.64	3.3045×10^{-5}	0.00642330
ERK1 activation	110056	21	10.07	20.12	4.277×10^{-5}	0.00642330
Cholesterol biosynthesis	191273	21	10.07	20.12	4.277×10^{-5}	0.00642330
G2 Phase	68911	21	10.07	20.12	4.277×10^{-5}	0.00642330
p130Cas linkage to MAPK signaling for integrins	372708	2	10.07	19.08	7.1872×10^{-5}	0.00651340
cGMP effects	418457	8	10.07	19.01	7.4597×10^{-5}	0.00651340
Regulation of cytoskeletal remodeling and cell spreading by IPP complex components	446388	8	10.07	19.01	7.4597×10^{-5}	0.00651340
Post-translational modification: synthesis of GPI-anchored proteins	163125	20	10.07	18.59	9.1878×10^{-5}	0.00651340
Fcgamma receptor (FCGR) dependent phagocytosis	2029480	3	10.07	17.95	0.00012676	0.00651340
A third proteolytic cleavage releases NICD	157212	7	10.07	17.90	0.00012995	0.00651340
Signalling by NGF	166520	7	10.07	17.90	0.00012995	0.00651340
Signaling by VEGF	194138	7	10.07	17.90	0.00012995	0.00651340
Regulation of thyroid hormone activity	350864	7	10.07	17.90	0.00012995	0.00651340
Nitric oxide stimulates guanylate cyclase	392154	7	10.07	17.90	0.00012995	0.00651340
Platelet homeostasis	418346	7	10.07	17.90	0.00012995	0.00651340
Termination of translesion DNA synthesis	5656169	20	10.07	17.46	0.00016155	0.00651340
PI3K events in ERBB4 signaling	1250342	4	10.07	17.26	0.00017862	0.00651340
PIP3 activates AKT signaling	1257604	4	10.07	17.26	0.00017862	0.00651340

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