Contents

List of Figures

List of Tables

Glossary

allele	A gene variant with a specific sequence and phenotype.
bioinformatics	Statistical or computational approaches to biological data or research tools.
cancer	A class of diseases, formally “malignant neoplasm”, of abnormal cellular growth and spread to other organs.
cancer gene	A gene which is involved in the malignancy of some cancers, encompassing oncogenes and tumour suppressors, which have molecular aberrations in cancer or variants which predispose individuals to cancer.
chemoprevention	The use of drugs to prevent early-stage cancers, generally applied to high-risk mutation carriers.
chemotherapy	The use of cytotoxic drugs to treat cancers, in combinations, generally applied to advanced stage cancers.
compound screen	A high-throughput screen performed using a library of chemical compounds.
computational biology	Applying computational or mathematical modelling to understanding biological systems and relationships.
driver mutation	A mutation which promotes cancer growth.
E-cadherin	Epithelial cadherin (calcium-dependent adhesion), a cell-adhesion protein encoded by *CDH1*.
edge or link	A relationship connecting a pair of elements of a graph structure or network, may be weighted or directional.

epistasis (biological)	The effects of a gene modifying or masking the phenotype of another gene.
epistasis (statistical)	A divergence of the observed double mutant phenotype from that expected based on the respective phenotypes of single mutant (Fisher, 1919).
essential	A gene which is required to be functional or expressed for a cell or organism to be viable, grow or develop.
familial	A trait recurrently occurring in families, not necessarily with a genetic cause.
functional redundancy	Genes which perform a common function, also known as genetic redundancy.
gene expression	A measure of the relative expression of each gene from the mRNA extracted from (pooled) cells.
genetic robustness	A system of biological pathways which (has evolved to) continue to function as a whole under various conditions, including the inactivation of various individual genes.
genome	All of the DNA sequence in the genome.
genomic	The use of data from all genes in the genome.
germline mutation	A mutation that occurred in germline cells and is passed between generation.
graph or network	A mathematical structure modelling or depicting the relationships between elements.
hereditary	A trait or disease which has a genetic cause and is inherited from family members.
high-throughput screen	An experimental procedure to perform a large scale series of chemical, genetic, or pharmacological tests.
hub	A central or highly connected component of a network.
induced essentiality	A gene becoming essential to viability under certain conditions, including inactivation of a synthetic lethal partner.

intrinsic subtype	Distinguishing cancer by molecular and genetic features.
metagene	A consistent signal of expression for a collection of genes such as a biological pathway, derived from singular value decomposition.
microarray	A high-throughput technique to measure presence or abundance of nucleic acid sequences from binding to probes.
molecular profile	A combination of genetic and biochemical measures which identifies characteristic traits of a tumour.
molecular subtype	A classification of cancers based on an identification using molecular properties.
mutant	A variant or dysfunctional phenotype arising from a mutation in a gene.
mutation	A change in DNA sequence that disrupts gene function.
non-oncogene addiction	The dependence of a cancer cell on functioning non-mutant genes.
'omics	A combination of approaches to generating biological data with high-throughput procedures such as genomics, proteomics or metabolomics.
oncogene	A gene that potentially causes cancer, typically by over-expression or mutant gene variants.
oncogene addiction	The dependence of a cancer cell on a specific oncogenic pathway.
pan cancer	A focus on the molecular and genetic features across cancers in different tissues.
passenger mutation	A mutation that occurs in cancers but does not affect the growth of cancers.
pathway	A series of biomolecules that produces a particular product or biological function.
pleiotropy	When a gene has multiple biological functions.
precision medicine	The application of prevention and treatment measures to target diseases by molecular and genetic features.

recurrent mutation	The repeated occurrence of mutations in a particular gene across cancers.
RNAi screen	A high-throughput screen performed using a RNA interference (RNAi).
RNA-Seq	The generation of transcriptome data from sequencing RNA.
scale-free	A property of a network which has a power law vertex degree distribution, that is several highly connected hub genes and many with very few connections.
shortest path	A path with the fewest possible edges which connects two particular vertices.
small world	A property of a network which is highly connected and has a low characteristic path length, derived from the mean shortest path length across all pairs of nodes.
somatic mutation	A mutation that occurs in somatic cells, during a patient's lifespan.
sporadic cancer	Cancers which do occur in patients with a family history or carry a high-risk genetic variant.
synthetic dosage lethal	A synthetic genetic interaction (SGI) analogous to synthetic lethality where where one gene is inactivated and the other overexpressed.
synthetic lethal	Genetic interactions where inactivation of multiple genes is inviable (or deleterious) which are viable if inactivated separately.
synthetic lethal screen	A high-throughput screen performed on isogenic cell lines to detect genes for which inhibition specifically deleterious to the null mutant genotype.
synthetic rescue	A synthetic genetic interaction when the combined mutations restores the wild-type the phenotype of one of the mutations.
synthetic sick	Genetic interactions where inactivation of multiple genes is deleterious which are viable if inactivated separately.

synthetic suppression	A synthetic genetic interaction when the combined mutations (partially) suppresses the mutant phenotype of one of the mutations.
targeted therapy	Cancer treatment that specifically acts against a molecular target, in contrast to standard chemotherapy.
transcriptome	All of the genes expressed in the genome.
treatment	Medical procedures for a disease to improve patient outcomes.
tumour	An abnormal lump of tissue or growth of cells, may be cancerous.
tumour suppressor	A gene potentially causes cancer, typically by disruption of functions which protect the cell from cancer.
vertex degree	A network metric of connectivity of vertices which uses the number of edges connected to each vertex or node.
vertex or node	An element of a graph structure or network.
wild-type	A natural phenotype of a trait or the normally functional allele which encodes it.

Acronyms

ADP	Adenosine Diphosphate.
ANOVA	Analysis of Variance.
AUROC	Area Under the Receiver Operating Characteristic (curve).
BiSEp	Bimodal Subsetting Expression.
CCLE	Cancer Cell Line Encyclopaedia.
cDNA	Complementary DNA (from mRNA).
CGP	Cancer Genome Project.
CNV	Copy Number Variation.
COSMIC	Catalogue Of Somatic Mutations In Cancer.
CpG	5′–C–phosphate–G–3′.
DAISY	Data Mining Synthetic Lethal Identification Pipeline.
DNA	Deoxyribonucleic Acid.
EMT	Epithelial-Mesenchymal Transition.
FDR	False Discovery Rate.
GO	Gene Ontology.
GPCR	G Crotein Coupled Receptor.
HDAC	Histone Deacetylase.
HDGC	Hereditary Diffuse Gastric Cancer.
HLRCC	Hereditary Leiomyomatosis and Renal Cell Carcinoma.
JAK	Janus Kinase.
microRNA	Micro RNA.
mRNA	Messenger RNA.
MSI	Microsatellite Instability.

mtSLIPT	Synthetic Lethal Interaction Prediction Tool (against mutation).
NGS	Next-Generation Sequencing.
PARP	Poly-ADP-Ribose Polymerase.
PCR	Polymerase Chain Reaction.
PI3K	Phosphoinositide 3-kinase.
PPI	Protein-Protein Interaction.
RNA	Ribonucleic Acid.
RNAi	RNA Interference.
ROC	Reciever Operating Characteristic (curve).
RSEM	RNA-Seq by Expectation Maximization (normalisation.
SGA	Synthetic Gene Array (technique).
SGI	Synthetic Genetic Interaction.
shRNA	Short Hairpin RNA.
siRNA	Short Interfering RNA.
SL	Synthetic Lethal.
SLIPT	Synthetic Lethal Interaction Prediction Tool.
SNP	Single Nucleotide Polymorphism.
SR	Synthetic Rescue (or viability).
SS	Synthetic Suppression.
SSL	Synthetic Sick.
TCGA	The Cancer Genome Atlas (genomics project).
WNT	Wingless-Related Integration Site.

Chapter 1

Introduction and Literature Review

This thesis presents research into genetic interactions using genomics data and bioinformatics approaches. Chapter 1 introduces 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 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 are expected to significantly impact on the clinical treatment of cancers, along with wider applications of genetics (Goodwin *et al.*, 2016; Roychowdhury and Chinnaiyan, 2016). These technologies enable focused genetic investigations on candidate genes selected from bioinformatic analyses of genomic 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 (Roychowdhury and Chinnaiyan, 2016; Waldron, 2016). Bioinformatics analysis for interpretation of genomic data is one of the main approaches to address this disparity (Goodwin *et al.*, 2016). Here, I outline the cancer genomics projects and findings which have led to the availability of genomics data used in this thesis, and recent findings in cancer research which demonstrate potential applications of using this data.

1.1.1 Cancer is a Global Health Issue

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). New Zealand has among the highest incidence (age-standardised per capita) of cancer in the world (Ferlay *et al.*, 2015).

While environmental factors often play a role, genetics is an important contributor to cancer risk. Most cancers occur more frequently with age and family history. Cancers arise from dysregulated cellular growth or differentiation. 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 and are sometimes offered preventative surgery (Guilford *et al.*, 2010; Scheuer *et al.*, 2002).

Chemotherapy is a treatment for many advanced stage cancers, 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 can severely 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).

1.1.1.1 The Genetics and Molecular Biology of Cancers

Cancers involve dysregulation of genes, including somatic and hereditary mutations, which may predispose individuals to high-risk cancers and familial cancer syndromes (American Cancer Society, 2017; Guilford *et al.*, 1998; Stratton *et al.*, 2009; NCI, 2015; Vogelstein *et al.*, 2013). The occurence of somatic mutation mutations increases 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) proposed the "hallmarks of cancer", molecular and cellular traits shared across cancers. These form the basis of a rational approach to categorising 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 and 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.

1.1.2 The Genomics Revolution in Cancer Research

Genomic technologies have transformed genetics research, including the study of health and disease (Goodwin *et al.*, 2016; Lander, 2011). Genomics enables systematic, unbiased studies across all of the genes in the genomes. 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). Genomes sequencing technologies continue to improve and become feasible in a wider range of applications.

Genomics has been used in many investigations (Goodwin *et al.*, 2016) but relatively few of the potential applications in healthcare have been realised yet (Roychowdhury and Chinnaiyan, 2016; Tran *et al.*, 2012). Cancer genomics, in particular, could have numerous benefits across diagnostics, prognosis, management, and treatment (Roychowdhury and Chinnaiyan, 2016). While direct impact of genomics 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

These investigations have been enabled by recent developments in genomics technologies, including microarrays and more recently “Next-Generation Sequencing” (NGS), which can both be used to generate high-throughput expression data. Microarray are a high-throughput molecular technique, reducing the cost, time, and labour required to study genes at the “genome” scale (Schena, 1996). Microarray can detect genotype or expression across many genes, making it feasible to perform on a statistically informative number of samples. Microarray 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 dinucleotide (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.

The introduction of massively parallel sequencing technologies has further expanded high-throughput molecular studies and the availability of genomics data. NGS enables rapid *de novo* genomes and transcriptome sequencing, in addition to gene expression studies (Goodwin *et al.*, 2016). However, the cost of sequencing for gene expression 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 (Tarazona *et al.*, 2011).

NGS is highly adaptable to different applications, including DNA sequencing (obtaining the base sequence for the exome or whole genome) or RNA-Seq (Goodwin *et al.*, 2016; Tran *et al.*, 2012; Waldron, 2016). RNA-Seq of the transcriptome is a common adaptation where 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. Subsets of the nucleic acid may be extracted for sequencing such as the coding regions of DNA (for the “exome”), mRNA, or micro RNA (microRNA). These “’omics” technologies (Roychowdhury and Chinnaiyan, 2016; Waldron, 2016) are applicable across a wide range of biomolecules to generate “molecular profiles” of a cell or sample (Perou *et al.*, 2000).

NGS technologies continue to be refined (Goodwin *et al.*, 2016) 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 yet improve (Goodwin *et al.*, 2016) with developments in speed and accuracy (such as semi-conductor platforms) or long reads, single molecule 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 anlaysis developed for RNA-Seq data could be applied to future genomics technologies.

1.1.2.2 Bioinformatics and Genomic Data

Genomic technologies have generated data at a scale which requires computational, mathematical, and statistical expertise to handle this data effectively (Markowetz, 2017; Tran *et al.*, 2012), 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.

Bioinformatics is often confused with the broader field "computational biology" (Markowetz, 2017), 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 focusing on 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

The Cancer Genome Project (CGP) was among the first genomics investigations into cancer (Dickson, 1999), using the human genomes 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 Genomes Project was to discover "cancer genes", which are frequently mutated in cancers by comparing cancer and normal tissue samples. These include both "oncogenes" (which drive cancer growth) and "tumour suppressors" (which protect against cancers) that 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 utility in 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. TCGA has 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: genotype, somatic mutations, gene expression, microRNA, DNA copy number, DNA methylation, and protein 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 (Koboldt *et al.*, 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 the *TP53*, *PIK3CA*, and *GATA3* genes. 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 the *CDH1*, *PTEN*, *TBX2*, and *FOXA1* genes.

TCGA stomach cancer analysis of 295 samples (Bass *et al.*, 2014) identified molecular subtypes of stomach cancers characterised by: the Epstein-Barr virus, microsatellite instability (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 has identified various genes as recurrent, driver mutations across cancer types which are likely to have a role in driving the development 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), TCGA has 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; Koboldt *et al.*, 2012; Linehan *et al.*, 2016; Muzny *et al.*, 2012; Weinstein *et al.*, 2014).

The "Pan Cancer" TCGA project (Hoadley *et al.*, 2014; Weinstein *et al.*, 2013) analysed 3527 samples across 12 tissue types. This project 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 MSI across cancer types. The Pan Cancer project 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 (Hoadley *et al.*, 2014) . The 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 substantial potential for impacts in cancer medicine: from diagnosis to treatment (Roychowdhury and Chinnaiyan, 2016; Tran *et al.*, 2012). Beyond direct use of genomes or RNA-Seq in clinical laboratories, genomic studies also generate biomarkers and inform development of treatments. These are likely to have a more immediate patient benefit considering the cost of routine genomes 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, with recurrent mutations that are typically concentrated to particular gene regions ("hotspots"). 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, DNA copy number variants, differential gene expression, or differential 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).

Distinguishing important "driver" mutations in cancer genes from "passenger mutation" mutations is challenging due to patient variation, tumour heterogeneity, and genomic instability producing many variant gene sequences (Stratton *et al.*, 2009; Tran *et al.*, 2012). 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 genomes maintenance (Vogelstein *et al.*, 2013). There remains a 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 population. Cancers could eventually be treated by their genomic features (Roychowdhury and Chinnaiyan, 2016), particularly grouping patients by the mutation, expression, or DNA methylation profiles of their cancers, which is already done in part (Parker *et al.*, 2009). Identifying actionable molecular targets is a key aspect of "precision medicine", the rationale to target molecular subtypes with separate treatment strategies (Glaire *et al.*, 2017). 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 "Pan Cancer" TCGA 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. As these molecular subtypes and genetic aberrations specific to cancers have been identified, there is an increasingly clear need for further development of treatments that target them.

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 tumour The 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.

1.1.4.4 Targeted Therapeutics and Pharmacogenomics

Targeted therapies with specificity against a molecular target are examples of precision cancer medicine. Molecular targets can be tested in laboratory conditions with RNA interference (RNAi) or pharmacological agents (Fece de la Cruz *et al.*, 2015). 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 can be directly inhibited, however, oncogenes with high homology to other genes or tumour suppressor genes are not amenable to direct targeting (Kaelin, Jr, 2009). Targeted anticancer therapeutics can exploit complex interactions to distinguish normal and cancerous cells which may benefit from studies of gene regulation or interaction networks (Hopkins, 2008). Targeted therapeutics have already been successfully applied as monoclonal antibodies against oncogenes, such as HER2 in breast cancer (Miles, 2001).

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 their random networks followed 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 Erdő 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 was 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 had a higher degree. These networks successfully captured 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).

High-throughput technologies such as siRNA screens, two-hybrid screens, microarrays and massively parallel sequencing have generated genomes-scale data and enabled analysis of biological networks (Barabási and Oltvai, 2004; Boone *et al.*, 2007; Goodwin *et al.*, 2016). 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. Many types of molecular networks can be constructed, depending on the biological application (). Synthetic 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 (Fece de la Cruz *et al.*, 2015). 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 genomes-scale.

1.2 Synthetic Lethal Cancer Medicine

Synthetic lethality has vast potential to improve cancer medicine by expanding application of targeted therapeutic 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 mutant and the wild-type (Fisher, 1919). Fisher's definition lends itself to quantitative traits and more broadly encompasses synthetic genetic interactions. 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 mutant. The double mutant does not necessarily have either of the single mutant phenotypes (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 SGI 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 SGI with less deviation from the wild-type than expected. For viability, "suppression" (SS) and "rescue" (SR) are positive SGIs giving at least partial restoration of wild-type growth from single mutant 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).

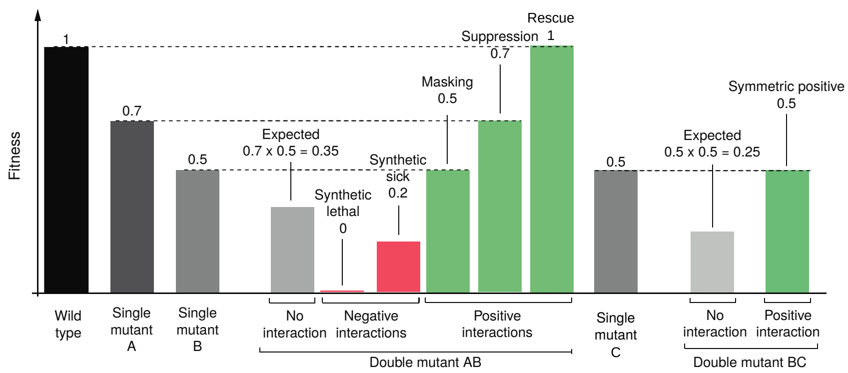

Figure 1.1: **Synthetic genetic interactions.** Impact of various negative and positive SGIs: negative interactions involve deleterious (sick) or inviable (lethal) phenotypes whereas positive interactions involve restoring viability by masking or suppressing the other mutation or complete rescue of thewild-type 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 (Boone *et al.*, 2007). 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 (Ashworth *et al.*, 2011; Kaelin, Jr, 2005).

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 genomes 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 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 interchangable, 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 (Lord *et al.*, 2015). 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, many existing computational methods are not suitable for queries of genomic data for interacting partners of a particular gene, as (1) they have been applied pairwise across the genomes, (2) they do not have software released to apply the methodology, or (3) they 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 genomes 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, four-fold more interactions were detected with synthetic gene array (SGA) mating screens than protein-protein interactions detected with yeast-2-hybrid (Tong *et al.*, 2004). The SGI network was scale-free and had a low average shortest path length, 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 that synthetic lethality may be comparably easier to detect than other SGIs.

Essential pathways are highly buffered, with five-fold 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, 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 therapeutic. 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 these genes are feasible anti-cancer drug targets.

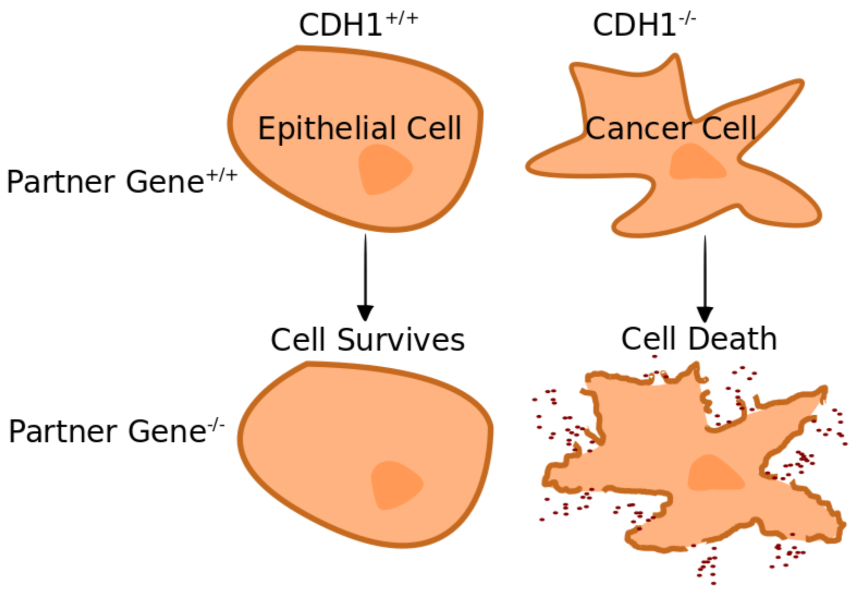

Figure 1.2: **Synthetic lethality in cancer.** Rationale of exploiting synthetic lethality for specificity against a tumour suppressor gene (e.g., *CDH1*) while other cells are spared under the inhibition of a partner gene.

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 therapeutic, including inactivated tumour suppressor genes. synthetic lethality may also enable distinguishing highly homologous oncogenes by functional differences by targeting their synthetic lethal partners.

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 and these discovery of these interactions has lead to translation to the clinic. These genetic interactions enable specific targeting of mutations in *BRCA1* or *BRCA2* tumour suppressor genes with Poly-ADP-ribose polymerase (PARP) inhibitors by inducing synthetic lethality in breast cancer (Farmer *et al.*, 2005). PARP inhibitors were one of the first targeted therapeutic against a tumour suppressor mutation to exhibit success in clinical trials.

BRCA1/*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. The combined loss of DNA repairs pathways gives a plausible mechanism for an effective anti-cancer treatment.

Thus PARP inhibitors could be applied with clinical use against *BRCA1* or *BRCA2* mutations in both hereditary and sporadic cancers (Ashworth, 2008; Kaelin, Jr, 2005). PARP inhibition has been found to be effective in ovarian cancer patients carrying

BRCA1 or *BRCA2* mutations and some patient without these mutations, suggesting synthetic lethality between PARP and other DNA repair pathways (Ström and Helleday, 2012). This supports the potential for PARP inhibition as a chemopreventative 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 chemopreventative 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 mutation 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 for studying 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 genomes 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 and cannot account for diverse genetic backgrounds or tumour heterogeneity (Fece de la Cruz *et al.*, 2015). Genetic screens can thus identify targets to develop, or can repurpose targeted therapies for disease, but alone will not directly identify a lead compound to develop for the market or for clinical translation.

Chemical screens are immediately applicable to the clinic, as they are 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 still many challenges to translating candidates into the clinic. Identifying specific target genes may contribute to overcoming such challenges, which can be approached with genetic screens 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

Synthetic lethal screens have been conducted for cancer genes in a variety of cancers. These have found synthetic lethality of *PIM1* over-expression with *PLK1* inhibition in prostate cells (van der Meer *et al.*, 2014), *FH* null mutations (involved in Hereditary leiomyomatosis and renal cell carcinoma (HLRCC)) with inhibition of adenylate cyclases (Boettcher *et al.*, 2014), and *WEE1* inhibitor treatment with knockdown of checkpoint kinases, Fanconi anaemia, and homologous recombination in colorectal cells (Aarts *et al.*, 2015). These results include genes that have been found to be co-expressed in cancers, are consistent with those identified in the literature, and that were successfully validated with RNAi and drug experiments. These findings demonstrate that synthetic lethal screening can identify partner genes with clinical relevance as biomarkers, therapeutic targets, or conferring sensitivity to existing treatments. These are of particular importance for familial cancer syndromes (Boettcher *et al.*, 2014; Telford *et al.*, 2015).

Hereditary diffuse gastric cancer (HDGC) is a cancer syndrome involving predisposition to early-onset malignant stomach and breast cancers that has been attributed to inactivating mutations in E-cadherin, encoded by *CDH1* (as discussed in Section 1.3). Telford *et al.* (2015) performed an RNAi screen on MCF10A breast cells for synthetic lethality with *CDH1*. In conjunction with a drug compound screen, inhibitors of Janus kinase (JAK), histone deacetylase (HDAC), phosphoinositide 3-kinase (PI3K), aurora kinase, and tyrosine kinases were demonstrated to be synthetic lethal with *CDH1*. Therefore the synthetic lethal strategy shows potential to achieve clinical impact against HDGC by the identification of compounds suitable for use in chemoprevention.

The examples above show that high-throughput screens are an effective approach to discover synthetic lethality in cancer. Screens have the power to test mode of action of drugs, find unexpected synthetic lethal interactions between pathways, or identify effective treatment strategies without prior knowledge of a mechanism. However, synthetic lethal screens are costly, labour-intensive, error-prone, and biased towards genes with effective RNAi knockdown libraries. Off-target effects and inconsistent replication synthetic lethality across different cell lines, tissues, or laboratories, are also problematic. Therefore there is a need for replication, validation, and alternative approaches to identify synthetic lethal candidates. In addition, varied conditions across experimental screens and differences between RNAi and drug screens makes meta-analysis extremely challenging.

Genome-scale synthetic lethal experiments (across gene pairs) are not feasible, even in model organisms, and these studies typically focus on specific gene candidates or the partners of a gene of interest. Therefore a computational approach is more suitable for this task and may also augment existing experimental screens.

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, and has both 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 PPIs, an important predictor of genetic interactions, is difficult given that high-throughput two-hybrid screens occur out of cellular context for multicellular organisms (Brückner *et al.*, 2009).

There are existing computational methods for predicting synthetic lethal gene pairs in humans, with a specific emphasis on 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 or replicate for new genes, or are reliant on data types not available for a wider range of genes to test.

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)	

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

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 SGI in yeast (Costanzo *et al.*, 2011) were matched to human orthologues, with over 500 involving a cancer gene (Futreal *et al.*, 2004). Under strict criteria of one-to-one orthologues, large effect size and significant interaction in yeast data, 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 syste,m and are difficult to replicate across gene backgrounds. This database centralises a wealth of knowledge scattered in the literature including cancer relevant genes, including the previously mentioned interactions of *BRCA1* and *BRCA2* with *PARP1*, and *TP53* with *WEE1* and *PLK1*, although the computational methodology was not released and was limited to 647 human genes. Their 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).

Machine learning approaches have also been explored for synthetic lethal discovery

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

(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). They focused on synthetic lethal gene pairs relevant to developing selective drugs against human cancer, building upon their previous database (Li *et al.*, 2014). Their "metaSL" approach utilises genomic, proteomic and annotation data and had a high statistical performance in yeast data with an area under receiver operating characteristic (AUROC) of 0.871 (as described in Section 2.3.5.1). They predicted orthologous synthetic lethal partners in human data were not experimentally validated but some were relevant to cancer such as *EGFR* with *PRKCZ*.

Computational approaches scale-up across the genomes at lower cost than experimental screen (Wu *et al.*, 2014). Wu *et al.* (2014) provided their most supported interactions online but the method is not available for analysis of other 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*. Accordingly, there remains a need to enable biological researchers to query further genes and do so 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 cells are not an ideal human cancer model because they do not have tissue specificity, multicellular gene regulation, or orthologues to several known cancer genes such as p53 (Guaragnella *et al.*, 2014). Although these studies have tried to anticipate these issues with stringent criteria such as requiring one-to-one orthologues (Deshpande *et al.*, 2013; Heiskanen and Aittokallio, 2012; Kranthi *et al.*, 2013), there remains the possibility that changes in gene function may affect whether these are solely redundant such as if functions had co-evolved 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 derived protein interactions and cancer gene databases (Futreal *et al.*, 2004; Higgins *et al.*, 2007; Keshava Prasad *et al.*, 2009), consisting 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. This demonstrated that network structure is important in the biological functions of cancers and could be exploited for targeting *TP53* loss of function mutations.

However, the approach of Kranthi *et al.* (2013) 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 cancers but this genome-wide approach did not detect many other cancer genes due to the extent of multiple testing. This enrichment may be due to the known drastic effect of removing p53 itself from the network as a highly connected, master regulator, and cancer driving tumour suppressor gene. 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 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 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) 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.

These novel results contained many *TP53* and AKT synthetic lethal partners (Zhang *et al.*, 2015), genes known to be important in many cancers. However, these genes also have a severe impact on the signalling pathways in an essentiality analysis of single gene disruptions and large phenotypic changes in cancer (Zhang *et al.*, 2015). This approach is amenable to detect functionally related pathways and protein complexes across the molecular function, cellular component, and biological process annotations provided by Gene Ontology. 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, the widespread availability which allows testing ofa broader 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 many samples and cancer types. Wang and Simon (2013) found differentially expressed genes (by the t-test, adjusted by False discovery rate (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 lacking 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 the 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, their analyses were limited to kinase genes and focused on currently druggable targets, 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 lethal pairs with colorectal cancer microarrays from a Han Chinese population with a sample size of 70 tumours and 12 normal tissue samples. Simultaneously differential expressed "tumour dependent" gene pairs (which includes co-expression) between cancer and normal tissue were used to rank 663 candidate synthetic lethal interactions identified in cell line siRNA experiments. Of the top 20 gene pairs, 17 were tested for 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 as a valuable proof-of-concept for integration of *in silico* approaches to synthetic lethal discovery in cancer, demonstrating its 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 to find those applicable to colorectal cancers (including in a non-Caucasian population).

1.2.7.5 Data Mining and Machine Learning

Recognising the utility of synthetic lethality to drug inhibition and specificity of anti-cancer 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-fold 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 a 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 and for essential genes in mice which could be harnessed for predicting prognosis and drug response.

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 were stringently compared by adjusting each for multiple tests with Bonferroni correction and only taking candidates identified in all analyses. 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. While DAISY performed well statistically, co-expression and shRNA functional examination contributed less to this than the mutation and copy number analysis (AUROC 0.683 alone). However, this methodology was very stringent, missing potentially valuable synthetic lethal candidates. Additionally, the software for the procedure has not been publicly released for replication.

Although the DAISY procedure performed 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 has not yet been widespread adoption of this approach. Co-expression analysis may exclude some synthetic lethal interactions, where inverse correlation could occur (Lu *et al.*, 2015). 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) 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 genomes 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 genomes 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 genomes evolution and gene expression patterns to the experimental data (Laufer *et al.*, 2013; Vizeacoumar *et al.*, 2013). This more simple model performed well, with an AUROC of 0.751 (lower than DAISY), and did not rely on data from cell lines which may not represent patient disease. Lu *et al.* (2015) predicted 591,000 human synthetic lethal partners with a probability score threshold of 0.81, giving a precision of 67% and 14-fold 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 genomes, 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 was 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 Mutual Exclusivity and Bimodality

Wappett *et al.* (2016) demonstrated a multi-omic approach to identify 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 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 RNA-Seq by Expectation Maximization (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. Wappett *et al.* (2016) 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 showed that this is applicable in both genetically homogeneous cell lines and heterogeneous cell population from patient samples. These approach are limited however, to genes that exhibit bimodal expression patterns which do not commonly occur, particularly in normalised gene expression data. Other approaches may need to be considered for gene such as *CDH1* which were not identified by BiSEp.

Srihari *et al.* (2015) used a computational analysis to identify synthetic lethal candidate genes from mutually exclusive alterations in cancers. This analysis focused on synthetic lethality with "DNA damage response genes" in cancers, including *CDH1*, using TCGA expression and copy number data across several cancers. The 718 genes that were identified as frequently altered in cancer were enriched among essential genes in cell lines deficient in DNA damage response genes which demonstrates "induced essentiality" in a cancer model. These were tested by `underMutExSL`, a hypergeometric test, for synthetic lethality. Of the DNA damage response genes examined, *CDH1* exhibited the most mutually exclusive alterations with other frequently altered genes in breast cancers, including related genes in focal adhesion such as *PTK2*. These results indicate that *CDH1* may be particularly subject to "non-oncogene addiction" in breast cancers and that computational analysis is valuable to rationally identify these putative synthetic lethal genes. These results were limited to frequently altered genes in cancer and could be improved with an approach that expands to consider synthetic lethal genes that are not themselves altered in cancers or further investigation into synthetic lethal pathways.

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 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 well-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 mutation 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. Drugs against *CDH1*-deficient cancers would be applicable for chemoprevention of HDGC and treatment of sporadic cancers carrying somatic *CDH1* mutations.

1.3.1 The *CDH1* gene and its Biological Functions

CDH1 is implicated in hereditary and sporadic lobular breast cancers and diffuse gastric cancers (Berx *et al.*, 1996; Berx and van Roy, 2009; De Leeuw *et al.*, 1997; Guilford *et al.*, 1998; Masciari *et al.*, 2007; Semb and Christofori, 1998; Vos *et al.*, 1997). E-cadherin is normally expressed in epithelial tissues 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).

The primary function of *CDH1* is cell-cell adhesion at 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 actin and microtubule cytoskeletons. The cytoskeletal role of E-cadherin includes maintaining 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).

As a transmembrane signalling protein E-cadherin also interacts with the extracellular environment and other cells, most notably forming adherens junctions between cells (Chen *et al.*, 2014; Tunggal *et al.*, 2005) and regulating of epithelial tissues by intercellular communication (Jeanes *et al.*, 2008). The signals mediated by adherens junctions are also passed on to intracellular signalling pathways . One such example is the regulation of β-catenin which interacts with both the actin cytoskeleton and acts as a transcription factor via the Wingless-related integration site (WNT) pathway (Jeanes *et al.*, 2008). Similarly, the Hippo and phosphoinositide 3-kinase (PI3K)/AKT pathways are also mediated in part by E-cadherin (De Santis *et al.*, 2009; Kim *et al.*, 2011). E-cadherin shares several downstream pathways with cell-surfacce proteins, such as integrins, and thus indirectly interacts with them.

The key roles of E-cadherin in maintaining cellular structure and regulating growth are consistent with *CDH1* being a tumour suppressor gene. E-cadherin loss in breast cancers has been shown to cause increased proliferation, lymph node invasion, and metastasis with poor cell-cell contact (Berx and van Roy, 2009). Thus the *CDH1* gene is an established invasion suppressor, with a key role in the epithelial-mesenchymal transition (EMT), an established mechanism of cancer progression (Hanahan and Weinberg, 2011). EMT is important during development and wound healing but such changes in cellular differentiation also occur in cancers.

1.3.2 Hereditary Diffuse Gastric (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 of these cancers, including hereditary diffuse gastric cancer (HDGC) (Guilford *et al.*, 1998). Due to carrying a dysfunctional allele, these individuals are prone to carcinogenic lesions in the breast or stomach if the remaining functional allele is inactivated, occurring more frequently and at an earlier age than in individuals with two functional *CDH1* alleles. The loss of the second allele is most often through hypermethylation suppressing expression rather than mutation (Grady *et al.*, 2000; Graziano *et al.*, 2003; Machado *et al.*, 2001; Oliveira *et al.*, 2009), although loss of heterozygousity may also occur (Guilford *et al.*, 2010).

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 (Hansford *et al.*, 2015). HDGC represents less than 1% of gastric cancers (Ferlay *et al.*, 2015), however, it is a serious health issue for several hundred families globally. E-cadherin is also mutated in 13% of sporadic gastric cancers (Ferlay *et al.*, 2015).

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 at the age of 20 years. 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 germline *CDH1* mutations either have surgery which has a significant impact on quality of life and risk of complications or remain at risk of developing stomach cancer. There are similar concerns for female mutations carriers for the management of high-risk lobular breast cancer. An effective treatment against *CDH1* mutant cancers would potentially have significant therapeutic and preventative applications for many patients.

1.3.3 Cell Line Models of *CDH1* Null Mutations

Previous work published by members of our research group 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 (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 siRNA and a complementary drug screen using 4057 known drugs to identify synthetic lethal partners to E-cadherin (Telford *et al.*, 2015). One of the strongest candidate pathways identified by the siRNA screen was the G protein coupled receptor (GPCR) signalling cascades, which were highly enriched in a Gene Ontology (GO) analysis. This was supported by validation with Pertussis toxin, known to target $G_{\alpha i}$ signalling (Clark, 2004).

1.4 Summary and Research Direction of Thesis

Genomic technologies 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 by the identification of driver 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 biomarker identification and cancer treatment. Synthetic lethal interactions 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.

Discovery of synthetic lethal partners of *CDH1* would contribute to an understanding of the molecular mechanisms driving 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*.

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 novel precision medicine targeted to particular molecular aberrations.

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 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 the R code used for simulation, prediction, and analysis will enable adoption of the methodology by the research community and comparison to existing methods. This thesis aims to develop predictions for synthetic lethal partner genes, with a focus on the *CDH1*, and to compare these to the findings of Telford *et al.* (2015). In addition, this thesis will 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

This thesis aims to identify synthetic lethal gene pairs using public gene expression data. Accordingly, Chapter 3 describes the methods developed to do so, including a synthetic lethal detection methodology (SLIPT) and the release of R software packages. This chapter also serves to document the original simulation and network analysis procedures developed to support the use of SLIPT and perform analyses throughout this thesis.

This thesis also aims to demonstrate the SLIPT methodology for analysis of RNA-Seq gene expression data. Chapter 4 does 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; Koboldt *et al.*, 2012). Chapter 4 demonstrates 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 to include graph structures in Chapter 5, which aimed to assess the importance of synthetic lethal genes within pathway structures. Chapter 5 also uses pathway structure to identify directional relationships between SLIPT and siRNA synthetic lethal candidates and explore the disparity between them. The SLIPT methodology is supported by simulation-based investigations in Chapters 3 and 6, which evaluate the ability of SLIPT to detect known synthetic lethal genes in simulated data. Graph structures were also used in Chapter 6 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 containing complex correlation structures within biological pathways.

Bibliography

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.

Ashworth, A., Lord, C.J., and Reis-Filho, J.S. (2011) Genetic interactions in cancer progression and treatment. *Cell*, **145**(1): 30–38.

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.

Brückner, A., Polge, C., Lentze, N., Auerbach, D., and Schlattner, U. (2009) Yeast two-hybrid, a powerful tool for systems biology. *Int J Mol Sci*, **10**(6): 2763–2788.

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.

Cardiff, R.D., Couto, S., and Bolon, B. (2011) Three interrelated themes in current breast cancer research: gene addiction, phenotypic plasticity, and cancer stem cells. *Breast Cancer Res*, **13**(5): 216.

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.

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.

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.

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.

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.

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. *Publ Math Inst Hung Acad Sci*, **5**(1): 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.

Friedman, N., Linial, M., Nachman, I., and Pe'er, D. (2000) Using Bayesian networks to analyze expression data. *J Comput Biol*, **7**(3-4): 601–620.

Fromental-Ramain, C., Warot, X., Lakkaraju, S., Favier, B., Haack, H., Birling, C., Dierich, A., Doll e, P., and Chambon, P. (1996) Specific and redundant functions of the paralogous Hoxa-9 and Hoxd-9 genes in forelimb and axial skeleton patterning. *Development*, **122**(2): 461–472.

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.

Glaire, M.A., Brown, M., Church, D.N., and Tomlinson, I. (2017) Cancer predisposition syndromes: lessons for truly precision medicine. *J Pathol*, **241**(2): 226–235.

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

Goodwin, S., McPherson, J.D., and McCombie, W.R. (2016) Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*, **17**(6): 333–351.

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.

Guaragnella, N., Palermo, V., Galli, A., Moro, L., Mazzoni, C., and Giannattasio, S. (2014) The expanding role of yeast in cancer research and diagnosis: insights into the function of the oncosuppressors p53 and BRCA1/2. *FEMS Yeast Res*, **14**(1): 2–16.

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

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

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

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.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.

Higgins, M.E., Claremont, M., Major, J.E., Sander, C., and Lash, A.E. (2007) CancerGenes: a gene selection resource for cancer genome projects. *Nucleic Acids Res*, **35**(Database issue): D721–726.

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.

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.

Hutchison, C.A., Chuang, R.Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H., Gill, J., Kannan, K., Karas, B.J., Ma, L., *et al.* (2016) Design and synthesis of a minimal bacterial genome. *Science*, **351**(6280): aad6253.

Imoto, S., Higuchi, T., Goto, T., Tashiro, K., Kuhara, S., and Miyano, S. (2004) Combining microarrays and biological knowledge for estimating gene networks via bayesian networks. *J Bioinform Comput Biol*, **2**(1): 77–98.

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

Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N.J., Chung, S., Emili, A., Snyder, M., Greenblatt, J.F., and Gerstein, M. (2003) A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science*, **302**(5644): 449–453.

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.

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.

Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., *et al.* (2009) Human Protein Reference Database–2009 update. *Nucleic Acids Res*, **37**(Database issue): D767–772.

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.

Koboldt, D.C., Fulton, R.S., McLellan, M.D., Schmidt, H., Kalicki-Veizer, J., McMichael, J.F., Fulton, L.L., Dooling, D.J., Ding, L., Mardis, E.R., *et al.* (2012) Comprehensive molecular portraits of human breast tumours. *Nature*, **490**(7418): 61–70.

Kockel, L., Zeitlinger, J., Staszewski, L.M., Mlodzik, M., and Bohmann, D. (1997) Jun in drosophila development: redundant and nonredundant functions and regulation by two mapk signal transduction pathways. *Genes & Development*, **11**(13): 1748–1758.

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, B., Ruotti, V., Stewart, R.M., Thomson, J.A., and Dewey, C.N. (2010) RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics*, **26**(4): 493–500.

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.

Markowetz, F. (2017) All biology is computational biology. *PLoS Biol*, **15**(3): e2002050.

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.

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.

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

Nowak, M.A., Boerlijst, M.C., Cooke, J., and Smith, J.M. (1997) Evolution of genetic redundancy. *Nature*, **388**(6638): 167–171.

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.

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.

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

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.

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.

Roychowdhury, S. and Chinnaiyan, A.M. (2016) Translating cancer genomes and transcriptomes for precision oncology. *CA Cancer J Clin*, **66**(1): 75–88.

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

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

Schena, M. (1996) Genome analysis with gene expression microarrays. *Bioessays*, **18**(5): 427–431.

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.

Srihari, S., Singla, J., Wong, L., and Ragan, M.A. (2015) Inferring synthetic lethal interactions from mutual exclusivity of genetic events in cancer. *Biology Direct*, **10**(1): 57.

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.

Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A., and Conesa, A. (2011) Differential expression in RNA-seq: a matter of depth. *Genome Res*, **21**(12): 2213–2223.

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) (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.

Tran, B., Dancey, J.E., Kamel-Reid, S., McPherson, J.D., Bedard, P.L., Brown, A.M., Zhang, T., Shaw, P., Onetto, N., Stein, L., *et al.* (2012) Cancer genomics: technology, discovery, and translation. *J Clin Oncol*, **30**(6): 647–660.

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

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.

University of California, Santa Cruz (UCSC) (2012) Ucsc cancer browser. Accessed 29/03/2012.

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.

Waldron, D. (2016) Cancer genomics: A multi-layer omics approach to cancer. *Nat Rev Genet*, **17**(8): 436–437.

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.

Wojtukiewicz, M.Z., Hempel, D., Sierko, E., Tucker, S.C., and Honn, K.V. (2016) Thrombin-unique coagulation system protein with multifaceted impacts on cancer and metastasis. *Cancer Metastasis Rev*, **35**(2): 213–233.

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.