Contents

List of Figures

List of Tables

Glossary

bioinformatics	Statistical or computational approaches to biological data or research tools.
bisulfite-Seq	Methylome data from sequencing bisulfite treated DNA.
CAGE-Seq	Transcriptome data from cap analysis of gene expression.
E-cadherin	Epithelial cadherin (calcium-dependent adhesion), a cell-adhesion protein encoded by the tumour suppressor gene, *CDH1*.
epigenome	An analysis of epigenetic modifications of all genes in the genome.
exome	A sequencing approach designed to generate data enriched for coding genes within the genome.
	.
metabolome	An analysis of all the metabolites and enzymes in the cell.
metagenomics	An analysis of all of the genes and genomes in a community.
microRNA	Short RNA molecules generally regarded to regulate gene expression by binding to mRNA.
omics	A rationale for applying molecular studies across all of the genes in the genome or biomolecules in the cell.
Pan cancer	Study or analysis orf cancers across tissue of origin.
pleiotropy	A gene which has multiple biological functions.

proteome	An analysis of all the proteins expressed from the genome.
RNA-Seq	Transcriptome data from sequencing RNA.
Sanger	A dideoxy chain termination method for DNA sequencing (named after Fred Sanger).
synthetic lethal	Genetic interactions where inactivation of multiple genes is inviable (or deleterious) which are viable if inactivated separately.
transcriptome	An analysis of all of the genes expressed in the genome.

Acronyms

ANOVA	Analysis of Variance.
AUROC	Area under the receiver operating characteristic (curve).
BiSEp	Bimodal Subsetting Expression.
CCLE	Cancer Cell Line Encyclopaedia.
cDNA	Complementary deoxyribonucleic acid (from mRNA).
ChIP-Seq	Chromatin immunopreciptation sequencing.
COSMIC	Catalogue Of Somatic Mutations In Cancer.
DAISY	Data mining synthetic lethal identification pipeline.
DDBJ	DNA Data Bank of Japan.
DNA	Deoxyribonucleic acid.
EMBL	European Molecular Biology Laboratory.
EMT	Epithelial-mesenchymal transition.
ENA	The European Nucleotide Archive.
GEO	Gene Expression Omnibus.
GO	Gene Ontology.
HDAC	Histone deacetylase.
HDGC	Hereditary diffuse gastric cancer.
HLRCC	Hereditary leiomyomatosis and renal cell carcinoma.
ICGC	International Cancer Genome Consortium.
InDel	Insertion or deletion (in DNA sequence).
JAK	Janus kinase.
lncRNA	Long non-coding ribonucleic acid.

miRNA	microRNA.
mRNA	Messenger ribonucleic acid.
NCBI	National Center for Biotechnology Information (in the USA).
NCI	National Cancer Institute (in the USA).
NGS	Next-generation sequencing.
NHGRI	National Human Genome Research Institute (in the USA).
NIG	National Institute of Genetics (in Japan).
NIH	National Institutes of Health (in the USA).
PCR	Polymerase chain reaction.
PI3K	Phosphoinositide 3-kinase.
qPCR	Quantitative (real-time) polymerase chain reaction.
RFLP	Restriction fragment length polymorphism.
RNA	Ribonucleic acid.
RNAi	Ribonucleic acid interference.
RRBS	Reduced representation bisulfite sequencing.
SGIs	synthetic genetic interaction.
siRNA	Short interfering ribonucleic acid.
SLIPT	Synthetic lethal interaction prediction tool.
SNP	Single nucleotide polymorphism.
TCGA	The Cancer Genome Atlas (genomics project).

Chapter 1

Introduction and Literature Review

The thesis presents research into genetic interactions using genomics data and bioinformatics approaches. This chapter introduces the recent developments in genomics and bioinformatics, particularly in their application to cancer research. Studies of synthetic lethal interactions, which have fundamental importance in genetics in both model organisms and renewed relevance in cancer biology specifically, will be discussed and reviewed in detail. A bioinformatic approach to synthetic lethal interactions enables a wider exploration of the function of genes and proteins in cancer cells, in contrast with candidate gene and experimental screening approaches. Synthetic lethal drug design aims to develop treatments with specificity against loss of function mutations in tumour suppressor genes, such as *CDH1* (which encodes E-cadherin) and was the focus of the analysis in this thesis. The role of *CDH1* in cellular and cancer biology is therefore also briefly reviewed.

1.1 Cancer Research in the Post-Genomic Era

Genomic technologies could significantly impacts on the clinical treatment of cancers along with wider applications of genetics. These technologies enable focused genetics investigations on candidates selected from bioinformatics analysis of genomics data. Facilitated by rapidly developing technologies, large-scale projects have investigated populations (1000 Genomes, 2010), cancers (Dickson, 1999; Zhang *et al.*, 2011), and functional genomics (Kawai *et al.*, 2001; ENCODE, 2004). However, genomic technologies have yet to be widely adopted in healthcare or oncology. Bioinformatics analysis for interpretation of genomic data is one of the main approaches to address this disparity. Here, I outline the cancer genomics projects and findings which have led to

availability of genomics data used in this thesis and recent findings in cancer research which demonstrates the potential applications of using this data.

1.1.1 Cancer is a Global Health Issue

Cancers are diseases of malignant cellular growth which typically involve tumour formation, invasion of tissues and spread to other organs. Cancers are the second leading cause of death globally (WHO, 2017), with an estimated annual incidence of 14.1 million cases and annual mortality of 8.2 million people (Ferlay *et al.*, 2015). Breast and stomach cancers are among the most prevalent cancers. Breast cancer is the most common cancer in women and has an estimated annual incidence of 1.6 million cases and mortality of 520,000 people. Stomach cancer has an estimated annual incidence of 950,000 cases and a mortality of 723,000 people. Cancer is also a major health concern here in New Zealand, with 19,100 people (including 2500 cases of breast cancer and 370 cases of stomach cancer) diagnosed annually (Hanna, 2003), near the highest incidence (age-standardised per capita) of cancer in the world (Ferlay *et al.*, 2015).

While environmental factors often also have a role, genetics has an important role in cancer risk. Most cancers occur more frequently with age and family history. Cancers arise from dysregulated cellular growth or differentiation from stem cells. These can occur through genetic mutations or alterations in gene regulation or expression which generally accumulate as the disease develops. Therefore, early diagnosis is important to ensure patient survival and quality of life. Identification of patients with genetic variants or family histories at a high-risk of particular cancers is an important health issue. These high-risk individuals are regularly monitored for some cancers are sometimes offered preventative surgery or treatment for pre-cancerous tissue (Guilford *et al.*, 2010; Scheuer *et al.*, 2002).

Chemotherapy is a treatment for many advanced stage cancer, designed to inhibit rapidly growing cells. However, this approach often has severe adverse effects, a narrow therapeutic window, and is not suitable for chemopreventative application in many cases (Kaelin, Jr, 2009). Patients at high-risk of cancers, are offered surveillence and preventative surgery but these approaches are not completely effective at preventing cancers and may impact on quality of life (Guilford *et al.*, 2010). Alternative chemoprevention and treatment strategies based on molecular biology and other fields are being investigated, including targeted molecular therapeutics (Bozovic-Spasojevic *et al.*, 2012; Kelly *et al.*, unpublished).

1.1.1.1 The Genetics and Molecular Biology of Cancers

Cancers involve dysregulation of genes including mutations which occur during a patient's lifetime and hereditary mutations which predispose them to high-risk cancers (American Cancer Society, 2017; Guilford *et al.*, 1998; NCI, 2015). Due to these familial cancer syndromes, hereditary risk factors, and the molecular changes occurring in cancers, they are in part a genetic disease involving many cancer genes Stratton *et al.* (2009); Vogelstein *et al.* (2013). The mutations increase the risk of cancer with age. An association of cancer incidence with the stem cell divisions in which mutations could occur across tissue types, suggests that cancers may be inseparably coupled with aging (Tomasetti and Vogelstein, 2015).

Hanahan and Weinberg (2000) propose the "hallmarks of cancer", molecular and cellular traits shared across cancers. These form the basis of a rational approach to the complex changes that occur in cancer. These traits include limitless replication potential, signals for indefinite growth, and invasive or metastatic capabilities. Cancers also evade apoptosis, the immune system, and sustain angiogenesis and energy metabolism (Hanahan and Weinberg, 2011). To achieve this, cancer cells change their genomes and the tumour microenvironment. Genomic instability has a role in the survival and proliferation of cancer cells and the progression of disease, as these malignant characteristics are acquired. Identifying the genetic mechanisms involved in the acquisition of these traits is important for understanding and effectively inhibiting cancer.

Gene expression can be used to characterise breast cancers. The "intrinsic subtypes" identified were characterised by estrogen receptor, *HER2*, and basal, epithelial signalling (Perou *et al.*, 2000). The expression profiles were similar across independent samples of the same tumour or the same patient and therefore represent the molecular state of a tumourThe molecular intrinsic subtypes "luminal A", "luminal B", "HER2-enriched", "basal-like", and "normal-like" have been replicated across microarray studies (Hu *et al.*, 2006), with their relevance to prognosis demonstrated, and a 50-gene subtype predictor developed (Parker *et al.*, 2009; Sørlie *et al.*, 2001). Despite specific differences in subtyping, there is widespread agreement that distinguishing luminal, HER2-enriched, and triple negative tumours has prognostic importance for patients (Dai *et al.*, 2015). The "Pan cancer" The Cancer Genome Atlas project (discussed in Section 1.1.3.2) demonstrates the importance of molecular similarities and differences between cancers across cancer tissue types (Weinstein *et al.*, 2013).

Gatza *et al.* (2010) used gene signatures for 18 cellular pathways in breast cancer to define subtypes with distinct molecular pathway activity. A "metagene" is a measure

pathway activation (derived from eigenvectors or principal components) which gives a consistent signal of gene expression (Anjomshoaa *et al.*, 2008; Huang *et al.*, 2003; Nagalla *et al.*, 2013). Unsupervised hierarchical clustering defined subtypes with common pathway activity, despite variation in mutations. These subtypes intrinsic subtypes and provide finer molecular stratification with a functional basis (Gatza *et al.*, 2014; Parker *et al.*, 2009). The subtypes with shared pathway activity have similar molecular characteristics (such as DNA copy number) and clinical properties including prognosis.

1.1.2 The Genomics Revolution in Cancer Research

Genomic technologies have transformed genetics research, including the study of health and disease (Lander, 2011; Peltonen and McKusick, 2001). Genomics enables systematic, unbiased studies across all of the genes in the genome. Cancer genomics investigations have been widely applied to different tissues across molecular profiles (Bamford *et al.*, 2004; Weinstein *et al.*, 2013; Zhang *et al.*, 2011). Genome sequencing technologies continue to improve and become feasible in a wider range of applications.

Genomics has been used in many investigations but relatively few of the potential applications in healthcare have been realised yet. Cancer genomics, in particular, could have numerous benefits across diagnostics, prognosis, management, and treatment. While direct impact of genomic on the clinic has been limited thus far, the cancer genes and therapeutic targets identified have begun to be introduced in the clinic (Stratton *et al.*, 2009).

1.1.2.1 High-Throughput Technologies

Microarrays are a high-throughput molecular technique, reducing the cost, time, and labour required to study genes at the "genome scale". Microarrays can detect genotype or expression across many genes, making it feasible to perform on a statistically informative number of samples. Microarrays are manufactured with probes which measure binding of nucleotides which either detect the presence of a sequence such as a single nucleotide polymorphism (SNP) or quantify sequences for DNA copy number, gene expression, or DNA CpG methylation. In addition to being more versatile with higher-throughput than polymerase chain reaction (PCR) based techniques, microarrays are considered cost-effective, particularly when scaled up to a large number of probes. These remained popular, especially in large-scale projects, during the introduction of newer technologies due to reliability and relatively lower cost for testing processing samples.

The introduction of massively parallel sequencing technologies has further expanded high-throughput molecular studies and the availability of genomics data. This "Next-Generation Sequencing" (NGS) enables rapid *de novo* genome and transcriptome sequencing, in addition to gene expression studies. However, the cost of sequencing for gene expression or DNA methylation studies is still considerably higher than a microarray study, limiting feasible sample sizes, and NGS studies have large compute requirements to handle the raw data. In many cases, the benefits of NGS technologies outweigh the additional cost. NGS technologies have the advantage of greater potential accuracy and sensitivity than microarrays. NGS has a wider dynamic range than microarrays and are not limited to genes with an already characterised sequence or functions.

NGS is highly adaptable to different applications, including DNA sequencing (whole genome or exome), DNA methylation (bisulfite-Seq), RNA-Seq, micro RNA (miRNA), long non-coding RNA (lncRNA), or chromatin immunopreciptation (ChIP). RNA-Seq of the transcriptome is a common adaptation where ribonucleic acid (RNA) is reverse transcribed and sequenced from the resulting complementary DNA (cDNA). This is utilised to quantify the levels of RNA and identify which regions of DNA are expressed. Similarly, bisulfite treatment converts cytosine residues to uracil, sparing methylated cytosine enabling them to be distinguished with bisulfite-Seq for detection of epigenetic modifications to generate an epigenome. Subsets of the nucleic acid may be extracted for sequencing such as the coding regions of DNA (for the "exome"), mRNA(RNA-Seq), microRNA, or capturing variable regions for DNA ("genotyping by sequencing") and methylation studies ("reduced-representation bisulfite sequencing). These "omics" technologies are applicable across a wide range of biomolecules to generate "molecular profiles" of a cell or sample.

NGS technologies continue to be refined with Illumina (the platform used to generate data in this project) and competitors continuing to improve products and decrease costs. As such, RNA-Seq for examining transcriptomes or expression studies is a growing field and will continue to be generated for a range of samples. The technology may further improve with developments in speed and accuracy (such as Ion Torrent platforms) or long reads, single molecules sequences (such as Pacific Biosciences, Oxford Nanopore, and Quantum Biosystems Japan). Due to the benefits of sequencing and the availability of public data, this thesis has focused on gene expression data generated by RNA-Seq. RNA-Seq data is publicly available from large-scale cancer

genomics projects and the methods developed for RNA-Seq data could be applied to future genomics technologies.

1.1.2.2 Bioinformatics and Genomics Data

Genomic technologies have generated data at a scale which requires computational, mathematical, and statistical expertise to handle this data effectively, in addition to an understanding of the biological context and research questions. The interdisciplinary field of "Bioinformatics", which draws upon these skills, focuses specifically on making inferences from genomics data or developing the tools to do so. Gene expression analysis is the focus of many bioinformatics research groups, drawing upon statistical approaches to appropriately handle microarray and RNA-Seq data along with making biological inferences from a large number of statistical tests.

This is often confused with the broader field "computational biology", which focuses on modelling and simulating aspects of biology and is not necessarily limited to genetics or data analysis. In practice, many researchers identify with both bioinformatics and computational biology or use techniques in both fields. This thesis uses many of these approaches, mainly in bioinformatics, to address biological research questions pertaining to synthetic lethal interactions.

1.1.3 Genomics Projects

Genomic projects have also been applied to various organisms, functional genetics (Kawai *et al.*, 2001; ENCODE, 2004), and human populations with an interest variability between individuals and health or disease risk (HapMap, 2003; 1000 Genomes, 2010). International projects and consortiums have begun to release data gathered using common agreed upon protocols across laboratories.These include many genomics projects including cancer genomics projects discussed below. The quality, consistency, and accessibility of these international projects is appealing, particularly for gene expression datasets where the more recent, larger projects have switched from microarray to RNA-Seq technologies.

1.1.3.1 The Cancer Genome Project

TheCancer Genome Project (CGP) was among the first genomics investigations into cancer (Dickson, 1999), using the human genome sequence (Collins and Barker, 2007; Lander *et al.*, 2001), the cancer research literature, and sequencing the genes of cancers themselves. The main aim of the Cancer Genome Project was to discover "cancer genes", which are frequently mutated in cancers by comparing cancer and normal tis-

sue samples. These include both "oncogenes" and "tumour suppressors" which are functionally activated and inactivated in cancers respectively. This project is ongoing and the continues to maintain the Catalogue Of Somatic Mutations In Cancer (COSMIC), a database of cancer genes (COSMIC, 2016). It includes 1,257,487 samples with 4,175,8787 gene mutations curated from 23,870 publications, including 29,112 whole genomes (COSMIC, 2016).

1.1.3.2 The Cancer Genome Atlas Project

The Cancer Genome Atlas (TCGA) network initially set out to demonstrate a pilot project on brain (McLendon *et al.*, 2008), ovarian (Bell *et al.*, 2011), and squamous cell lung (Hammerman *et al.*, 2012) cancers. The project then expanded, aiming to analyse 500 samples each for 20-25 tumour tissue types. They have since exceeded that goal, with data available for 33 cancer types including 10 "rare" cancers, a total of over 10,000 samples (TCGA, 2017). The TCGA projects set out to generate a molecular "profile" of the tumour (and some matched normal tissue) samples: the genotype, somatic mutations, gene expression, DNA copy number, and RNA methylation levels. Data which cannot be used to identify the patients is are publicly available

The Cancer Genome Atlas pilot projects (Bell *et al.*, 2011; Hammerman *et al.*, 2012; McLendon *et al.*, 2008) serve to demonstrate the power of applying genomic technologies to cancer research at such as scale. TCGA demonstrated the potential discovery of the molecular basis of cancer with these tissues, including the describing recurrently mutated genes in each cancer, identifying differentially methylated regions, and proposing transcriptional subtypes for ovarian cancers. The molecular aberrations in each cancer represent potential therapeutic targets in some cases and some were shown to have an impact on patient survival.

The TCGA breast cancer analysis (TCGA, 2012) consisted of 802 samples with exomes, copy number variants, RPPA protein quantification, and DNA methylation, mRNA, and microRNA arrays with 97 whole genomes sequenced. Four main molecular classes were identified to subtype the samples, despite considerable heterogeneity between samples. Recurrent mutations across more than 10% of samples were identified in *TP53*,*PIK3CA*, and *GATA*. TCGA further suggests subtypes by HER2 and EGFR protein levels. In a further analysis of 817 breast cancer samples including 127 invasive lobular breast and 88 mixed type samples (Ciriello *et al.*, 2015), 3 molecular subtypes of lobular breast cancer were identified. Lobular breast cancer was also characterised by recurrent mutations in *CDH1*, *PTEN*, *TBX2*, and *FOXA1*.

The TCGA stomach cancer analysis of 295 samples (Bass *et al.*, 2014) identified 4 molecular subtypes of stomach cancers characterised by: the Epstein-Barr virus, MSI, genomic instability, and chromosomal instability. Abberrations in *PD-L1*, *PIK3CA*, and *JAK2* were also identified in stomach cancers which may present therapeutic targets.

TCGA have identified various genes as recurrent, driver mutations across cancer types which are likely to have a role in driving the proliferation of these cancers and present a molecular target that could be applied across tissue types. In addition to disregarding the tissue-based distinction between colon and rectal cancers based on molecular similarlity (Muzny *et al.*, 2012), the TCGA project have observed differences within tumour types and proposed molecular subtyping for breast, clear cell renal, papillary renal, stomach, skin, bladder, and prostate cancers (Abeshouse *et al.*, 2015; Akbani *et al.*, 2015; Bass *et al.*, 2014; Ciriello *et al.*, 2015; Creighton *et al.*, 2013; Hammerman *et al.*, 2012; Linehan *et al.*, 2016; Muzny *et al.*, 2012; TCGA, 2012; Weinstein *et al.*, 2014).

The "Pan cancer" project (Hoadley *et al.*, 2014; Weinstein *et al.*, 2013) analysed 3527 samples across 12 tissue types. This project was performed a comprehensive analysis of molecular data across cancer types to identify molecular simliarities and differences. These included recurrent *TP53*, *BRCA1* and *BRCA2* mutations , HER2 over-expression, and microsatellite instability across cancer types. The pan cancer project (Hoadley *et al.*, 2014) has identified 11 molecular subtypes across these tissues, with only 5 of these corresponding to tissue cancer types due to molecular similarities shared across cancer types. This project further supports the genomic stratification of cancer patients, demonstrated in breast cancer (Parker *et al.*, 2009; Pereira *et al.*, 2016; Perou *et al.*, 2000), and there being core molecular characteristics across cancers (Hanahan and Weinberg, 2000, 2011).

While these findings contribute to further understanding cancer biology within and across tissue types, the main objective of such projects is to publicly release data to analyse in future investigations (McLendon *et al.*, 2008; TCGA, 2017; Weinstein *et al.*, 2013). These serve as a vast resource of common and rare cancer types and are publicly available for further analysis (cBioPortal, 2017; TCGA, 2017; Zhang *et al.*, 2011).

1.1.4 Genomic Cancer Medicine

Cancer genomics has significant potential for impacts in cancer medicine: from diagnosis to treatment. Beyond direct use of genome or RNA-Seq in clinical laboratories,

genomic studies also generate the biomarkers and inform development of treatments. These are likely to have a more immediate patient benefit considering the cost of routine genome sequencing for diagnostics.

1.1.4.1 Cancer Genes and Driver Mutations

There are two main classes of "cancer genes" (Futreal *et al.*, 2001). Oncogenes are activated in cancers either by gain of function mutations in proto-oncogenes, amplification of DNA, or elevated gene expression. Their normal functions are typically to regulate stem cells or to promote cellular growth and recurrent mutations are typically concentrated to particular gene regions. Conversely, tumour suppressor genes are those inactivated in cancer either by loss of function mutations, deletion of DNA copies, or reduced of gene expression, including hypermethylation. Their normal functions are typically to regulate cell division, DNA repair, and cell signalling.

Detecting these cancer genes has accelerated with genomic technologies, as demonstrated by COSMIC and TCGA (COSMIC, 2016; Weinstein *et al.*, 2013). Recurrent mutations, or DNA copy number variants and differential gene expression or DNA methylation are all indicative of cancer genes (Mattison *et al.*, 2009), which can be detected in genomics data (Pereira *et al.*, 2016; Weinstein *et al.*, 2013). However, many cancer genes have been identified or replicated in genomics data.

Distinguishing important "driver" mutations in cancer genes from "passenger" mutations is challenging due to patient variation, tumour heterogeneity, and genomic instability producing many variant gene sequences (Stratton *et al.*, 2009). Driver mutations can be identified by whether they co-occur or are mutually exclusive with mutations in other genes in cancers, are recurrently mutated across a significant proportion of samples for a specific tissue type, or if mutations are recurrent across different cancer tissue types (cBioPortal, 2017; Pereira *et al.*, 2016; COSMIC, 2016; Weinstein *et al.*, 2013; Zhang *et al.*, 2011). Approximately 140 driver mutations have been identified, including many novel genes in particular cancers from genomic studies, with 2–8 in typically occurring in each tumour usually affecting cell fate, survival, or genome maintenance (Vogelstein *et al.*, 2013). There need to translate the identification of many cancer genes and driver mutations to patient benefit by repurposing or designing of therapeutic interventions against these molecular targets.

1.1.4.2 Precision Cancer Medicine

The importance of genomics is emphasised in translational cancer research in contrast with current strategies of healthcare based on what works well for the most of the pop-

ulation. Cancers could eventually be treated by their genomic features, particularly grouping patients by the mutation, expression, or DNA methylation profiles of their cancers, which is already done in part. Identifying actionable molecular targets is a key aspect of "precision medicine", the rationale to target molecular subtypes with separate treatment strategies. To this end many driver mutations and gene expression signatures for distinguishing cancers have been identified. Some oncogenic driver mutations have effective pharmacological inhibitors designed against them but there remain many cancer genes and mutations, particularly tumour suppressors, for which there is not yet a targeted therapy.

1.1.4.3 Molecular Diagnostics and Pan-Cancer Medicine

Molecular features such as mutations or gene expression signatures have been proposed to diagnose tumour subtypes. In breast cancer, several distinct "intrinsic subtypes" have been identified, distinguished by molecular mechanisms, with differences in malignancy and patient outcome (Parker *et al.*, 2009; Perou *et al.*, 2000). Conversely, common molecular mechanisms may be shared between cancers across tissue types as discovered by the TCGA "Pan cancer" project, which combined molecular profiles across tissue types Weinstein *et al.* (2013). Molecular subtypes could feasibly be included in clinical testing as a panel of biomarkers for diagnosis, monitoring drug response, or predicting risk of recurrence. Having identified these molecular subtypes and genetic aberrations specific to cancers, there is clear need for further development of treatments that target them.

1.1.4.4 Targeted Therapeutics and Pharmacogenomics

Targeted therapies with specificity against a molecular target as examples of precision cancer medicine. Molecular targets can be tested in laboratory conditions with RNA interference (RNAi) or pharmacological agents. Identification of molecular targets is important for developing novel anti-cancer treatments along with validation and drug testing. For oncogenic mutations, the recurrent mutant variant or over-expressed gene is directly inhibited using structure-aided drug design or compound screening. However, oncogenes with high homology to other genes or tumour suppressor genes are not amenable to direct targeting (Kaelin, Jr, 2009).

Targeted therapeutics have been applied as monoclonal antibodies against oncogenes (such as *HER2*) with relative success in clinical trials (Miles, 2001). Such targeted therapies have potential applications across tissue types and specificity against tumour cells (even in advanced disease or as a chemopreventative in high-risk individ-

uals), with wide therapeutic windows and use in combination therapies. However, they have already begun to manifest problems with resistance, recurrence, tissue specificity, and design of inhibitors specific to oncogenic variants. Targeted anticancer therapeutics can exploit complex interactions to distinguish normal and cancerous cells which may benefit from studies of gene regulation or interaction networks.

The unexpected synergy between inhibitors of the oncogenes $BRAF^{\mathrm{V600E}}$ and *EGFR* in colorectal cancer is an example of such a system Prahallad *et al.* (2012). Despite successful application of vemurafenib against $BRAF^{\mathrm{V600E}}$ in melanomas Dienstmann and Tabernero (2011); Ravnan and Matalka (2012), colorectal cancers with $BRAF^{\mathrm{V600E}}$ mutations have poor prognosis and lack drug response. Prahallad *et al.* (2012) used an RNAi screen and found that *EGFR* inhibition is synergistic with vemurafenib against $BRAF^{\mathrm{V600E}}$ in colon cell lines and xenografts due to feedback activation of *EGFR*. Vemurafenib induced rapid reactivation of MAPK/ERK signalling via *EGFR* in colorectal cell lines in a tissue-specific manner Corcoran *et al.* (2012) and may be be relevant to acquired resistance in melanoma Sun *et al.* (2014). Thus combination therapies against several molecular pathways may be necessary to anticipate acquired resistance Ravnan and Matalka (2012) and targeted therapeutics may be further refined from understanding the pathway structure and functional interactions across cancer cells.

1.1.5 Systems and Network Biology

Driver mutations in oncogenes and tumour suppressor genes do not occur in isolation. The genetic interactions, regulatory and cellular signalling, and metabolic reactions are inter-related and may each be perturbed by aberrations in gene function occurring in cancers. These relationships can be represented by biological networks of connected pairs of genes with a relationship. Due to the complexity of a cell, these molecular networks are very large consisting of thousands of nodes comprised by genes or proteins.

The properties of large networks were first studied by constructing random networks by randomly linking a fixed number of nodes (Erdős and Rényi, 1959, 1960). Despite the random nature of these networks, properties such as their connectivity were well characterised. The vertex degree (number of partners for each node) of random network follows a Poisson distribution, however this property does not hold in nature. Thus natural networks are non-random or not formed in this way Barabási and Oltvai (2004).

This work formed the foundation for studying complex networks (van Steen, 2010), which model features of observed networks not found in Erdős and Rényi's random networks (Erdős and Rényi, 1959, 1960). The small world property, made popular

by findings in social networks (Travers and Milgram, 1969), is the remarkably short path lengths between any nodes in a small world network. A small world network is well-connected with a characteristic path length (the average length of shortest paths between all pairs of nodes) proportional to the logarithm of the number of nodes. Watts and Strogatz (1998) developed a model of random rewiring of a regular network to construct random networks with the small world property and a high clustering coefficient. While these properties are more representative of networks occurring in nature, their model is limited by the degree distribution which converges to a Poisson distribution as it is rewired Barrat and Weigt (2000). The vertex degree distribution of naturally occurring networks often follows a power law distribution with most nodes having far fewer connections than average and a small subset of highly connected network 'hubs' Barabási and Albert (1999).

Barabási and Albert (1999) constructed a network model in an entirely different way to randomly generate scale-free networks which have a power law degree distribution. They constructed random networks by preferential attachment, modelling growth of a network by sequentially adding nodes with links to existing nodes. The scale-free nature of the random networks was ensured by adding new nodes with an increasing probability of attachment to an existing node if it has higher degree. These networks successfully capture the scale-free nature of many observed networks with short characteristic path length and low eccentricity resulting in super small worlds Barabási and Albert (1999). Scale-free networks are limited by a low clustering coefficient and lack of modular structure; however, they have enabled the study of scale-free network topology and served as a basis for modified scale-free models (Dorogovtsev and Mendes, 2003; Holme and Kim, 2002).

High-throughput technologies such as siRNA screens, two-hybrid screens, microarrays and massively parallel sequencing have generated genome-scale data and enabled analysis of biological networks. Many types of molecular networks can be constructed, depending on the biological application. Genetic interactions are relatively unexplored within molecular networks and may lead to better understanding of the role of gene functions in cellular function and disease. High-throughput screens in humans, mammals, and non-model organisms are costly and labour-intensive. Computational approaches with effective predictive models are therefore a more feasible alternative to study the connectivity of a biological network in a complex metazoan cell at the genome-scale.

1.1.5.1 Network Medicine and Polypharmacology

Molecular networks are biological networks consisting of biological molecules including genes, transcripts (with non-coding and microRNAs), or proteins related by known interactions and gene regulatory or metabolic pathways. Targeted therapeutics have had some success for drug discovery, particularly in anticancer applications, including exploiting these molecular networks by designing combination therapies and applying a network pharmacology framework Hopkins (2008). Rational design of drugs selective to a single target has often failed to deliver clinical efficacy. Many existing effective drugs modulate multiple proteins, having been selected for biological effects or clinical outcome rather than molecular targets. Proponents of network biology and polypharmacology (specific binding to multiple targets) recommend developing drugs with a desired target profile designed for the target topology Barabási and Oltvai (2004); Hopkins (2008). Multi-target treatments aim to achieve a clinical outcome through modulation of molecular networks since the genetic robustness of a cell often compensates for loss of a single molecular target.

While multi-target drugs may be more difficult to design, they are faster to test clinically than drug combinations which are usually required to be tested separately first Hopkins (2008). Synthetic lethal treatments for cancer, drug combinations and multi-target drugs to combat resistance to chemotherapy and antibiotics can be informed by biological networks Barabási and Oltvai (2004); Hopkins (2008). Further optimisation of timing and dosing of drug combinations may increase efficacy of treatments with low efficacy applied separately. Low doses and drug holidays are other counter intuitive approaches which may increase clinical efficacy, reduce adverse effects, and reduce drug resistance (Sun *et al.*, 2014; Tsai *et al.*, 2012).

A molecular map of the interactions and pathways in the mammalian cellular network has the potential to impact upon drug design and clinical practice, particularly in treatment of cancer and infectious disease. Characterisation of the target system and impact of existing treatments, such as $BRAF^{\mathrm{V600E}}$ and *EGFR* inhibitors, enable wider application of the mechanisms for such interventions exploiting genetic interactions or pathways. This could lead to development of more effective treatment interventions for these systems and prediction of similar molecular systems for development of novel drug targets and combinations.

1.2 A Synthetic Lethal Approach to Cancer Medicine

Synthetic lethality has vast potential to improve cancer medicine by expanding application of targeted therapeutics to include inactivation of tumour suppressors and genes that are difficult to target directly. Synthetic lethal interactions are also studied for gene function and drug mode-of-action in model organisms. This section introduces the concept of synthetic lethality as it was originally conceived and how it has been adopted conceptually in cancer research. Detecting these interactions at scale and interpreting them is the focus of this thesis, hence we start with an overview of the concepts involved, initial work on the interaction, and the rationale for applications to cancer. Specific investigations into synthetic lethality in cancer, detection by experimental screening, and prediction by computational analysis will then be reviewed.

1.2.1 Synthetic Lethal Genetic Interactions

Genetic interactions are a core concept of molecular biology, discovered among earliest investigations of Mendelian genetics, and have received revived interest with new technologies and potential applications. Biological epistasis is the effect of an allele at one locus "masking" the phenotype of another locus (Bateson and Mendel, 1909). Statistical epistasis is where there is significant disparity between the observed and expected phenotype of a double mutant, compared to the respective phenotypes of single mutants and the wild-type (Fisher, 1919). Fisher's definition lends itself to quantitative traits and more broadly encompasses synthetic genetic interaction (SGIs). These have become popular for studies in yeast genetics and cancer drug design (Boone *et al.*, 2007; Kaelin, Jr, 2005).

SGIs are substantial deviations of growth or viability from the expected null mutant phenotype (of an organism or cell) assuming additive (deleterious) effects of the single mutants. The double mutant does not necessarily have either single mutant phenotype (as shown for cellular growth phenotypes in Figure 1.1). Most SGIs are more viable than either single mutant or less viable than the expected double mutant. Mutations are "synergistic" in negative SGIs with more deviation from the wild-type than expected. Formally, "synthetic sick" (SSL) and "synthetic lethal" (SL) interactions are negative SGIs giving growth inhibition and complete inviability respectively. In cancer research, synthetic lethality more broadly describes any negative SGI with specific inhibition of a mutant cell, including SSL interactions. Mutations are "alleviating" in positive SGIs with less deviation from the wild-type than expected. For viability, "suppression" and "rescue" are positive SGIs giving at least partial restoration of wild-type

growth from single mutants with growth impairment and lethal phenotypes respectively. Negative SGIs were markedly more common than positive SGIs in a number of studies in model systems Boucher and Jenna (2013); Tong *et al.* (2004).

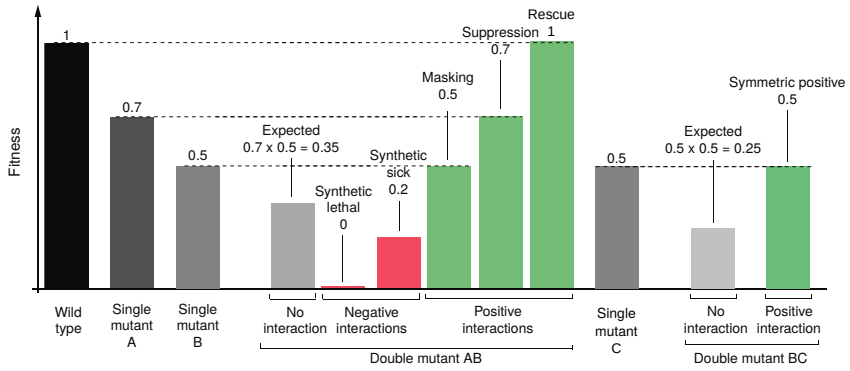

Figure 1.1: **Synthetic genetic interactions.** Impact of various negative and positive SGIs: negative interactions involve deleterious (sick) or inviable (lethal) phenotypes whereas positive interactions involve restoring viability by masking or suppressing the other mutation or complete rescue of the wildtype phenotype. Figure adapted from (Costanzo *et al.*, 2011) concerning growth viability fitness in yeast.

1.2.2 Synthetic Lethal Concepts in Genetics

Synthetic lethal genes are generally regarded to arise due to functional redundancy. Due to the functional level of SGIs, synthetic lethal genes do not need to directly interact, nor be expressed in the same cell or at the same developmental stage: serving related functions is sufficient to affect cell (or organism) viability and be relevant to drug-mode-of-action cancer biology. Combined loss of genes performing an essential or important function in a cell are therefore deleterious. Synthetic lethal gene pairs are therefore pairwise essential with "induced essentiality": each synthetic lethal gene becomes essential to the cell upon loss of the other.

Since synthetic lethal gene partners can be affected by extracellular stimuli such as chemicals, essentiality of synthetic lethal genes can be induced by the environment of a cell. An environmental stress condition may inhibit one or the other synthetic lethal gene, such as exposure to chemicals, in which case the synthetic lethal partner gene is "conditionally essential" (Hillenmeyer, 2008). Thus the evolutionary rationale for the abundance of SGIs (compared to the surprisingly low number of essential genes) in a Eukaryotic genome can be attributed to genetic functional redundancy and network robustness of a cell which are advantageous to survival.

Biological functions are typically performed by a pathway of genes (or their products). Synthetic lethal genes are occur within the same biological pathway and between them (Boone *et al.*, 2007; Costanzo *et al.*, 2010; Kelley and Ideker, 2005). Many genes of the same pathway may be functionally inter-changable, synthetic lethal partners of a particular gene. Therefore biological pathways can exhibit induced essentiality with loss of the synthetic lethal partner gene and synthetic lethality may occur at pathway level or in a gene regulation network.

1.2.3 Synthetic Lethality in Model Systems

Genetic high-throughput screens have identified unexpected, functionally informative, and clinically relevant synthetic lethal interactions; including synthetic lethal partners of genes recurrently mutated in cancer or attributed to familial early-onset cancers. While screening presents an appealing strategy for synthetic lethal discovery, computational approaches are becoming popular as an alternative or complement to experimental methods to overcome inherent bias and limitations of experimental screens. An array of recently developed computational methods (Jerby-Arnon *et al.*, 2014; Lu *et al.*, 2015; Tiong *et al.*, 2014; Wang and Simon, 2013; Wappett, 2014) show the need for synthetic lethal discovery in the fundamental genetics and translational cancer research community. However, existing computational methods are not suitable for queries of genomic data for interacting partners of a particular gene: they have been applied pairwise across the genome, do not have software released to apply the methodology, or lack statistical measures of error for further analysis. A robust prediction of gene interactions is an effective and practical approach at a scale of the entire genome for ideal translational applications, analysis of biological systems, and constructing functional gene networks.

1.2.3.1 Synthetic Lethal Pathways and Networks

SGIs are common in genomes, 4× more interactions were detected with synthetic gene array (SGA) mating screens than protein-protein (yeast-2-hybrid) interactions (Tong *et al.*, 2004). The SGI network is scale-free with power-law vertex degree distribution and low average shortest path length (3.3) as expected for a complex biological network (Barabási and Oltvai, 2004). Highly connected "hub" genes with the highest number of links (vertex degree) are functionally important with many negative SGI hubs involved in cell cycle regulation and many positive SGI hubs involved in translation (Baryshnikova *et al.*, 2010b; Costanzo *et al.*, 2010). Negative SGIs were far more common than positive SGIs, with synthetic gene loss being more likely to be deleterious to cell

than advantageous which indicates than synthetic lethality may be comparably easier to detect than other SGIs.

Essential pathways are highly buffered with 5× more interactions than other SGIs, consistent with strong selection for survival, as found with conditional and partial mutations in essential genes (Davierwala *et al.*, 2005). This SGI network had scale-free topology and rarely shared interactions with the protein-protein interaction network. These networks are related by an "orthogonal" relationship: shared partners in one network tend to be themselves connected directly in the other network. Essential genes were likely to have closely related functions, whereas non-essential networks were relatively more inclined to have SGIs between distinct biological pathways.

1.2.3.2 Evolution of Synthetic Lethality

There is poor conservation of specific SGIs between *S. cerevisiae* and *S. pombe* with 29% of the interactions tested in both distantly related species being conserved between them (Dixon *et al.*, 2008). The remaining interactions show high species-specific differences; however, many of the species-specific interactions were still conserved between biological pathways, protein complexes, or protein-protein interaction modules. Similarly, conservation of pathway redundancy was also found between Eukaryotes (*S. cerevisiae*) and prokaryotes (*E. coli*) (Butland *et al.*, 2008). Negative SGIs were more likely to be conserved between biological pathways, whereas positive SGIs were more likely to be conserved within a pathway or protein complex (Roguev *et al.*, 2008).

A modest 5% of interactions were conserved between unicellular (*S. cerevisiae*) and multicellular (*C. elegans*) organisms. However, the nematode SGI network had similar scale-free topology and modularity despite differences in methodology: metazoan synthetic lethal screens with RNA interference (RNAi) are incomplete knockouts, whereas screening null mutations is feasible in yeast (Bussey *et al.*, 2006). The nematode SGI screen identified network hubs with important interactions to orthologues of known human disease genes (Lehner *et al.*, 2006). Despite the lack of direct conservation of SGIs between yeasts and nematode worms, genetic redundancy was consistent with an "induced essentiality" model of SGIs where gene functions are conserved with network restructuring over evolutionary change (Tischler *et al.*, 2008).

While nematode models are more closely related to human cells which are also screened with RNAi, cancer cells can present growth and viability phenotypes more comparable to yeast models. Therefore findings from both SGA and RNAi models are relevant to understanding human and cancer cells. RNAi has also been applied to human and mouse cancer cells with short interfering RNA (siRNA) in cell culture and

genetic screening experiments. These findings suggest that SGI network "rewiring" is a concern for identifying specific synthetic lethal interactions in cancer as specific synthetic lethal genes may vary between genetic backgrounds. Thus it is expected at a pathway approach will be more robust in the context of evolution, patient variation, tumour heterogeneity, or disease progression.

1.2.4 Synthetic Lethality in Cancer

Loss of function occurs in many genes in cancers including tumour suppressors and yet few interventions target such mutations compared to targeted therapies for gain of function mutation in oncogenes (Kaelin, Jr, 2005). Synthetic lethality is a powerful design strategy for therapies selective against loss of gene function with potential for application against a range of genes and diseases (Fece de la Cruz *et al.*, 2015; Kaelin, Jr, 2009). When genes are disrupted in cancers, the induced essentiality of synthetic lethal partners presents a vulnerability that may be exploited for anti-cancer therapy. Since synthetic lethality affects cellular viability by indirect functional relationships between genes, it is suitable for indirectly targeting mutations in cancers via synthetic lethal partners with targeted therapeutics. These have could be highly specific against cancer cells (with the target mutation) over non-cancer cells (with a functional compensating gene). Analogous to "oncogene addiction", where cancer cells adapt to particular oncogenic growth signals and become reliant on them to remain viable (Luo *et al.*, 2009; Weinstein, 2000), synthetic lethal partners of inactivated tumour suppressors are required to maintain cancer cell viability and proliferation. As such cancers are subject to "non-oncogene addiction" and are feasible anti-cancer drug targets.

The synthetic lethal approach to cancer medicine is most amenable to loss of function mutations in tumour suppressor genes, where it would feasibly be effective against any loss of function mutation across the tumour suppressor with a viable synthetic lethal partner gene (as shown in Figure 1.2). However, the approach may also be suitable for cases where cancer cells have mutations where the normal function of the gene is disrupted such as if it were over-expressed ("synthetic dosage lethality") or if an oncogenic mutation interfered with the function of the proto-oncogene. Thus synthetic lethality makes it feasible to target a range of cancer-specific mutations with targeted therapeutics, including inactivated tumour suppressor genes. synthetic lethality may also enable distinguishing highly homologous oncogenes by functional differences by targeting their synthetic lethal partners.

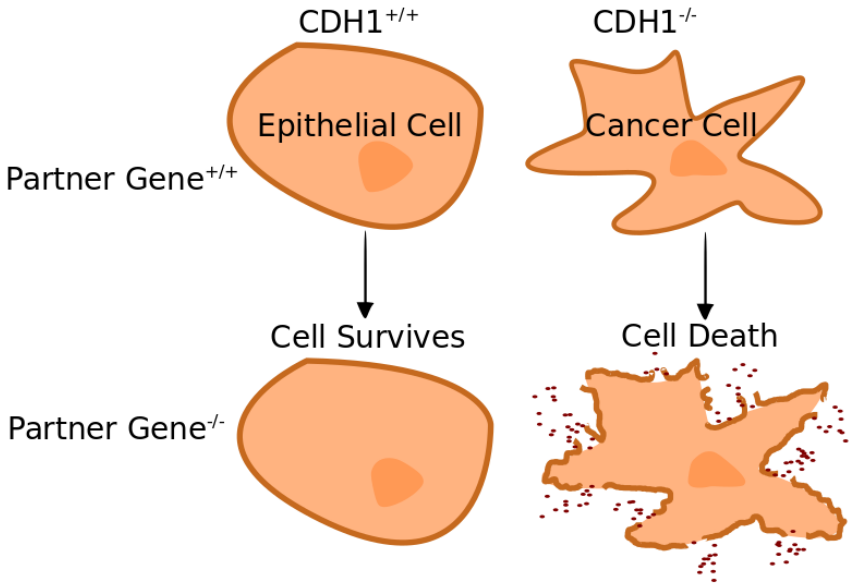

Figure 1.2: **Synthetic lethality in cancer.** Rationale of exploiting synthetic lethality for specificity against a tumour suppressor gene (e.g., *CDH1*) while other cells are spared under the inhibition of a partner gene.

1.2.5 Clinical Impact of Synthetic Lethality in Cancer

The synthetic lethal interaction of *BRCA1* or *BRCA2* with *PARP1* in breast cancer is an example of how gene interactions are important in cancer, including translation to the clinic. These genetic interactions enable specific targeting of mutations in *BRCA1* or *BRCA2* tumour suppressor genes with PARP inhibitors by inducing synthetic lethality in breast cancer (Farmer *et al.*, 2005). PARP inhibitors are one of the first targeted therapeutics against a tumour suppressor mutation with success in clinical trials.

BRCA1 or *BRCA2* and *PARP1* genes demonstrate the application of the synthetic lethal approach to cancer therapy (Ashworth, 2008; Kaelin, Jr, 2005). *BRCA1* and *BRCA2* are homologous DNA repair genes, widely known as tumour suppressors; mutation carriers have substantially increased risk of breast (risk by age 70 of 57% for *BRCA1* and 59% for *BRCA2*) and ovarian cancers (risk by age 70 of 40% for *BRCA1* and 18% for *BRCA2*) (Chen and Parmigiani, 2007). The *BRCA1* or *BRCA2* genes, which usually repair DNA or destroy the cell if it cannot be repaired, have inactivating somatic mutations in some familial and sporadic cancers. Poly-ADP-ribose polymerase (PARP) genes are tumour suppressor genes involved in base excision DNA repair. Loss of PARP activity results in single-stranded DNA breaks. However, $PARP1^{-/-}$ knock-

out mice are viable and healthy indicating low toxicity from PARP inhibition (Bryant *et al.*, 2005).

Bryant *et al.* (2005) showed that *BRCA2* cells were sensitive to PARP inhibition by siRNA of *PARP1* or drug inhibition (which targets *PARP1* and *PARP2*) using Chinese hamster ovary cells, MCF7 and MDA-MB-231 breast cell lines. This effect was sufficient to kill mouse tumour xenografts and showed high specificity to *BRCA2* deficient cells in culture and xenografts. Farmer *et al.* (2005) replicated these results in embryonic stem cells and showed that *BRCA1* cells were also sensitive to PARP inhibition relative to the wild-type with siRNA and drug experiments in cell culture and drug activity against *BRCA1* or *BRCA2* deficient embryonic stem cell mouse xenografts. They found evidence that PARP inhibition causes DNA lesions, usually repaired in wild-type cells, which lead to chromosomal instability, cell cycle arrest, and induction of apoptosis in *BRCA1* or *BRCA2* deficient cells. Therefore, the pathways cooperate to repair DNA giving a plausible mechanism for combined loss as an effective anti-cancer treatment.

Thus PARP inhibitors have potential for clinical use against *BRCA1* or *BRCA2* mutations in hereditary and sporadic cancers (Ashworth, 2008; Kaelin, Jr, 2005). PARP inhibition has been found to be effective in cancer patients carrying *BRCA1* or *BRCA2* mutations and some other ovarian cancers, suggesting synthetic lethality between PARP and other DNA repair pathways (Ström and Helleday, 2012). This supports the potential for PARP inhibition as a chemo-preventative alternative to prophylactic surgery for high-risk individuals with *BRCA1* or *BRCA2* mutations (Ström and Helleday, 2012). Hormone-based therapy has also been suggested as a chemo-preventative in such high-risk individuals and aromatase inhibitors have completed phase I clinical trials for this purpose (Bozovic-Spasojevic *et al.*, 2012). Ström and Helleday (2012) also postulate increased efficacy of PARP inhibitors in the hypoxic DNA-damaging tumour micro-environment.

A PARP inhibitor, olaparib, showed fewer adverse effects than cytotoxic chemotherapy and anti-tumour activity in various clinical trials against *BRCA1* or *BRCA2* deficient familial or sporadic breast, ovarian, and prostate cancers (Audeh *et al.*, 2010; Fong *et al.*, 2009, 2010; Tutt *et al.*, 2010). This treatment has a favourable therapeutic window and similarly low toxicity between mutation carriers of *BRCA1* or *BRCA2* mutations and sporadic cases. These PARP inhibitors have been FDA approved for some cancers McLachlan *et al.* (2016), are effective against germline and sporadic *BRCA1* or

BRCA2 mutations, and are a potential prevention alternative to prophylactic surgery for high-risk mutation carriers Ström and Helleday (2012).

This demonstrates the clinical impact of a well characterised system of synthetic lethality with known cancer risk genes. Synthetic lethality has the benefit of being effective against inactivation of tumour suppressor genes by any means, broader than targeting a specific oncogenic mutation (Kaelin, Jr, 2005). The targeted therapy is effective in both sporadic and hereditary *BRCA1* or *BRCA2* deficient tumours acting against an oncogenic molecular aberration across several tissues.

1.2.6 High-throughput Screening for Synthetic Lethality

RNA interference (RNAi) technologies have enabled extensive investigations of genetic redundancy in mammalian experimental models including testing experimentally for synthetic lethality (Fraser, 2004). Synthetic lethal RNAi screens are performed, using short interfering RNA (siRNA) or short hairpin RNA (shRNA) to target specific genes is isogenic cells. Identifying synthetic lethality is crucial to study gene function, drug mechanisms, and design novel therapies (Lum *et al.*, 2004). Candidate selection of synthetic lethal gene pairs relevant to cancer has shown some success but is limited because interactions are difficult to predict; they can occur between seemingly unrelated pathways in model organisms (Costanzo *et al.*, 2011). While biologically informed hypotheses have had some success in synthetic lethal discovery (Bitler *et al.*, 2015; Bryant *et al.*, 2005; Farmer *et al.*, 2005), interactions occurring indirectly between distinct pathways would be missed (Boone *et al.*, 2007; Costanzo *et al.*, 2011). Scanning the entire genome for interactions against a clinically relevant gene is an emerging strategy being explored with high-throughput screens (Fece de la Cruz *et al.*, 2015) and computational approaches (Boucher and Jenna, 2013; van Steen, 2012).

Experimental screening for synthetic lethality is an appealing strategy for wider discovery of functional interactions *in vivo* despite many potential sources of error which must be considered. The synthetic lethal concept has both genetic and pharmacological screening applications to cancer research. Genetic screens, with RNAi to discover the specific genes involved, inform development of targeted therapies with a known mode of action, anticipated mechanisms of resistance, and biomarkers for treatment response. RNAi is a transient knockdown of gene expression more similar to the effect of drugs than complete gene loss and is more representative of disease than model organisms (Bussey *et al.*, 2006). The RNAi gene knockdown process has inherent toxicity to some cells, potential off-target effects, and issues with a high false positive rate. Therefore, it

is important to validate any candidates in a secondary screen and replicate knockdown experiments with a number of independent shRNAs. Genetic screens have potential for quantitative gene disruption experiments to selectively target over-expressed genes in cancer via synthetic dosage lethality. While powerful for understanding fundamental cellular function, analysis of isogenic cell lines is inherently limited by assuming only a single mutation differs between them despite susceptibility to "genetic drift" and cannot account for diverse genetic backgrounds or tumour heterogeneity (Fece de la Cruz *et al.*, 2015). Genetic screens thus identify targets to develop or repurpose targeted therapies for disease but alone will not directly identify a lead compound to develop for the market or clinical translation.

Chemical screens are immediately applicable to the clinic by directly screening for selective lead compounds with suitable pharmacological properties. However chemical screens lack a known mode of action, may affect many targets, and screen a narrow range of genes with existing drugs. With either approach there are many challenges translating candidates into the clinic such as finding targets relevant to a range of patients, validation of targets, accounting for a range of genetic (and epigenetic) contexts or tumour micro-environment, identifying effective synergistic combinations, enhancers of existing radiation or cytotoxic treatments, avoiding inherent or acquired drug resistance, and developing biomarkers for patients which will respond to synthetic lethal treatment, including integrating these into clinical trials and clinical practice. Identifying specific target genes is an effective way to anticipate such challenges, which can be approached with genetic screens, so we will focus on these and computational alternatives. Screening methods have proven a fruitful area of research, despite being costly, laborious, and having many different sources of error. These limitations suggest a need for complementary computational approaches to synthetic lethal discovery.

1.2.6.1 Synthetic Lethal Screens

Overexpression of genes is another suitable application for synthetic lethality since over-expressed genes cannot be distinguished from the wild-type by direct sequence specific targeted therapy. Overexpression of oncogenes, such as *EGFR*, *MYC*, and *PIM1*, has been found to drive many cancers. *PIM1* is a candidate for synthetic lethal drug design in lymphomas and prostate cancers, where it interacts with *MYC* to drive cancer growth. van der Meer *et al.* (2014) performed an RNAi screen for synthetic lethality between *PIM1* over-expression and gene knockdown in RWPE prostate cancer cell lines. *PLK1* gene knockdown and drug inhibition was effective as a specific inhibitor of *PIM1* over-expressing prostate cells in cell culture and mouse tumour xenografts.

PLK1 inhibition reduced *MYC* expression in pre-clinical models, consistent with expression in human tumours in which *PIM1* and *PLK1* are co-expressed and correlated with tumour grade. Thus RNAi screening was valuable to identify therapeutic targets and biomarkers for patient response as demonstrated with the finding of *PLK1* as a candidate drug target against prostate cancer progression.

Hereditary leiomyomatosis and renal cell carcinoma (HLRCC) is a cancer syndrome of predisposition to benign tumours in the uterus and risk of malignant cancer of the kidney attributed to inherited mutations in fumarate hydratase (*FH*). Boettcher *et al.* (2014) performed an RNAi screen on HEK293T renal cells for synthetic lethality with *FH*. They found enrichment of haem metabolism (consistent with the literature) and adenylate cyclase pathways (consistent with cAMP dysregulation in *FH* mutant cells). Synthetic lethality between *FH* mutation and adenylate cyclases were validated with gene knockdown, drug experiments, and replicated across both HEK293T renal cells and VOK262 cells derived from a HLRCC patient, suggesting new therapeutic targets against the disease.

Similarly, hereditary diffuse gastric cancer (HDGC) is a cancer syndrome of predisposition to early-onset malignant stomach and breast cancers attributed to mutations in E-cadherin (*CDH1*). Telford *et al.* (2015) performed an RNAi screen on MCF10A breast cells for synthetic lethality with *CDH1*. They found enrichment of G-protein coupled receptors (GPCRs) and cytoskeletal gene functions. The results were consistent with a concurrent drug compound screen with several candidates validated by lentiviral shRNA gene knockdown and drug testing including inhibitors of Janus kinase, histone deacetylases, phosphoinositide 3-kinase, aurora kinase, and tyrosine kinases. Therefore the synthetic lethal strategy has potential for clinical impact against HDGC, with an interest in interventions with low adverse effects for chemo-prevention, including repurposing existing approved drugs for activity against *CDH1* deficient cancers.

RNAi screening for synthetic lethality is also useful for functional genetics to understand drug sensitivity. Aarts *et al.* (2015) screened the WiDr colorectal cell line for synthetic lethality between *WEE1* inhibitor treatment and an RNAi library of 1206 genes with functions known to be amenable to drug treatment or important in cancer such as kinases, phosphatases, tumour suppressors, and DNA repair (a pathway *WEE1* regulates). Screening identified several synthetic lethal candidates including genes involved in cell cycle regulation, DNA replication, repair, homologous recombination, and Fanconi anaemia. Synthetic lethality with cell-cycle and DNA repair genes was consistent with the literature and validation in a panel of breast and colorectal cell

lines supported checkpoint kinases, Fanconi anaemia, and homologous recombination as synthetic lethal partners of *WEE1*. These results show that synthetic lethality can be used to improve drug sensitivity as a combination treatment, especially to exploit genomic instability and DNA repair, which are known to be clinically applicable from previous results with *BRCA1* or *BRCA2* genes and PARP inhibitors (Lord *et al.*, 2015). Therefore, *WEE1* inhibitors are an example of treatment which could be repurposed with the synthetic lethal strategy. Similar findings would be valuable to clinicians as a source of biomarkers and novel treatments. While using a panel of cell lines to replicate findings across genetic background is a promising approach to ensure wide clinical application of validated synthetic lethal partners, a computational approach may be more effective as it could account for wider patient variation than scaling up intensive experiments on a wide array of cell lines and could screen beyond limited candidates from an RNAi library.

Chemical genetic screens are also a viable strategy to identify therapeutically relevant synthetic lethal interactions. Bitler *et al.* (2015) investigated *ARID1A* mutations, aberrations in chromatin remodelling known to be common in ovarian cancers, for drug response. Ovarian RMG1 cells were screened for drug response specific to *ARID1A* knockdown cells. They used *ARID1A* gene knockdown for consistent genetic background, with control experiments and 3D cell culture to ensure relevance to drug activity in the tumour micro-environment. Screening a panel of commercially available drugs targeting epigenetic regulators found *ESH2* methyltransferase inhibitors effective and specific against *ARID1A* mutation with validation in a panel of ovarian cell lines. Synthetic lethality between *ARID1A* and *ESH2* was supported by decreases in H3K27me3 epigenetic marks and markers of apoptosis in response to *ESH2* inhibitors. This was mechanistically supported with differential expression of *PIK3IP1* and association of both synthetic lethal genes with the *PIK3IP1* promoter identifying the PI3K-AKT signalling pathway as disrupted when both genes were inhibited.

This successfully demonstrates the importance of synthetic lethality in epigenetic regulators, identifies a therapeutically relevant synthetic lethal interaction, and shows that chemical genetic screens could model drug response and combination therapy in cancer cells. However this approach is limited to finding synthetic lethal interactions between genes with known similar function, which may not be the most suitable for treatment. Further limiting experiments to genes with existing targeted drugs reduces the number of synthetic lethal interactions detected, assumes the specificity of drugs to a particular target, and many of these drugs are not yet clinically available yet anyway,

although they are clinical trials for other diseases or limited to access by patients from a particular countries.

The examples above show that high-throughput screens are an effective approach to discover synthetic lethality in cancer with a wide range of applications. Screens are more comprehensive than hypothesis-driven candidate gene approaches and successfully find known and novel synthetic lethal interactions with potential for rapid clinical application. They have the power to test mode of action of drugs, find unexpected synthetic lethal interactions between pathways, or identify effective treatment strategies without needing a clear mechanism. However, synthetic lethal screens are costly, labour-intensive, error-prone, and biased towards genes with effective RNAi knockdown libraries. Limited genetic background, lethality to wild-type cell during gene knockdown, off-target effects, and difficultly replicating synthetic lethality across different cell lines, tissues, laboratories, or conditions stems from a high false positive rate and a lack of standardised thresholds to identify synthetic lethality in a high-throughput screen. Therefore there is a need for replication, validation, and alternative approaches to identify synthetic lethal candidates. Varied conditions between experimental screens and differences between RNAi and drug screens renders meta-analysis ineffective.

Genome-scale synthetic lethal experiments (across gene pairs) are not feasible, even in model organisms, and they typically focus on specific gene candidates or the partners of a gene of interest (such as importance in health). Therefore a computational approach is more suitable for this task and may further augment experimental screening to replicate screen candidates beyond experimental models.

1.2.7 Computational Prediction of Synthetic Lethality

1.2.7.1 Bioinformatics Approaches to Genetic Interactions

Prediction of gene interaction networks is a feasible alternative to high-throughput screening with biological importance and clinical relevance. There are many existing methods to predict gene networks, as reviewed by van Steen (2012) and Boucher and Jenna (2013) and summarised in Table 1.1. However, many of these methods have limitations including the requirement for existing SGI data, several data inputs, and reliability of gene function annotation. Many of the existing methods also assume conservation of individual interactions between species, which has been found not to hold in yeast studies (Dixon *et al.*, 2008). Tissue specificity is important in gene regulation and gene expression, which are used as predictors of genetic interaction. However, tissue specificity of genetic interactions cannot be explored in yeast studies

and has not been considered in many studies of multicellular model organisms, human networks, or cancers. Similarly, investigation into tissue specificity of protein-protein interactions (PPIs) , an important predictor of genetic interactions, is difficult given that high-throughput two-hybrid screens occur out of cellular context for multicellular organisms.

Table 1.1: Methods for Predicting Genetic Interactions

Method	Input Data	Species	Source	Tool Offered
Between Pathways Model	PPI, SGI	*S. cerevisiae*	Kelley and Ideker (2005)	
Within Pathways Model	PPI, SGI	*S. cerevisiae*	Kelley and Ideker (2005)	
Decision Tree	PPI, expression, phenotype	*S. cerevisiae*	Wong *et al.* (2004)	2 Hop
Logistic Regression	SGI, PPI, co-expression, phenotype	*C. elegans*	Zhong and Sternberg (2006)	Gene Orienteer
Network Sampling	SGI, PPI, GO	*S. cerevisiae*	Le Meur and Gentleman (2008) Le Meur *et al.* (2014)	SLGI(R)
Random Walk	GO, PPI, expression	*S. cerevisiae* *C. elegans*	Chipman and Singh (2009)	
Shared Function	Co-expression, PPI, text mining, phylogeny	*C. elegans*	Lee *et al.* (2010b)	WormNet
Logistic Regression	Co-expression, PPI, phenotype	*C. elegans*	Lee *et al.* (2010a)	GI Finder
Jaccard Index	GO, SGI, PPI, phenotype	Eukarya	Hoehndorf *et al.* (2013)	
Machine Learning			Pandey *et al.* (2010)	MNMC
Machine Learning Meta-Analysis			Wu *et al.* (2014)	MetaSL
Flux Variability Analysis Flux Balance Analysis Network Simulation	Metabolism	*E. coli* *M. pneumoniae*	Güell *et al.* (2014)	

There are a number of existing computational methods for predicting synthetic lethal gene pairs in humans with a specific interest in cancer (as summarised in 1.2). While these demonstrate the power and need for predictions of synthetic lethality in human and cancer contexts, limitations of previous methods could be met with a different approach. Existing computational approaches to synthetic lethal prediction are often difficult to interpret, replicate for new genes, or are reliant on data types not available for a wider range of genes to test.

1.2.7.2 Comparative Genomics

A comparative genomic approach by Deshpande *et al.* (2013) used the results of well characterised high-throughput mutation screens in *S. cerevisiae* as candidates for synthetic lethality in humans (Baryshnikova *et al.*, 2010a; Costanzo *et al.*, 2010, 2011; Tong *et al.*, 2001, 2004). Yeast synthetic lethal partners were compared to human orthologues to find cancer relevant synthetic lethal candidate pairs with direct therapeutic potential. Proposed as a complementary approach to siRNA screens, approximately 24,000 of the 116,000 negative SGIs in yeast (Costanzo *et al.*, 2011) were matched to human orthologues, with over 500 involving a cancer gene (Futreal *et al.*, 2004). Under strict

Table 1.2: Methods for Predicting Synthetic Lethality in Cancer

Method	Input Data	Source	Tool Offered
Network Centrality	protein-protein interactions	Kranthi *et al.* (2013)	
Differential Expression	Expression Mutation	Wang and Simon (2013)	
Comparative Genomic Chemical-Genomic	Yeast synthetic gene interactions Homology	Heiskanen and Aittokallio (2012)	
Comparative Genomic	Yeast synthetic gene interactions Homology	Deshpande *et al.* (2013)	
Machine Learning		Discussed by Babyak (2004) and Lee and Marcotte (2009)	
Differential Expression	Expression	Tiong *et al.* (2014)	
Literature Database		Li *et al.* (2014)	Syn-Lethality
Meta-Analysis	Meta-Analysis Machine Learning	Wu *et al.* (2014)	MetaSL
Pathway Analysis		Zhang *et al.* (2015)	
Protein Domains	Homology	Kozlov *et al.* (2015)	
Data-Mining Machine Learning	Expression Somatic mutation and DNA CNV siRNA in cell lines	Jerby-Arnon *et al.* (2014) Ryan *et al.* (2014) Crunkhorn (2014) Lokody (2014)	DAISY (method)
Genome Evolution Hypothesis Test Machine Learning	Expression DNA CNV Known SL	Lu *et al.* (2013) Lu *et al.* (2015)	
Bimodality	Expression DNA CNV Somatic Mutation	Wappett (2014) Wappett *et al.* (2016)	BImodal Subsetting ExPression (BiSEp)
Directional Chi-Square	Expression (microarray) Somatic mutation	Kelly, S. T., Guilford, P. J., and Black, M. A. Dissertation (Kelly, 2013) and developed here	SLIPT

criteria of one-to-one orthologues, large effect size and significant interaction in yeast data ($\epsilon < -0.2$, $p < 0.05$), 1522 interactions were identified with 70 involving cancer genes. Of the 21 gene interactions tested with pairs of siRNA in IMR1 fibroblast cells, 6 exhibited synthetic lethal effects. The two strongest interactions (*SMARCB1* with *PSMA4* and *ASPSCR1* with *PSMC2*) were successfully validated by protein analysis of human cells and replication with tetrad analysis for yeast orthologues.

Another approach to systematic synthetic lethality discovery specific to human cancer (in contrast to the plethora of yeast synthetic lethality data) was to build a database as done by Li *et al.* (2014). In their relational database, called "Syn-lethality", they have curated both known experimentally discovered synthetic lethal pairs in humans (113 pairs) from the literature and those predicted from synthetic lethality between orthologous genes in *S. cerevisiae* yeast (1114 pairs). This knowledge-based database is the first dedicated to human cancer synthetic lethal interactions and integrates gene function annotation, pathway and molecular mechanism data with experimental and predicted synthetic lethal gene pairs. This combination of data sources is intended to tackle the trade-off between more conclusive synthetic lethal experiments in yeast and more clinically relevant synthetic lethal experiments in human cancer models,

such as RNAi, especially when high-throughput screens are costly and prone to false positives in either system and difficult to replicate across gene backgrounds. This database centralises a wealth of knowledge scattered in the literature including cancer relevant genes (*BRCA1*, *BRCA2*, *PARP1*, *PTEN*, *VHL*, *MYC*, *EGFR*, *MSH2*, *KRAS*, and *TP53*) and is publicly available as a Java App. These included the previously mentioned interactions of *BRCA1* and *BRCA2* with *PARP1* and *TP53* with *WEE1* and *PLK1*. However, the computational methodology was not released, so it is not possible to replicate their results, nor to add to the findings with new datasets, which are limited to 647 human genes. Suggested future directions were promising, such as constructing networks of known synthetic lethality, applying known synthetic lethality to cancer treatment, data mining, replicating the approach for synthetic lethality in model organisms, signalling pathways, and developing a complete global network in human cancer or yeast (both of which are still incomplete with experimental data), some of which has been implemented in "SynLethDB" (Guo *et al.*, 2016).

Table 1.3: Machine Learning Methods used by Wu *et al.* (2014)

Method	**Source**	**Tool Offered**
Random Forest	Breiman (2001)	
Random Forest	Hall *et al.* (2009)	WEKA
J48 (decision tree)		
Bayes (Log Regression)		
Bayes (Network)		
PART (Rule-based)		
RBF Network		
Bagging / Bootstrap		
Classification via Regression		
Support Vector Machine (Linear)	Vapnik (1995)	
Support Vector Machine (RBF – Gaussian)	Joachims (1999)	
Multi-Network Multi-Class (MNMC)	Pandey *et al.* (2010)	
MetaSL (Meta-Analysis)	Wu *et al.* (2014)	MetaSL

Machine learning approaches have also been proposed for synthetic lethal discovery (Babyak, 2004; Lee and Marcotte, 2009). Due to concerns that these may be subject to overfitting or noise, Wu *et al.* (2014) developed a meta-analysis method (based on the machine learning methods in Table 1.3) for synthetic lethal gene pairs relevant to developing selective drugs against human cancer, building upon their previous database

(Li *et al.*, 2014). They used training data of 10,885 synthetic lethal interactions from yeast experiments of which 7347 occurred between the 5,504 non-essential genes. Their "metaSL" approach utilises genomic, proteomic and annotation data (including GO terms Ashburner *et al.* (2000), PPI, protein complexes, and biological pathway) with strong statistical performance in yeast data (area under the receiver operating characteristic (AUROC) of 0.871). The predicted orthologous synthetic lethal partners in human data were not experimentally validated but several would be relevant to cancer such as *EGFR* with *PRKCZ*. They note that computational approaches scale-up across the genome at lower cost than experimental screen and share their most supported interactions online. However, the method is not available for analysis of other genes studied by the cancer research community. While machine learning has great potential as a predictor, the results vary greatly depending on the predictive features selected and it is not clear which threshold should be used to report reliably detected genes. Syn-Lethality (Li *et al.*, 2014) and MetaSL (Wu *et al.*, 2014) demonstrate the value of computational approaches to synthetic lethality but omit many genes of importance in cancer, such as *CDH1*, and there remains a need to enable biological researchers to query such genes in a particular tissue or genetic background.

There is also concern for analyses based on yeast data that many synthetic lethal interactions may not be conserved between species Dixon *et al.* (2009), although interactions between pathways may be more comparable. It is unsurprising that many of the interactions identified were not experimentally validated. There have been many gene duplications in the separate evolutionary histories of humans and yeast which may lead to differences in genetic redundancy. Yeast are not an ideal human cancer model because they are do not have tissue specificity, multicellular gene regulation, or orthologues to a number of known cancer genes such as p53. Although these studies have tried to anticipate these issues with stringent criteria such as requiring one-to-one orthologues, there remains the possibility that changes in gene function may affect whether these are solely redundant such as if functions had coevolved without sequence homology. Many genes will also be excluded since they lack homologues in yeast, the corresponding experimental data, or having paralogues in either species. Thus conservation of yeast interactions is not an ideal strategy and analysis of human data directly for comparison with human experimental data will be the focus of this thesis.

1.2.7.3 Analysis and Modelling of Protein Data

Kranthi *et al.* (2013) took a network approach to discovery of synthetic lethal candidate selection applying the concept to "centrality" to a human PPI network involving

interacting partners of known cancer genes. The effect of removing pairs of genes on connectivity of the network was used as a surrogate for viability which is supported by observations that the PPI and synthetic lethal networks are orthogonal in *S. cerevisiae* studies (Tong *et al.*, 2004). They showed that the human cancer protein interaction network (of 1539 proteins and 6471 interactions) exhibits the power law distribution expected of a scale-free synthetic lethal network with high connectivity (average vertex degree of 23.67 and network efficiency of 0.2952). Their top 100 candidate interactions included interactions of the tumour suppressor *TP53* with *BRCA1*, *CDKNA1*, *CDKNA2*, *MET*, and *RB1* which have been detected by prior studies. The gene pairs were often observed to be in the same or a plausible compensatory pathway. Thus the network structure is important in the biological functions of cancers and could be exploited for targeting *TP53* loss of function mutations.

However, their approach was limited to known cancer genes and is not applicable to genes that do not have PPI data. Other nucleotide sequencing data types are more commonly available for cancer studies at a genomic scale. Of further concern is that the results were enriched for p53 synthetic lethal partners which is relevant to many cancer researchers but this genome-wide approach did not detect many other cancer genes due to multiple testing. This enrichment may be due to the known drastic effect of removing p53 itself from the network as a master regulator, cancer driving tumour suppressor gene, and highly connected network "hub". The focus on cancer genes is useful for translation into therapeutics but does not account for variable genetic backgrounds or effect of protein removal on the whole cellular network.

Focusing on the potential for synthetic lethality to be an effective anti-cancer drug target, Zhang *et al.* (2015) used modelling of signalling pathways to identify synthetic lethal interactions between known drug targets and cancer genes by simulating gene knockdowns. A computational approach was applied to avoid the limitations of experimental RNAi screens such as scale, instability of knockdown, and off-target effects. This 'hybrid' method of a data-driven model and known signalling pathways showed potential as a means to predict cell death in single and combination gene knockouts. They used time series protein phosphorylation data (Lee *et al.*, 2012) for 28 signalling proteins and Gene Ontology (GO) (GO) pathways Ashburner *et al.* (2000); Blake *et al.* (2015). This approach successfully detected many known essential genes in the human gene essentiality database, known synthetic lethal partners in the Syn-Lethality database (Li *et al.*, 2014), and predicted novel synthetic lethal gene pairs. The strongest essential genes in single knockdowns were *AKT*, *TP53*, *CHK1*, *S6K1*, and *CYCLIND1*.

Pairwise knockdowns identified 252 candidate synthetic lethal interactions including *TP53* with *CHK1*, *S6K1*, *WEE1*, *CYCLIND1*, and *CASP9*; *AKT* with *WEE1*; and *CDK1* with *CYCLIND1*. These novel results contained many *TP53* and AKT synthetic lethal partners, genes known to be important in many cancers, however these also have a severe impact on the signalling pathways in their essentiality analysis of single gene disruptions and large phenotypic changes in cancer. This approach is amenable to detect functionally related pathways and protein complexes across the molecular function, cellular component, and biological process annotations provided by GO. The results were consistent with the experimental results in the literature but the novel synthetic lethal interactions have yet to be validated. While the mathematical reasoning and algorithms are given, the code was not released to replicate the findings or apply the methodology beyond the signalling pathways analysed by Zhang *et al.* (2015). While this is an interesting approach, the analysis of this thesis will focus on gene expression and RNAi data which is available to test a wider range of candidate gene pairs.

1.2.7.4 Differential Gene Expression

Differential gene expression has been explored to predict synthetic lethal pairs in cancer which would be widely applicable due to the availability of public gene expression data for a large number of samples and cancer types. Wang and Simon (2013) found differentially expressed genes (by the t-test, adjusted by FDR) between tumours with or without functional p53 mutations in TCGA (McLendon *et al.*, 2008) and Cancer Cell Line Encyclopaedia (CCLE) (Barretina *et al.*, 2012) RNA-Seq gene expression data as candidate synthetic lethal partner pathways of p53. They identified 2, 8, and 21 candidate synthetic lethal partner genes in 3 microarray datasets from the NCI60 cell lines, 31 partner genes from the CCLE RNA-Seq data (Barretina *et al.*, 2012), and 50 in TCGA RNA-Seq data (Muzny *et al.*, 2012). *PLK1* was replicated across 4 of these analyses and 17 other genes were replicated across 2 analyses (including *MTOR*, *PLK4*, *MAST2*, *MAP3K4*, *AURKA*, *BUB1* and 6 CDK genes) with many playing a role in cell cycle regulation. This was supported by a drug sensitivity experiment on the NCI60 cell lines which found that cells which lacked functional p53 were more sensitive to paclitaxel (which targets *PLK1*, *AURKA*, and *BUB1*). This demonstrated the potential of gene expression as a surrogate for gene function and use of public genomic data to predict synthetic lethal gene pairs in cancer. Wang and Simon (2013) advocated for pre-screening of expression profiles to augment future RNAi screens. However, the analyses were limited to kinase genes and focused on currently druggable genes, lacking

wider application of synthetic lethal prediction methodology. This approach may not be feasible or applicable in cancer genes with a lower mutation rate than p53.

Tiong *et al.* (2014) also investigated gene expression as a predictor of synthetic lescale-freethal pairs with colorectal cancer microarrays from a Han Chinese population with a sample size of 70 tumour and 12 normal tissue samples. Simultaneously differentially expression of "tumour dependent" gene pairs (which includes co-expression) between cancer and normal tissue was used to rank 663 candidate synthetic lethal interactions identified in cell line siRNA experiments. Of the top 20 genes, 17 were tested for testing differential expression at the protein level with immunohistochemistry staining and correlation with clinical characteristics, with 11 pairs exhibiting synergistic effects. Some of the predicted synthetic lethal pairs were consistent with the literature (including *TP53* with *S6K1* and partners of *KRAS*, *PTEN*, *BRCA1*, and *BRCA2*) and two novel synthetic lethal interactions (*TP53* with *CSNK1E* and *CTNNB1*) were validated in pre-clinical models. This serves a valuable proof-of-concept for integration of *in silico* approaches to synthetic lethal discovery in cancer demonstrating it's utility to triage and identify synthetic lethal partners of p53 applicable to colorectal tissues. Although the experimental work was the focus of the paper, these findings show that bioinformatics synthetic lethal candidates can be validated in patient tissue samples (from a non-caucasian population) to find those applicable to colorectal cancers.

1.2.7.5 Data Mining and Machine Learning

Recognising the utility of synthetic lethality to drug inhibition and specificity of anticancer treatments, Jerby-Arnon *et al.* (2014) also saw the need for effective prediction of gene essentiality and synthetic lethality to augment experimental studies of SL. They developed the DAta mIning SYnthetic lethal identification pipeline (DAISY), a data-driven approach for genome-wide analysis of synthetic lethality in public cancer genomics data from TCGA and CCLE (Barretina *et al.*, 2012). DAISY is intended to predict the candidate synthetic lethal partners of a query gene such as genes recurrently mutated in cancer.

Jerby-Arnon *et al.* (2014) combined a computational approach to triage candidates with a conventional RNAi screen to validate synthetic lethal partners. They screened a selection of computationally predicted candidates and randomly selected genes with RNAi against *VHL* loss of function mutation in RCC4 renal cell lines. The computational method had a high AUROC of 0.779 and predictions were enriched 4× for validated RNAi hits over randomly selected genes. This approach detected known synthetic lethal pairs such as *BRCA1* or *BRCA2* genes with *PARP1* and *MSH2* with

DHFR. The synthetic lethal candidates identified with both RNAi screening and computational prediction formed an extensive network of 2077 genes with 2816 synthetic lethal interactions and similar network of 3158 genes with 3635 synthetic dosage lethal interactions (for synthetic lethality with over-expression). Each network was scale-free as expected of a biological network and was enriched for known cancer genes, essential genes in mice, and could be harnessed for predicting prognosis and drug response. While demonstrating the feasibility of combining experimental and computational approaches to synthetic lethality in cancer, there remain challenges in predicting synthetic lethal genes, novel drug targets, and translation into the clinic.

The DAISY methodology (Jerby-Arnon *et al.*, 2014) compares the results of analysis of several data types to predict synthetic lethality, namely: DNA copy number and somatic mutation for TCGA patient samples and CCLE cell lines. The cell lines were also analysed with gene expression and gene essentiality (shRNA screening) profiles. Genes were classed as inactivated by copy number deletion, somatic loss of function mutation, or low expression and tested for synthetic lethal gene partners which are either essential in screens or not deleted with copy number variants. Co-expression is also used for synthetic lethality prediction based on studies in yeast (Costanzo *et al.*, 2010; Kelley and Ideker, 2005). Copy number, gene expression and, essentiality analyses are stringently compared by adjusting each for multiple tests with Bonferroni correction and only taking hits which occur in all analyses. This methodology was also adapted for synthetic dosage lethality by testing for partner genes where genes are overactive with high copy number or expression. As discussed above, the predictions performed well and an RNAi screen for the example of *VHL* in renal cancer validated predicted synthetic lethal partners of *VHL* demonstrating the feasibility of combining approaches to synthetic lethal discovery in cancer and using computational predictions to enable more efficient high-throughput screening. DAISY performs well statistically with a AUROC of 0.779 on a set of gene pairs with experimental screen data, although co-expression and shRNA functional examination contributes much less of this than the mutation and copy number analysis (AUROC 0.683 alone). However, this methodology is very stringent, missing potentially valuable synthetic lethal candidates, may not be applicable to genes of interest to other groups and the software for the procedure has not been publicly released for replication.

Although the DAISY procedure performs well and has been well received by the scientific community (Crunkhorn, 2014; Lokody, 2014; Ryan *et al.*, 2014), showing a need for such methodology, there is no indication of adoption of the methodology in the

community yet. The co-expression analysis may not be the most effective way to test gene expression for directional synthetic lethal interactions (where inverse correlation would be expected). In the interests of a large sample size, tissue types were not tested separately despite tissue-specific synthetic lethality being likely since gene function (and by extension expression, isoforms, and clinical characteristics) in cancers may often be tissue-dependent. Some data forms and analyses used, such as gene essentiality, may not be available for all cancers, genes, or tissues, and may not be reproduced.

Lu *et al.* (2015) critique the assumption of co-expression in the DAISY methodology and propose an alternative computational prediction of synthetic lethality based on machine learning methods and a "cancer genome evolution" hypothesis. Using DNA copy number and gene expression data from TCGA patient samples, a cancer genome evolution model assumes that synthetic lethal gene pairs behave in two distinct ways in response to an inactive synthetic lethal partner gene, either a "compensation" pattern where the other synthetic lethal partner is overactive or a "co-loss underrepresentation" pattern where the other synthetic lethal partner is less likely to be lost, since loss of both genes would cause death of the cancer cell. During the genome evolution of cancers, the cell becomes addicted to the remaining synthetic lethal partner due to induced gene essentiality. These patterns would explain why DAISY detects only a small number of synthetic lethal pairs, compared to the large number expected based on model organism studies (Boone *et al.*, 2007), and the disparity between screening and computationally predicted synthetic lethal candidates due to testing different classes of synthetic lethal gene pairs.

Lu *et al.* (2015) compared a genome-wide computational model of genome evolution and gene expression patterns to the experimental data of Vizeacoumar *et al.* (2013) and Laufer *et al.* (2013). This simpler model performing well with an AUROC of 0.751 but was less than DAISY, although it did not rely on data from cell lines which may not represent patient disease. They predict a larger comprehensive list of 591,000 human synthetic lethal partners with a probability score threshold of 0.81, giving a precision of 67% and 14× enrichment of synthetic lethal true positives compared to randomly selected gene pairs. Discovery of such a vast number of cancer-relevant synthetic lethal interactions in humans would not be feasible experimentally and is a valuable resource for research and clinical applications. These predictions are not limited by assuming co-expression of synthetic lethal partners or evolutionary conservation with model organisms enabling wider synthetic lethal discovery. However, there remains a lack of basis for an expectation of how many synthetic lethal partners a particular gene

will have, how many pairs there are in the human genome, and whether pathways or correlation structure would influence predicted synthetic lethal partners.

Large scale, computational approaches have yet to determine whether synthetic lethal interactions are tissue-specific since Lu *et al.* (2015) used pan-cancer data for 14136 patients with 31 cancer types. Experimental data used for comparison was a small training dataset specific to colorectal cancer, and based on screens for other phenotypes, which may limit performance of the model or application to other cancers. Proposed expansion of the computational approach to mutation, microRNA, or epigenetic modulation of gene function and tumour micro-environment or heterogeneity suggests that synthetic lethal discovery could be widely applied to the current challenges in cancer genomics. This approach was also based on machine learning methodology and not supported by a software release for the community to develop, contribute to, or reproduce beyond the gene pairs given in the supplementary results.

1.2.7.6 Bimodality

Wappett *et al.* (2016) demonstrate a multi-omic approach to identification of synthetic lethality in cancer with a strategy to detect bimodal patterns in molecular profiles. They released this solution as the BiSEp R package Wappett (2014) which aims to detect subtle bimodal and non-normal patterns in expression data. Since loss of gene function is not consistently genetic, Wappett *et al.* (2016) advocate the use of gene expression (loss of messenger RNA (mRNA)) and deletion (loss of copy number) data in addition to mutation. The BiSEp procedure was demonstrated on an analysis of 881 cell lines from CCLE (Barretina *et al.*, 2012), 442 cell lines from COSMIC (Forbes *et al.*, 2015), and RSEM normalised RNA-Seq data for 178 TCGA lung patient samples (Collisson *et al.*, 2014). BiSEp was demonstrated to have significant enrichment of validated tumour suppressor, synthetic lethal gene pairs (detecting 76 experimentally supported gene pairs) and was improved (detecting 420) with expression data rather than relying on detecting loss of gene function by mutation or deletion. They identified interactions with genes relevant to cancer with support in experimental screens including *ERCC4* with *XRCC1*, *BRCA1* with *PARP3*, and *SMARCA1* with *SMARCA4*.

Wappett *et al.* (2016) demonstrated that analysis of genomics data, particularly expression data, is relevant to augment the identification of synthetic lethal interactions with screening experiments. They further show that this is applicable in both genetically homogeneous cell lines and heterogeneous cell population from patient samples. This approach is limited however to genes which exhibit bimodal expression patterns which do not commonly occur, particularly in normalised gene expression data, and

other approaches may need to be considered for gene such as *CDH1* which were not identified by BiSEp.

1.2.7.7 Rationale for Further Development

Many of the approaches discussed here aimed to identify the strongest synthetic lethal pairs across the yeast or human genomes (Deshpande *et al.*, 2013; Lu *et al.*, 2015; Wappett *et al.*, 2016; Wu *et al.*, 2014), which may not be an ideal strategy to identify interactions in particular functions or relevance to particular cancers. These demonstrate a need for computational approaches to prioritise candidate gene pairs for validation but this thesis will focus on the interactions with *CDH1* with particular importance in breast and stomach cancers, although these partners may be applicable in other cancers. As such, this thesis presents a query-based method, amenable to identification of candidate partners for a selected gene of functional or translational importance such as *CDH1*.

1.3 E-cadherin as a Synthetic Lethal Target

E-cadherin is a transmembrane protein (encoded by *CDH1*) with several characterised functions in the cytoskeleton and cell-to-cell signalling. Here we outline the characterised functions of E-cadherin and its importance in cancer biology. *CDH1* is a tumour suppressor gene, with loss of function occurring in both familial (germline mutations) and sporadic (somatic mutations) cancers. As such, *CDH1* inactivation is a prime example of a genetic event that could be targeted by synthetic lethality for anti-cancer treatments. Most notably this includes patients at risk of developing hereditary breast and stomach cancers for which conventional surgical or cytotoxic chemotherapy is not ideal and who have a known genetic aberration in their familial syndromic cancers. Effective treatments against *CDH1* inactivation would also benefit patients with sporadic diffuse gastric cancers since they often present with symptoms at a late stage.

1.3.1 The *CDH1* gene and its Biological Functions

The tumour suppressor gene *CDH1* is implicated in hereditary and sporadic lobular breast cancers (Berx *et al.*, 1996; Berx and van Roy, 2009; De Leeuw *et al.*, 1997; Masciari *et al.*, 2007; Semb and Christofori, 1998; Vos *et al.*, 1997). The *CDH1* gene encodes the E-cadherin protein and is normally expressed in epithelial tissues, where it has also been identified as an invasion suppressor and loss of *CDH1* function has been implicated in breast cancer progression and metastasis (Becker *et al.*, 1994; Berx *et al.*, 1995; Christofori and Semb, 1999).

1.3.1.1 Cytoskeleton

The primary function of *CDH1* is cell-cell adhesion forming the adherens junction, maintaining the cytoskeleton and mediating molecular signals between cells. The function of the adherens complex is particularly important for cell structure and regulation because it interacts with cytoskeletal actins and microtubules. The cytoskeletal role of E-cadherin maintains healthy cellular viability and growth in epithelial tissues including cellular polarity (Jeanes *et al.*, 2008). E-cadherin is not essential to cellular viability but loss in epithelial cells does lead to defects in cytoskeletal structure and proliferation. In addition to a central role in the adherens complex, E-cadherin is involved in many other cellular functions and thus *CDH1* is regarded as a highly pleiotropic gene (Kroepil *et al.*, 2012).

1.3.1.2 Extracellular and Tumour Micro-environment

As a transmembrane signalling protein E-cadherin also interacts with the extracellular environment and other cells, most notably forming tight junctions between cells (Chen *et al.*, 2014; Tunggal *et al.*, 2005). These junctions serve to both regulate movement of ion signals between cells and separate membrane proteins on the apical and basal surfaces of a cell, maintaining cell polarity. Thus E-cadherin is an important regulator of epithelial tissues by intercellular communication (Jeanes *et al.*, 2008). It also has important roles in the extracellular matrix, including fibrin clot formation. The role of intercellular interactions and the tissue micro-environment are important themes in cancer research, being a potential mechanism for cancer progression and malignancy in a addition to it's potential for specifically targeting tumour cells.

1.3.1.3 Cell-Cell Adhesion and Signalling

The signals mediated by tight junctions are also passed on to intracellular signalling pathways and thus E-cadherin also has a role in maintaining cellular function and growth. One such example is the regulation of β-catenin which interacts with both the actin cytoskeleton and acts as a transcription factor via the WNT pathway (Jeanes *et al.*, 2008). Similarly, the Hippo and PI3K/AKT pathways are implicated in being mediated by E-cadherin (De Santis *et al.*, 2009; Kim *et al.*, 2011), having roles in promoting cell survival, proliferation, and repressing apoptosis. E-cadherin shares several downstream pathways with signalling pathways such as integrins and thus indirectly interacts with them, particularly since feedback loops may occur in such pathways. Conversely, the multifaceted roles of E-cadherin have been shown with over-expression

in ovarian cells promoting tumour growth, while it maintains healthy cellular functions in other cells (Brouxhon *et al.*, 2014; Dong *et al.*, 2012).

1.3.2 *CDH1* as a Tumour (and Invasion) Suppressor

E-cadherin has key roles in maintaining cellular structure and regulating growth, consistent with *CDH1* being a tumour suppressor gene. Loss of *CDH1* in epithelial tissues leads to disrupted cell polarity, differentiation, and migration. E-cadherin loss has been identified as a recurrent driver tumour suppressor mutation in sporadic cancers of many tissues including breast, stomach, lung, colon, and pancreas tissue.

1.3.2.1 Breast Cancers and Invasion

E-cadherin loss in breast cancers has been shown to cause increased proliferation, lymph node invasion, and metastasis with poor cell-cell contact. Thus *CDH1* gene has also been implicated as an invasion suppressor, with a key role in the epithelial-mesenchymal transition (EMT), an established mechanism of cancer progression (Hanahan and Weinberg, 2011). The epithelial-mesenchymal transition is important during development and wound healing but such changes in cellular differentiation also occur in cancers. If *CDH1* is inactivated by mutation or DNA methylation (Berx *et al.*, 1996; Guilford, 1999; Machado *et al.*, 2001), it is likely that EMT will drive growth of E-cadherin deficient cancers (Berx and van Roy, 2009; Graziano *et al.*, 2003; Polyak and Weinberg, 2009). While loss of E-cadherin is not sufficient to cause EMT or tumourigenesis, it is an important step in this mechanism of tumour progression and a potential therapeutic intervention may therefore also impede cancer progression and have activity against advanced stage cancers.

1.3.3 Hereditary Diffuse Gastric Cancer and Lobular Breast Cancer

CDH1 loss of function mutations also causes familial cancers, including diffuse gastric cancer and lobular breast cancer (Graziano *et al.*, 2003; Guilford *et al.*, 2010, 1999; Oliveira *et al.*, 2009). Individuals carrying a null mutation in *CDH1* have a syndromic predisposition to early-onset these cancers, known as hereditary diffuse gastric cancer (HDGC) (Guilford *et al.*, 1998). Due to the loss of an allele, these individuals are prone to carcinogenic lesion in the breast and stomach when the other allele is inactivated, occurring much more frequently and thus younger than in individuals without a second functional allele of *CDH1*. The loss of the second allele is most often hypermethylation suppressing expression rather than mutation (Grady *et al.*, 2000; Machado *et al.*, 2001;

Oliveira *et al.*, 2009), although loss of heterozygousity may also occur (Guilford *et al.*, 2010). Therefore HDGC is an autosomal dominant cancer syndrome with incomplete penetrance. The "lifetime" (until age 80 years) risk for mutation carriers of diffuse gastric cancer is 70% in males and 56% in females (Hansford *et al.*, 2015; van der Post *et al.*, 2015). In addition, the lifetime risk of lobular breast cancer is 42% in female mutation carriers.

HDGC affects less than 1 in a million people globally (Ferlay *et al.*, 2015) and less than 1% of gastric cancers. However, HDGC is documented to affect several hundred families globally. E-cadherin mutations in the germline is implicated in 1-3% of gastric cancers presenting with a family history, varing between high and low incidence populations. E-cadherin is also mutated in 13% of sporadic gastric cancers.

While diagnostic testing for *CDH1* genotype has enabled more effective management of HDGC and improved patient outcomes, there are still limited options for clinical interventions (Guilford *et al.*, 2010). Individuals with a family history of HDGC are recommended to be tested for *CDH1* mutations in late adolescence and are offered prophylactic stomach surgery before the risk of developing cancers increases with age. Another option is annual endoscopic screening to diagnose early stage stomach cancers with surgical intervention once they are detected (Oliveira *et al.*, 2013). However, these early stage cancers are difficult to detect and may be missed in regular screening. Thus patients carrying *CDH1* mutations either have surgical interventions with a significant impact on quality of life and risk of complications or remain at risk of developing advanced stage stomach cancers. Due to the lower mortality rate due to stomach cancers, there is increasing concerns among these HDGC families on the elevated risk of lobular breast cancers for women later in life.

The current clinical management of HDGC still has significant risks for patients and therefore a greater understanding of the molecular and cellular function of *CDH1* is important for its role in these cancers. Such studies may lead to alternative treament strategies such as pharmacological treatments with specificity against *CDH1* null cells, once they lose the second allele. While a loss of gene function cannot be targeted directly, designing a treatment with specificity against *CDH1* may also have activity in sporadic cancers in a range of epithelial cancers. Thus an effective treatment against *CDH1* mutant cancers would potentially have significant therapeutic and preventative applications in a large number of patients.

1.3.4 Cell Line Models of *CDH1* Null Mutations

Previous work our research group has published used a model of homozygous $CDH1^{-/-}$ null mutation in non-malignant MCF10A breast cells to show that loss of *CDH1* alone was not sufficient to induce EMT with compensatory changes in the expression of other cell adhesion genes occurring (Chen *et al.*, 2014). However, *CDH1* deficient cells did manifest changes in morphology, migration, and weaker cell adhesion (Chen *et al.*, 2014).

This $CDH1^{-/-}$ MCF10A model has been used in a genome-wide screen of 18,120 genes using short interfering RNA (siRNA) and a complementary drug screen using 4057 compounds to identify synthetic lethal partners to E-cadherin (Telford *et al.*, 2015). One of the strongest candidate pathways identified by Telford *et al.* (2015) were the GPCR signalling cascades, which were highly enriched by GO analysis of the candidate synthetic lethal partners the primary siRNA screen. This was supported by validation with Pertussis toxin, known to target $G_{\alpha i}$ signalling (Clark, 2004), as were various candidate cytoskeletal pathways by inhibition of Janus kinase (JAK) and aurora kinase. The drug screen also produced candidates in histone deacetylase (HDAC) and phosphoinositide 3-kinase (PI3K) which were supported by validation and time course experiments.

1.4 Summary and Research Direction of Thesis

Genomic technologies and the data available from them have immense potential for understanding of genetics and improving healthcare, including identification of genes altered in cancer for molecular diagnosis, prognostic biomarkers, and therapeutic targets. This has been demonstrated with the identification of cancer genes in many cancers, distinguishing tumour subtypes by expression profiles, and the development of targeted therapies against oncogenes (such as *BRAF*) and tumour suppressors (such as *BRCA1*). Synthetic lethality is an important genetic interaction to study fundamental cellular functions and exploit them for biomarkers and cancer treatment. They present a means to target loss of function mutations and genetic dysregulation in tumour suppressor genes by identifying interacting partners with redundant or compensating molecular functions.

CDH1 (encoding E-cadherin) is an example of a tumour suppressor gene implicated in sporadic breast and stomach cancers. Germline mutations in *CDH1* are also found in many patients with familial early onset cancers (HDGC). Discovery of synthetic lethal partners would be contribute to an understanding on the molecular mechanisms driv-

ing the growth of *CDH1* deficient tumours and identification of potential therapeutic targets or chemopreventative agents for management of HDGC. The clinical potential of the synthetic lethal approach has been demonstrated with the application of olaparib against *BRCA1* and *BRCA2* mutations Lord *et al.* (2015) but there remains the need to systematically identify synthetic lethal partner genes for other tumour suppressors such as *CDH1*. A synthetic lethal screen has been conducted on breast cell lines Telford *et al.* (2015) but computational approaches to identification of synthetic lethal partners of *CDH1* remains to be done.

While there are a wide range of experimental and computational approaches to synthetic lethal discovery, many are limited to particular applications, prone to false positives, inconsistent across independent approaches, or enriched for particular genes of interest. Therefore synthetic lethal interactions are difficult to replicate or apply in the clinic. Computational approaches to synthetic lethality are not widely adopted by the cancer research community and experimental approaches cannot be combined to study synthetic lethality at a genome-wide scale. However, these show interest in synthetic lethal discovery in the community and the need for robust predictions of synthetic lethal interactions in cancer and human tissues.

Effective screening, prediction, and analysis of synthetic lethal interactions are a crucial part of developing next generation anti-cancer strategies. Therefore, we propose developing a computational statistical procedure to identify synthetic lethal interactions and construct gene networks. This will enable the development of personalised medicine targeted to particular molecular aberrations. Genetic tests and genomic have the potential to revolutionise cancer screening, diagnosis, and prognostics; targeted therapeutics, similarly, have applications in prevention and therapy of sporadic or hereditary cancers with known molecular properties.

To address the concerns raised by recent computational approaches to synthetic lethal discovery in cancer (Jerby-Arnon *et al.*, 2014; Lu *et al.*, 2015; Wappett *et al.*, 2016), I present similar analysis using solely gene expression data which is widely available for a large number of samples in many different cancers. This uses a statistical methodology the Synthetic Lethal Interaction Prediction Tool (SLIPT) developed for this purpose. To further determine the limitations and implications of synthetic lethal predictions, modelling and simulation was performed upon the statistical behaviour of synthetic lethal gene pairs in genomics data. Comparison of synthetic lethal gene candidates from public data analysis and experimental candidates, pathway analysis, and networks structure will also be presented to investigate the relationships between

synthetic lethal candidates. Release of R codes used for simulation, prediction, and analysis will enable adoption of the methodology in the cancer research community and comparison to existing methods. Therefore my thesis aims to develop such predictions for synthetic lethal partner genes with a focus on the example of E-cadherin to compare to the findings of Telford *et al.* (2015), develop of network approaches for pathway structure, and simulate gene expression on pathway structure with the bioinformatics and computational biology investigations.

1.4.1 Thesis Aims

Understanding synthetic lethality is important in cancers, having shown an impact clinical practice and patient outcomes for certain genes already. Thus this thesis aims to identify synthetic lethal gene pairs using public gene expression data. Accordingly, chapter 3 aims to develop the methods to do so including a synthetic lethal detection methodology (SLIPT) and its release as an R software package. This chapter also serves to document the original simulation and network analysis procedures developed to support the use of SLIPT and perform analyses in throughout this thesis.

This thesis also aims to demonstrate SLIPT methodology for analysis of RNA-Seq gene expression data. Chapter 4 aimed to do so by performing an analysis to identify candidate synthetic lethal gene partners of *CDH1* in public breast and stomach cancer data (Bass *et al.*, 2014; TCGA, 2012). Chapter 4 also aimed to demonstrate the biological relevance of these candidate synthetic lethal partners by identifying synthetic lethal pathways and comparing them with the results of an experimental siRNA screen (Telford *et al.*, 2015).

Pathway analysis was extended with graph structures in chapter 5 which aimed to assess the importance of synthetic lethal genes within pathway structures. Chapter 5 also aimed to use pathway structure to identify directional relationships between SLIPT and siRNA synthetic lethal candidates and explore the disparity between them. The SLIPT methodology was also supported by simulations investigations in chapters 3 and 6 which aimed to evaluate the SLIPT method on detecting known synthetic lethal genes in simulated data. Graph structures were also used in chapter 6 which also aimed to determine the effect of pathway structures of synthetic lethal detection with SLIPT in simulated data and ascertain that the simulation results were comparable to expression data with contains complex correlation structures within biological pathways.

References

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

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

Adler, D. (2005) *vioplot: Violin plot.* R package version 0.2.

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

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

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

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

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

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, **25**(1): 25–29.

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

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

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., Gulbahce, N., and Loscalzo, J. (2011) Network medicine: a network-based approach to human disease. *Nat Rev Genet*, **12**(1): 56–68.

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

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

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

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

Baryshnikova, A., Costanzo, M., Dixon, S., Vizeacoumar, F.J., Myers, C.L., Andrews, B., and Boone, C. (2010a) Synthetic genetic array (sga) analysis in saccharomyces cerevisiae and schizosaccharomyces pombe. *Methods Enzymol*, **470**: 145–79.

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

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

Bates, D. and Maechler, M. (2016) *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version 1.2-7.1.

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

Becker, K.F., Atkinson, M.J., Reich, U., Becker, I., Nekarda, H., Siewert, J.R., and Hfler, H. (1994) E-cadherin gene mutations provide clues to diffuse type gastric carcinomas. *Cancer Research*, **54**(14): 3845–3852.

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

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

Berx, G., Cleton-Jansen, A.M., Nollet, F., de Leeuw, W.J., van de Vijver, M., Cornelisse, C., and van Roy, F. (1995) E-cadherin is a tumour/invasion suppressor gene mutated in human lobular breast cancers. *EMBO J*, **14**(24): 6107–15.

Berx, G., Cleton-Jansen, A.M., Strumane, K., de Leeuw, W.J., Nollet, F., van Roy, F., and Cornelisse, C. (1996) E-cadherin is inactivated in a majority of invasive human

lobular breast cancers by truncation mutations throughout its extracellular domain. *Oncogene*, **13**(9): 1919–25.

Berx, G. and van Roy, F. (2009) Involvement of members of the cadherin superfamily in cancer. *Cold Spring Harb Perspect Biol*, **1**: a003129.

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

Blake, J.A., Christie, K.R., Dolan, M.E., Drabkin, H.J., Hill, D.P., Ni, L., Sitnikov, D., Burgess, S., Buza, T., Gresham, C., *et al.* (2015) Gene Ontology Consortium: going forward. *Nucleic Acids Res*, **43**(Database issue): D1049–1056.

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

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

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

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

Bozovic-Spasojevic, I., Azambuja, E., McCaskill-Stevens, W., Dinh, P., and Cardoso, F. (2012) Chemoprevention for breast cancer. *Cancer treatment reviews*, **38**(5): 329–339.

Breiman, L. (2001) Random forests. *Machine Learning*, **45**(1): 5–32.

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

Brouxhon, S.M., Kyrkanides, S., Teng, X., Athar, M., Ghazizadeh, S., Simon, M., O'Banion, M.K., and Ma, L. (2014) Soluble E-cadherin: a critical oncogene modulating receptor tyrosine kinases, MAPK and PI3K/Akt/mTOR signaling. *Oncogene*, **33**(2): 225–235.

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

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

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

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

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

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

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

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

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

Christofori, G. and Semb, H. (1999) The role of the cell-adhesion molecule E-cadherin as a tumour-suppressor gene. *Trends in Biochemical Sciences*, **24**(2): 73 – 76.

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

Clark, M.J. (2004) Endogenous Regulator of G Protein Signaling Proteins Suppress G o-Dependent -Opioid Agonist-Mediated Adenylyl Cyclase Supersensitization. *Journal of Pharmacology and Experimental Therapeutics*, **310**(1): 215–222.

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

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

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

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

Corcoran, R.B., Ebi, H., Turke, A.B., Coffee, E.M., Nishino, M., Cogdill, A.P., Brown, R.D., Della Pelle, P., Dias-Santagata, D., Hung, K.E., *et al.* (2012) Egfr-mediated 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.

Courtney, K.D., Corcoran, R.B., and Engelman, J.A. (2010) The PI3K pathway as drug target in human cancer. *J Clin Oncol*, **28**(6): 1075–1083.

Creighton, C.J., Morgan, M., Gunaratne, P.H., Wheeler, D.A., Gibbs, R.A., Robertson, A., Chu, A., 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.

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

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

De Leeuw, W.J., Berx, G., Vos, C.B., Peterse, J.L., Van de Vijver, M.J., Litvinov, S., Van Roy, F., Cornelisse, C.J., and Cleton-Jansen, A.M. (1997) Simultaneous loss of E-cadherin and catenins in invasive lobular breast cancer and lobular carcinoma in situ. *J Pathol*, **183**(4): 404–11.

De Santis, G., Miotti, S., Mazzi, M., Canevari, S., and Tomassetti, A. (2009) E-cadherin directly contributes to PI3K/AKT activation by engaging the PI3K-p85 regulatory subunit to adherens junctions of ovarian carcinoma cells. *Oncogene*, **28**(9): 1206–1217.

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

Deshpande, R., Asiedu, M.K., Klebig, M., Sutor, S., Kuzmin, E., Nelson, J., Piotrowski, J., Shin, S.H., Yoshida, M., Costanzo, M., *et al.* (2013) A comparative genomic approach for identifying synthetic lethal interactions in human cancer. *Cancer Res*, **73**(20): 6128–36.

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

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

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

Dixon, S.J., Andrews, B.J., and Boone, C. (2009) Exploring the conservation of synthetic lethal genetic interaction networks. *Commun Integr Biol*, **2**(2): 78–81.

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

Dong, L.L., Liu, L., Ma, C.H., Li, J.S., Du, C., Xu, S., Han, L.H., Li, L., and Wang, X.W. (2012) E-cadherin promotes proliferation of human ovarian cancer cells in vitro via activating MEK/ERK pathway. *Acta Pharmacol Sin*, **33**(6): 817–822.

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

Dorsam, R.T. and Gutkind, J.S. (2007) G-protein-coupled receptors and cancer. *Nat Rev Cancer*, **7**(2): 79–94.

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

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

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

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

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

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

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F. (2015) Cancer incidence and mortality worldwide:

sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*, **136**(5): E359–386.

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

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

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

Forbes, S.A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S., *et al.* (2015) COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res*, **43**(Database issue): D805–811.

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

Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004) A census of human cancer genes. *Nat Rev Cancer*, **4**(3): 177–183.

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

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

Gatza, M.L., Kung, H.N., Blackwell, K.L., Dewhirst, M.W., Marks, J.R., and Chi, J.T. (2011) Analysis of tumor environmental response and oncogenic pathway activation identifies distinct basal and luminal features in HER2-related breast tumor subtypes. *Breast Cancer Res*, **13**(3): R62.

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

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

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

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

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

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

Grady, W.M., Willis, J., Guilford, P.J., Dunbier, A.K., Toro, T.T., Lynch, H., Wiesner, G., Ferguson, K., Eng, C., Park, J.G., *et al.* (2000) Methylation of the CDH1 promoter as the second genetic hit in hereditary diffuse gastric cancer. *Nat Genet*, **26**(1): 16–17.

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

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

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

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

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

Guilford, P.J., Hopkins, J.B., Grady, W.M., Markowitz, S.D., Willis, J., Lynch, H., Rajput, A., Wiesner, G.L., Lindor, N.M., Burgart, L.J., *et al.* (1999) E-cadherin germline mutations define an inherited cancer syndrome dominated by diffuse gastric cancer. *Hum Mutat*, **14**(3): 249–55.

Guo, J., Liu, H., and Zheng, J. (2016) SynLethDB: synthetic lethality database toward discovery of selective and sensitive anticancer drug targets. *Nucleic Acids Res*, **44**(D1): D1011–1017.

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

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I.H. (2009) The weka data mining software: an update. *SIGKDD Explor Newsl*, **11**(1): 10–18.

Hammerman, P.S., Lawrence, M.S., Voet, D., Jing, R., Cibulskis, K., Sivachenko, A., Stojanov, P., McKenna, A., Lander, E.S., Gabriel, S., *et al.* (2012) Comprehensive genomic characterization of squamous cell lung cancers. *Nature*, **489**(7417): 519–525.

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

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

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

Hansford, S., Kaurah, P., Li-Chang, H., Woo, M., Senz, J., Pinheiro, H., Schrader, K.A., Schaeffer, D.F., Shumansky, K., Zogopoulos, G., *et al.* (2015) Hereditary Diffuse Gastric Cancer Syndrome: CDH1 Mutations and Beyond. *JAMA Oncol*, **1**(1): 23–32.

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

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

Hell, P. (1976) Graphs with given neighbourhoods i. problémes combinatorics at theorie des graphes. *Proc Coil Int CNRS, Orsay*, **260**: 219–223.

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

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

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

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

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

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

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

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

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

Jeanes, A., Gottardi, C.J., and Yap, A.S. (2008) Cadherins and cancer: how does cadherin dysfunction promote tumor progression? *Oncogene*, **27**(55): 6920–6929.

Jerby-Arnon, L., Pfetzer, N., Waldman, Y., McGarry, L., James, D., Shanks, E., Seashore-Ludlow, B., Weinstock, A., Geiger, T., Clemons, P., *et al.* (2014) Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. *Cell*, **158**(5): 1199–1209.

Joachims, T. (1999) Making large-scale support vector machine learning practical. In S. Bernhard, lkopf, J.C.B. Christopher, and J.S. Alexander (editors), *Advances in kernel methods*, 169–184. MIT Press.

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

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

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

Kamada, T. and Kawai, S. (1989) An algorithm for drawing general undirected graphs. *Information Processing Letters*, **31**(1): 7–15.

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

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

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

Kelly, S.T., Single, A.B., Telford, B.J., Beetham, H.G., Godwin, T.D., Chen, A., Black, M.A., and Guilford, P.J. (unpublished) Towards HDGC chemoprevention:

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

Kim, N.G., Koh, E., Chen, X., and Gumbiner, B.M. (2011) E-cadherin mediates contact inhibition of proliferation through Hippo signaling-pathway components. *Proc Natl Acad Sci USA*, **108**(29): 11930–11935.

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

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

Kroepil, F., Fluegen, G., Totikov, Z., Baldus, S.E., Vay, C., Schauer, M., Topp, S.A., Esch, J.S., Knoefel, W.T., and Stoecklein, N.H. (2012) Down-regulation of CDH1 is associated with expression of SNAI1 in colorectal adenomas. *PLoS ONE*, **7**(9): e46665.

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

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

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

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

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

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

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

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

Lee, A.Y., Perreault, R., Harel, S., Boulier, E.L., Suderman, M., Hallett, M., and Jenna, S. (2010a) Searching for signaling balance through the identification of genetic interactors of the rab guanine-nucleotide dissociation inhibitor gdi-1. *PLoS ONE*, **5**(5): e10624.

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

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

Lee, M.J., Ye, A.S., Gardino, A.K., Heijink, A.M., Sorger, P.K., MacBeath, G., and Yaffe, M.B. (2012) Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell*, **149**(4): 780–94.

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

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

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

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

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

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

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

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

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

Machado, J., Olivera, C., Carvalh, R., Soares, P., Berx, G., Caldas, C., Sercuca, R., Carneiro, F., and Sorbrinho-Simoes, M. (2001) E-cadherin gene (*CDH1*) promoter methylation as the second hit in sporadic diffuse gastric carcinoma. *Oncogene*, **20**: 1525–1528.

Masciari, S., Larsson, N., Senz, J., Boyd, N., Kaurah, P., Kandel, M.J., Harris, L.N., Pinheiro, H.C., Troussard, A., Miron, P., *et al.* (2007) Germline E-cadherin mutations in familial lobular breast cancer. *J Med Genet*, **44**(11): 726–31.

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

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

McLendon, R., Friedman, A., Bigner, D., Van Meir, E.G., Brat, D.J., 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.

Novomestky, F. (2012) *matrixcalc: Collection of functions for matrix calculations*. R package version 1.0-3.

Oliveira, C., Senz, J., Kaurah, P., Pinheiro, H., Sanges, R., Haegert, A., Corso, G., Schouten, J., Fitzgerald, R., Vogelsang, H., *et al.* (2009) Germline *CDH1* deletions in hereditary diffuse gastric cancer families. *Human Molecular Genetics*, **18**(9): 1545–1555.

Oliveira, C., Seruca, R., Hoogerbrugge, N., Ligtenberg, M., and Carneiro, F. (2013) Clinical utility gene card for: Hereditary diffuse gastric cancer (HDGC). *Eur J Hum Genet*, **21**(8).

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

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

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

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

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

Polyak, K. and Weinberg, R.A. (2009) Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nat Rev Cancer*, **9**(4): 265–73.

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.

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.

Scheuer, L., Kauff, N., Robson, M., Kelly, B., Barakat, R., Satagopan, J., Ellis, N., Hensley, M., Boyd, J., Borgen, P., *et al.* (2002) Outcome of preventive surgery and screening for breast and ovarian cancer in BRCA mutation carriers. *J Clin Oncol*, **20**(5): 1260–1268.

Semb, H. and Christofori, G. (1998) The tumor-suppressor function of E-cadherin. *Am J Hum Genet*, **63**(6): 1588–93.

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

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

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

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

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

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

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

Telford, B.J., Chen, A., Beetham, H., Frick, J., Brew, T.P., Gould, C.M., Single, A., Godwin, T., Simpson, K.J., and Guilford, P. (2015) Synthetic lethal screens identify

vulnerabilities in gpcr signalling and cytoskeletal organization in E-cadherin-deficient cells. *Mol Cancer Ther*, **14**(5): 1213–1223.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Tunggal, J.A., Helfrich, I., Schmitz, A., Schwarz, H., Gunzel, D., Fromm, M., Kemler, R., Krieg, T., and Niessen, C.M. (2005) E-cadherin is essential for in vivo epidermal barrier function by regulating tight junctions. *EMBO J*, **24**(6): 1146–1156.

Tutt, A., Robson, M., Garber, J.E., Domchek, S.M., Audeh, M.W., Weitzel, J.N., Friedlander, M., Arun, B., Loman, N., Schmutzler, R.K., *et al.* (2010) Oral poly(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 der Post, R.S., Vogelaar, I.P., Carneiro, F., Guilford, P., Huntsman, D., Hoogerbrugge, N., Caldas, C., Schreiber, K.E., Hardwick, R.H., Ausems, M.G., *et al.* (2015) Hereditary diffuse gastric cancer: updated clinical guidelines with an emphasis on germline CDH1 mutation carriers. *J Med Genet*, **52**(6): 361–374.

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

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

Vapnik, V.N. (1995) *The nature of statistical learning theory*. Springer-Verlag New York, Inc.

Vizeacoumar, F.J., Arnold, R., Vizeacoumar, F.S., Chandrashekhar, M., Buzina, A., Young, J.T., Kwan, J.H., Sayad, A., Mero, P., Lawo, S., *et al.* (2013) A negative genetic interaction map in isogenic cancer cell lines reveals cancer cell vulnerabilities. *Mol Syst Biol*, **9**: 696.

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

Vos, C.B., Cleton-Jansen, A.M., Berx, G., de Leeuw, W.J., ter Haar, N.T., van Roy, F., Cornelisse, C.J., Peterse, J.L., and van de Vijver, M.J. (1997) E-cadherin inactivation in lobular carcinoma in situ of the breast: an early event in tumorigenesis. *Br J Cancer*, **76**(9): 1131–3.

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

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

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

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

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

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

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

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

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

Wickham, H. and Chang, W. (2016) *devtools: Tools to Make Developing R Packages Easier*. R package version 1.12.0.

Wickham, H., Danenberg, P., and Eugster, M. (2017) *roxygen2: In-Line Documentation for R*. R package version 6.0.1.

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

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

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

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

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

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