

Contents

1	Introduction	1
1.1	Cancer Research in the Post-Genomic Era	1
1.1.1	Cancer as a Global Health Concern	2
1.1.1.1	Genetics and Molecular Biology in Cancers	3
1.1.2	The Human Genome Revolution	5
1.1.2.1	The First Human Genome Sequence	5
1.1.2.2	Impact of Genomics	6
1.1.3	Technologies to Enable Genetics Research	7
1.1.3.1	DNA Sequencing and Genotyping Technologies	7
1.1.3.2	Microarrays and Quantitative Technologies	8
1.1.3.3	Massively Parallel “Next Generation” Sequencing	9
1.1.3.3.1	Molecular Profiling with Genomics Technology	10
1.1.3.3.2	Established Sequencing Technologies	11
1.1.3.3.3	Emerging Sequencing Technologies	12
1.1.3.4	Bioinformatics as Interdisciplinary Genomic Analysis	14
1.1.4	Follow-up Large-Scale Genomics Projects	14
1.1.5	Cancer Genomes	15
1.1.5.1	The Cancer Genome Atlas Project	16
1.1.5.2	The International Cancer Genome Consortium	17
1.1.5.2.1	Findings from Cancer Genomes	17
1.1.5.2.2	Genomic Comparisons Across Cancer Tissues	19
1.1.5.2.3	Cancer Genomic Data Resources	20
1.1.6	Genomic Cancer Medicine	20
1.1.6.1	Cancer Genes and Driver Mutations	21
1.1.6.2	Personalised or Precision Cancer Medicine	22
1.1.6.2.1	Molecular Diagnostics and Pan-Cancer Medicine	22
1.1.6.3	Targeted Therapeutics and Pharmacogenomics	23
1.1.6.3.1	Targeting Oncogenic Driver Mutations	23
1.1.6.4	Systems and Network Biology	24
1.1.6.4.1	Network Medicine, and Polypharmacology	27
1.2	A Synthetic Lethal Approach to Cancer Medicine	28
1.2.1	Synthetic Lethal Genetic Interactions	28
1.2.2	Synthetic Lethal Concepts in Genetics	29
1.2.3	Studies of Synthetic Lethality	30
1.2.3.1	Synthetic Lethal Pathways and Networks	30

1.2.3.1.1	Evolution of Synthetic Lethality	31
1.2.4	Synthetic Lethal Concepts in Cancer	32
1.2.5	Clinical Impact of Synthetic Lethality in Cancer	33
1.2.6	High-throughput Screening for Synthetic Lethality	35
1.2.6.1	Synthetic Lethal Screens	37
1.2.7	Computational Prediction of Synthetic Lethality	40
1.2.7.1	Bioinformatics Approaches to Genetic Interactions . .	40
1.2.7.2	Comparative Genomics	41
1.2.7.3	Analysis and Modelling of Protein Data	44
1.2.7.4	Differential Gene Expression	46
1.2.7.5	Data Mining and Machine Learning	47
1.2.7.6	Bimodality	50
1.2.7.7	Rationale for Further Development	50
1.3	E-cadherin as a Synthetic Lethal Target	51
1.3.1	The <i>CDH1</i> gene and it's Biological Functions	51
1.3.1.1	Cytoskeleton	51
1.3.1.2	Extracellular and Tumour Micro-Environment	52
1.3.1.3	Cell-Cell Adhesion and Signalling	52
1.3.2	<i>CDH1</i> as a Tumour (and Invasion) Suppressor	52
1.3.2.1	Breast Cancers and Invasion	53
1.3.3	Hereditary Diffuse Gastric Cancer and Lobular Breast Cancer .	53
1.3.4	Somatic Mutations	54
1.3.4.1	Mutation Rate	54
1.3.4.2	Co-occurring Mutations	55
1.3.5	Models of <i>CDH1</i> loss in cell lines	56
1.4	Summary and Research Direction of Thesis	56
2	Methods and Resources	5
2.1	Bioinformatics Resources for Genomics Research	5
2.1.1	Public Data and Software Packages	5
2.1.1.1	Cancer Genome Atlas Data	6
2.1.1.2	Reactome and Annotation Data	7
2.2	Data Handling	8
2.2.1	Normalisation	8
2.2.2	Sample Triage	8
2.2.3	Metagenes and the Singular Value Decomposition	10
2.2.3.1	Candidate Triage and Integration with Screen Data . .	10
2.3	Techniques	11
2.3.1	Statistical Procedures and Tests	11
2.3.2	Gene Set Over-representation Analysis	12
2.3.3	Clustering	12
2.3.4	Heatmap	13
2.3.5	Modeling and Simulations	13
2.3.5.1	Receiver Operating Characteristic (Performance) . . .	14
2.3.6	Resampling Analysis	15
2.4	Pathway Structure Methods	16

2.4.1	Network and Graph Analysis	16
2.4.2	Sourcing Graph Structure Data	17
2.4.3	Constructing Pathway Subgraphs	17
2.4.4	Network Analysis Metrics	17
2.5	Implementation	18
2.5.1	Computational Resources and Linux Utilities	18
2.5.2	R Language and Packages	19
2.5.3	High Performance and Parallel Computing	22
3	Methods Developed During Thesis	24
3.1	A Synthetic Lethal Detection Methodology	25
3.2	Synthetic Lethal Simulation and Modelling	27
3.2.1	A Model of Synthetic Lethality in Expression Data	27
3.2.2	Simulation Procedure	31
3.3	Detecting Simulated Synthetic Lethal Partners	34
3.3.1	Binomial Simulation of Synthetic lethality	34
3.3.2	Multivariate Normal Simulation of Synthetic lethality	36
3.3.2.1	Multivariate Normal Simulation with Correlated Genes	39
3.3.2.2	Specificity with Query-Correlated Pathways	46
3.3.2.2.1	Importance of Directional Testing	46
3.4	Graph Structure Methods	47
3.4.1	Upstream and Downstream Gene Detection	47
3.4.1.1	Permutation Analysis for Statistical Significance	49
3.4.1.2	Ranking Based on Biological Context	50
3.4.2	Simulating Gene Expression from Graph Structures	51
3.5	Customised Functions and Packages Developed	54
3.5.1	Synthetic Lethal Interaction Prediction Tool	54
3.5.2	Data Visualisation	56
3.5.3	Extensions to the iGraph Package	58
3.5.3.1	Sampling Simulated Data from Graph Structures	58
3.5.3.2	Plotting Directed Graph Structures	59
3.5.3.3	Computing Information Centrality	60
3.5.3.4	Testing Pathway Structure with Permutation Testing	60
3.5.3.5	Metapackage to Install iGraph Functions	60
4	Synthetic Lethal Analysis of Gene Expression Data	6
4.1	Synthetic lethal genes in breast cancer	8
4.1.1	Synthetic lethal pathways in breast cancer	10
4.1.2	Expression profiles of synthetic lethal partners	10
4.1.2.1	Subgroup pathway analysis	11
4.2	Comparison of synthetic lethal gene candidates	13
4.2.1	Comparison with differential expression	13
4.2.2	Comparison with correlation	13
4.2.3	Comparison with primary siRNA screen candidates	13
4.2.3.1	Comparison of screen at pathway level	16
4.2.3.1.1	Resampling of genes for pathway enrichment	16

4.2.4	Comparison with secondary screen siRNA screen candidates . .	20
4.2.4.1	Comparison of candidate SL Pathways	20
4.3	Mutation, Copy Number, and Methylation	20
4.3.1	Synthetic lethality by DNA copy number	22
4.3.2	Synthetic lethality by somatic mutation	22
4.3.2.1	Mutation analysis	22
4.3.3	ANOVA of Expression Predictors	22
4.4	Global Synthetic Lethality	23
4.4.1	Hub Genes	23
4.5	Metagene Analysis	23
4.5.1	Pathway expression	23
4.5.2	Somatic mutation	23
4.5.3	Synthetic lethal metagenes	23
4.6	Replication in stomach cancer	23
4.7	Replication in cell line encyclopaedia	24
4.8	Summary	26
5	Synthetic Lethal Pathway Structure	142
5.1	Reactome Network structure and Information Centrality as a measure of gene essentiality	143
5.2	Synthetic lethal genes in synthetic lethal pathways	143
5.3	Centrality and connectivity of synthetic lethal genes	143
5.4	Upstream or downstream synthetic lethal candidates	143
5.5	Hierachical approach	143
5.6	Discussion	143
5.7	Conclusion	143
6	Simulation and Modeling of Synthetic Lethal Pathways	144
6.1	Simulations and Modelling Synthetic Lethality in Expression Data . . .	146
6.2	Simulations over simple graph structures	147
6.2.1	Performance	147
6.2.2	Synthetic lethality across graph stuctures	147
6.2.3	Performance with inhibition links	147
6.2.4	Performance with 20,000 genes	147
6.3	Simulations over pathway-based graphs	147
6.4	Comparing methods	147
6.4.1	SLIPT and Chi-Squared	147
6.4.1.1	Correlated query genes	147
6.4.2	Correlation	147
6.4.3	Bimodality with BiSEp	147
7	Discussion	148
7.1	Significance	150
7.2	Future Directions	151
7.3	Conclusion	152

8 Conclusion	153
A Sample Correlation	154
B Replicate Samples in TCGA Breast	156
C Software Used for Thesis	160
D Secondary Screen Data	169
E Mutation Analysis in Breast Cancer	171
F Expression Analysis in Stomach Cancer	179
G Mutation Analysis in Stomach Cancer	180

List of Figures

1.1	Synthetic genetic interactions	29
1.2	Synthetic lethality in cancer	33
2.1	Read count density	8
2.2	Read count sample mean	9
3.1	Framework for synthetic lethal prediction	25
3.2	Synthetic lethal prediction adapted for mutation	26
3.3	A model of synthetic lethal gene expression	28
3.4	Modeling synthetic lethal gene expression	29
3.5	Synthetic lethality with multiple genes	30
3.6	Simulating gene function	32
3.7	Simulating synthetic lethal gene function	33
3.8	Simulating synthetic lethal gene expression	33
3.9	Performance of binomial simulations	35
3.10	Comparison of statistical performance	35
3.11	Performance of multivariate normal simulations	37
3.12	Simulating expression with correlated gene blocks	40
3.13	Simulating expression with correlated gene blocks	41
3.14	Synthetic lethal prediction across simulations	42
3.15	Performance with correlations	43
3.16	Comparison of statistical performance with correlation structure	44
3.17	Performance with query correlations	45
3.18	Statistical evaluation of directional criteria	47
3.19	Performance of directional criteria	48
3.20	Simulated graph structures	52
3.21	Simulating expression from a graph structure	53
3.22	Simulating expression from graph structure with inhibitions	55
3.23	Demonstration of violin plots with custom features	57
3.24	Demonstration of annotated heatmap	57
3.25	Simulating graph structures	59
4.1	Synthetic lethal expression profiles of analysed samples	12
4.2	Comparison of SLIPT to siRNA	15
4.3	Resampled intersection of SLIPT and siRNA candidates	18
4.4	Synthetic lethal expression profiles of stomach samples	31
4.5	Comparison of SLIPT in stomach to siRNA	32

A.1	Correlation profiles of removed samples	154
A.2	Correlation analysis and sample removal	155
B.1	Replicate excluded samples	156
B.2	Replicate samples with all remaining	157
B.3	Replicate samples with some excluded	158
B.3	Replicate samples with some excluded	159
E.1	Synthetic lethal expression profiles of analysed samples	177
E.2	Comparison of mtSLIPT to siRNA	178
G.1	Synthetic lethal expression profiles of stomach samples	183
G.2	Comparison of mtSLIPT in stomach to siRNA	185

List of Tables

1.1	Methods for Predicting Genetic Interactions	40
1.2	Methods for Predicting Synthetic Lethality in Cancer	41
1.3	Methods used by ?	43
2.1	Excluded Samples by Batch and Clinical Characteristics.	9
2.2	Computers used during Thesis	19
2.3	Linux Utilities and Applications used during Thesis	20
2.4	R Installations used during Thesis	20
2.6	R Packages used during Thesis	20
2.5	R Packages Developed during Thesis	23
4.1	Candidate synthetic lethal genes against E-cadherin from SLIPT	9
4.2	Pathways for <i>CDH1</i> partners from SLIPT	10
4.3	Pathway composition for clusters of <i>CDH1</i> partners from SLIPT	14
4.4	Pathway composition for <i>CDH1</i> partners from SLIPT and siRNA screen- ing	17
4.5	Pathways for <i>CDH1</i> partners from SLIPT	19
4.6	Pathways for <i>CDH1</i> partners from SLIPT and siRNA primary screen .	21
4.7	Candidate synthetic lethal metagenes against <i>CDH1</i> from SLIPT	24
4.8	Candidate synthetic lethal genes against E-cadherin from SLIPT in stomach cancer	25
4.9	Pathways for <i>CDH1</i> partners from SLIPT in stomach cancer	26
4.10	Pathway composition for clusters of <i>CDH1</i> partners in stomach SLIPT	27
4.11	Pathway composition for <i>CDH1</i> partners from SLIPT and siRNA screen- ing	28
4.12	Pathways for <i>CDH1</i> partners from SLIPT in stomach cancer	29
4.13	Pathways for <i>CDH1</i> partners from SLIPT in stomach and siRNA screen	30
4.14	Candidate synthetic lethal metagenes against <i>CDH1</i> from SLIPT in stomach cancer	33
4.15	Candidate synthetic lethal genes against E-cadherin from SLIPT in CCLE	34
4.16	Pathways for <i>CDH1</i> partners from SLIPT in CCLE	35
4.17	Candidate synthetic lethal genes against E-cadherin from SLIPT in breast CCLE	36
4.18	Pathways for <i>CDH1</i> partners from SLIPT in breast CCLE	37
4.19	Candidate synthetic lethal genes against E-cadherin from SLIPT in stomach CCLE	38
4.20	Pathways for <i>CDH1</i> partners from SLIPT in stomach CCLE	39

C.1	R Packages used during Thesis	160
D.1	Comparing SLIPT genes against Secondary siRNA Screen in breast cancer	169
D.2	Comparing mtSLIPT genes against Secondary siRNA Screen in breast cancer	170
D.3	Comparing SLIPT genes against Secondary siRNA Screen in stomach cancer	170
E.1	Candidate synthetic lethal genes against E-cadherin from mtSLIPT . .	171
E.2	Pathways for <i>CDH1</i> partners from mtSLIPT	172
E.3	Pathway composition for clusters of <i>CDH1</i> partners from mtSLIPT . .	173
E.4	Pathway composition for <i>CDH1</i> partners from mtSLIPT and siRNA . .	174
E.5	Pathways for <i>CDH1</i> partners from mtSLIPT	175
E.6	Pathways for <i>CDH1</i> partners from mtSLIPT and siRNA primary screen	176
E.7	Candidate synthetic lethal metagenes against <i>CDH1</i> from mtSLIPT . .	178
G.1	Candidate synthetic lethal genes against E-cadherin from mtSLIPT in stomach cancer	181
G.2	Pathways for <i>CDH1</i> partners from mtSLIPT in stomach cancer	182
G.3	Pathway composition for clusters of <i>CDH1</i> partners in stomach mtSLIPT	184
G.4	Pathway composition for <i>CDH1</i> partners from mtSLIPT and siRNA . .	186
G.5	Pathways for <i>CDH1</i> partners from mtSLIPT in stomach cancer	187
G.6	Pathways for <i>CDH1</i> partners from mtSLIPT in stomach and siRNA screen	188
G.7	Candidate synthetic lethal metagenes against <i>CDH1</i> from mtSLIPT in stomach cancer	189