## Contents

[Glossary](#Q1-1-4)   
[Acronyms](#Q1-1-6)   
4 [Synthetic Lethal Analysis of Gene Expression Data](#x1-60004)   
 4.1 [Synthetic lethal genes in breast cancer](#x1-70001)   
  4.1.1 [Synthetic lethal pathways in breast cancer](#x1-80001)   
  4.1.2 [Expression profiles of synthetic lethal partners](#x1-90002)   
   4.1.2.1 [Subgroup pathway analysis](#x1-100001)   
 4.2 [Comparison of synthetic lethal gene candidates](#x1-110002)   
  4.2.1 [Comparison with siRNA screen candidates](#x1-120001)   
   4.2.1.1 [Comparison with correlation](#x1-130001)   
   4.2.1.2 [Comparison with viability](#x1-140002)   
   4.2.1.3 [Comparison with secondary siRNA screen candidates](#x1-150003)   
   4.2.1.4 [Comparison of screen at pathway level](#x1-160004)   
 4.3 [Metagene Analysis](#x1-180003)   
  4.3.1 [Pathway expression](#x1-190001)   
  4.3.2 [Somatic mutation](#x1-200002)   
  4.3.3 [Mutation locus](#x1-210003)   
  4.3.4 [Synthetic lethal metagenes](#x1-220004)   
 4.4 [Replication in stomach cancer](#x1-230004)   
  4.4.1 [Synthetic Lethal Genes and Pathways](#x1-240001)   
  4.4.2 [Synthetic Lethal Expression Profiles](#x1-250002)   
  4.4.3 [Comparison to Primary Screen](#x1-260003)   
   4.4.3.1 [Resampling Analysis](#x1-270001)   
  4.4.4 [Metagene Analysis](#x1-280004)   
 4.5 [Global Synthetic Lethality](#x1-290005)   
  4.5.1 [Hub Genes](#x1-300001)   
  4.5.2 [Hub Pathways](#x1-310002)   
 4.6 [Replication in cell line encyclopaedia](#x1-320006)   
 4.7 [Discussion](#x1-330007)   
  4.7.1 [Strengths of the SLIPT Methodology](#x1-340001)   
  4.7.2 [Syntheic Lethal Pathways for E-cadherin](#x1-350002)   
  4.7.3 [Replication and Validation](#x1-360003)   
   4.7.3.1 [Integration with siRNA Screening](#x1-370001)   
   4.7.3.2 [Replication across Tissues and Cell lines](#x1-380002)   
 4.8 [Summary](#x1-390008)   
   [Aims](#x1-400008)   
   [Summary](#x1-410008)

[References](#Q1-1-80)

## List of Figures

4.1 [Synthetic lethal expression profiles of analysed samples](#x1-90011)  
4.2 [Comparison of SLIPT to siRNA](#x1-120012)  
4.3 [Compare SLIPT and siRNA genes with correlation](#x1-130013)  
4.4 [Compare SLIPT and siRNA genes with correlation](#x1-130024)  
4.5 [Compare SLIPT and siRNA genes with siRNA viability](#x1-140015)  
4.6 [Compare SLIPT and siRNA genes with viability](#x1-140026)  
4.7 [Compare SLIPT and siRNA genes with siRNA viability](#x1-140037)  
4.8 [Resampled intersection of SLIPT and siRNA candidates](#x1-170018)  
4.9 [Pathway metagene expression profiles](#x1-190019)  
a [PIK3CA mutation](#x1-20001r1)  
b [PIK3CA or PIK3R1 mutation](#x1-20002r2)  
c [CDH1 mutation](#x1-20003r3)  
d [TP53 mutation](#x1-20004r4)  
4.11 [Somatic mutation against PI3K metagene](#x1-2000511)  
a [PI3KCA gene](#x1-21001r1)  
b [PI3KCA metagene](#x1-21002r2)  
c [PIK3R1 gene](#x1-21003r3)  
d [PIK3CA protein](#x1-21004r4)  
e [AKT protein](#x1-21005r5)  
f [CDH1 protein](#x1-21006r6)  
g [TP53 protein](#x1-21007r7)  
4.13 [Somatic mutation locus against expression](#x1-2100813)  
4.14 [Synthetic lethal expression profiles of stomach samples](#x1-2500114)  
4.15 [Synthetic lethal partners across query genes](#x1-2900115)

## List of Tables

4.1 [Candidate synthetic lethal gene partners of CDH1 from SLIPT](#x1-70011)  
4.2 [Pathways for CDH1 partners from SLIPT](#x1-80012)  
4.3 [Pathway composition for clusters of CDH1 partners from SLIPT](#x1-100013)  
4.4 [Pathway composition for CDH1 partners from SLIPT and siRNA screening](#x1-160014)  
4.5 [Pathways for CDH1 partners from SLIPT](#x1-170025)  
4.6 [Pathways for CDH1 partners from SLIPT and siRNA primary screen](#x1-170036)  
4.7 [Candidate synthetic lethal metagenes against CDH1 from SLIPT](#x1-220017)  
4.8 [Pathways for CDH1 partners from SLIPT in stomach cancer](#x1-240018)  
4.9 [Query synthetic lethal genes with the most SLIPT partners](#x1-300019)  
4.10 [Pathways for genes with the most SLIPT partners](#x1-3100110)  
4.11 [Pathways for CDH1 partners from SLIPT in CCLE](#x1-3200111)  
4.12 [Pathways for CDH1 partners from SLIPT in breast CCLE](#x1-3200212)

## Glossary

synthetic lethal

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

## Acronyms

ANOVA

Analysis of Variance.

siRNA

Short interfering ribonucleic acid.

SLIPT

Synthetic lethal interaction prediction tool.

## Chapter 4 Synthetic Lethal Analysis of Gene Expression Data

Having developed a statistical synthetic lethal detection methodology ([Synthetic Lethal Interaction Prediction Tool (SLIPT)](#glo: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 ([TCGA](#XTCGA2012), [2012](#XTCGA2012)) which covers a range of clinical subtypes and is more closely modelled by [short interfering ribonucleic acid (siRNA)](#glo:siRNA) data ([Telford *et al.*](#XTelford2015), [2015](#XTelford2015)) 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) ([Barretina *et al.*](#XBarretina2012), [2012](#XBarretina2012)) 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](#glo: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.*](#XTelford2015) ([2015](#XTelford2015)) on isogenic cell lines. For comparison, an alternative [SLIPT](#glo: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.*](#XChen2014), [2014](#XChen2014)). Pathways will be analysed by over-representation analysis (with resampling for comparisons with [siRNA](#glo: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](#glo:SLIPT) directly on metagenes to detect synthetic lethal pathways.

Together these results will demonstrate the wide range of applications for [SLIPT](#glo: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](#glo: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](#x1-60005).

Some of the findings presented in this Chapter have also been included in manuscripts submitted for publication ([Kelly *et al.*](#XKellyBMC), [2017a](#XKellyBMC),[b](#XKellyHDGC)) 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](#glo:SLIPT) methodology (as described in Section 3.1) was applied to the normalised TCGA breast cancer gene expression dataset (n = 1168). As shown in Table [4.1](#x1-70011), the most significant genes had strong evidence of expression-based association with CDH1 (high χ2 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 neccessary 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 cruicial 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](#glo:SLIPT) candidate genes. Furthermore, few known biological functions of CDH1 were among the strongest SL candidates so the remaining candidate genes may also be informative since they are likely to contain these expected functions in addition to novel relationships for CDH1. Thus further pathway level analyses were also conducted to examine biological functions over-represented among synthetic candidate genes and identify synthetic lethal pathways.

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

Gene

Observed

Expected

χ2 value

p-value

p-value (FDR)

TRIP10

62

130

162

5.65 × 10-34

1.84 × 10-31

EEF1B2

56

130

158

3.10 × 10-33

9.45 × 10-31

GBGT1

61

131

156

1.08 × 10-32

3.14 × 10-30

ELN

81

130

149

3.46 × 10-31

8.82 × 10-29

TSPAN4

78

130

146

1.63 × 10-30

3.79 × 10-28

GLIPR2

72

130

146

1.68 × 10-30

3.86 × 10-28

RPS20

73

131

145

1.89 × 10-30

4.28 × 10-28

RPS27A

80

130

143

5.53 × 10-30

1.18 × 10-27

EEF1A1P9

63

130

141

1.91 × 10-29

3.74 × 10-27

C1R

73

130

141

2.05 × 10-29

3.97 × 10-27

LYL1

73

130

140

2.99 × 10-29

5.74 × 10-27

RPLP2

71

130

139

4.88 × 10-29

9.07 × 10-27

C10orf10

73

130

138

6.72 × 10-29

1.20 × 10-26

DULLARD

74

131

138

9.29 × 10-29

1.61 × 10-26

PPM1F

64

130

136

1.61 × 10-28

2.65 × 10-26

OBFC2A

69

130

136

2.49 × 10-28

3.93 × 10-26

RPL11

70

130

136

2.56 × 10-28

3.97 × 10-26

RPL18A

70

130

135

3.08 × 10-28

4.70 × 10-26

MFNG

76

131

133

7.73 × 10-28

1.12 × 10-25

RPS17

77

131

133

8.94 × 10-28

1.29 × 10-25

MGAT1

73

130

132

1.44 × 10-27

2.03 × 10-25

RPS12

72

130

128

8.57 × 10-27

1.12 × 10-24

C10orf54

73

130

127

1.37 × 10-26

1.75 × 10-24

LOC286367

72

130

126

2.20 × 10-26

2.70 × 10-24

GMFG

70

130

126

2.20 × 10-26

2.70 × 10-24

Strongest candidate SL partners for CDH1 by [SLIPT](#glo:SLIPT) with observed and expected samples with low expression of both genes

The modified mtSLIPT methodology (as described in Section 3.1) was also applied to the normalised TCGA breast cancer gene expression dataset, against somatic loss of function mutations in CDH1. As shown in Table D.1, the most significant genes also had strong evidence of expression associated with CDH1 mutations (high χ2 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](#x1-70011)) nor were as many genes detected. This is unsurprising due to the lower sample size with matching somatic mutation data and the lower frequency of CDH1 mutations compared to low expression by 1 ∕ 3 quantiles.

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

#### 4.1.1 Synthetic lethal pathways in breast cancer

Translational pathways were strongly over-represented in [SLIPT](#glo:SLIPT) partners, as shown in Table [4.2](#x1-80012). These include ribosomal subunits, initiation, peptide elongation, and termination. Regulatory processes involving mRNA including 3’ untranslated region (UTR) binding, L13a-mediated translational silencing, and nonsense-mediated decay were also implicated. These are consistent with protein translation being subject to “non-oncogene addiction” ([Luo *et al.*](#XLuo2009), [2009](#XLuo2009)), as a core process that is dysregulated to sustain cancer proliferation and survival ([Gao and Roux](#XGao2015), [2015](#XGao2015)).

Immune pathways, including the adaptive immune system and responses to infectious diseases were also strongly implicated as synthetic lethal with loss of E-cadherin. This is consistent with the alterations of immune response being a hallmark of cancer [Hanahan and Weinberg](#XHanahan2000) ([2000](#XHanahan2000)), 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.*](#XTelford2015), [2015](#XTelford2015)) and the

differences between to findings in patient data will be described in more detail in Section [4.2.1.4](#x1-160004).

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](#glo:SLIPT) partners for CDH1

It is also notable that the pathways over-represented in [SLIPT](#glo:SLIPT) candidate genes have strongly significant over-representation of Reactome pathways from the hypergeometric test (as described in Section [2.3.2](#x1-180002)). 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 D.2). This analysis more closely represents the null CDH1 mutations in HDGC ([Guilford *et al.*](#XGuilford1998), [1998](#XGuilford1998)) and the experimental MCF10A cell model ([Chen *et al.*](#XChen2014), [2014](#XChen2014)). 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.*](#XTelford2015) ([2015](#XTelford2015)) 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](#x1-90011) (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.*](#XTelford2015) ([2015](#XTelford2015)) were also identified by this approach. In addition, we identified novel gene candidates, which had little effect on viability in isogenic cell line experiments.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

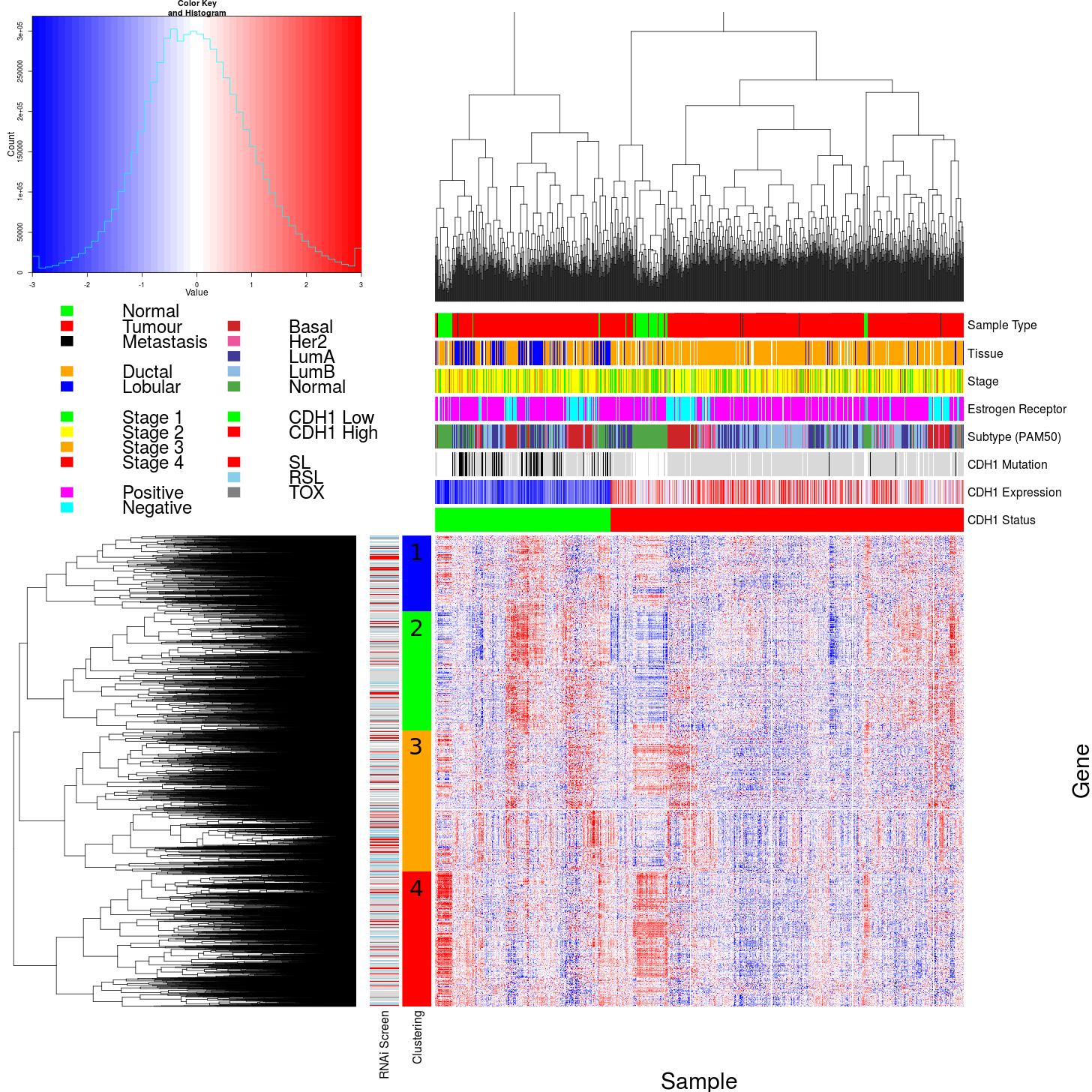


Figure 4.1: **Synthetic lethal expression profiles of analysed samples.** Gene expression profile heatmap (correlation distance) of all samples (separated by the 1∕3 quantile of CDH1 expression) analysed in TCGA breast cancer dataset for gene expression of 5,165 candidate partners of E-cadherin (CDH1) from [SLIPT](#glo: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](#glo:siRNA) screen [Telford *et al.*](#XTelford2015) ([2015](#XTelford2015)). 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.

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.*](#XDai2015) ([2015](#XDai2015)); [Eroles *et al.*](#XEroles2012) ([2012](#XEroles2012)); [Parker *et al.*](#XParker2009) ([2009](#XParker2009)) 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 D.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](#glo:SLIPT)) as it requires highly expressing CDH1 samples for comparison.

The CDH1 mutant samples (in Figure [4.1](#x1-90011)) were predominantly among the CDH1 lowly expressing samples and distributed throughout CDH1 samples with clustering analysis. Thus the molecular profiles of CDH1 low samples a 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 D.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 [6.1](#x1-70001)). 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](#x1-90011) or D.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

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/1 (Rhodopsin-like receptors)

258

50

6.7 × 10-41

Phosphorylation of CD3 and TCR zeta chains

18

14

1.3 × 10-40

Interferon gamma signalling

74

24

5.0 × 10-39

GPCR ligand binding

326

57

1.8 × 10-38

Cytokine Signaling in Immune system

268

48

8.9 × 10-37

Downstream TCR signalling

45

18

1.8 × 10-35

Gαi signalling events

167

33

2.2 × 10-33

Cell surface interactions at the vascular wall

99

21

1.3 × 10-26

Interferon Signalling

164

28

1.7 × 10-26

Extracellular matrix organisation

238

35

2.7 × 10-25

Antigen activates B Cell Receptor leading to generation of second messengers

32

12

7.2 × 10-25

**Pathways Over-represented in Cluster 4**

**Pathway Size**

**Cluster Genes**

**p-value (FDR)**

Extracellular matrix organisation

238

48

8.0 × 10-41

Class A/1 (Rhodopsin-like receptors)

258

47

2.8 × 10-36

GPCR ligand binding

326

54

2.1 × 10-34

Gαs 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 Ca2+

82

19

3.0 × 10-26

Developmental Biology

420

57

3.5 × 10-26

Gαi 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

Gαq 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

Synthetic lethal gene candidates for CDH1 from [SLIPT](#glo:SLIPT) performed on RNA-Seq expression data were also used for pathway over-representation analyses (as described in Section [2.3.2](#x1-180002)). 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.*](#XParker2009), [2009](#XParker2009)) shown in Figure [4.1](#x1-90011).

As shown by the most over-represented pathways in Table [4.3](#x1-100013), each correlated cluster of candidate synthetic lethal partners of CDH1 contains functionally different genes. Cluster 1 contains genes with less evidence of over-represented pathways than other clusters, corresponding to less correlation between genes within the cluster, and to it being a relatively small group. While there is some indication that collagen biosynthesis, microfibril elastic fibres, extracellular matrix, and metabolic pathways may be over-represented in Cluster 1, these results are mainly based on small pathways containing few synthetic lethal genes. Genes in Cluster 2 exhibited low expression in normal tissue samples compared to tumour samples (see Figure [4.1](#x1-90011)) 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 Gαi 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](#glo:SLIPT) partners with respect to CDH1 mutation, shown in Table D.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 GCPR 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](#x1-100013) and D.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](#glo:SLIPT) in TCGA breast cancer data) and experimental (the primary [siRNA](#glo:siRNA) screen performed by [Telford *et al.*](#XTelford2015) ([2015](#XTelford2015))) approaches in Figure [4.2](#x1-120012). 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 exlcuded 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](#x1-120012)).

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

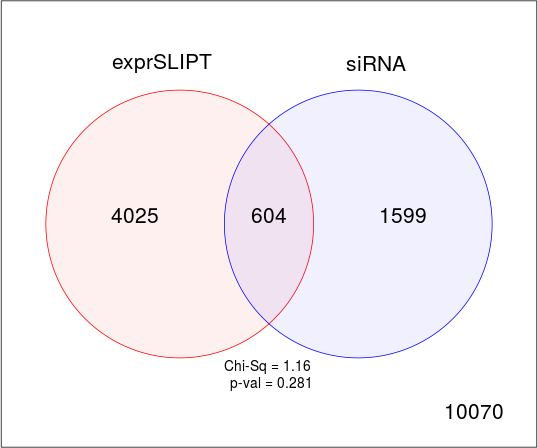


Figure 4.2: **Comparison of** [**SLIPT**](#glo:SLIPT) **to** [**siRNA**](#glo:siRNA)**.** Testing the overlap of gene candidates for E-cadherin synthetic lethal partners between computational ([SLIPT](#glo:SLIPT)) and experimental screening ([siRNA](#glo:siRNA)) approaches. The χ2 test suggests that the overlap is no more than would be expected by chance (p = 0.281).

##### 4.2.1.1 Comparison with correlation

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

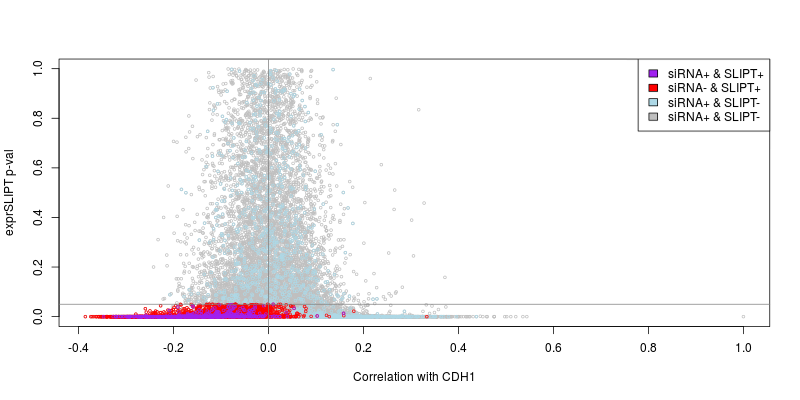


Figure 4.3: **Compare** [**SLIPT**](#glo:SLIPT) **and** [**siRNA**](#glo:siRNA) **genes with correlation.** The χ2 p-values for genes tested by [SLIPT](#glo:SLIPT) (in TCGA breast cancer) expression analysis were compared against Pearson’s correlation of gene expression with CDH1. Genes detected by [SLIPT](#glo:SLIPT) or [siRNA](#glo:siRNA) are coloured according to the legend.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

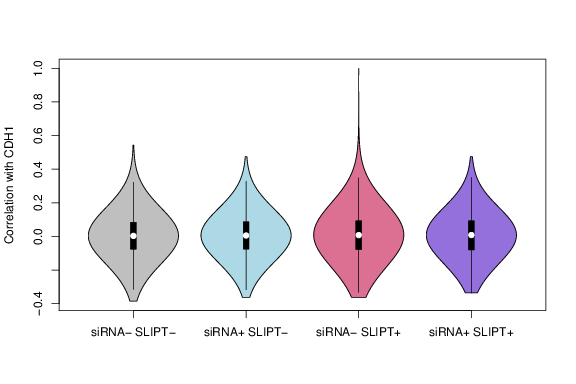


Figure 4.4: **Compare** [**SLIPT**](#glo:SLIPT) **and** [**siRNA**](#glo:siRNA) **genes with correlation.** Genes detected as candidate synthetic lethal partners by [SLIPT](#glo:SLIPT) (in TCGA breast cancer) expression analysis and experimental screening (with [siRNA](#glo:siRNA)) were compared against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

Another potential means to triage drug target candidates is correlation of expression profiles with CDH1. Correlation with CDH1 was compared to [SLIPT](#glo:SLIPT) and [siRNA](#glo:siRNA) results in Figure [4.3](#x1-130013). The genes not detected by [SLIPT](#glo:SLIPT) (including [siRNA](#glo:siRNA) candidates) included genes with high (insignificant) [SLIPT](#glo: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](#glo:SLIPT) p-values, although there were exceptions to this trend and larger positive correlations were negative correations.

The majority of [SLIPT](#glo: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 [6.1](#x1-70001).

There were not strong postive correlations with CDH1 among [siRNA](#glo:siRNA) candidates, consistent with previous findings that co-expression is not predictive of synthetic lethality ([Jerby-Arnon *et al.*](#XJerby2014), [2014](#XJerby2014); [Lu *et al.*](#XLu2015), [2015](#XLu2015)). Negative correlation may not be indicative of synthetic lethality either as many [siRNA](#glo:siRNA) candidates also had positive correlations. The [SLIPT](#glo: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 D.3), although the range of the χ2 p-values differ due to lower sample size for mutation analysis.

However, the apparent tendancy for genes detected by [SLIPT](#glo:SLIPT) or [siRNA](#glo: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](#x1-130024) and D.4). Therefore further triage of gene candidates by correaltion is not suitable, nor is use of correlation itself to preduct synthetic lethal partners in the first place.

##### 4.2.1.2 Comparison with viability

A similar comparison of [SLIPT](#glo:SLIPT) results was made with the viability ratio (of CDH1 mutant to wildtype) in the primary [siRNA](#glo:siRNA) screen performed by [Telford *et al.*](#XTelford2015) ([2015](#XTelford2015)). The significance and viability thresholds used for [SLIPT](#glo:SLIPT) and [siRNA](#glo:siRNA) detection of synthetic lethal candidate partners of CDH1 are clear in Figure [4.5](#x1-140015). 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](#glo:SLIPT) (see Section 3.1) and minimum wildtype viability for [siRNA](#glo:siRNA) ([Telford *et al.*](#XTelford2015), [2015](#XTelford2015)).

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

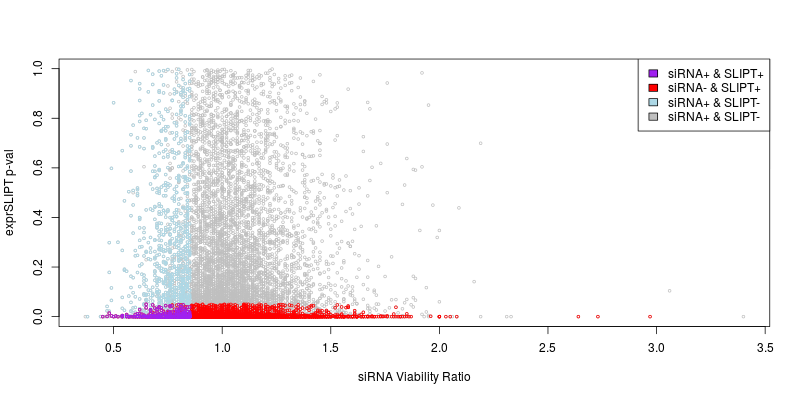


Figure 4.5: **Compare** [**SLIPT**](#glo:SLIPT) **and** [**siRNA**](#glo:siRNA) **genes with** [**siRNA**](#glo:siRNA) **viability.** The χ2 p-values for genes tested by [SLIPT](#glo:SLIPT) (in TCGA breast cancer) expression analysis were compared against the viability ratio of CDH1 mutant and wildtype cells in the primary [siRNA](#glo:siRNA) screen. Genes detected by [SLIPT](#glo:SLIPT) or [siRNA](#glo:siRNA) are coloured according to the legend.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

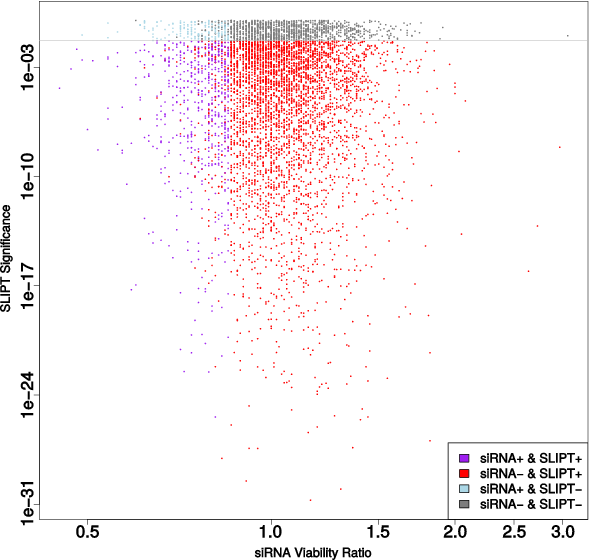


Figure 4.6: **Compare** [**SLIPT**](#glo:SLIPT) **and** [**siRNA**](#glo:siRNA) **genes with viability.** The χ2 p-values for genes tested by [SLIPT](#glo:SLIPT) (in TCGA breast cancer) expression analysis were compared (on a log-scale) against the viability ratio of CDH1 mutant and wildtype cells in the primary [siRNA](#glo:siRNA) screen. Genes detected by [SLIPT](#glo:SLIPT) or [siRNA](#glo: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](#glo:SLIPT) and [siRNA](#glo:siRNA) candidates. Many genes not detected by both approaches were numerous in Figures [4.2](#x1-120012) and D.2. These genes detected by either are not neccessarily near the thresholds for the other. In this respect the [SLIPT](#glo: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 neccessarily more strongly supported by either, the genes with a viability closer to 1 (no synthetic lethal effect) in [siRNA](#glo:siRNA) included those with more significant [SLIPT](#glo:SLIPT) p-values whereas more extreme viability ratios tended to be less significant (as shown by a logarithmic plot in Figure [4.6](#x1-140026)). Although it should be noted that genes with more moderate viability ratios were more common and [SLIPT](#glo: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](#glo:SLIPT) candidates or those detected by both approaches having considerably different viability ratios (as shown in Figures [4.7](#x1-140037) and D.5). The difference between the gene groups stems largely from the viability thresholds used by [Telford *et al.*](#XTelford2015) ([2015](#XTelford2015)) to detect synthetic lethal candidates in the primary screen, rather than more extreme viability ratios for genes identified by [SLIPT](#glo:SLIPT).

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

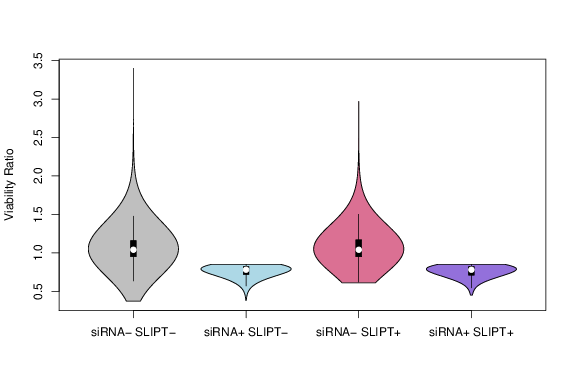


Figure 4.7: **Compare** [**SLIPT**](#glo:SLIPT) **and** [**siRNA**](#glo:siRNA) **genes with** [**siRNA**](#glo:siRNA) **viability.** Genes detected as candidate synthetic lethal partners by [SLIPT](#glo:SLIPT) (in TCGA breast cancer) expression analysis and experimental screening (with [siRNA](#glo:siRNA)) were compared against the viability ratio of CDH1 mutant and wildtype cells in the primary [siRNA](#glo:siRNA) screen. There were clear no differences in viability between genes detected by [SLIPT](#glo:SLIPT) and those not with the differences being primarily due to viability thresholds being used to detect synthetic lethality by [Telford *et al.*](#XTelford2015) ([2015](#XTelford2015)).

##### 4.2.1.3 Comparison with secondary 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](#glo:siRNA)) 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](#glo:SLIPT) results from gene expression and somatic mutation analysis. A secondary screen was also performed by [Telford *et al.*](#XTelford2015) ([2015](#XTelford2015)) on the isogenic MCF10A breast cell lines to individually validate the [siRNAs](#glo:siRNA) separately, with the strongest candidates being those exhibiting synthetic lethal viability ratios replicated across independently targeting [siRNAs](#glo:siRNA). 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](#glo:SLIPT) in breast cancer (and the 486 genes tested by [SLIPT](#glo:SLIPT) in stomach cancer).

The secondary screen results are given in Appendix C which show that [SLIPT](#glo:SLIPT) candidate genes are more significantly (p = 7.49 × 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 the individual genes detected by either approach do not neccessarily 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](#glo: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

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

Predicted only by [SLIPT](#glo: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

Gαq signalling events

159

26

6.7 × 10-23

Gastrin-CREB signalling pathway via PKC and MAPK

180

27

2.0 × 10-21

Gαi 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](#glo:SLIPT) and siRNA screen (604 genes)

Pathway Size

Genes Identified

p-value (FDR)

Visual phototransduction

54

9

6.9 × 10-10

Gαs 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 P450 - 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 Ca2+

66

6

4.2 × 10-5

These pathway over-representation analyses (performed as described in Section [2.3.2](#x1-180002)) correspond to genes separated into [SLIPT](#glo:SLIPT) or siRNA screen candidates unique to either method or detected by both (Table [4.4](#x1-160014)). The [SLIPT](#glo: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](#x1-80012) and in the clustering analysis (Table [4.3](#x1-100013)). 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.*](#XTelford2015) ([2015](#XTelford2015)). 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 × 10-10) and Gαs signalling events (p = 1.7 × 10-7), than other signalling pathways.

The pathway analysis for mtSLIPT against CDH1 mutations (in Table D.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](#glo:SLIPT) against low CDH1 expression, signalling pathways including GPCRs, NOTCH, EERB2, PDGF, and SCF-KIT. These findings indicate the signalling pathways are among the most suitable vulnerability to exploit in targeting CDH1 deficient tumours as they can be detected in both a patient cohort (with TCGA expression data) and tested in a laboratory system. However, it is possible that the isolated experimental system is set up to preferentially detect kinase singalling pathways (which are amenable to pharmacological inhibition and translation to the clinic) and the other pathways identified by [SLIPT](#glo: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.

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 more robustly conserved ([Dixon *et al.*](#XDixon2008), [2008](#XDixon2008)) and the

computational approach using patient samples from complex tumour microenvironment has considerably different strengths to an experimental screen ([Telford *et al.*](#XTelford2015), [2015](#XTelford2015)) 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.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

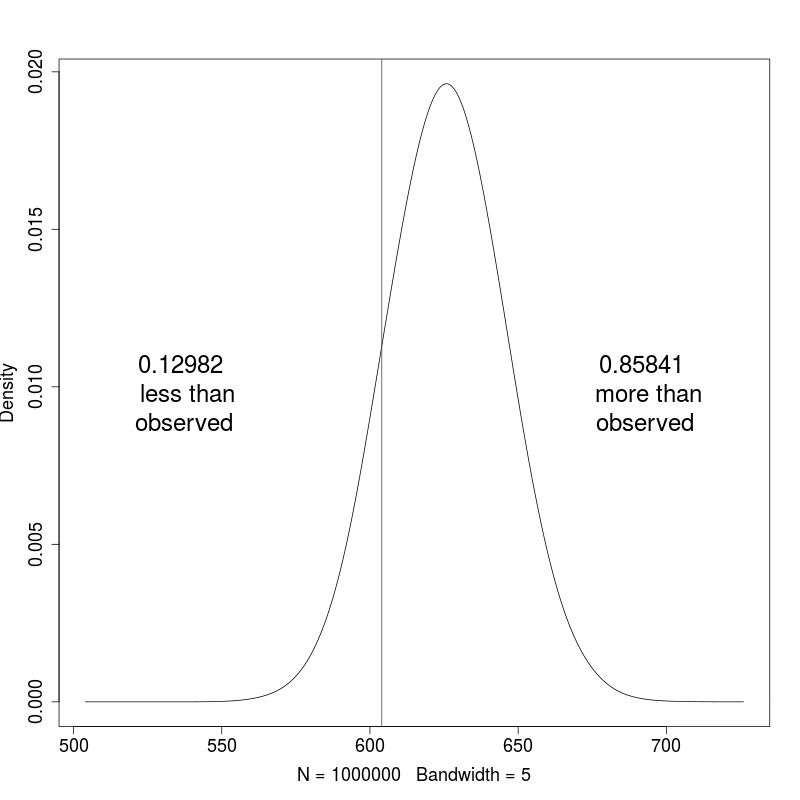


Figure 4.8: **Resampled intersection of** [**SLIPT**](#glo:SLIPT) **and siRNA candidates.** Resampling analysis of intersect size from genes detected by [SLIPT](#glo:SLIPT) and siRNA screening approaches over 1 million replicates. The proportion of expected intersection sizes for random samples below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

The overlap between synthetic lethal from bioinformatics [SLIPT](#glo:SLIPT) predictions and siRNA screening has raised other questions including whether the pathways over-represented would be expected by chance. This of particular concern since the siRNA candidate genes themselves are highly over-represented for particular pathways (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](#glo:SLIPT) candidates and their intersection with the siRNA hits shows whether we detect these pathways more than we expect by chance (as described in Section [2.3.6](#x1-230006)).

Of particular concern are the over-represented pathways in genes detected by both methods. Pathway over-representation alone does not detect whether [SLIPT](#glo:SLIPT) predicted genes or siRNA candidates are enriched within each other. This resampling analysis therefore detects whether over-represented pathways were detected by [SLIPT](#glo: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](#x1-170018), 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. This is consistent with previous findings (see Figure [4.2](#x1-120012)) and does not preclude pathway relationships being supported by resampling.

A permutation analysis was performed to resample the genes tested by both approaches to investigate whether the observed pathway over-representation could have occurred in a randomly selected sample of genes from the experimental candidates, that is, whether the pathway predictions from [SLIPT](#glo:SLIPT) could be expected by chance (as described in sections [2.2.3.1](#x1-150001) and [2.3.6](#x1-230006)). While the number of siRNA candidate genes detected by [SLIPT](#glo: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](#x1-160014). The resampling analysis for pathways was compared to the pathway over-representation for [SLIPT](#glo:SLIPT) predicted synthetic lethal partners in Table [4.5](#x1-170025). Similarly, the pathway resampling

for intersection between [SLIPT](#glo:SLIPT) predictions and experimental screen candidates was compared to pathway over-representation in Table [4.6](#x1-170036) for intersection with siRNA data.

The pathway resampling approach for [SLIPT](#glo:SLIPT)-specific gene candidates (Table [4.5](#x1-170025)) replicates the gene set over-representation analysis for all [SLIPT](#glo:SLIPT) genes, detecting evidence of synthetic lethal pathways for CDH1 in translational, immune, and cell signalling pathways including Gαi 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](#glo:SLIPT)-specific pathways in Table [4.5](#x1-170025). In particular, some of the most significantly over-represented pathways had higher observed χ2 values than any of the 1 million random permutations. Similar pathways were also replicated by permutation analysis for mt[SLIPT](#glo:SLIPT) candidate partners against CDH1 mutation (shown in Table D.5). This shows that many of the pathways detected specfically by [SLIPT](#glo:SLIPT) are replicated by permutation procedures and that the permutation approach is capable of detecting many of the most strongly over-represented pathways.

Table 4.5: Pathways for CDH1 partners from SLIPT

Reactome Pathway

Over-representation

Permutation

**Eukaryotic Translation Elongation**

1.3 × 10-207

< 1.241 × 10-5

**Peptide chain elongation**

5.6 × 10-201

< 1.241 × 10-5

**Viral mRNA Translation**

1.2 × 10-196

< 1.241 × 10-5

**Eukaryotic Translation Termination**

1.2 × 10-196

< 1.241 × 10-5

**Formation of a pool of free 40S subunits**

3.7 × 10-194

< 1.241 × 10-5

**Nonsense Mediated Decay independent of the Exon Junction Complex**

5.3 × 10-187

< 1.241 × 10-5

**L13a-mediated translational silencing of Ceruloplasmin expression**

9.6 × 10-183

< 1.241 × 10-5

**3’ -UTR-mediated translational regulation**

9.6 × 10-183

< 1.241 × 10-5

**GTP hydrolysis and joining of the 60S ribosomal subunit**

1.9 × 10-181

< 1.241 × 10-5

**Nonsense-Mediated Decay**

6.2 × 10-176

< 1.241 × 10-5

**Nonsense Mediated Decay enhanced by the Exon Junction Complex**

6.2 × 10-176

< 1.241 × 10-5

Adaptive Immune System

6.5 × 10-174

0.15753

**Eukaryotic Translation Initiation**

5.7 × 10-173

< 1.241 × 10-5

**Cap-dependent Translation Initiation**

5.7 × 10-173

< 1.241 × 10-5

**SRP-dependent cotranslational protein targeting to membrane**

2.0 × 10-171

< 1.241 × 10-5

**Translation**

6.1 × 10-170

< 1.241 × 10-5

Infectious disease

1.6 × 10-166

0.23231

**Influenza Infection**

1.9 × 10-163

< 1.241 × 10-5

**Influenza Viral RNA Transcription and Replication**

1.9 × 10-160

< 1.241 × 10-5

**Influenza Life Cycle**

2.5 × 10-156

< 1.241 × 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

**Gαi signalling events**

8.9 × 10-100

< 1.241 × 10-5

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

The permutation approach was then also applied to the intersection between computational and experimental candidates. Where the permutation anlaysis is testing for consistent detection of pathways independent of their pre-existing status as experimental candidates. The pathway results for these candidate partners (in Table [4.6](#x1-170036)) 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 Gαs signalling events were supported by both over-representation and resampling analyses. Other pathways supported by both analyses were cytoplasmic elastic fibre formation, associated HS-GAG protein modification pathways, energy metabolism, and the fibrin clotting cascade.

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

**Gαs signalling events**

1.6 × 10-7

0.012988

Retinoid metabolism and transport

1.7 × 10-7

0.20487

Transcriptional regulation of white adipocyte differentiation

6.5 × 10-6

0.38197

Acyl chain remodelling of PS

6.5 × 10-6

0.58485

Chemokine receptors bind chemokines

6.5 × 10-6

0.97255

Defective EXT2 causes exostoses 2

6.9 × 10-6

0.056437

Defective EXT1 causes exostoses 1, TRPS2 and CHDS

6.9 × 10-6

0.056437

Signalling by NOTCH4

6.9 × 10-6

0.15497

Platelet activation, signalling and aggregation

6.9 × 10-6

0.53358

Phase 1 - Functionalisation of compounds

1.3 × 10-5

0.24836

Amine ligand-binding receptors

1.7 × 10-5

0.3195

Acyl chain remodelling of PE

2.4 × 10-5

0.7307

Signalling by GPCR

2.4 × 10-5

0.9939

**Molecules associated with elastic fibres**

2.6 × 10-5

0.0072929

DAP12 interactions

2.6 × 10-5

0.78273

Cytochrome P450 - arranged by substrate type

3.2 × 10-5

0.87019

GPCR ligand binding

3.8 × 10-5

0.99417

Acyl chain remodelling of PC

4.0 × 10-5

0.65415

Response to elevated platelet cytosolic Ca2+

4.2 × 10-5

0.55461

Arachidonic acid metabolism

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).

Many of the pathways supported in the intersection by permutation analysis were also replicated in the mtSLIPT anlaysis of partners tested with CDH1 mutation (in Table D.6), including Gαs, elastic fibres, HS-GAG, and energy metabolism. While there were differences between the pathways Identified by over-representation anlaysis, those replicated by permutation were highly concordant supported the combined use of these pathways approaches to identify synthetic lethal gene functions and targets.

While this indicates that Gαs 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](#glo:SLIPT), which may be driven by these more specific constituent pathways.

However, several pathways, including some immune functions and neurotransmitters, were supported by the resampling analysis (in Tables [4.6](#x1-170036) and D.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.

Therefore computational and experimental approaches to synthetic lethal screening for CDH1 lead to a broader functional characterisation and many candidate partners, when combined, despite different strengths and limitations. Compared to candidate gene approaches, experimental genome-wide screens are an appealing unbiased strategy for identifying synthetic lethal interactions. Since these screens are costly, laborious, and subject to genetic background, computational analysis can augment candidate triage to either reduce the initial panel of screened genes or prioritise validation.

GPCR pathways were detected among both computational and experimental synthetic lethal candidates, with more support in the experimental screen (Table [4.6](#x1-170036)). The homogeneous cell line model may be more likely to detect particular pathways. For instance, [SLIPT](#glo:SLIPT) identified immune pathways, not expected to be detected in isolated cell culture. GPCR signalling was supported in experimental models [Telford *et al.*](#XTelford2015) ([2015](#XTelford2015)) with some of these pathways replicated in varied genetic

backgrounds of patient samples. These pathways require further investigation such as identification of more specific pathways, higher order interactions, and modes of resistance.

The pathway composition across computational and experimental synthetic lethal candidates was informative with over-representation (Table [4.4](#x1-160014)) and supported by resampling analysis (Table [4.6](#x1-170036)), despite a modest intersection of genes between them (Figure [4.2](#x1-120012)). Either approach may be significant for a pathway in this intersection without being supported by the other: resampling analysis may support pathways that were not over-represented due to small effect sizes, thus both tests are required for a candidate pathway. The pathways detected by both over-representation and resampling are the strongest candidates for further investigation, such as Gαs signalling, a strong candidate in prior analyses with a role in the regulation of translation in cancer [Gao and Roux](#XGao2015) ([2015](#XGao2015)), another function supported by [SLIPT](#glo:SLIPT) analysis.

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

### 4.3 Metagene Analysis

Metagenes serve as a consistent signal of pathway activity. The direction of metagenes is generally arbitrary but care has been taken to ensure that these occur in a direction which reflect overall activation of the pathway (as described in Section [2.2.3](#x1-140003)). This will be supported by examining the pathway expression of gene signatures in breast cancer to ensure they behave as expected in TCGA expression data. These metagenes were also compared to somatic mutation to show the limitations of mutation as a measure of gene activity. Having established that metagenes generated with this procedure reflect gene activity, these were then applied to the Reactome pathways for synthetic lethal analysis of pathways directly to provide an alternative approach to identifying synthetic lethal pathways with CDH1.

#### 4.3.1 Pathway expression

Pathway metagenes (generated as described in Section [2.2.3](#x1-140003)) for gene signatures of key processes in breast cancer ([Gatza *et al.*](#XGatza2011), [2011](#XGatza2011)) were used to check that metagenes were generated in the correct direction to indicate pathway activation. These gene signatures were plotting in Figure [4.9](#x1-190019) for comparison with clinical factors and somatic mutations. The “intrinsic subtype” was computed by performing the PAM50 procedure [Parker *et al.*](#XParker2009) ([2009](#XParker2009)) for RNASeq data which was highly concordant with the subtypes provided by UCSC for TCGA samples previously analysed by microarrays ([TCGA](#XTCGA2012), [2012](#XTCGA2012)). Somatic mutations were reported for recurrently mutated genes in breast cancer, as reported by TCGA ([TCGA](#XTCGA2012), [2012](#XTCGA2012)), related genes, and those previously discussed to be important in hereditary breast cancers (BRCA1, BRCA2, and CDH1 ).

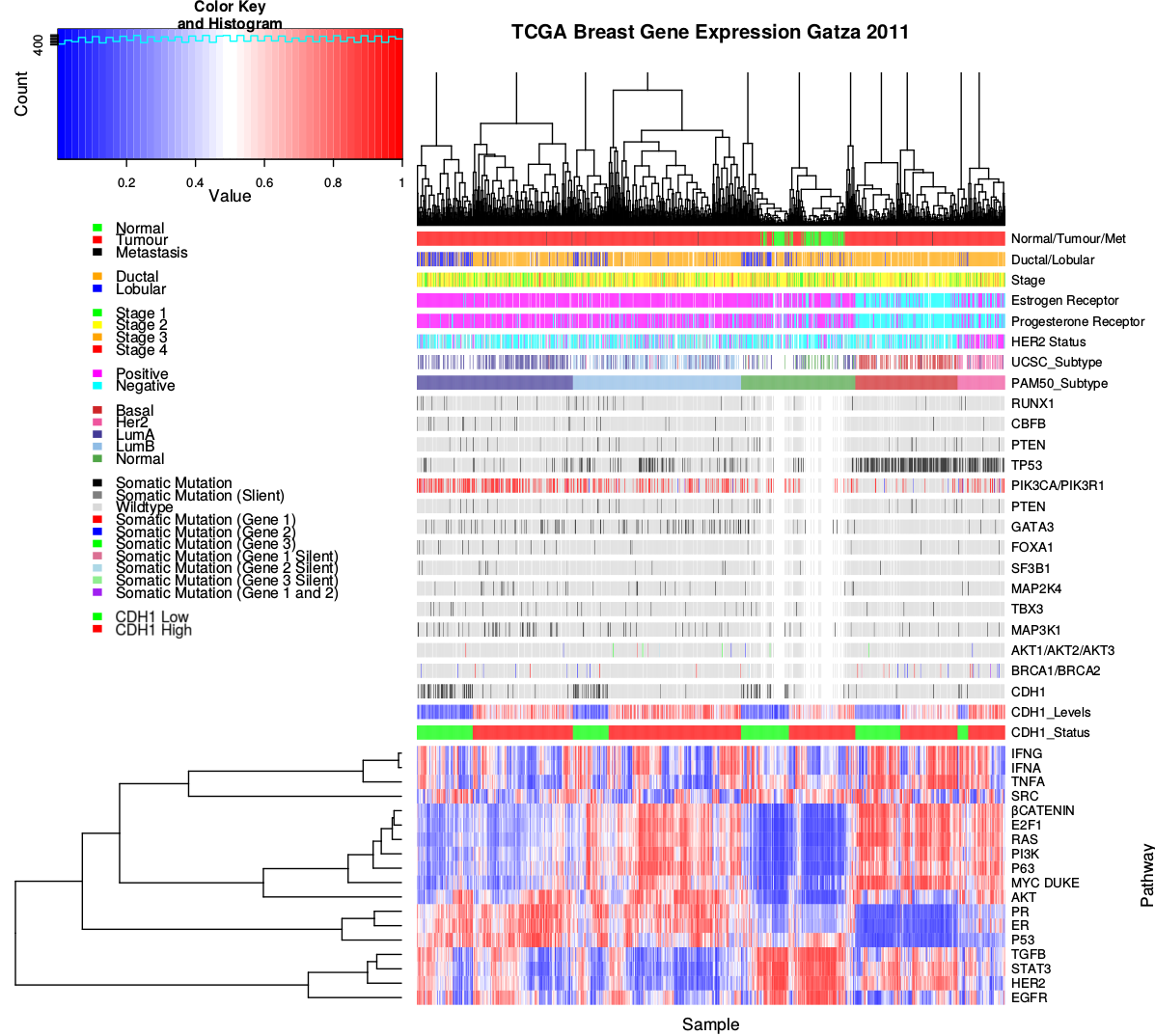


Figure 4.9: **Pathway metagene expression profiles.** Expression profiles for metagene signatures from [Gatza *et al.*](#XGatza2011) ([2011](#XGatza2011)) in TCGA breast data, annotated for clinical factors and cancer gene mutations. Samples were clustered independently for each intrinsic subtype and by CDH1 expression status.

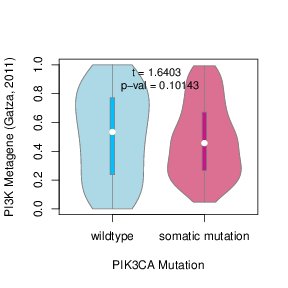
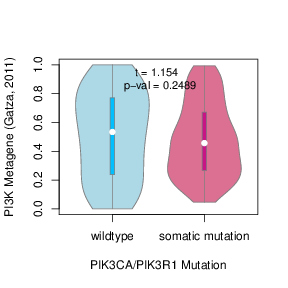
These gene signatures reflect intrinsic subtypes as expected. In particular, the estrogen and progesterone receptor signatures are low in the ER- and PR- basal subtype tumours. These tumours also had the highest frequency of TP53 mutations and a corresponding reduction of p53 metagene activity, as expected for loss of a tumour suppressor. The luminal A and luminal B tumour subtypes are the most similar, which is reflected in these metagenes signatures, although they are distinguishable molecular subtypes as shown by elevated PI3K, AKT, RAS, and β-catenin signalling in luminal B tumours. Although, these pathways were also elevated in Basal and HER2-enriched subtypes and lowly expressed in the “normal-like” subtype (which contained the normal samples). These intrinsic subtype specific gene signature profiles were further supported with metagenes for an extended set of signatures ([Gatza *et al.*](#XGatza2014), [2014](#XGatza2014)), as shown in Figure E.1.

TP53 mutations were the most frequent and more common in basal subtype. Similarly, GATA3 mutations were more common in luminal subtype tumours. PI3K mutations were more frequent across breast tumours, although these were less common in the basal subtype despite an elevated metagene (this discrepancy will the discussed further in Section [4.3.2](#x1-200002)). CDH1 mutations similarly occurred across molecular subtypes with the exception of the basal subtype (as observed in gene expression with Figure [4.1](#x1-90011)). CDH1 low samples occurred in all subtypes but were predominantly lobular subtype. Apart from these genes, mutations did not show clear specficity to a particular subtype and the variation between samples was reflects the range of molecular cascades that can result in tumours with similar molecular profiles, supporting the use of expression for cancer diagnostics and identification of molecular targets.

The direction of the metagenes were also consistent with the clincal characteristics and formed a consensus of gene activity as shown in Figures E.2–E.5. In each of the examples for gene signatures for PI3K (Figure E.2), p53 (Figure E.3), estrogen receptor (Figure E.4), and BRCA (Figure E.5) genes ([Gatza *et al.*](#XGatza2011), [2011](#XGatza2011), [2014](#XGatza2014)), the expression of the majority of the genes were highly concordant with the metagene, being either positively or negatively correlated. These were generally consistent with established clinical and molecular subtypes of breast cancer and the recurrent mutations shown. However, the PIK3CA and PIK3R1 mutant samples did not necessarily have elevated PI3K pathway metagene activity (as shown in Figure E.2).

#### 4.3.2 Somatic mutation

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

 PIK3CA mutation PIK3CA or PIK3R1 mutation

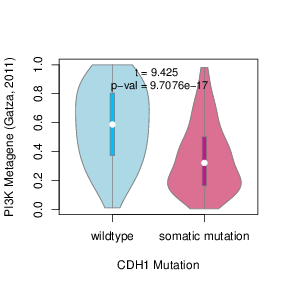
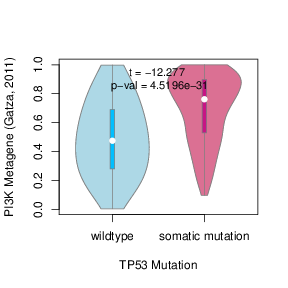
 CDH1 mutation TP53 mutation

Figure 4.11: **Somatic mutation against PI3K metagene.** Mutations in PIK3CA, PIK3R1, CDH1, and TP53 were examined in TCGA breast cancer for their effect on the PI3K ([Gatza *et al.*](#XGatza2011), [2011](#XGatza2011)) pathway metagene. The tumour suppressors CDH1 and TP53 showed an increase and decrease in the metagene respectively, whereas PIK3CA and PIK3R1 mutations had little effect on the metagene levels.

It should be noted that metagenes, while consistent with the consensus of constituent expressed genes, were not neccessarily reflecting the somatic mutation status. The PI3K ([Gatza *et al.*](#XGatza2011), [2011](#XGatza2011)) metagene levels in particular, were not statistically significant in between mutant and wildtype PIK3CA samples (shown in Figure [4.11](#x1-2000511)). Although the PI3K metagene differed across CDH1 and TP53 mutations, remarkably in opposite directions considering that PI3K is an oncogenic growth pathway and these are both most frequently tumour suppressors inactivated in cancers. This shows that CDH1 and TP53 deficient tumours have distinct molecular growth pathways and that synthetic lethal inactivations against CDH1 inactivation may not be applicable to other cancers with driver mutations such as TP53, although these were kept in the analysis for comparison. These differences may be related to these mutations being more frequent in tumours with difference clinical characteristics (as observed in Section [4.3.1](#x1-190001)). Thus mutations do not neccessarily have corresponding changes in pathway expression, particularly for oncogenes which may change in function rather than being upregulated.

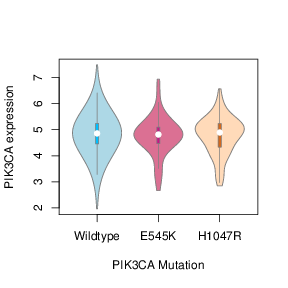
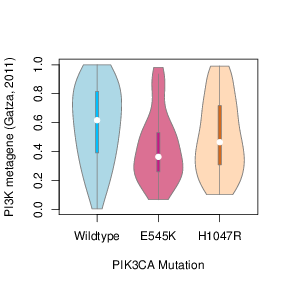
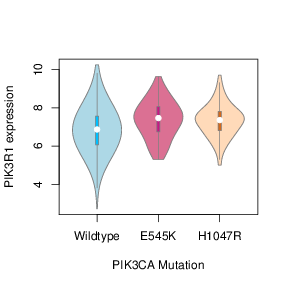
While the more specific PIK3CA ([Gatza *et al.*](#XGatza2014), [2014](#XGatza2014)) metagene showed significant differences with PIK3CA and PIK3R1 mutations (as shown in Figure D.7), this metagene replicated stronger differences for CDH1 and TP53. These differences were less pronounced in the protein levels of p110α (enocded by PIK3CA) and the downstream AKT gene (shown in Figures D.8 and D.9 respectively). Although this may be due to this regulatory cascade (kinases) being transmitted as a change in protein state (phosphorylation) rather than changes in expression levels. Another consideration is that mutations at different loci have different effects on protein function, particularly for oncogenes.

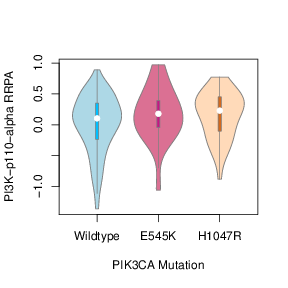
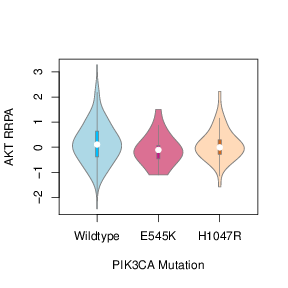
#### 4.3.3 Mutation locus

The gene locus distribution of PIK3CA and it’s receptor PIK3R1 were consistent with oncogenic and tumour suppressor mutations, as shown in Figure D.6. PIK3CA is an has recurrent mutations in 2 hotspots, centered around the E545K and H1047R (shown in Figure D.6a), as expected for an oncogene. This contrasts with the tumour suppressors, PIK3R1, and CDH1 (shown in Figures D.6b and D.6c respectively), which have low frequency inactivating mutations spread across them. A notable exception is TP53 (shown in Figure D.6d) which displays both inactivating mutations throughout and

recurrent (oncogenic) mutations at high frequency, consistent with the complex role of TP53 in cancer biology which is outside of the scope of this thesis and shown for comparison.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

 PI3KCA gene  PI3KCA metagene PIK3R1 gene

 PIK3CA protein AKT protein

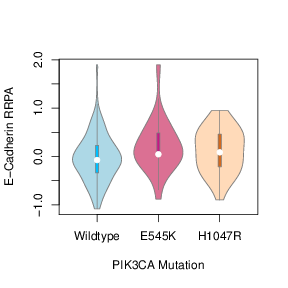
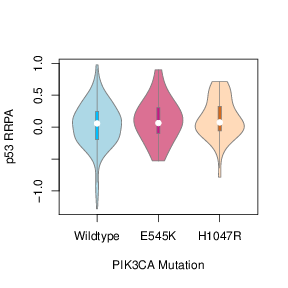
 CDH1 protein TP53 protein

Figure 4.13: **Somatic mutation locus against expression.** The recurrent E545K and H1047R oncogene mutations in PIK3CA were examined in TCGA breast cancer to show the effect of mutation locus on gene, pathway, and protein expression. While neither of these mutations had an impact of PIK3CA mRNA expression, E545K had specifically lower PI3K ([Gatza *et al.*](#XGatza2011), [2011](#XGatza2011)) metagene levels and both mutations had higher PIK3R1 mRNA expression. However, these differences were not reflected in the protein expression levels.

These differences in gene locus may explain why mutations do not necessarily have corresponding changes in gene or metagene expression. Specfically, the recurrent E545K and H1047R oncogene mutations in PIK3CA did not affect PIK3CA mRNA expression but E545K had specifically lower PI3K ([Gatza *et al.*](#XGatza2011), [2011](#XGatza2011)) metagene levels. Both mutations had higher PIK3R1 mRNA expression but these differences differences were not reflected in the protein expression levels of p110α protein (encoded by PIK3CA), it’s downstream target AKT, E-cadherin (encoded by CDH1), or p53 (as ashown in Figure [4.13](#x1-2100813)).

While the complex effects of mutation in oncogenes such as PIK3CA are not neccessarily detected in a pathway metagene, these do capture the consensus of patheay gene expression and account for other potential means of pathway activation. Thus metagenes are sufficient as a measure of gene activity for the purposes of synthetic lethal detection with [SLIPT](#glo:SLIPT). This approach is more applicable to tumour suppressor genes with a relationship between gene expression and activity (rather than activation at the protein level) but this is not a major concern since synthetic lethality is more clinically relevant for targeting tumour suppressor mutations than oncogenes.

#### 4.3.4 Synthetic lethal metagenes

Pathway metagenes for Reactome pathways (generated as described in Section [2.2.3](#x1-140003)) were also used for testing synthetic lethal partner pathways with CDH1 by [SLIPT](#glo:SLIPT). Since the metagenes have are higher when the pathway as a whole is activated, they are amenable to [SLIPT](#glo:SLIPT) analysis using low metagene levels for inactivated pathways. These synthetic lethal metagenes differed to the over-represented pathways among synthetic lethal gene candidates. However, there were some similarities to previous findings, as shown in Tables [4.7](#x1-220017). In particular, translational pathways were replicated as observed in Table [4.2](#x1-80012). While the specific pathways differ, immune pathways (such as NF-κB) were also supported by metagene synthetic lethal analysis.

Table 4.7: Candidate synthetic lethal metagenes against CDH1 from SLIPT

Pathway

ID

Observed

Expected

χ2value

p-value

p-value (FDR)

Glycogen storage diseases

3229121

68

130

176

6.62 × 10-37

1.82 × 10-34

Myoclonic epilepsy of Lafora

3785653

68

130

176

6.62 × 10-37

1.82 × 10-34

Diseases of carbohydrate metabolism

5663084

68

130

176

6.62 × 10-37

1.82 × 10-34

Arachidonic acid metabolism

2142753

81

130

157

8.13 × 10-33

1.49 × 10-30

Translation initiation complex formation

72649

70

130

152

7.08 × 10-32

1.17 × 10-29

Synthesis of 5-eicosatetraenoic acids

2142688

68

130

151

1.25 × 10-31

1.88 × 10-29

SRP-dependent cotranslational protein targeting to membrane

1799339

69

130

150

2.01 × 10-31

2.76 × 10-29

L13a-mediated translational silencing of Ceruloplasmin expression

156827

72

130

148

5.91 × 10-31

6.44 × 10-29

3’ -UTR-mediated translational regulation

157279

72

130

148

5.91 × 10-31

6.44 × 10-29

|  |
| --- |
| Activation of the mRNA upon binding of the cap-binding complex and eIFs, |
| and subsequent binding to 43S |

72662

70

130

147

1.14 × 10-30

9.28 × 10-29

Formation of the ternary complex, and subsequently, the 43S complex

72695

70

130

147

1.14 × 10-30

9.28 × 10-29

Ribosomal scanning and start codon recognition

72702

70

130

147

1.14 × 10-30

9.28 × 10-29

Eukaryotic Translation Elongation

156842

72

130

146

1.19 × 10-30

9.28 × 10-29

Nonsense Mediated Decay independent of the Exon Junction Complex

975956

71

130

146

1.24 × 10-30

9.28 × 10-29

Viral mRNA Translation

192823

70

130

146

1.51 × 10-30

1.04 × 10-28

Eukaryotic Translation Termination

72764

70

130

146

1.51 × 10-30

1.04 × 10-28

NF-kB is activated and signals survival

209560

71

130

145

1.90 × 10-30

1.19 × 10-28

Peptide chain elongation

156902

72

130

145

1.91 × 10-30

1.19 × 10-28

Influenza Life Cycle

168255

70

130

145

1.95 × 10-30

1.19 × 10-28

Formation of a pool of free 40S subunits

72689

73

130

145

2.01 × 10-30

1.19 × 10-28

Nonsense-Mediated Decay

927802

71

130

145

2.44 × 10-30

1.34 × 10-28

Nonsense Mediated Decay enhanced by the Exon Junction Complex

975957

71

130

145

2.44 × 10-30

1.34 × 10-28

GTP hydrolysis and joining of the 60S ribosomal subunit

72706

72

130

145

2.58 × 10-30

1.37 × 10-28

Influenza Viral RNA Transcription and Replication

168273

72

130

144

4.01 × 10-30

2.07 × 10-28

Signaling by NOTCH1 HD Domain Mutants in Cancer

2691230

79

130

143

5.99 × 10-30

2.82 × 10-28

Strongest candidate SL partners for CDH1 by [SLIPT](#glo:SLIPT) with observed and expected samples with low expression of both genes

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

The metagene analyses differ more between expresssion and CDH1 mutation than previous analyses, with more specific signalling pathways identified in the mutation analysis. This supports the usage of a complete null mutant model in experimental testing for synthetic lethality of signalling pathways against CDH1 inactivation rather than a knockdown in expression. However, low expression of partners has been used in either case to be applicable to dose-dependent pharmacological inhibition and across genes where mutations have different functional consequences, including variants of unknown significance.

These results show an independent pathway approach to detecting synthetic lethal gene functions interacting with CDH1. Synthetic lethal metagenes, replicates support for these pathways independent of pathway size (as genes are weighted equally). The synthetic lethal analysis against low CDH1 expression support prior findings in translational and immune pathways even if they were not able to detected in an experimental screen ([Telford *et al.*](#XTelford2015), [2015](#XTelford2015)). Together these findings support the role of CDH1 loss in cancer disrupting cell signalling with wider effects on protein translation and metabolism necessary for the proliferation of cancer cells. This is consistent with the GPCR pathways such as G˙αs signalling being supported by [SLIPT](#glo:SLIPT) gene candidates and the experimental primary siRNA screen, as shown by resampling in Section [4.2.1.4.1](#x1-170001).

### 4.4 Replication in stomach cancer

The synthetic lethal analysis of genes and pathways (previously described for TCGA breast cancer data) was replicated in TCGA stomach cancer. The accompanying data for [SLIPT](#glo:SLIPT) and mtSLIPT analyses against CDH1 expression and mutation are in Appendices F and G respectively.

The experimental screen ([Telford *et al.*](#XTelford2015), [2015](#XTelford2015)) was conducted in MCF10A breast cells so it may not be as comparable to stomach cancer. Nevertheless, CDH1 is also important in stomach cancer biology as a driver tumour suppressor gene and as a germline mutation in many cases of hereditary diffuse gastric cancer.

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

#### 4.4.1 Synthetic Lethal Genes and Pathways

The strongest [SLIPT](#glo:SLIPT) genes for stomach cancer (shown in Table F.1) did not neccessarily directly correspond to those observed in breast cancer (shown in Table [4.1](#x1-70011)). However, several gene functions were replicated in stomach cancer. Cell membrane genes including EMP3, GYPC, LGALS1, PRR24, and FUNCD2 were among the strongest SL candidates. Similarly, cell signalling genes including PLEKHO1, RARRES2, VEGFB, HSPB2, and CREM were detected in stomach cancers. It is notable that several of these genes (EMP3, PLEKHO1, and FUNCD2) have a known role in cancer. Together these genes support the roles of CDH1 in cell membrane and signalling functions (of epithelial tissues) which are perturbed in both breast and stomach cancers.

The strongest mtSLIPT genes tested against CDH1 mutatoin for stomach cancer (shown in Table G.1) supported similar gene functions. Membrane and cell-adhesion genes including KFBP6,THY1,CLELC2B, NISCH, TSPAN1,and KCTD12 and signalling genes including ZEB2, CCND2, NEURL1B, KFBP6, and OGN were detected. Similarly, these include cancer genes such as VIM,ZEB2,BCL2,THY1, and RUNX1T1. The mtSLIPT procedure also replicated several of the strongest candidates in breast cancer (shown in Table D.1) such as NRIP2 and NISCH.

Together, these gene candidates indicate widespread functions of CDH1 and strongly detectable synthetic lethality with many genes from a strategy that can be applied across cancer types. More specifically, the signalling genes included GPCR signalling genes (such as

GNG11, GNAI1, DZIP1, PTGFR, and KCTD12), a growth signalling pathway which was one of the most supported synthetic lethal pathways in breast cancer analysis, the experimental screen ([Telford *et al.*](#XTelford2015), [2015](#XTelford2015)), and has many actionable drug targets which have been applied to other diseases.

These findings were further supported by the pathways over-represented in [SLIPT](#glo:SLIPT) candidates from TCGA stomach cancer (shown in Table [4.8](#x1-240018)) which were replicated the translational and immune pathways observed in TCGA breast cancer (shown in Tabel [4.2](#x1-80012)). Further support for GCPR signalling pathways including the class A/1 receptors. The extracellular matrix was also detected at the pathway level in stomach cancer [SLIPT](#glo:SLIPT) candidates and replicated in mtSLIPT analysis for CDH1 mutation (shown in Table G.2), including elastic fibres, glycosylation, collagen, and integrin cell-surface interactions. Thus there was strong evidence for the role of extracellular matrix pathways and the tumour microenvironment in CDH1 deficient stomach cancers, in addition to cell signalling and translation pathways important in tumour growth across breast and stomach cancer.

Table 4.8: 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](#glo:SLIPT) partners for CDH1

#### 4.4.2 Synthetic Lethal Expression Profiles

The expression profiles of candidate synthetic lethal partners dtected by [SLIPT](#glo:SLIPT) and mtSLIPT in stomach cancer were plotted against clinical characteristics as described in section for breast cancer data in Section [4.1.2](#x1-90002) (shown in Figures [4.14](#x1-2500114) and G.1 respectively). As expected the majority of CDH1 mutant samples had low expression of CDH1 and were the diffuse type of stomach cancer.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

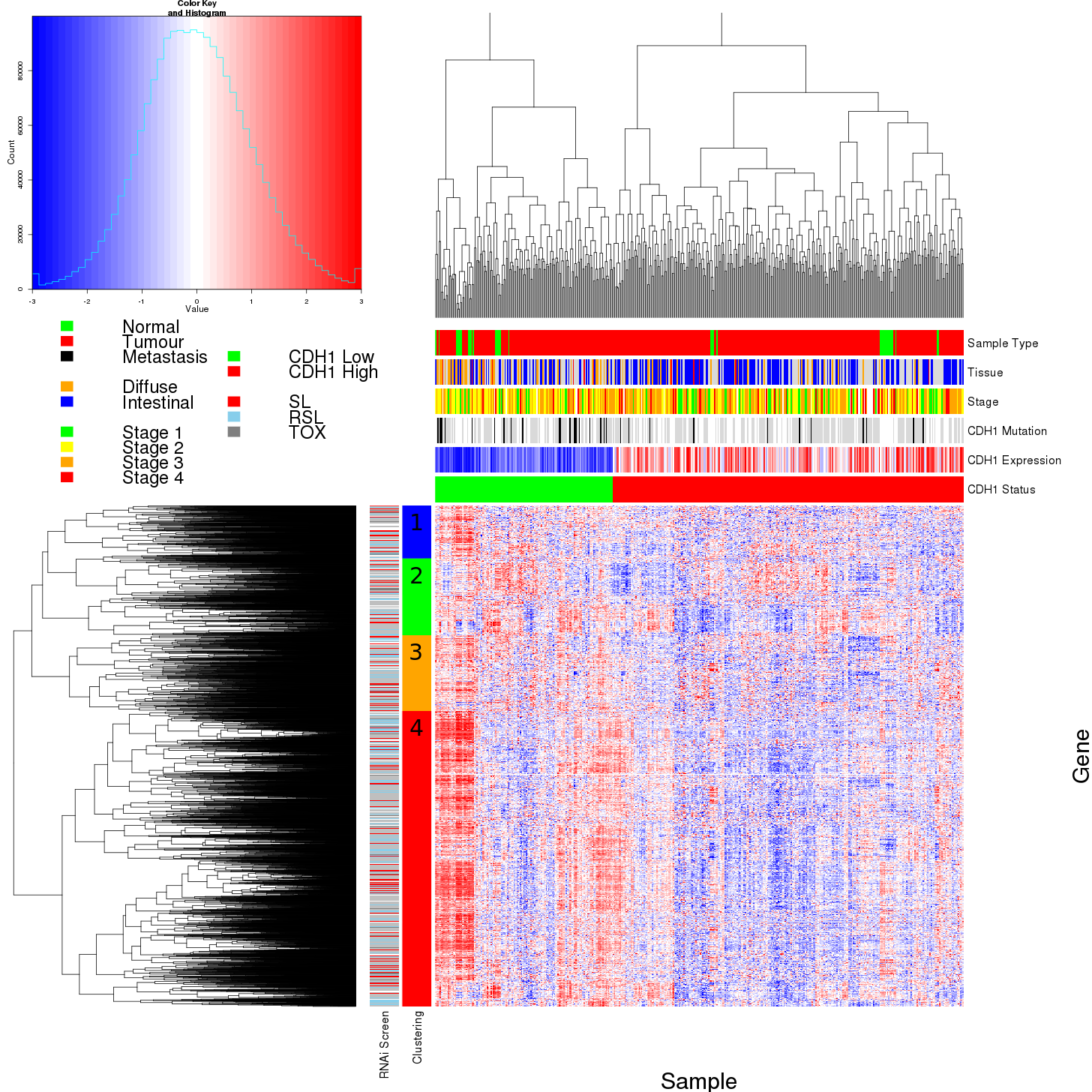


Figure 4.14: **Synthetic lethal expression profiles of analysed samples.** Gene expression profile heatmap (correlation distance) of all samples (separated by the 1 ∕ 3 quantile of CDH1 expression) analysed in TCGA stomach cancer dataset for gene expression of 4,365 candidate partners of E-cadherin (CDH1) from [SLIPT](#glo: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.*](#XTelford2015) ([2015](#XTelford2015)). Clusters had different sample groups highly expressing the synthetic lethal candidates in CDH1 low samples, notably diffuse and CDH1 mutant samples have elevated expression in one or more distinct clusters, although there was less complexity and variation among candidate synthetic lethal partners than in breast data. CDH1 low samples also contained most of samples with CDH1 mutations.

The [SLIPT](#glo:SLIPT) partners of CDH1 exhibited similar clustering in staomch cancer to breast cancer, replicating the diverse roles of elevated partner genes in different clinical samples. Specifically (in Figure [4.14](#x1-2500114)), the diffuse type stomach cancers had higher expression of the candidate synthetic lethal partners (where CDH1 has a role as a driver mutation), despite an unbiased clustering. This is consistent with compensating expression of synthetic lethal partners under loss of CDH1, as suggested by [Lu *et al.*](#XLu2015) ([2015](#XLu2015)). The pathway composition of gene clusters for stomach cancer (shown in Table F.2) was also highly concordant with breast cancer findings (shown in Table [4.3](#x1-100013)). These included replicated of translation (Cluster 1), immune functions (Cluster 2), G˙αs signalling (Cluster 3), and further support for the roles of GPCRs and the extracellular matrix (Cluster 4) in the synthetic lethal partners and functions of CDH1, replicated across stomach and breast cancers. Clusters 1 and 4, which had particularly high expression of [SLIPT](#glo:SLIPT) candidate partner genes in the diffuse subtype, also had the most significant over-representation of pathways.

There was less variation between the expression profiles of mtSLIPT partners of CDH1 in stomach cancer, although clusters were still detectable (as shown in Figure G.1). While the genes and pathways detected was lewss significant (due to lower sample size), the composition of clusters was further indicative for the roles of extracellular matrix (including elastic fibres), immune functions, and the cell signalling.

#### 4.4.3 Comparison to Primary Screen

The number of genes detected by both [SLIPT](#glo:SLIPT) in TCGA stomach cancer data and siRNA in breast cell lines (shown in Figure F.1) was also not a significant overlap (as observed for breast cancer in Figure [4.2](#x1-120012)). This was particularly the case of mtSLIPT against CDH1 mutation in stomach cancer which detected very few genes (as shown in Figure G.2) due to low sample size and mutation frequency.

This smaller overlap can also be attributed to the tissue-specific differences between the stomach cancers and the breast cells used for the experimental model ([Chen *et al.*](#XChen2014), [2014](#XChen2014)). Nevertheless, many genes were detected across [SLIPT](#glo:SLIPT) in stomach cancers and the experimental screen ([Telford *et al.*](#XTelford2015), [2015](#XTelford2015)) and the pathways detected were consistent with prior observations in breast cancer. Despite differences in the specific

genes detected, the functions of CDH1 were conserved across epithetial cancers in different tissues and synthetic lethal inhibition of interacting pathways may be effective against molecular targets such as CDH1 inactivation across tissue types.

However, the pathway composition of [SLIPT](#glo:SLIPT)-specific genes and those replicated with the siRNA primary screen ([Telford *et al.*](#XTelford2015), [2015](#XTelford2015)) were highly concordant between the pathways identified by [SLIPT](#glo:SLIPT) in TCGA stomach cancer (shown in Table F.3) and pathways previously identified in TCGA breast cancer (shown in Table [4.4](#x1-160014)). In both cases, translation and immune pathways were highly over-represented in [SLIPT](#glo:SLIPT)-specific genes, which we would not expect to be detected by siRNA screening in cell lines, as discussed in Section [4.2.1.4](#x1-160004). In addition, the extracellular matrix was supported by in stomach cancer. While the pathways identified by specifically by [SLIPT](#glo:SLIPT) in stomach cancer or siRNA screening were similar to those observed for breast cancer (in Table [4.4](#x1-160014)), the pathways over-represented in the intersection for stomach cancer [SLIPT](#glo:SLIPT) candidates and the siRNA primary screen ([Telford *et al.*](#XTelford2015), [2015](#XTelford2015)) also had a clear over-representation of signalling pathways, although they differed from those observed in breast cancer [SLIPT](#glo:SLIPT) candidates. GPCR signalling was supported in genes detected in both TCGA stomach cancer and screening, including G˙αq, G˙αs, serotonin receptors, and class A signalling (shown in more detail in Table F.5). In addition MAPK and NOTCH signalling pathways were detected. These replicate the findings in breast cancer and show consistent detection of signalling pathways in stomach cancer despite less genes being detected by [SLIPT](#glo:SLIPT) and patient samples differing from the tissue in which the experiments were conducted.

Similarly, the [SLIPT](#glo:SLIPT)-specific gene candidates against CDH1 mutation (shown in Table G.4) replicated pathways observed in breast cancer (shown in Table D.4), despite a lower number of genes detected. In particular, the extracellular matrix and elastic fibres were over-represented. While the number of genes overlapping with the siRNA was too low to be amenable to pathway analysis, there is further indication that members of these genes replicated across mutation [SLIPT](#glo:SLIPT) analyses include cell-membrane, elastic fibre, and GPCR signalling genes.

##### 4.4.3.1 Resampling Analysis

Similarly, resampling for [SLIPT](#glo:SLIPT) specific candidates (shown in Tables F.4 and G.5) replicated many of the most highly over-represented pathways in stomach cancer. These

include translational, immune, GPCR signalling, and elastic fibres, consistent with previous analyses in breast cancer (shown in Tables [4.5](#x1-170025) and D.5).

While fewer pathways were supported by resampling for the intersection of [SLIPT](#glo:SLIPT) and experimental screen ([Telford *et al.*](#XTelford2015), [2015](#XTelford2015)) candidate partners in stomach cancer than breast cancer, many of those detected (shown in Table F.5) replicate those detected in breast cancer (shown in Tables [4.6](#x1-170036) and D.6). The pathways detected by both permutation and over-representation were more likely to be replicated across stomach and breast cancer than those detected by over-representation alone, supporting the use of this procedure to detect synthetic lethal pathways applicable across cancer types. The include G˙αs signalling and elastic fibre formation as discussed for breast cancer (in Section [4.2.1.4.1](#x1-170001))

While many pathways were detected by resampling for mtSLIPT against CDH1 mutation in stomach cancer (shown in Table G.6), there were not enough genes detected by both mtSLIPT and the siRNA primary screen to determine over-represented pathways. Therefore this may be due to small numbers of genes which does not constitute support for pathway composition. However, this under-powered analysis does not preclude the replicated synthetic lethal pathways detected across [SLIPT](#glo:SLIPT) expression analyses in TCGA breast and stomach cancer data with an accompanying siRNA primary screen ([Telford *et al.*](#XTelford2015), [2015](#XTelford2015)). Rather this further supports the use of [SLIPT](#glo:SLIPT) to test against low expression of query genes as measure of gene inactivation to avoid this issue, despite mutation (which often produces similar results) being more indicative of complete gene inactivation.

#### 4.4.4 Metagene Analysis

Metagene analysis (as conducted in Section [4.3.4](#x1-220004)) was also performed for TCGA stomach cancer expression data, using Reactome pathways. These results (as shown in Table F.6) provided further support for signalling and extracellular processes as synthetic lethal pathways across stomach and breast cancer. Namely, cell-cell communication, VEGF signalling, and various GPCR pathways were detected.

Signalling and immune pathways were also supported by mtSLIPT analysis of metagene pathway expression against CDH1 mutation, as shown in Table G.7. Although these results were generally less statistically significant than expression analyses. Signalling pathways detected as synthetic lethal metagenes include prostacyclin, SCF-KIT, ERK, MAPK, NGF, VEGF, and PI3K/AKT. The innate immune response, the inflammasome, and integrin

signaling were also implicated to be synthetic lethal with CDH1 mutations. Cell surface interactions, cholesterol biosynthesis, and platelet homeostasis also support the role of extracellular processes in proliferation of CDH1 deficient cancers and interactions of CDH1 with the extracellular environment that was not tested in the cell line experimental screen.

### 4.5 Global Synthetic Lethality

Global levels of synthetic lethality were analysed to address concerns raised by the high numbers of synthetic lethal candidates for CDH1. The [SLIPT](#glo:SLIPT) procedure (as described in Section 3.1) was performed with each possible query gene from the TCGA breast cancer RNA-Seq dataset. Due to the computational demands of this procedure, it was performed on the New Zealand eScience Infrastructure Intel Pan supercomputer (as described in Section [2.5.3](#x1-320003)).

The observed number of [SLIPT](#glo:SLIPT) appears to be typical for most genes in the TCGA breast RNA- Seq dataset as shown in Figure [4.15](#x1-2900115). This figure was actually lower than the majority (95%) of genes tested, although CDH1 was ranked higher for a similar in [SLIPT](#glo:SLIPT) analysis of TCGA stomach cancer data, shown in Figure H.1. The differences in sample size make these analyses difficult to compare but (in either case), the number of partners detected for CDH1 is not unexpected, eeven when adjust for multiple comparisons across candidate partners.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

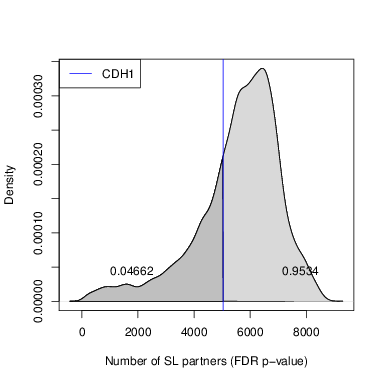


Figure 4.15: **Synthetic lethal partners across query genes.** Global synthetic lethal pairs were examined across the genome in TCGA breast expression data by applying [SLIPT](#glo:SLIPT) across query genes. The high number of predicted partners for CDH1 was typical for a human gene and lower than many other genes.

The number of detected candidates reported here is higher than in Figures [4.2](#x1-120012) and F.1 because these exlcuded genes not tested by the siRNA primary screen ([Telford *et al.*](#XTelford2015), [2015](#XTelford2015)) for comparison with it. For an statistically rigorous measure of global synthetic lethality, multiple comparison procedures would need to be performed for all pairs of genes tested. However, only partner genes for each query [SLIPT](#glo:SLIPT) analysis were performed for the purposes of comparing the number of partners predicted with those observed for CDH1 throughout this thesis.

#### 4.5.1 Hub Genes

The genes with the most synthetic lethal interactions by this [SLIPT](#glo:SLIPT) analysis are the “hub” genes of a synthetic lethal network. These genes with the highest number of candidate partners detected by [SLIPT](#glo:SLIPT) in TCGA breast cancer expression data are summarised in Table [4.9](#x1-300019). These include several genes involved in cellular signalling such as TGFBR2, PDGFRA, FAM126A, KCTD12, MAML2, and CAV1. Gene regulation including chromatin, DNA, and RNA bindings genes were also observed as hub genes such as CELF2, PLAGL1, TSHZ2, FOXO1, and SVEP1. Genes involved in the cellular membrane such as ANXA1 and FAM171A1 were also observed in addition to genes specifically implicated in cell adhesion and tight junctions such as TNS1, BOC, AMOTL1, FAT4, and EPB41L2.

Table 4.9: Query synthetic lethal genes with the most SLIPT partners

Gene

Direction

raw p-value

p-value (FDR)

[SLIPT](#glo:SLIPT) raw p-value

[SLIPT](#glo:SLIPT) (FDR)

TGFBR2

8134

17982

17973

8007

8006

A2M

8571

17605

17583

8345

8339

TNS1

8019

17949

17934

7874

7873

PROS1

8539

17668

17642

8317

8310

ANXA1

9085

17330

17302

8689

8682

CELF2

8665

17406

17368

8370

8355

BOC

8694

17371

17348

8384

8381

PLAGL1

8792

17361

17327

8448

8436

PDGFRA

8296

17650

17621

8095

8087

FAM171A1

8874

17560

17533

8567

8562

FAM126A

8510

17383

17356

8184

8178

TSHZ2

7942

17983

17976

7787

7786

KCTD12

8366

17651

17621

8115

8108

MAML2

8336

17537

17503

8069

8061

FOXO1

8027

17753

17737

7840

7836

AMOTL1

8425

17388

17347

8147

8139

FAT4

8111

17750

17732

7925

7919

CAV1

8645

17491

17464

8342

8331

SVEP1

7945

17859

17842

7791

7784

EPB41L2

8415

17327

17296

8097

8092

Genes with the most candidate SL partners [SLIPT](#glo:SLIPT) in TCGA breast expression data with the number of partner genes predicted by direction criteria and χ2 testing separately and combined as a [SLIPT](#glo:SLIPT) analysis. Where specified, the p-values for the χ2 test were adjusted for multiple tests (FDR).

Genes involved in adhesion and tight junctions were also hub genes in stomach cancer (shown in Table H.1) such as HEG1, FAT4, NFASC, LAMA4, LAMC1, TNS1, and AMOTL1. These also included cytoskeletal genes such as ANK2, TTC28, and MACF1. Cancer genes were also among hub genes across breast and stomach cancer such as BOC, FAT4, and MRVI1.

It is therefore unsurprising that signalling and regulatory genes have been detected throughout this thesis. Not only are they suitable targets for anti-cancer therapy, they are also highly interacting genes themselves and so it is plausible that their interactions would be detectable by [SLIPT](#glo:SLIPT). This is consistent with the established role of abberant signalling and gene regulation in proliferation and survival of tumours and the importance of these pathways in development with highly redundant functions across many genes under complex regulation. These are also highly amenable to detection by [SLIPT](#glo:SLIPT) analysis of expression data since their functions are dynamically regulated with corresponding changes in expression.

Cytoskeletal, membrane bound, and extracellular matrix genes are also among highly interacting synthetic lethal hubs, including focal adhesion, tight junctions, microtubules, and fibronectin. These support the use of synthetic lethal interactions to target CDH1, as a tumour suppressor gene involved in these functions. Cellular structure and cell-cell interactions are thus important functions with highly redundant genes for which there are many feasible synthetic lethal interactions by which to understand regulation of cellular functions. These functions may also be exploited as vulnerabilities in cancer as they are frequently disruped in cancers, including HDGC where loss of CDH1 is a driver of cancer proliferation and malignancy.

#### 4.5.2 Hub Pathways

Pathways over-represented among TCGA breast cancer hub genes (as shown in Table [4.10](#x1-3100110)) particularly support the importance of signalling pathways, such as the PI3K/AKT pathway, as synthetic lethal hubs. The highly redundant natures of cell-cell interaction and the extracellular matrix functions was also further supported.

Table 4.10: 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 Ca2+

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](#glo:SLIPT) analysis of TCGA breast expression data

Pathway over-representation for synthetic lethal hub genes was replicated in TCGA stomach cancer expression data. However, these pathways differ considerably from breast cancer, as shown in Table H.2. Cell-cell interactions and extracellular matrix pathways, including elastic fibres, were also among the hub genes for stomach cancer. The signalling pathways differ as expected in a different tissue type, although BMP and PAK signalling were detected as hub gene functions.

### 4.6 Replication in cell line encyclopaedia

Table 4.11: 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](#glo:SLIPT) partners for CDH1

As breast cancer cell lines are the experimental system in which many cancer genetics and drug targets are investigated, these were analysed in addition to patient samples from TCGA. The cancer cell line encyclopaedia (CCLE) is a resource for genomics profiles across a range of cell lines. These have also been used to generate synthetic lethal candidates for comparison to those in experimental screen and predictions from TCGA expression data.

The cancer cell line encyclopaedia provides further support for synthetic lethal genes and pathways that may be applicable across cell types and reproducible in experimental systems. In contrast to the homogeneous pooled cell samples of patients, the cell lines provide a genetically homogeneous cell population in which to examine molecular functions and as a preclinical model of cancerous disease. The complete set of 1037 cell lines was tested for synthetic lethality across tissues, in addition to the 59 breast cell lines and 38 stomach cell lines being tested separately for partners of CDH1. Synthetic lethal genes were detected by [SLIPT](#glo:SLIPT) (as described in Section 3.1) and over-represented synthetic lethal Reactome pathways (as described in Section [2.3.2](#x1-180002)).

Synthetic lethal gene candidates were detectable by [SLIPT](#glo:SLIPT) across each of these sample sets of cells lines (as shown in Tables I.1–I.3. Although these were most highly significant across the samples in the CCLE expression dataset (as shown in Table I.1) and included genes detected in prior analyses such as VIM, ZEB2, EMP3. Pathways were also highly over-represented among synthetic lethal candidates for the full CCLE dataset (as shown in Table [4.11](#x1-3200111)) including Rho GTPase (GPCRs), immmune, and gene regulation (chromatin and transcription). This is unexpected since immune pathways would not be expected to be detectable in isolated cell lines, although this could be attributed to cytokine and integrin signalling occuring the cancer cells in addition to interactions with immune cells in the tumour microenvironment (which could not be distinguished in patient samples). Cell cycle and mitosis were among the highest synthetic lethal pathways across cell lines supporting CDH1 deficient cells having abberant cell signalling and consequences for proliferation such as cancer cells. However, cell cycle genes were not as strongly supported in TCGA patient samples or the siRNA screen ([Telford *et al.*](#XTelford2015), [2015](#XTelford2015)) and they may not be applicable to epithelial tissues such as breast or stomach cancer or amenable to selective inhibition in experimental models.

Table 4.12: 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 biosythesis

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](#glo:SLIPT) partners for CDH1

Synthetic lethal pathways specfic to [SLIPT](#glo:SLIPT) candidates from breast cell lines (as shown in Table [4.12](#x1-3200212)) were more consistent with previous obervations, particularly the established role of E-cadherin in cell junctions and the Adherens complex. Although the number of [SLIPT](#glo:SLIPT) candidate genes detected in stomach cell lines was insufficient to replicate the findings in breast cell lines to TCGA patient samples. However, [SLIPT](#glo:SLIPT) candidates across breast and stomach CCLE cell lines were over-represented (as shown in Table I.5) for similar pathways to breast cell lines with additional support for extracellular matrix pathways including elastic fibres which were replicated with resampling across breast and stomach TCGA analyses and the primary siRNA screen [Telford *et al.*](#XTelford2015) ([2015](#XTelford2015)).

### 4.7 Discussion

#### 4.7.1 Strengths of the SLIPT Methodology

Synthetic lethal discovery with [SLIPT](#glo:SLIPT) used established statistical procedures identify putative partner genes from gene expression data. Such use of the χ2-value is amenable to pathway or permutation analyses and could feasibly be applied to other disease gene or pair-wise across the genome. Although genome-wide approaches were unable to find informative candidate genes for E-cadherin [Lu *et al.*](#XLu2015) ([2015](#XLu2015)). Synthetic lethal discovery in cancer has focused on genes with severe cellular mutant phenotypes, such as essential genes or the oncogenes TP53 and AKT [Lu *et al.*](#XLu2015) ([2015](#XLu2015)); [Tiong *et al.*](#XTiong2014) ([2014](#XTiong2014)); [Wang and Simon](#XWang2013) ([2013](#XWang2013)), with other cancer genes, such as CDH1, requiring more focused investigations. Prior computational approaches for synthetic lethal discovery, in cancer, vary widely ([Jerby-Arnon *et al.*](#XJerby2014), [2014](#XJerby2014); [Lu *et al.*](#XLu2015), [2015](#XLu2015); [Tiong *et al.*](#XTiong2014), [2014](#XTiong2014); [Wappett *et al.*](#XWappett2016), [2016](#XWappett2016)). There is no consensus for which for the approach is more appropriate and they are difficult to compare as they either do not have a released code implementation or do not make predictions solely from normalised expression data.

However, the query-based approach demonstrated by [SLIPT](#glo:SLIPT) analysis is suitable for wider application on expression data and augmenting experimental studies such as high-throughput screens. This approach has identified biologically plausible synthetic lethal

pathways for CDH1, triaged candidates from experimenal screening ([Telford *et al.*](#XTelford2015), [2015](#XTelford2015)), and replicates genes and pathways across breast and somtach cancers datasets. In addition, [SLIPT](#glo:SLIPT) avoids critical assumptions underlying the design of some prior approaches such as coexpression of synthetic candidates or that they will have known (annotated) similarities in function.

The DAISY methodology [Jerby-Arnon *et al.*](#XJerby2014) ([2014](#XJerby2014)), which took a similar query-based approach with the tumour suppressor VHL, has been critiqued for being too stringent [Lu *et al.*](#XLu2015) ([2015](#XLu2015)) which impedes pathway analysis. Since functional redundancy does not require genes to be expressed at the same time, the [SLIPT](#glo:SLIPT) approach does not assume co-expression of synthetic lethal genes which may enrich for synthetic lethal genes in established coregulated pathways. Rather, the interpretation of synthetic lethality for [SLIPT](#glo:SLIPT) was similar to other computational methods based on ‘co-loss under-representation’, ‘compensation’, or ‘simultaneous differential expression’ [Lu *et al.*](#XLu2015) ([2015](#XLu2015)); [Tiong *et al.*](#XTiong2014) ([2014](#XTiong2014)); [Wang and Simon](#XWang2013) ([2013](#XWang2013)).

Genomics analyses are prone to false-positives and require statistical caution, particularly where working with gene-pairs scale up the number of multiple tests drastically, at the expense of statistical power. Experimental screens for synthetic lethality are also error-prone, 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. Thus analyses throughout this thesis have focused on querying for partners of a particular gene of interest. Statistical modelling and simulations (in Section 3.3 and Chapter [6](#x1-60006)) will further support the design decisions underlying [SLIPT](#glo:SLIPT) analysis and it’s strengths over other approaches.

#### 4.7.2 Syntheic Lethal Pathways for E-cadherin

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-representation) mutual loss ([Jerby-Arnon *et al.*](#XJerby2014), [2014](#XJerby2014); [Kelly](#XKelly2013), [2013](#XKelly2013); [Lu *et al.*](#XLu2015), [2015](#XLu2015)). 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 lethal detection is error-prone and identifying genes relevant for clinical application will be difficult

without a supporting biological pathway rationale. As such, investigations into the genes identified by [SLIPT](#glo:SLIPT), correlation structure between them, and those which were validated by experimental screening ([Telford *et al.*](#XTelford2015), [2015](#XTelford2015)) focused on the pathway level throughout this Chapter. Similarly, comparisons across analyses were largely made at the pathway level, including comparisons between expression and mutation, breast and stomach TCGA datasets, and patient sample data with cell line expression profiles.

Potential synthetic lethal partners of CDH1 identified by [SLIPT](#glo:SLIPT) had many distinct functions, with each gene cluster highly expressed in different patient subgroups (Figure [4.1](#x1-90011)). The expression profiles of the SL partners of CDH1 predicted from the TCGA breast cancer RNA-Seq data (expected to have compensating high or stable expression) and their corresponding functional enrichment found subgroups of genes with functional organisation particularly among CDH1 low breast tumours. 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 estrogen receptor negative tumours, although this is independent of tumour stage and consistent with poor prognosis in these patients and could inform other treatment strategies or prevent ineffective treatment further impacting quality of life in these patients. These results suggest that synthetic lethal partner expression varies between patients; that these different tumour classes would react differently to the same treatment; that treatment of different pathways and combinations in different patients is the most effective approach to target genes compensating for CDH1 gene loss; and the expression of synthetic partners could be a clinically important biomarker.

The pathways that synthetic lethal partners of CDH1 identified by [SLIPT](#glo:SLIPT) were involved in a diverse range of biological functions and differed to those detected experimentally. This discrepancy may be accounted for by gene expression analyses detecting both synthetic lethal partners (as screened for experimentally [Telford *et al.*](#XTelford2015) ([2015](#XTelford2015))) and their downstream targets (not detected by siRNA), capturing the wider pathways and mechanisms involved in synthetic lethality with CDH1 inactivation. In particular, GPCR phosphorylation cascades (which regulates gene expression and translation in cancers [Gao and Roux](#XGao2015) ([2015](#XGao2015))) were predicted to be synthetic lethal with CDH1. The predicted synthetic lethal partners occurred across functionally distinct pathways, including characterised functions of CDH1. The most consistently supported pathways include elastic fibres in the extracelullar matrix, GPCR signalling, and translation presenting vulnerabilities for CDH1

deficient cancer cells from extracellular stimuli to the core growth mechanisms of a cell.

This diversity in synthetic lethal functions is consistent with the wide ranging role of CDH1 in cell-cell adhesion, cell signalling, and the cytoskeletal structure of epithelial tissues. Pathway structure may be relevant to identifying potential drug targets from gene expression signatures, indicating downstream effector genes and mechanisms leading to cell inviability. Identification of distinct synthetic lethal gene clusters may further lead to the elucidation of drug resistance mechanisms. While these pathways are indicative of the main functions of E-cadherin and synthetic lethal partners, it remains to identify the genes within these pathways that are the most actionable or supported across [SLIPT](#glo:SLIPT) analysis in patient samples and detected by experiments in preclincal models ([Chen *et al.*](#XChen2014), [2014](#XChen2014); [Telford *et al.*](#XTelford2015), [2015](#XTelford2015)). The specific genes within key pathways will be be discussed in Chapter [5](#x1-60005), along with further investigations into their relation to pathway structure. While these are important clinical implications, the synthetic lethal predictions lack enough confidence for direct translation into pre-clinical models or clinical applications leading to a need for statistical modelling and simulation of synthetic lethality in genomics expression data.

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

#### 4.7.3 Replication and Validation

##### 4.7.3.1 Integration with siRNA Screening

The pathway composition across computational and experimental synthetic lethal candidates was informative with over-representation (Table [4.4](#x1-160014)) and supported by resampling analysis (Table [4.6](#x1-170036)), despite a modest intersection of genes between them (Figure [4.2](#x1-120012)). Either approach may be significant for a pathway in this intersection without being supported by the other: resampling analysis may support pathways that were not

over-represented due to small effect sizes, thus both tests are required for a candidate pathway.

The pathways detected by both over-representation and resampling are the strongest candidates for further investigation and the pathway structure analyses in Chapter [5](#x1-60005) will focus on these pathways detected by both over-representation and resampling. Particularly, those replicated across datasets or with pathway metagenes. In addition to GCPR pathways detected across these analyses, the PI3K cascade will also be investigated in Chapter [5](#x1-60005), this signalling pathway is a well characterised mediator between GCPR receptors and regulation of translation ([Gao and Roux](#XGao2015), [2015](#XGao2015)) (both detected throughout this Chapter) and exhibited unexpected behaviour with pathway the metagenes (in Section [4.3](#x1-180003)). This pathway is activated by protein Phosphorylation states and thus inactivatino may not be detectable with expression.

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

##### 4.7.3.2 Replication across Tissues and Cell lines

Furthermore, synthetic lethal partners identified by [SLIPT](#glo:SLIPT) were replicated across breast and stomach cancer. These were particularly concordant at the pathway level, as expected between tissues since synthetic lethal pathways have higher conservation between species ([Dixon *et al.*](#XDixon2008), [2008](#XDixon2008)). These findings support gene functions conserved across CDH1 deficient cancers in breast and stomach tissues, presenting vulnerabilities that could be applied against molecular targets in both cancers. In addition, these analyses serve as a replication across independent patient cohorts from breast and stomach cancers, decreasing the likelihood of the synthetic lethal pathways detected being false positives or artifacts of either dataset.

Synthetic lethal pathways were also replicated across expression analyses of TCGA patient samples in heterogeneous tumours and homogeneous cell line isolates. This further supports that the subset of synthetic lethal functions detectable in experimental models ([Chen *et al.*](#XChen2014), [2014](#XChen2014); [Telford *et al.*](#XTelford2015), [2015](#XTelford2015)) would be applicable tumours of patients with CDH1 deficient cancers.

There are many gene functions replicated across breast cancer gene expression analyses. Many of these were also replicated with mutation analysis and with stomach cancer or cell line expression data. These pathways were more consistent across replication analyses than previous investigations with TCGA microarray data [Kelly](#XKelly2013) ([2013](#XKelly2013)).

### 4.8 Summary

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

This approach has been applied to robustly detect synthetic lethal pathways for the E-cadherin (CDH1) in TCGA breast cancer molecular profiles with comparisons to experimenal screening ([Telford *et al.*](#XTelford2015), [2015](#XTelford2015)) in cell lines, and replication in TCGA stomach cancer molecular profiles and across cell types in the cancer cell line encyclopaedia. The pathway replicated across several analyses included extracellular matrix pathways (such as elastic fibres formation), cell signalling (including GPCRs), and core gene regulation and translation processes crucial for the growth and proliferation of cancer cells. These pathways show evidence of non-oncogene addiction for CDH1 deficient cells and present vulnerabilities which may be exploited for specific treatment against CDH1 mutations in HCGC and sporadic cancers. There was also support for synthetic lethal pathways with CDH1 in cell adhesion and cytoskeletal processes to which CDH1 belongs, supporting the finding that synthetic lethality occurs within biological pathways ([Boone *et al.*](#XBoone2007), [2007](#XBoone2007); [Kelley and Ideker](#XKelley2005), [2005](#XKelley2005)).

While translational and immune pathways detected by [SLIPT](#glo:SLIPT) were not supported by primary siRNA screening ([Telford *et al.*](#XTelford2015), [2015](#XTelford2015)), these were replicated across various analyses. Due to the differences between an experimenal cell line model ([Barretina *et al.*](#XBarretina2012), [2012](#XBarretina2012); [Chen *et al.*](#XChen2014), [2014](#XChen2014); [Fece de la Cruz *et al.*](#XFece2015), [2015](#XFece2015)) and patient molecular profiles ([Bass *et al.*](#XTCGA2014GC), [2014](#XTCGA2014GC); [TCGA](#XTCGA2012), [2012](#XTCGA2012)), these would not be expected to be completely concordant. Furthermore, many pathways are difficult to test in an isolated experimental system. Nevertheless, many of the genes and pathways detected by [SLIPT](#glo:SLIPT) are suitable to inform further investigations and triage of therapeutic targets against CDH1 deficient tumours in combination with experimenal screening.

A characteristic of gene interaction networks is a scale-free topology leading to highly interacting hub genes, these represent important genes in a functional network. Cell surface

interactions, the extracellular matrix, and cell signalling (particularly PI3K/AKT signalling) were also found to be synthetic lethal hubs with more interactions detected than other genes. This indicates that these pathways are functionally important to survival of cancer cells since they are subject to high functional redundancy, despite frequent disruptions in cancer. These pathways being involved in a disproportionate number of synthetic lethal interactions is also consistent with their detection for CDH1.

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

Aims

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

Summary

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

## References

   Aarts, M., Bajrami, I., Herrera-Abreu, M.T., Elliott, R., Brough, R., Ashworth, A., Lord, C.J., and Turner, N.C. (2015) Functional genetic screen identifies increased sensitivity to wee1 inhibition in cells with defects in fanconi anemia and hr pathways. *Mol Cancer Ther*, **14**(4): 865–76.

   Abeshouse, A., Ahn, J., Akbani, R., Ally, A., Amin, S., Andry, C.D., Annala, M., Aprikian, A., Armenia, J., Arora, A., *et al.* (2015) The Molecular Taxonomy of Primary Prostate Cancer. *Cell*, **163**(4): 1011–1025.

   Adamski, M.G., Gumann, P., and Baird, A.E. (2014) A method for quantitative analysis of standard and high-throughput qPCR expression data based on input sample quantity. *PLoS ONE*, **9**(8): e103917.

   Agarwal, S., Deane, C.M., Porter, M.A., and Jones, N.S. (2010) Revisiting date and party hubs: Novel approaches to role assignment in protein interaction networks. *PLoS Comput Biol*, **6**(6): e1000817.

   Agrawal, N., Akbani, R., Aksoy, B.A., Ally, A., Arachchi, H., Asa, S.L., Auman, J.T., Balasundaram, M., Balu, S., Baylin, S.B., *et al.* (2014) Integrated genomic characterization of papillary thyroid carcinoma. *Cell*, **159**(3): 676–690.

   Akbani, R., Akdemir, K.C., Aksoy, B.A., Albert, M., Ally, A., Amin, S.B., Arachchi, H., Arora, A., Auman, J.T., Ayala, B., *et al.* (2015) Genomic Classification of Cutaneous Melanoma. *Cell*, **161**(7): 1681–1696.

   Akobeng, A.K. (2007) Understanding diagnostic tests 3: receiver operating characteristic curves. *Acta Pdiatrica*, **96**(5): 644–647.

   American Cancer Society (2017) Genetics and cancer. [https://www.cancer.org/cancer/cancer-\_causes/genetics.html](https://www.cancer.org/cancer/cancer-causes/genetics.html). Accessed: 22/03/2017.

   American Society for Clinical Oncology (ASCO) (2017) The genetics of cancer. [http://www.cancer.net/navigating-*cancer-*care/cancer-*basics/genetics/genetics-*cancer](http://www.cancer.net/navigating-cancer-care/cancer-basics/genetics/genetics-cancer). Accessed: 22/03/2017.

   Anjomshoaa, A., Lin, Y.H., Black, M.A., McCall, J.L., Humar, B., Song, S., Fukuzawa, R., Yoon, H.S., Holzmann, B., Friederichs, J., *et al.* (2008) Reduced expression of a gene proliferation signature is associated with enhanced malignancy in colon cancer. *Br J Cancer*, **99**(6): 966–973.

   Araki, H., Knapp, C., Tsai, P., and Print, C. (2012) GeneSetDB: A comprehensive meta-database, statistical and visualisation framework for gene set analysis. *FEBS Open Bio*, **2**: 76–82.

   Ashworth, A. (2008) A synthetic lethal therapeutic approach: poly(adp) ribose polymerase inhibitors for the treatment of cancers deficient in dna double-strand break repair. *J Clin Oncol*, **26**(22): 3785–90.

   Audeh, M.W., Carmichael, J., Penson, R.T., Friedlander, M., Powell, B., Bell-McGuinn, K.M., Scott, C., Weitzel, J.N., Oaknin, A., Loman, N., *et al.* (2010) Oral poly(adp-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: a proof-of-concept trial. *Lancet*, **376**(9737): 245–51.

   Babyak, M.A. (2004) What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom Med*, **66**(3): 411–21.

   Bamford, S., Dawson, E., Forbes, S., Clements, J., Pettett, R., Dogan, A., Flanagan, A., Teague, J., Futreal, P.A., Stratton, M.R., *et al.* (2004) The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer*, **91**(2): 355–358.

   Barabási, A.L. and Albert, R. (1999) Emergence of scaling in random networks. *Science*, **286**(5439): 509–12.

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

   Barrat, A. and Weigt, M. (2000) On the properties of small-world network models. *The European Physical Journal B - Condensed Matter and Complex Systems*, **13**(3): 547–560.

   Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehar, J., Kryukov, G.V., Sonkin, D., *et al.* (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, **483**(7391): 603–607.

   Barry, W.T. (2016) *safe: Significance Analysis of Function and Expression*. R package version 3.14.0.

   Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K., Youn, J.Y., Ou, J., San Luis, B.J., Bandyopadhyay, S., *et al.* (2010) Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat Meth*, **7**(12): 1017–1024.

   Bass, A.J., Thorsson, V., Shmulevich, I., Reynolds, S.M., Miller, M., Bernard, B., Hinoue, T., Laird, P.W., Curtis, C., Shen, H., *et al.* (2014) Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*, **513**(7517): 202–209.

   Bateson, W. and Mendel, G. (1909) *Mendel’s principles of heredity, by W. Bateson*. University Press, Cambridge [Eng.].

   Beck, T.F., Mullikin, J.C., and Biesecker, L.G. (2016) Systematic Evaluation of

Sanger Validation of Next-Generation Sequencing Variants. *Clin Chem*, **62**(4): 647–654.

   Bell, D., Berchuck, A., Birrer, M., Chien, J., Cramer, D., Dao, F., Dhir, R., DiSaia, P., Gabra, H., Glenn, P., *et al.* (2011) Integrated genomic analyses of ovarian carcinoma. *Nature*, **474**(7353): 609–615.

   Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)*, **57**(1): 289–300.

   Bitler, B.G., Aird, K.M., Garipov, A., Li, H., Amatangelo, M., Kossenkov, A.V., Schultz, D.C., Liu, Q., Shih Ie, M., Conejo-Garcia, J.R., *et al.* (2015) Synthetic lethality by targeting ezh2 methyltransferase activity in arid1a-mutated cancers. *Nat Med*, **21**(3): 231–8.

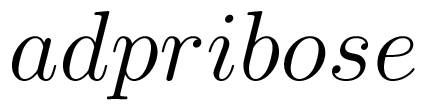
   Boettcher, M., Lawson, A., Ladenburger, V., Fredebohm, J., Wolf, J., Hoheisel, J.D., Frezza, C., and Shlomi, T. (2014) High throughput synthetic lethality screen reveals a tumorigenic role of adenylate cyclase in fumarate hydratase-deficient cancer cells. *BMC Genomics*, **15**: 158.

   Boone, C., Bussey, H., and Andrews, B.J. (2007) Exploring genetic interactions and networks with yeast. *Nat Rev Genet*, **8**(6): 437–49.

   Borgatti, S.P. (2005) Centrality and network flow. *Social Networks*, **27**(1): 55 – 71.

   Boucher, B. and Jenna, S. (2013) Genetic interaction networks: better understand to better predict. *Front Genet*, **4**: 290.

   Brin, S. and Page, L. (1998) The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, **30**(1): 107 – 117.

   Bryant, H.E., Schultz, N., Thomas, H.D., Parker, K.M., Flower, D., Lopez, E., Kyle, S., Meuth, M., Curtin, N.J., and Helleday, T. (2005) Specific killing of BRCA2 deficient tumours with inhibitors of poly polymerase. *Nature*, **434**(7035): 913–7.

   Burk, R.D., Chen, Z., Saller, C., Tarvin, K., Carvalho, A.L., Scapulatempo-Neto, C., Silveira, H.C., Fregnani, J.H., Creighton, C.J., Anderson, M.L., *et al.* (2017) Integrated genomic and molecular characterization of cervical cancer. *Nature*, **543**(7645): 378–384.

   Bussey, H., Andrews, B., and Boone, C. (2006) From worm genetic networks to complex human diseases. *Nat Genet*, **38**(8): 862–3.

   Butland, G., Babu, M., Diaz-Mejia, J.J., Bohdana, F., Phanse, S., Gold, B., Yang, W., Li, J., Gagarinova, A.G., Pogoutse, O., *et al.* (2008) esga: E. coli synthetic genetic array analysis. *Nat Methods*, **5**(9): 789–95.

   Cancer Research UK (2017) Family history and cancer genes.

[http://www.cancerresearchuk.org/about-*cancer/causes-*of-*cancer/inherited-*cancer-*genes-*and-*increased-*cancer-*risk/family-*history-*and-*inherited-*cancer-*genes](http://www.cancerresearchuk.org/about-cancer/causes-of-cancer/inherited-cancer-genes-and-increased-cancer-risk/family-history-and-inherited-cancer-genes). Accessed: 22/03/2017.

   Cancer Cell Line Encyclopedia (CCLE) (2014) Broad-Novartis Cancer Cell Line Encyclopedia. <http://www.broadinstitute.org/ccle>. Accessed: 07/11/2014.

   cBioPortal for Cancer Genomics (cBioPortal) (2017) cBioPortal for Cancer Genomics. <http://www.cbioportal.org/>. Accessed: 26/03/2017.

   Cerami, E.G., Gross, B.E., Demir, E., Rodchenkov, I., Babur, O., Anwar, N., Schultz, N., Bader, G.D., and Sander, C. (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res*, **39**(Database issue): D685–690.

   Chen, A., Beetham, H., Black, M.A., Priya, R., Telford, B.J., Guest, J., Wiggins, G.A.R., Godwin, T.D., Yap, A.S., and Guilford, P.J. (2014) E-cadherin loss alters cytoskeletal organization and adhesion in non-malignant breast cells but is insufficient to induce an epithelial-mesenchymal transition. *BMC Cancer*, **14**(1): 552.

   Chen, S. and Parmigiani, G. (2007) Meta-analysis of BRCA1 and BRCA2 penetrance. *J Clin Oncol*, **25**(11): 1329–1333.

   Chen, X. and Tompa, M. (2010) Comparative assessment of methods for aligning multiple genome sequences. *Nat Biotechnol*, **28**(6): 567–572.

   Cherniack, A.D., Shen, H., Walter, V., Stewart, C., Murray, B.A., Bowlby, R., Hu, X., Ling, S., Soslow, R.A., Broaddus, R.R., *et al.* (2017) Integrated Molecular Characterization of Uterine Carcinosarcoma. *Cancer Cell*, **31**(3): 411–423.

   Chipman, K. and Singh, A. (2009) Predicting genetic interactions with random walks on biological networks. *BMC Bioinformatics*, **10**(1): 17.

   Ciriello, G., Gatza, M.L., Beck, A.H., Wilkerson, M.D., Rhie, S.K., Pastore, A., Zhang, H., McLellan, M., Yau, C., Kandoth, C., *et al.* (2015) Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell*, **163**(2): 506–519.

   Clough, E. and Barrett, T. (2016) The Gene Expression Omnibus Database. *Methods Mol Biol*, **1418**: 93–110.

   Collingridge, D.S. (2013) A primer on quantitized data analysis and permutation testing. *Journal of Mixed Methods Research*, **7**(1): 81–97.

   Collins, F.S. and Barker, A.D. (2007) Mapping the cancer genome. Pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies. *Sci Am*, **296**(3): 50–57.

   Collins, F.S., Morgan, M., and Patrinos, A. (2003) The Human Genome Project: lessons from large-scale biology. *Science*, **300**(5617): 286–290.

   Collisson, E., Campbell, J., Brooks, A., Berger, A., Lee, W., Chmielecki, J., Beer, D., Cope, L., Creighton, C., Danilova, L., *et al.* (2014) Comprehensive molecular profiling of lung adenocarcinoma. *Nature*, **511**(7511): 543–550.

   Corcoran, R.B., Ebi, H., Turke, A.B., Coffee, E.M., Nishino, M., Cogdill, A.P., Brown, R.D., Della Pelle, P., Dias-Santagata, D., Hung, K.E., *et al.* (2012) Egfr-mediated reactivation of mapk signaling contributes to insensitivity of BRAF -mutant colorectal cancers to raf inhibition with vemurafenib. *Cancer Discovery*, **2**(3): 227–235.

   Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L., Toufighi, K., Mostafavi, S., *et al.* (2010) The genetic landscape of a cell. *Science*, **327**(5964): 425–31.

   Costanzo, M., Baryshnikova, A., Myers, C.L., Andrews, B., and Boone, C. (2011) Charting the genetic interaction map of a cell. *Curr Opin Biotechnol*, **22**(1): 66–74.

   Creighton, C.J., Morgan, M., Gunaratne, P.H., Wheeler, D.A., Gibbs, R.A., Robertson, A., Chu, A., Beroukhim, R., Cibulskis, K., Signoretti, S., *et al.* (2013) Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature*, **499**(7456): 43–49.

   Croft, D., Mundo, A.F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., Kamdar, M.R., *et al.* (2014) The Reactome pathway knowledgebase. *Nucleic Acids Res*, **42**(database issue): D472D477.

   Crunkhorn, S. (2014) Cancer: Predicting synthetic lethal interactions. *Nat Rev Drug Discov*, **13**(11): 812.

   Csardi, G. and Nepusz, T. (2006) The igraph software package for complex network research. *InterJournal*, **Complex Systems**: 1695.

   Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., *et al.* (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, **486**(7403): 346–352.

   Dai, X., Li, T., Bai, Z., Yang, Y., Liu, X., Zhan, J., and Shi, B. (2015) Breast cancer intrinsic subtype classification, clinical use and future trends. *Am J Cancer Res*, **5**(10): 2929–2943.

   Davierwala, A.P., Haynes, J., Li, Z., Brost, R.L., Robinson, M.D., Yu, L., Mnaimneh, S., Ding, H., Zhu, H., Chen, Y., *et al.* (2005) The synthetic genetic interaction spectrum of essential genes. *Nat Genet*, **37**(10): 1147–1152.

   Demir, E., Babur, O., Rodchenkov, I., Aksoy, B.A., Fukuda, K.I., Gross, B., Sumer, O.S., Bader, G.D., and Sander, C. (2013) Using biological pathway data with Paxtools. *PLoS Comput Biol*, **9**(9): e1003194.

   Deshpande, R., Asiedu, M.K., Klebig, M., Sutor, S., Kuzmin, E., Nelson, J., Piotrowski, J., Shin, S.H., Yoshida, M., Costanzo, M., *et al.* (2013) A comparative

genomic approach for identifying synthetic lethal interactions in human cancer. *Cancer Res*, **73**(20): 6128–36.

   Dickson, D. (1999) Wellcome funds cancer database. *Nature*, **401**(6755): 729.

   Dienstmann, R. and Tabernero, J. (2011) BRAF as a target for cancer therapy. *Anticancer Agents Med Chem*, **11**(3): 285–95.

   Dijkstra, E.W. (1959) A note on two problems in connexion with graphs. *Numerische Mathematik*, **1**(1): 269–271.

   Dixon, S.J., Fedyshyn, Y., Koh, J.L., Prasad, T.S., Chahwan, C., Chua, G., Toufighi, K., Baryshnikova, A., Hayles, J., Hoe, K.L., *et al.* (2008) Significant conservation of synthetic lethal genetic interaction networks between distantly related eukaryotes. *Proc Natl Acad Sci U S A*, **105**(43): 16653–8.

   Dorogovtsev, S.N. and Mendes, J.F. (2003) *Evolution of networks: From biological nets to the Internet and WWW*. Oxford University Press, USA.

   Erds, P. and Rényi, A. (1959) On random graphs I. *Publ Math Debrecen*, **6**: 290–297.

   Erds, P. and Rényi, A. (1960) On the evolution of random graphs. In *Publ. Math. Inst. Hung. Acad. Sci*, volume 5, 17–61.

   Eroles, P., Bosch, A., Perez-Fidalgo, J.A., and Lluch, A. (2012) Molecular biology in breast cancer: intrinsic subtypes and signaling pathways. *Cancer Treat Rev*, **38**(6): 698–707.

   Ezkurdia, I., Juan, D., Rodriguez, J.M., Frankish, A., Diekhans, M., Harrow, J., Vazquez, J., Valencia, A., and Tress, M.L. (2014) Multiple evidence strands suggest that there may be as few as 19 000 human protein-coding genes. *Human Molecular Genetics*, **23**(22): 5866.

   Farmer, H., McCabe, N., Lord, C.J., Tutt, A.N., Johnson, D.A., Richardson, T.B., Santarosa, M., Dillon, K.J., Hickson, I., Knights, C., *et al.* (2005) Targeting the dna repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*, **434**(7035): 917–21.

   Fawcett, T. (2006) An introduction to ROC analysis. *Pattern Recognition Letters*, **27**(8): 861 – 874. {ROC} Analysis in Pattern Recognition.

   Fece de la Cruz, F., Gapp, B.V., and Nijman, S.M. (2015) Synthetic lethal vulnerabilities of cancer. *Annu Rev Pharmacol Toxicol*, **55**: 513–531.

   Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F. (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*, **136**(5): E359–386.

   Fisher, R.A. (1919) Xv.the correlation between relatives on the supposition of mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, **52**(02): 399–433.

   Fong, P.C., Boss, D.S., Yap, T.A., Tutt, A., Wu, P., Mergui-Roelvink, M., Mortimer, P., Swaisland, H., Lau, A., O’Connor, M.J., *et al.* (2009) Inhibition of poly(adp-ribose) polymerase in tumors from BRCA mutation carriers. *N Engl J Med*, **361**(2): 123–34.

   Fong, P.C., Yap, T.A., Boss, D.S., Carden, C.P., Mergui-Roelvink, M., Gourley, C., De Greve, J., Lubinski, J., Shanley, S., Messiou, C., *et al.* (2010) Poly(adp)-ribose polymerase inhibition: frequent durable responses in BRCA carrier ovarian cancer correlating with platinum-free interval. *J Clin Oncol*, **28**(15): 2512–9.

   Fraser, A. (2004) Towards full employment: using RNAi to find roles for the redundant. *Oncogene*, **23**(51): 8346–52.

   Futreal, P.A., Kasprzyk, A., Birney, E., Mullikin, J.C., Wooster, R., and Stratton, M.R. (2001) Cancer and genomics. *Nature*, **409**(6822): 850–852.

   Gao, B. and Roux, P.P. (2015) Translational control by oncogenic signaling pathways. *Biochimica et Biophysica Acta*, **1849**(7): 753–65.

   Gatza, M.L., Kung, H.N., Blackwell, K.L., Dewhirst, M.W., Marks, J.R., and Chi, J.T. (2011) Analysis of tumor environmental response and oncogenic pathway

activation identifies distinct basal and luminal features in HER2-related breast tumor subtypes. *Breast Cancer Res*, **13**(3): R62.

   Gatza, M.L., Lucas, J.E., Barry, W.T., Kim, J.W., Wang, Q., Crawford, M.D., Datto, M.B., Kelley, M., Mathey-Prevot, B., Potti, A., *et al.* (2010) A pathway-based classification of human breast cancer. *Proc Natl Acad Sci USA*, **107**(15): 6994–6999.

   Gatza, M.L., Silva, G.O., Parker, J.S., Fan, C., and Perou, C.M. (2014) An integrated genomics approach identifies drivers of proliferation in luminal-subtype human breast cancer. *Nat Genet*, **46**(10): 1051–1059.

   Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*, **5**(10): R80.

   Genz, A. and Bretz, F. (2009) Computation of multivariate normal and t probabilities. In *Lecture Notes in Statistics*, volume 195. Springer-Verlag, Heidelberg.

   Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2016) *mvtnorm: Multivariate Normal and t Distributions*. R package version 1.0-5. URL.

   Gilbert, W. and Maxam, A. (1973) The nucleotide sequence of the lac operator. *Proceedings of the National Academy of Sciences*, **70**(12): 3581–3584.

   Git, A., Dvinge, H., Salmon-Divon, M., Osborne, M., Kutter, C., Hadfield, J., Bertone, P., and Caldas, C. (2010) Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. *RNA*, **16**(5): 991–1006.

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

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

   Guilford, P., Hopkins, J., Harraway, J., McLeod, M., McLeod, N., Harawira, P., Taite, H., Scoular, R., Miller, A., and Reeve, A.E. (1998) E-cadherin germline mutations in familial gastric cancer. *Nature*, **392**(6674): 402–5.

   Guilford, P., Humar, B., and Blair, V. (2010) Hereditary diffuse gastric cancer: translation of CDH1 germline mutations into clinical practice. *Gastric Cancer*, **13**(1): 1–10.

   Hajian-Tilaki, K. (2013) Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian J Intern Med*, **4**(2): 627–635.

   Hammerman, P.S., Lawrence, M.S., Voet, D., Jing, R., Cibulskis, K., Sivachenko,

A., Stojanov, P., McKenna, A., Lander, E.S., Gabriel, S., *et al.* (2012) Comprehensive genomic characterization of squamous cell lung cancers. *Nature*, **489**(7417): 519–525.

   Han, J.D.J., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J.M., Cusick, M.E., Roth, F.P., *et al.* (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*, **430**(6995): 88–93.

   Hanahan, D. and Weinberg, R.A. (2000) The hallmarks of cancer. *Cell*, **100**(1): 57–70.

   Hanahan, D. and Weinberg, R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**(5): 646–674.

   Hanna, S. (2003) Cancer incidence in new zealand (2003-2007). In D. Forman, D. Bray F Brewster, C. Gombe Mbalawa, B. Kohler, M. Piñeros, E. Steliarova-Foucher, R. Swaminathan, and J. Ferlay (editors), *Cancer Incidence in Five Continents*, volume X, 902–907. International Agency for Research on Cancer, Lyon, France. Electronic version <http://ci5.iarc.fr> Accessed 22/03/2017.

   Heiskanen, M., Bian, X., Swan, D., and Basu, A. (2014) caArray microarray database in the cancer biomedical informatics gridTM (caBIGTM). *Cancer Research*, **67**(9 Supplement): 3712–3712.

   Heiskanen, M.A. and Aittokallio, T. (2012) Mining high-throughput screens for cancer drug targets-lessons from yeast chemical-genomic profiling and synthetic lethality. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, **2**(3): 263–272.

   Herschkowitz, J.I., Simin, K., Weigman, V.J., Mikaelian, I., Usary, J., Hu, Z., Rasmussen, K.E., Jones, L.P., Assefnia, S., Chandrasekharan, S., *et al.* (2007) Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol*, **8**(5): R76.

   Hillenmeyer, M.E. (2008) The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science*, **320**: 362–365.

   Hoadley, K.A., Yau, C., Wolf, D.M., Cherniack, A.D., Tamborero, D., Ng, S., Leiserson, M.D., Niu, B., McLellan, M.D., Uzunangelov, V., *et al.* (2014) Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*, **158**(4): 929–944.

   Hoehndorf, R., Hardy, N.W., Osumi-Sutherland, D., Tweedie, S., Schofield, P.N., and Gkoutos, G.V. (2013) Systematic analysis of experimental phenotype data reveals gene functions. *PLoS ONE*, **8**(4): e60847.

   Holm, S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**(2): 65–70.

   Holme, P. and Kim, B.J. (2002) Growing scale-free networks with tunable clustering. *Physical Review E*, **65**(2): 026107.

   Hopkins, A.L. (2008) Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol*, **4**(11): 682–690.

   Hu, Z., Fan, C., Oh, D.S., Marron, J.S., He, X., Qaqish, B.F., Livasy, C., Carey, L.A., Reynolds, E., Dressler, L., *et al.* (2006) The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics*, **7**: 96.

   Huang, E., Cheng, S., Dressman, H., Pittman, J., Tsou, M., Horng, C., Bild, A., Iversen, E., Liao, M., Chen, C., *et al.* (2003) Gene expression predictors of breast cancer outcomes. *Lancet*, **361**: 1590–1596.

   Illumina, Inc (Illumina) (2017) Sequencing and array-based solutions for genetic research. <https://www.illumina.com/>. Accessed: 26/03/2017.

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

   Internationl Human Genome Sequencing Consortium (IHGSC) (2004) Finishing the euchromatic sequence of the human genome. *Nature*, **431**(7011): 931–945.

   Jerby-Arnon, L., Pfetzer, N., Waldman, Y., McGarry, L., James, D., Shanks, E.,

Seashore-Ludlow, B., Weinstock, A., Geiger, T., Clemons, P., *et al.* (2014) Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. *Cell*, **158**(5): 1199–1209.

   Ju, Z., Liu, W., Roebuck, P.L., Siwak, D.R., Zhang, N., Lu, Y., Davies, M.A., Akbani, R., Weinstein, J.N., Mills, G.B., *et al.* (2015) Development of a robust classifier for quality control of reverse-phase protein arrays. *Bioinformatics*, **31**(6): 912.

   Kaelin, Jr, W. (2005) The concept of synthetic lethality in the context of anticancer therapy. *Nat Rev Cancer*, **5**(9): 689–98.

   Kaelin, Jr, W. (2009) Synthetic lethality: a framework for the development of wiser cancer therapeutics. *Genome Med*, **1**: 99.

   Kandoth, C., Schultz, N., Cherniack, A.D., Akbani, R., Liu, Y., Shen, H., Robertson, A.G., Pashtan, I., Shen, R., Benz, C.C., *et al.* (2013) Integrated genomic characterization of endometrial carcinoma. *Nature*, **497**(7447): 67–73.

   Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., *et al.* (2001) Functional annotation of a full-length mouse cDNA collection. *Nature*, **409**(6821): 685–690.

   Kelley, R. and Ideker, T. (2005) Systematic interpretation of genetic interactions using protein networks. *Nat Biotech*, **23**(5): 561–566.

   Kelly, S., Chen, A., Guilford, P., and Black, M. (2017a) Synthetic lethal interaction prediction of target pathways in E-cadherin deficient breast cancers. Submitted to BMC Genomics.

   Kelly, S.T. (2013) *Statistical Predictions of Synthetic Lethal Interactions in Cancer*. Dissertation, University of Otago.

   Kelly, S.T., Single, A.B., Telford, B.J., Beetham, H.G., Godwin, T.D., Chen, A., Black, M.A., and Guilford, P.J. (2017b) Towards HDGC chemoprevention: vulnerabilities in E-cadherin-negative cells identified by genome-wide interrogation of isogenic cell lines and whole tumors. Submitted to Cancer Prev Res.

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

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

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

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

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

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

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

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

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

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

   Lee, A.Y., Perreault, R., Harel, S., Boulier, E.L., Suderman, M., Hallett, M., and Jenna, S. (2010a) Searching for signaling balance through the identification of genetic

interactors of the rab guanine-nucleotide dissociation inhibitor gdi-1. *PLoS ONE*, **5**(5): e10624.

   Lee, I., Lehner, B., Vavouri, T., Shin, J., Fraser, A.G., and Marcotte, E.M. (2010b) Predicting genetic modifier loci using functional gene networks. *Genome Research*, **20**(8): 1143–1153.

   Lee, I. and Marcotte, E.M. (2009) Effects of functional bias on supervised learning of a gene network model. *Methods Mol Biol*, **541**: 463–75.

   Lehner, B., Crombie, C., Tischler, J., Fortunato, A., and Fraser, A.G. (2006) Systematic mapping of genetic interactions in caenorhabditis elegans identifies common modifiers of diverse signaling pathways. *Nat Genet*, **38**(8): 896–903.

   Li, X.J., Mishra, S.K., Wu, M., Zhang, F., and Zheng, J. (2014) Syn-lethality: An integrative knowledge base of synthetic lethality towards discovery of selective anticancer therapies. *Biomed Res Int*, **2014**: 196034.

   Linehan, W.M., Spellman, P.T., Ricketts, C.J., Creighton, C.J., Fei, S.S., Davis, C., Wheeler, D.A., Murray, B.A., Schmidt, L., Vocke, C.D., *et al.* (2016) Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. *N Engl J Med*, **374**(2): 135–145.

   Lokody, I. (2014) Computational modelling: A computational crystal ball. *Nature Reviews Cancer*, **14**(10): 649–649.

   Lord, C.J., Tutt, A.N., and Ashworth, A. (2015) Synthetic lethality and cancer therapy: lessons learned from the development of PARP inhibitors. *Annu Rev Med*, **66**: 455–470.

   Lu, X., Kensche, P.R., Huynen, M.A., and Notebaart, R.A. (2013) Genome evolution predicts genetic interactions in protein complexes and reveals cancer drug targets. *Nat Commun*, **4**: 2124.

   Lu, X., Megchelenbrink, W., Notebaart, R.A., and Huynen, M.A. (2015) Predicting human genetic interactions from cancer genome evolution. *PLoS One*, **10**(5): e0125795.

   Lum, P.Y., Armour, C.D., Stepaniants, S.B., Cavet, G., Wolf, M.K., Butler, J.S., Hinshaw, J.C., Garnier, P., Prestwich, G.D., Leonardson, A., *et al.* (2004) Discovering modes of action for therapeutic compounds using a genome-wide screen of yeast heterozygotes. *Cell*, **116**(1): 121–137.

   Luo, J., Solimini, N.L., and Elledge, S.J. (2009) Principles of Cancer Therapy: Oncogene and Non-oncogene Addiction. *Cell*, **136**(5): 823–837.

   Mattison, J., van der Weyden, L., Hubbard, T., and Adams, D.J. (2009) Cancer gene discovery in mouse and man. *Biochim Biophys Acta*, **1796**(2): 140–161.

   Maxam, A.M. and Gilbert, W. (1977) A new method for sequencing DNA. *Proceedings of the National Academy of Science*, **74**(2): 560–564.

   McCourt, C.M., McArt, D.G., Mills, K., Catherwood, M.A., Maxwell, P., Waugh, D.J., Hamilton, P., O’Sullivan, J.M., and Salto-Tellez, M. (2013) Validation of next generation sequencing technologies in comparison to current diagnostic gold standards for BRAF, EGFR and KRAS mutational analysis. *PLoS ONE*, **8**(7): e69604.

   McLachlan, J., George, A., and Banerjee, S. (2016) The current status of parp inhibitors in ovarian cancer. *Tumori*, **102**(5): 433–440.

   McLendon, R., Friedman, A., Bigner, D., Van Meir, E.G., Brat, D.J., Mastrogianakis, G.M., Olson, J.J., Mikkelsen, T., Lehman, N., Aldape, K., *et al.* (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, **455**(7216): 1061–1068.

   Miles, D.W. (2001) Update on HER-2 as a target for cancer therapy: herceptin in the clinical setting. *Breast Cancer Res*, **3**(6): 380–384.

   Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*, **5**(7): 621–628.

   Muzny, D.M., Bainbridge, M.N., Chang, K., Dinh, H.H., Drummond, J.A., Fowler, G., Kovar, C.L., Lewis, L.R., Morgan, M.B., Newsham, I.F., *et al.* (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, **487**(7407): 330–337.

   Nagalla, S., Chou, J.W., Willingham, M.C., Ruiz, J., Vaughn, J.P., Dubey, P., Lash, T.L., Hamilton-Dutoit, S.J., Bergh, J., Sotiriou, C., *et al.* (2013) Interactions between immunity, proliferation and molecular subtype in breast cancer prognosis. *Genome Biol*, **14**(4): R34.

   Neeley, E.S., Kornblau, S.M., Coombes, K.R., and Baggerly, K.A. (2009) Variable slope normalization of reverse phase protein arrays. *Bioinformatics*, **25**(11): 1384.

   Pandey, G., Zhang, B., Chang, A.N., Myers, C.L., Zhu, J., Kumar, V., and Schadt, E.E. (2010) An integrative multi-network and multi-classifier approach to predict genetic interactions. *PLoS Comput Biol*, **6**(9).

   Parker, J., Mullins, M., Cheung, M., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., *et al.* (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology*, **27**(8): 1160–1167.

   Peltonen, L. and McKusick, V.A. (2001) Genomics and medicine. Dissecting human disease in the postgenomic era. *Science*, **291**(5507): 1224–1229.

   Pereira, B., Chin, S.F., Rueda, O.M., Vollan, H.K., Provenzano, E., Bardwell, H.A., Pugh, M., Jones, L., Russell, R., Sammut, S.J., *et al.* (2016) Erratum: The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. *Nat Commun*, **7**: 11908.

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

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

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

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

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

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

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

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

   Rung, J. and Brazma, A. (2013) Reuse of public genome-wide gene expression data. *Nat Rev Genet*, **14**(2): 89–99.

   Rustici, G., Kolesnikov, N., Brandizi, M., Burdett, T., Dylag, M., Emam, I., Farne, A., Hastings, E., Ison, J., Keays, M., *et al.* (2013) ArrayExpress update–trends in database growth and links to data analysis tools. *Nucleic Acids Res*, **41**(Database issue): D987–990.

   Ryan, C., Lord, C., and Ashworth, A. (2014) Daisy: Picking synthetic lethals from cancer genomes. *Cancer Cell*, **26**(3): 306–308.

   Sander, J.D. and Joung, J.K. (2014) Crispr-cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*, **32**(4): 347–55.

   Sanger, F. and Coulson, A. (1975) A rapid method for determining sequences in dna by primed synthesis with dna polymerase. *Journal of Molecular Biology*, **94**(3): 441 – 448.

   Scheuer, L., Kauff, N., Robson, M., Kelly, B., Barakat, R., Satagopan, J., Ellis, N.,

Hensley, M., Boyd, J., Borgen, P., *et al.* (2002) Outcome of preventive surgery and screening for breast and ovarian cancer in BRCA mutation carriers. *J Clin Oncol*, **20**(5): 1260–1268.

   Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T. (2005) Rocr: visualizing classifier performance in r. *Bioinformatics*, **21**(20): 7881.

   Slurm development team (Slurm) (2017) Slurm workload manager. <https://slurm.schedmd.com/>. Accessed: 25/03/2017.

   Sørlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., *et al.* (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA*, **98**(19): 10869–10874.

   Stajich, J.E. and Lapp, H. (2006) Open source tools and toolkits for bioinformatics: significance, and where are we? *Brief Bioinformatics*, **7**(3): 287–296.

   Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009) The cancer genome. *Nature*, **458**(7239): 719–724.

   Ström, C. and Helleday, T. (2012) Strategies for the use of poly(adenosine diphosphate ribose) polymerase (parp) inhibitors in cancer therapy. *Biomolecules*, **2**(4): 635–649.

   Sun, C., Wang, L., Huang, S., Heynen, G.J.J.E., Prahallad, A., Robert, C., Haanen, J., Blank, C., Wesseling, J., Willems, S.M., *et al.* (2014) Reversible and adaptive resistance to BRAF(v600e) inhibition in melanoma. *Nature*, **508**(7494): 118–122.

   Taylor, I.W., Linding, R., Warde-Farley, D., Liu, Y., Pesquita, C., Faria, D., Bull, S., Pawson, T., Morris, Q., and Wrana, J.L. (2009) Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nat Biotechnol*, **27**(2): 199–204.

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

   The 1000 Genomes Project Consortium (1000 Genomes) (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**(7319): 1061–1073.

   The Cancer Genome Atlas Research Network (TCGA) (2012) Comprehensive molecular portraits of human breast tumours. *Nature*, **490**(7418): 61–70.

   The Cancer Genome Atlas Research Network (TCGA) (2017) The Cancer Genome Atlas Project. <https://cancergenome.nih.gov/>. Accessed: 26/03/2017.

   The Cancer Society of New Zealand (Cancer Society of NZ) (2017) What is cancer? [https://otago-*southland.cancernz.org.nz/en/cancer-*information/other-*links/what-*is-*cancer-*3/](https://otago-southland.cancernz.org.nz/en/cancer-information/other-links/what-is-cancer-3/). Accessed: 22/03/2017.

   The Catalogue Of Somatic Mutations In Cancer (COSMIC) (2016) Cosmic: The catalogue of somatic mutations in cancer. <http://cancer.sanger.ac.uk/cosmic>. Release 79 (23/08/2016), Accessed: 05/02/2017.

   The Comprehensive R Archive Network (CRAN) (2017) Cran. [https://cran.r-\_project.org/](https://cran.r-project.org/). Accessed: 24/03/2017.

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

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

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

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

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

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

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

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

   Tong, A.H., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G.F., Brost, R.L., Chang, M., *et al.* (2004) Global mapping of the yeast genetic interaction network. *Science*, **303**(5659): 808–13.

   Travers, J. and Milgram, S. (1969) An experimental study of the small world problem. *Sociometry*, **32**(4): 425–443.

   Tsai, H.C., Li, H., Van Neste, L., Cai, Y., Robert, C., Rassool, F.V., Shin, J.J., Harbom, K.M., Beaty, R., Pappou, E., *et al.* (2012) Transient low doses of dna-demethylating agents exert durable antitumor effects on hematological and epithelial tumor cells. *Cancer Cell*, **21**(3): 430–46.

   Tutt, A., Robson, M., Garber, J.E., Domchek, S.M., Audeh, M.W., Weitzel, J.N., Friedlander, M., Arun, B., Loman, N., Schmutzler, R.K., *et al.* (2010) Oral

poly(adp-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and advanced breast cancer: a proof-of-concept trial. *Lancet*, **376**(9737): 235–44.

   van der Meer, R., Song, H.Y., Park, S.H., Abdulkadir, S.A., and Roh, M. (2014) RNAi screen identifies a synthetic lethal interaction between PIM1 overexpression and PLK1 inhibition. *Clinical Cancer Research*, **20**(12): 3211–3221.

   van Steen, K. (2012) Travelling the world of genegene interactions. *Briefings in Bioinformatics*, **13**(1): 1–19.

   van Steen, M. (2010) *Graph Theory and Complex Networks: An Introduction*. Maarten van Steen, VU Amsterdam.

   Vargas, J.J., Gusella, G., Najfeld, V., Klotman, M., and Cara, A. (2004) Novel integrase-defective lentiviral episomal vectors for gene transfer. *Hum Gene Ther*, **15**: 361–372.

   Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., and Kinzler, K.W. (2013) Cancer genome landscapes. *Science*, **339**(6127): 1546–1558.

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

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

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

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

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

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

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

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

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

   Wong, S.L., Zhang, L.V., Tong, A.H.Y., Li, Z., Goldberg, D.S., King, O.D., Lesage, G., Vidal, M., Andrews, B., Bussey, H., *et al.* (2004) Combining biological networks to predict genetic interactions. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(44): 15682–15687.

   World Health Organization (WHO) (2017) Fact sheet: Cancer. <http://www.who.int/mediacentre/factsheets/fs297/en/>. Updated February 2017, Accessed: 22/03/2017.

   Wu, M., Li, X., Zhang, F., Li, X., Kwoh, C.K., and Zheng, J. (2014) In silico prediction of synthetic lethality by meta-analysis of genetic interactions, functions, and pathways in yeast and human cancer. *Cancer Inform*, **13**(Suppl 3): 71–80.

   Yu, H. (2002) Rmpi: Parallel statistical computing in r. *R News*, **2**(2): 10–14.

   Zhang, F., Wu, M., Li, X.J., Li, X.L., Kwoh, C.K., and Zheng, J. (2015) Predicting essential genes and synthetic lethality via influence propagation in signaling pathways of cancer cell fates. *J Bioinform Comput Biol*, **13**(3): 1541002.

   Zhang, J., Baran, J., Cros, A., Guberman, J.M., Haider, S., Hsu, J., Liang, Y., Rivkin, E., Wang, J., Whitty, B., *et al.* (2011) International cancer genome

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

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

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