

Library Declaration Form



University of Otago Library

Author's full name and year of birth: Simon Thomas Kelly,
(for cataloguing purposes) 24 February 1992

Title of thesis: A Bioinformatics Approach to Synthetic Lethal Interactions in Breast Cancer with Gene Expression Data

Degree: Doctor of Philosophy

Department: Department of Biochemistry

Permanent Address: 710 Cumberland Street, Dunedin, NZ

I agree that this thesis may be consulted for research and study purposes and that reasonable quotation may be made from it, provided that proper acknowledgement of its use is made.

I consent to this thesis being copied in part or in whole for

- i) a library
- ii) an individual

at the discretion of the University of Otago.

Signature:

Date:

A Bioinformatics Approach to
Synthetic Lethal Interactions in
Breast Cancer with Gene
Expression Data

S. Thomas Kelly

a thesis submitted for the degree of
Doctor of Philosophy
at the University of Otago, Dunedin,
New Zealand.

31 March 2017

Abstract

Background

Synthetic lethal genetic interactions are re-emerging in the post-genomics era due to their potential for precision medicine against cancers. Synthetic lethal drug design exploits functional redundancy with genes disrupted in cancers (including tumour suppressors) to develop specific treatments. E-cadherin, encoded by *CDH1*, is a tumour suppressor gene with loss of function in breast and stomach cancers. Experimental screens have identified candidate synthetic lethal interactions for drug target triage which can be further supported with bioinformatics analysis. Furthermore, gene expression data is amenable to investigation of the pathway composition and structure of synthetic lethal candidates.

Approach

A computational methodology, the Synthetic Lethal Prediction Tool (SLIPT) was developed to detect synthetic lethal interactions in gene expression data. This methodology was demonstrated on interactions with *CDH1* in breast and stomach cancer data from The Cancer Genome Atlas (TCGA) project. Synthetic lethal genes and pathways were further investigated with unsupervised clustering, gene set over-representation analysis, metagenes, and permutation resampling. In particular, analyses focused on comparing SLIPT gene candidates to an experimental siRNA screen Telford *et al.* (2015). Graph theory methods were also applied to the most supported pathways to test for pathway structure among synthetic lethal candidates. Simulation and modelling was used to assess the statistical performance of SLIPT, including simulated data with correlation structures derived from graph structures.

Findings

Many candidate synthetic lethal partners of *CDH1* were detected in both TCGA breast cancer. These genes clustered into several distinct groups, with distinct biological functions and elevated expression in different clinical subtypes. While the number of genes detected by both approaches was not significant, these contained significantly enriched pathways. In particular $G_{\alpha i}$ signalling, cytoplasmic microfibres, and extracellular fibrin clotting were robustly supported by both approaches which is consistent with the known cytoskeletal and cell signalling roles of E-cadherin and validation of GPCRs performed by Telford *et al.* (2015). Many of these pathways were replicated in stomach cancer. The pathways supported only by SLIPT included regulation of immune signalling and translational elongation which were unlikely to be detected in an isogenic cell line model but are still candidates for further investigation.

Synthetic lethal candidates detected by SLIPT and siRNA were compared within graph structures of the candidate synthetic lethal pathways. These genes did not differ by network measures of importance or connectivity in the pathway. There was also little support for SLIPT gene candidates being upstream or downstream of siRNA gene candidates within a pathway, consistently across pathways.

A model of synthetic lethality used to simulate gene expression data with synthetic lethal partners of a gene. The SLIPT methodology had high statistical performance detecting few synthetic lethal partners, although this diminishes with more synthetic lethal partners or lower sample size. The SLIPT methodology performs better than Pearson correlation or the χ^2 -test. In particular, it performs well with high specificity for datasets containing thousands of genes or genes positively correlated with the query gene (as expected in human expression data). It was also robust across correlation structures, including those derived from complex pathway structures and often distinguished synthetic lethal genes from those positively or negatively correlated with them. Therefore SLIPT is appropriate to identify synthetic lethal genes within pathways and use candidate synthetic lethal genes (and their correlates) to identify synthetic lethal pathways.

Summary

Thus my thesis has developed, evaluated and refined a bioinformatics approach to discovery of synthetic lethal genes solely from gene expression data. This approach has been demonstrated to detect biologically informative and clinically relevant candidate partners for *CDH1* in breast and stomach cancers. These investigations have also involved the development of network analysis and simulation procedures which may be more widely applicable.

Research Contributions During Candidature

Publications

Kelly, S. T. and Spencer, H. G. (2017) Population-Genetics Models of Sex-Limited Genomic Imprinting. *Theoretical Population Biology* **115**:35-44
doi:10.1016/j.tpb.2017.03.004

Manuscripts Submitted

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

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

Conference Presentations

Consortium of Biological Sciences 2017 (Kobe) December TBC

eResearch 2017 (Queenstown) February 20th-22nd

Research Bazaar 2016 (Dunedin) February 2nd-4th

eResearch 2016 (Queenstown) February 9th-11th

Genetics Otago Symposium 2016 (Dunedin) March 7th-8th

DunDead: Zombie Science and Culture Festival 2014 (Dunedin) August 16th-17th

eResearch 2014 (Hamilton) June 30th-July 2nd (Supported by Google)

Poster Presentations

Next Generation Sequencing Asia 2016 (Singapore) October 11th-12th (Supported by the University of Otago Division of Health Sciences; Maurice and Phyllis Paykel Trust)

Research Bazaar 2015 (Melbourne, Australia) February 16th-18th (Supported by the New Zealand eScience Infrastructure)

Otago School of Medical Sciences Postgraduate Symposium 2015 (Dunedin, New Zealand) April 28th-29th

QMB Cancer Drugs Satellite 2014 (Queenstown, New Zealand) August 24th-25th

Seminar Presentations

University of Otago Department of Biochemistry 2017 (Dunedin) November TBC

Tohoku University 2016 (Sendai) November 11th

Okinawa Institute of Science and Technology 2016 (Onna) November 1st

Sokendai Graduate University 2016 (Hayama) October 25th

Tokyo University Institute of Medical Science 2016 (Shirokanedai) October 24th

National Institute of Genetics 2016 (Mishima) October 21st

RIKEN Division of Genomic Technologies 2016 (Yokohama) October 20th

Software Packages

Software packages in the R language have been released. Please refer to the appropriate GitHub repository for more information (including documentation, vignettes, and installation instructions), on the following account: <https://github.com/TomKellyGenetics>

- `slipt` to accompany the synthetic lethal publication above and release SLIPT (Synthetic Lethal Interaction Prediction Tool)
- `vioplotx` to provide enhanced violin plots
- `heatmap.2x` to provide annotated heatmaps
- `igraph.extensions` metapackage for the packages for iGraph objects:
 - `plot.igraph` to provide plotting for directed graphs
 - `info.centraliity` to compute network analysis metrics
 - `pathway.structure.permutation` for resampling within pathways
 - `graphsim` to simulate expression (`mvtnorm`) from pathway structures

The `slipt-app` GitHub repository also hosts an application for Synthetic Lethal Interaction Prediction Tool (SLIPT) developed in the R `shiny` environment as part of a related project. There is also a digital copy of this thesis, including high resolution full-colour figures, hosted at:

<https://github.com/TomKellyGenetics/thesis/blob/master/thesis.pdf>

Acknowledgements

I thank my supervisors A/Prof. Mik Black and Prof. Parry Guilford for their support and guidance throughout this my postgraduate studies. It has been a great experience, I look forward to seeing what your research groups produce in the future, may this not be the end.

I am also thankful for the guidance and mentorship of Prof. Hamish Spencer for career advice throughout my studies and time in his research group.

I am also grateful to the past and current members of these research groups, and my peers at the laboratory benches and computers across campus. The peer support, comraderie, and guidance to newer students has been an incredible part of my time at Otago and has made my thesis studies not just easier but possible at all. The postgraduate community is very special here and have truly made some lifelong friends from all over the world, you are talented researchers and amazing people. May we meet again some day. Whereever you may end up, there's always time to catch up and I'd be delighted to host some visits while working abroad.

I cannot thank my friends, flatmates and family enough for their patience and support during such as massive, challenging, and (I'm sure you've heard too many times) stressful undertaking during both my PhD and the study leading up to it. There are too many of you to name everyone here without leaving someone out, so thank you all for everything you've done, both the good times and the tough. Thank you for at least pretending to understand complex math oft brough up at the wrong moment. Thank you for checking my writing or slides, even when sprung on you last minute. Thank for your time when what I really needed was a chat over a walk or a drink and a moment to think clearly.

I thank the various organisations that supported this research project:

- This thesis was supported by the Postgraduate Tassell Scholarship in Cancer Research, a University of Otago Doctoral Scholarship.
- The New Zealand eScience Infrastructure (NeSI) provided access to the Intel Pan high-performance computing cluster, support, and training to use it effectively. Various aspects of this thesis would not have been possible without access to such a resource.
- The Health Research Council (HRC) of New Zealand provided funding for experimental research in the Cancer Genetics Laboratory. Again some aspects of this project would not have been possible without access to the data and findings funded by this grant.
- The Allan Wilson Centre and Otago School of Biomedical Sciences provided funding for summer research placements which was a valuable opportunity to gain experience and training used in this thesis project.

I thank the following organisations for support towards presenting findings in this thesis at conference and seminars:

- Google (eResearch 2014 conference, Hamilton)
- NeSI (Software Carpentry training and Research Bazaar 2015, Melbourne)
- REANNZ, NZGL, and NeSI (eResearch 2016 conference, Queenstown)
- Otago Division of Health Sciences, Oxford Global, and Maurice and Phyllis Paykel Trust (NGS Asia 2016, Singapore)
- RIKEN Division of Genomics Technologies and the Okinawa Institute of Science and Technology (funding seminar visits in Japan)

Thanks most of all to my fiancé, Dr Yui Kawagishi, you've been an inspiration. Thank you for your support and encouragement, every day, even from afar: it has always made a difference. It's been incredible to see you flourish in your career and I look forward to joining you again soon. May the next chapter of our adventures involve a bit less Skype across timezones.

どうもありがとう由ちゃん。また来月。頑張った！行きます！

Contents

| | |
|---|-------------|
| Glossary | xxi |
| Acronyms | xxii |
| 1 Introduction and Literature Review | 1 |
| 1.1 Cancer Research in the Post-Genomic Era | 1 |
| 1.1.1 Cancer is a Global Health Issue | 2 |
| 1.1.1.1 The Genetics and Molecular Biology of Cancers | 3 |
| 1.1.2 The Genomics Revolution in Cancer Research | 4 |
| 1.1.2.1 High-Throughput Technologies | 4 |
| 1.1.2.2 Bioinformatics and Genomics Data | 6 |
| 1.1.3 Genomics Projects | 6 |
| 1.1.3.1 The Cancer Genome Project | 6 |
| 1.1.3.2 The Cancer Genome Atlas Project | 7 |
| 1.1.4 Genomic Cancer Medicine | 8 |
| 1.1.4.1 Cancer Genes and Driver Mutations | 9 |
| 1.1.4.2 Precision Cancer Medicine | 10 |
| 1.1.4.3 Molecular Diagnostics and Pan-Cancer Medicine | 10 |
| 1.1.4.4 Targeted Therapeutics and Pharmacogenomics | 10 |
| 1.1.5 Systems and Network Biology | 11 |
| 1.1.5.1 Network Medicine and Polypharmacology | 13 |
| 1.2 A Synthetic Lethal Approach to Cancer Medicine | 14 |
| 1.2.1 Synthetic Lethal Genetic Interactions | 14 |
| 1.2.2 Synthetic Lethal Concepts in Genetics | 15 |
| 1.2.3 Synthetic Lethality in Model Systems | 16 |
| 1.2.3.1 Synthetic Lethal Pathways and Networks | 16 |
| 1.2.3.2 Evolution of Synthetic Lethality | 17 |
| 1.2.4 Synthetic Lethality in Cancer | 18 |
| 1.2.5 Clinical Impact of Synthetic Lethality in Cancer | 19 |
| 1.2.6 High-throughput Screening for Synthetic Lethality | 21 |
| 1.2.6.1 Synthetic Lethal Screens | 22 |
| 1.2.7 Computational Prediction of Synthetic Lethality | 25 |
| 1.2.7.1 Bioinformatics Approaches to Genetic Interactions | 25 |
| 1.2.7.2 Comparative Genomics | 26 |
| 1.2.7.3 Analysis and Modelling of Protein Data | 29 |
| 1.2.7.4 Differential Gene Expression | 31 |

| | | |
|----------|---|-----------|
| 1.2.7.5 | Data Mining and Machine Learning | 32 |
| 1.2.7.6 | Mutually Exclusive Bimodality | 35 |
| 1.2.7.7 | Rationale for Further Development | 35 |
| 1.3 | E-cadherin as a Synthetic Lethal Target | 36 |
| 1.3.1 | The <i>CDH1</i> gene and its Biological Functions | 36 |
| 1.3.1.1 | Cytoskeleton | 36 |
| 1.3.1.2 | Extracellular and Tumour Micro-environment | 37 |
| 1.3.1.3 | Cell-Cell Adhesion and Signalling | 37 |
| 1.3.2 | <i>CDH1</i> as a Tumour (and Invasion) Suppressor | 37 |
| 1.3.2.1 | Breast Cancers and Invasion | 38 |
| 1.3.3 | Hereditary Diffuse Gastric Cancer and Lobular Breast Cancer | 38 |
| 1.3.4 | Cell Line Models of <i>CDH1</i> Null Mutations | 39 |
| 1.4 | Summary and Research Direction of Thesis | 40 |
| 1.4.1 | Thesis Aims | 42 |
| 2 | Methods and Resources | 43 |
| 2.1 | Bioinformatics Resources for Genomics Research | 43 |
| 2.1.1 | Public Data and Software Packages | 43 |
| 2.1.1.1 | Cancer Genome Atlas Data | 44 |
| 2.1.1.2 | Reactome and Annotation Data | 45 |
| 2.2 | Data Handling | 45 |
| 2.2.1 | Normalisation | 45 |
| 2.2.2 | Sample Triage | 46 |
| 2.2.3 | Metagenes and the Singular Value Decomposition | 48 |
| 2.2.3.1 | Candidate Triage and Integration with Screen Data | 48 |
| 2.3 | Techniques | 49 |
| 2.3.1 | Statistical Procedures and Tests | 49 |
| 2.3.2 | Gene Set Over-representation Analysis | 50 |
| 2.3.3 | Clustering | 50 |
| 2.3.4 | Heatmap | 51 |
| 2.3.5 | Modeling and Simulations | 51 |
| 2.3.5.1 | Receiver Operating Characteristic (Performance) | 52 |
| 2.3.6 | Resampling Analysis | 52 |
| 2.4 | Pathway Structure Methods | 53 |
| 2.4.1 | Network and Graph Analysis | 53 |
| 2.4.2 | Sourcing Graph Structure Data | 54 |
| 2.4.3 | Constructing Pathway Subgraphs | 55 |
| 2.4.4 | Network Analysis Metrics | 55 |
| 2.5 | Implementation | 56 |
| 2.5.1 | Computational Resources and Linux Utilities | 56 |
| 2.5.2 | R Language and Packages | 57 |
| 2.5.3 | High Performance and Parallel Computing | 60 |

| | | |
|----------|---|------------|
| 3 | Methods Developed During Thesis | 62 |
| 3.1 | A Synthetic Lethal Detection Methodology | 62 |
| 3.2 | Synthetic Lethal Simulation and Modelling | 65 |
| 3.2.1 | A Model of Synthetic Lethality in Expression Data | 65 |
| 3.2.2 | Simulation Procedure | 69 |
| 3.3 | Detecting Simulated Synthetic Lethal Partners | 72 |
| 3.3.1 | Binomial Simulation of Synthetic lethality | 72 |
| 3.3.2 | Multivariate Normal Simulation of Synthetic lethality | 74 |
| 3.3.2.1 | Multivariate Normal Simulation with Correlated Genes | 77 |
| 3.3.2.2 | Specificity with Query-Correlated Pathways | 84 |
| 3.3.2.3 | Importance of Directional Testing | 84 |
| 3.4 | Graph Structure Methods | 86 |
| 3.4.1 | Upstream and Downstream Gene Detection | 86 |
| 3.4.1.1 | Permutation Analysis for Statistical Significance | 87 |
| 3.4.1.2 | Hierarchy Based on Biological Context | 88 |
| 3.4.2 | Simulating Gene Expression from Graph Structures | 89 |
| 3.5 | Customised Functions and Packages Developed | 93 |
| 3.5.1 | Synthetic Lethal Interaction Prediction Tool | 93 |
| 3.5.2 | Data Visualisation | 94 |
| 3.5.3 | Extensions to the iGraph Package | 97 |
| 3.5.3.1 | Sampling Simulated Data from Graph Structures | 97 |
| 3.5.3.2 | Plotting Directed Graph Structures | 97 |
| 3.5.3.3 | Computing Information Centrality | 98 |
| 3.5.3.4 | Testing Pathway Structure with Permutation Testing | 98 |
| 3.5.3.5 | Metapackage to Install iGraph Functions | 99 |
| 4 | Synthetic Lethal Analysis of Gene Expression Data | 100 |
| 4.1 | Synthetic Lethal Genes in Breast Cancer | 101 |
| 4.1.1 | Synthetic Lethal Pathways in Breast Cancer | 103 |
| 4.1.2 | Expression Profiles of Synthetic Lethal Partners | 104 |
| 4.1.2.1 | Subgroup Pathway Analysis | 107 |
| 4.2 | Comparing Synthetic Lethal Gene Candidates | 110 |
| 4.2.1 | Primary siRNA Screen Candidates | 110 |
| 4.2.2 | Comparison with Correlation | 111 |
| 4.2.3 | Comparison with Primary Screen Viability | 113 |
| 4.2.4 | Comparison with Secondary siRNA Screen Validation | 114 |
| 4.2.5 | Comparison to Primary Screen at Pathway Level | 116 |
| 4.2.5.1 | Resampling Genes for Pathway Enrichment | 118 |
| 4.2.6 | Integrating Synthetic Lethal Pathways and Screens | 121 |
| 4.3 | Metagene Analysis | 123 |
| 4.3.1 | Pathway Expression | 124 |
| 4.3.2 | Somatic Mutation | 126 |
| 4.3.3 | Synthetic Lethal Pathway Metagenes | 130 |
| 4.3.4 | Synthetic Lethality in Breast Cancer | 131 |
| 4.4 | Replication in Stomach Cancer | 132 |
| 4.5 | Discussion | 133 |

| | | |
|----------|--|------------|
| 4.5.1 | Strengths of the SLIPT Methodology | 133 |
| 4.5.2 | Synthetic Lethal Pathways for E-cadherin | 134 |
| 4.5.3 | Replication and Validation | 136 |
| 4.5.3.1 | Integration with siRNA Screening | 136 |
| 4.5.3.2 | Replication across Tissues | 137 |
| 4.6 | Summary | 137 |
| 5 | Synthetic Lethal Pathway Structure | 139 |
| 5.1 | Synthetic Lethal Genes in Reactome Pathways | 139 |
| 5.1.1 | The PI3K/AKT Pathway | 140 |
| 5.1.2 | The Extracellular Matrix | 142 |
| 5.1.3 | G Protein Coupled Receptors | 145 |
| 5.1.4 | Gene Regulation and Translation | 145 |
| 5.2 | Network Analysis of Synthetic Lethal Genes | 146 |
| 5.2.1 | Gene Connectivity and Vertex Degree | 147 |
| 5.2.2 | Gene Importance and Centrality | 148 |
| 5.2.2.1 | Information Centrality | 148 |
| 5.2.2.2 | PageRank Centrality | 150 |
| 5.3 | Relationships between Synthetic Lethal Genes | 152 |
| 5.3.1 | Hierarchical Pathway Structure | 152 |
| 5.3.1.1 | Contextual Hierarchy of PI3K | 152 |
| 5.3.1.2 | Testing Contextual Hierarchy of Synthetic Lethal Genes | 152 |
| 5.3.2 | Upstream or Downstream Synthetic Lethality | 156 |
| 5.3.2.1 | Measuring Structure of Candidates within PI3K | 156 |
| 5.3.2.2 | Resampling for Synthetic Lethal Pathway Structure | 158 |
| 5.4 | Discussion | 160 |
| 5.5 | Summary | 162 |
| 6 | Simulation and Modeling of Synthetic Lethal Pathways | 164 |
| 6.1 | Synthetic Lethal Detection Methods | 165 |
| 6.1.1 | Performance of SLIPT and χ^2 across Quantiles | 165 |
| 6.1.1.1 | Correlated Query Genes affects Specificity | 169 |
| 6.1.2 | Alternative Synthetic Lethal Detection Strategies | 171 |
| 6.1.2.1 | Correlation for Synthetic Lethal Detection | 171 |
| 6.1.2.2 | Testing for Bimodality with BiSEp | 173 |
| 6.2 | Simulations with Graph Structures | 174 |
| 6.2.1 | Performance over a Graph Structure | 175 |
| 6.2.1.1 | Simple Graph Structures | 175 |
| 6.2.1.2 | Constructed Graph Structures | 177 |
| 6.2.2 | Performance with Inhibitions | 180 |
| 6.2.3 | Synthetic Lethality across Graph Structures | 185 |
| 6.2.4 | Performance within a Simulated Human Genome | 189 |
| 6.3 | Simulations in More Complex Graph Structures | 193 |
| 6.3.1 | Simulations over Pathway-based Graphs | 194 |
| 6.3.2 | Pathway Structures in a Simulated Human Genome | 197 |
| 6.4 | Discussion | 200 |

| | | |
|----------|---|------------|
| 6.4.1 | Simulation Procedure | 200 |
| 6.4.2 | Comparing Methods with Simulated Data | 201 |
| 6.4.3 | Design and Performance of SLIPT | 202 |
| 6.4.4 | Simulations from Graph Structures | 204 |
| 6.5 | Summary | 205 |
| 7 | Discussion | 207 |
| 7.1 | Synthetic Lethality and <i>CDH1</i> Biology | 207 |
| 7.1.1 | Established Functions of <i>CDH1</i> | 208 |
| 7.1.2 | The Molecular Role of <i>CDH1</i> in Cancer | 208 |
| 7.2 | Significance | 209 |
| 7.2.1 | Synthetic Lethality in the Genomic Era | 209 |
| 7.2.2 | Clinical Interventions based on Synthetic Lethality | 211 |
| 7.3 | Future Directions | 212 |
| 7.4 | Conclusions | 214 |
| | References | 216 |
| A | Sample Quality | 240 |
| A.1 | Sample Correlation | 240 |
| A.2 | Replicate Samples in TCGA Breast | 243 |
| B | Software Used for Thesis | 247 |
| C | Mutation Analysis in Breast Cancer | 256 |
| C.1 | Synthetic Lethal Genes and Pathways | 256 |
| C.2 | Synthetic Lethal Expression Profiles | 259 |
| C.3 | Comparison to Primary Screen | 262 |
| C.3.1 | Resampling Analysis | 264 |
| C.4 | Compare SLIPT genes | 266 |
| C.5 | Metagene Analysis | 268 |
| C.6 | Expression of Somatic Mutations | 269 |
| C.7 | Metagene Expression Profiles | 272 |
| D | Intrinsic Subtyping | 275 |
| E | Stomach Expression Analysis | 277 |
| E.1 | Synthetic Lethal Genes and Pathways | 277 |
| E.2 | Comparison to Primary Screen | 281 |
| E.2.1 | Resampling Analysis | 283 |
| E.3 | Metagene Analysis | 285 |
| F | Synthetic Lethal Genes in Pathways | 286 |
| G | Pathway Connectivity for Mutation SLIPT | 294 |
| H | Information Centrality for Gene Essentiality | 298 |

| | | |
|----------|--|------------|
| I | Pathway Structure for Mutation SLIPT | 301 |
| J | Performance of SLIPT and χ^2 | 304 |
| J.1 | Correlated Query Genes affects Specificity | 310 |
| K | Graph Structures | 316 |
| K.1 | Simulations from Simple Graph Structures | 316 |
| K.1.1 | Simulations from Inhibiting Graph Structures | 318 |
| K.2 | Simulation across Graph Structures | 321 |
| K.3 | Simulations from Complex Graph Structures | 325 |
| K.3.1 | Simulations from Complex Inhibiting Graphs | 328 |
| K.4 | Simulations from Pathway Graph Structures | 334 |

List of Tables

| | | |
|-----|--|-----|
| 1.1 | Methods for Predicting Genetic Interactions | 26 |
| 1.2 | Methods for Predicting Synthetic Lethality in Cancer | 27 |
| 1.3 | Methods used by Wu <i>et al.</i> (2014) | 28 |
| 2.1 | Excluded Samples by Batch and Clinical Characteristics. | 46 |
| 2.2 | Computers used during Thesis | 57 |
| 2.3 | Linux Utilities and Applications used during Thesis | 57 |
| 2.4 | R Installations used during Thesis | 58 |
| 2.5 | R Packages used during Thesis | 58 |
| 2.6 | R Packages Developed during Thesis | 60 |
| 4.1 | Candidate synthetic lethal gene partners of <i>CDH1</i> from SLIPT | 102 |
| 4.2 | Pathways for <i>CDH1</i> partners from SLIPT | 104 |
| 4.3 | Pathway composition for clusters of <i>CDH1</i> partners from SLIPT | 108 |
| 4.4 | Analysis of variance (ANOVA) for Synthetic Lethality and Correlation with <i>CDH1</i> | 112 |
| 4.5 | Comparing SLIPT genes against secondary siRNA screen in breast cancer | 115 |
| 4.6 | Pathway composition for <i>CDH1</i> partners from SLIPT and siRNA screen- ing | 117 |
| 4.7 | Pathways for <i>CDH1</i> partners from SLIPT | 120 |
| 4.8 | Pathways for <i>CDH1</i> partners from SLIPT and siRNA primary screen . | 122 |
| 4.9 | Candidate synthetic lethal metagenes against <i>CDH1</i> from SLIPT | 131 |
| 5.1 | ANOVA for Synthetic Lethality and Vertex Degree | 148 |
| 5.2 | ANOVA for Synthetic Lethality and Information Centrality | 150 |
| 5.3 | ANOVA for Synthetic Lethality and PageRank Centrality | 152 |
| 5.4 | ANOVA for Synthetic Lethality and PI3K Hierarchy | 155 |
| 5.5 | Resampling for pathway structure of synthetic lethal detection methods | 159 |
| B.1 | R Packages used during Thesis | 247 |
| C.1 | Candidate synthetic lethal gene partners of <i>CDH1</i> from mtSLIPT | 257 |
| C.2 | Pathways for <i>CDH1</i> partners from mtSLIPT | 258 |
| C.3 | Pathway composition for clusters of <i>CDH1</i> partners from mtSLIPT . . . | 261 |
| C.4 | Pathway composition for <i>CDH1</i> partners from mtSLIPT and siRNA . . | 263 |
| C.5 | Pathways for <i>CDH1</i> partners from mtSLIPT | 264 |
| C.6 | Pathways for <i>CDH1</i> partners from mtSLIPT and siRNA primary screen | 265 |
| C.7 | Candidate synthetic lethal metagenes against <i>CDH1</i> from mtSLIPT . . | 268 |

| | | |
|-----|--|-----|
| D.1 | Comparison of Intrinsic Subtypes | 275 |
| E.1 | Synthetic lethal gene partners of <i>CDH1</i> from SLIPT in stomach cancer | 277 |
| E.2 | Pathways for <i>CDH1</i> partners from SLIPT in stomach cancer | 278 |
| E.3 | Pathway composition for clusters of <i>CDH1</i> partners in stomach SLIPT | 280 |
| E.4 | Pathway composition for <i>CDH1</i> partners from SLIPT and siRNA screen- ing | 282 |
| E.5 | Pathways for <i>CDH1</i> partners from SLIPT in stomach cancer | 283 |
| E.6 | Pathways for <i>CDH1</i> partners from SLIPT in stomach and siRNA screen | 284 |
| E.7 | Candidate synthetic lethal metagenes against <i>CDH1</i> from SLIPT in stomach cancer | 285 |
| G.1 | ANOVA for Synthetic Lethality and Vertex Degree | 297 |
| G.2 | ANOVA for Synthetic Lethality and Information Centrality | 297 |
| G.3 | ANOVA for Synthetic Lethality and PageRank Centrality | 297 |
| H.1 | Information centrality for genes and molecules in the Reactome network | 299 |
| I.1 | ANOVA for Synthetic Lethality and PI3K Hierarchy | 301 |
| I.2 | Resampling for pathway structure of synthetic lethal detection methods | 303 |

List of Figures

| | | |
|------|--|-----|
| 1.1 | Synthetic genetic interactions | 15 |
| 1.2 | Synthetic lethality in cancer | 19 |
| 2.1 | Read count density | 47 |
| 2.2 | Read count sample mean | 47 |
| 3.1 | Framework for synthetic lethal prediction | 63 |
| 3.2 | Synthetic lethal prediction adapted for mutation | 64 |
| 3.3 | A model of synthetic lethal gene expression | 66 |
| 3.4 | Modeling synthetic lethal gene expression | 67 |
| 3.5 | Synthetic lethality with multiple genes | 68 |
| 3.6 | Simulating gene function | 70 |
| 3.7 | Simulating synthetic lethal gene function | 70 |
| 3.8 | Simulating synthetic lethal gene expression | 71 |
| 3.9 | Performance of binomial simulations | 73 |
| 3.10 | Comparison of statistical performance | 73 |
| 3.11 | Performance of multivariate normal simulations | 75 |
| 3.12 | Simulating expression with correlated gene blocks | 78 |
| 3.13 | Simulating expression with correlated gene blocks | 79 |
| 3.14 | Synthetic lethal prediction across simulations | 80 |
| 3.15 | Performance with correlations | 81 |
| 3.16 | Comparison of statistical performance with correlation structure | 82 |
| 3.17 | Performance with query correlations | 83 |
| 3.18 | Statistical evaluation of directional criteria | 84 |
| 3.19 | Performance of directional criteria | 85 |
| 3.20 | Simulated graph structures | 89 |
| 3.21 | Simulating expression from a graph structure | 91 |
| 3.22 | Simulating expression from graph structure with inhibitions | 92 |
| 3.23 | Demonstration of violin plots with custom features | 95 |
| 3.24 | Demonstration of annotated heatmap | 95 |
| 3.25 | Simulating graph structures | 98 |
| 4.1 | Synthetic lethal expression profiles of analysed samples | 106 |
| 4.2 | Comparison of SLIPT to siRNA | 110 |
| 4.3 | Compare SLIPT and siRNA genes with correlation | 111 |
| 4.4 | Compare SLIPT and siRNA genes with correlation | 112 |
| 4.5 | Compare SLIPT and siRNA genes with viability | 113 |

| | | |
|------|---|-----|
| 4.6 | Compare SLIPT genes with siRNA viability | 114 |
| 4.7 | Resampled intersection of SLIPT and siRNA candidates | 118 |
| 4.8 | Pathway metagene expression profiles | 125 |
| 4.9 | Expression profiles for constituent genes of PI3K | 127 |
| 4.10 | Expression profiles for estrogen receptor related genes | 128 |
| 4.11 | Somatic mutation against the PI3K metagene | 129 |
| 5.1 | Synthetic Lethality in the PI3K Cascade | 141 |
| 5.2 | Synthetic Lethality in the Elastic Fibre Formation Pathway | 143 |
| 5.3 | Synthetic Lethality in the Fibrin Clot Formation | 144 |
| 5.4 | Synthetic Lethality and Vertex Degree | 147 |
| 5.5 | Synthetic Lethality and Centrality | 150 |
| 5.6 | Synthetic Lethality and PageRank | 151 |
| 5.7 | Hierarchical Structure of PI3K | 153 |
| 5.8 | Hierarchy Score in PI3K against Synthetic Lethality in PI3K | 154 |
| 5.9 | Structure of Synthetic Lethality in PI3K | 156 |
| 5.10 | Structure of Synthetic Lethality Resampling in PI3K | 157 |
| 6.1 | Performance of χ^2 and SLIPT across quantiles | 167 |
| 6.2 | Performance of χ^2 and SLIPT across quantiles with more genes | 168 |
| 6.3 | Performance of χ^2 and SLIPT across quantiles with query correlation | 169 |
| 6.4 | Performance of χ^2 and SLIPT across quantiles with query correlation and more genes | 170 |
| 6.5 | Performance of negative correlation and SLIPT | 172 |
| 6.6 | Simple graph structures | 175 |
| 6.7 | Performance of simulations on a simple graph | 176 |
| 6.8 | Performance of simulations is similar in simple graphs | 178 |
| 6.9 | Performance of simulations on a pathway | 179 |
| 6.10 | Performance of simulations on a simple graph with inhibition | 181 |
| 6.11 | Performance is higher on a simple inhibiting graph | 182 |
| 6.12 | Performance of simulations on a constructed graph with inhibition | 183 |
| 6.13 | Performance is affected by inhibition in graphs | 185 |
| 6.14 | Detection of Synthetic Lethality within a Graph Structure with Inhibitions | 187 |
| 6.15 | Performance of simulations including a simple graph | 190 |
| 6.16 | Performance on a simple graph improves with more genes | 191 |
| 6.17 | Performance on an inhibiting graph improves with more genes | 193 |
| 6.18 | Performance of simulations on the PI3K cascade | 196 |
| 6.19 | Performance of simulations including the PI3K cascade | 198 |
| 6.20 | Performance on pathways improves with more genes | 199 |
| A.1 | Correlation profiles of removed samples | 241 |
| A.2 | Correlation analysis and sample removal | 242 |
| A.3 | Replicate excluded samples | 243 |
| A.4 | Replicate samples with all remaining | 244 |
| A.5 | Replicate samples with some excluded | 245 |
| C.1 | Synthetic lethal expression profiles of analysed samples | 260 |

| | | |
|------|---|-----|
| C.2 | Comparison of mtSLIPT to siRNA | 262 |
| C.3 | Compare mtSLIPT and siRNA genes with correlation | 266 |
| C.4 | Compare mtSLIPT and siRNA genes with correlation | 266 |
| C.5 | Compare mtSLIPT and siRNA genes with siRNA viability | 267 |
| C.6 | Somatic mutation against PIK3CA metagene | 269 |
| C.7 | Somatic mutation against PI3K protein | 270 |
| C.8 | Somatic mutation against AKT protein | 271 |
| C.9 | Pathway metagene expression profiles | 272 |
| C.10 | Expression profiles for p53 related genes | 273 |
| C.11 | Expression profiles for BRCA related genes | 274 |
| E.1 | Synthetic lethal expression profiles of stomach samples | 279 |
| E.2 | Comparison of SLIPT in stomach to siRNA | 281 |
| F.1 | Synthetic Lethality in the PI3K/AKT Pathway | 286 |
| F.2 | Synthetic Lethality in the PI3K/AKT Pathway in Cancer | 287 |
| F.3 | Synthetic Lethality in the Extracellular Matrix | 288 |
| F.4 | Synthetic Lethality in the GPCRs | 289 |
| F.5 | Synthetic Lethality in the GPCR Downstream | 290 |
| F.6 | Synthetic Lethality in the Translation Elongation | 291 |
| F.7 | Synthetic Lethality in the Nonsense-mediated Decay | 292 |
| F.8 | Synthetic Lethality in the 3' UTR | 293 |
| G.1 | Synthetic Lethality and Vertex Degree | 294 |
| G.2 | Synthetic Lethality and Centrality | 295 |
| G.3 | Synthetic Lethality and PageRank | 296 |
| H.1 | Information centrality distribution | 300 |
| I.1 | Synthetic Lethality and Heirarchy Score in PI3K | 301 |
| I.2 | Heirarchy Score in PI3K against Synthetic Lethality in PI3K | 302 |
| I.3 | Structure of Synthetic Lethality in PI3K | 302 |
| I.4 | Structure of Synthetic Lethality Resampling | 303 |
| J.1 | Performance of χ^2 and SLIPT across quantiles | 304 |
| J.2 | Performance of χ^2 and SLIPT across quantiles | 306 |
| J.3 | Performance of χ^2 and SLIPT across quantiles with more genes | 308 |
| J.4 | Performance of χ^2 and SLIPT across quantiles with query correlation | 310 |
| J.5 | Performance of χ^2 and SLIPT across quantiles with query correlation | 312 |
| J.6 | Performance of χ^2 and SLIPT across quantiles with query correlation and more genes | 314 |
| K.1 | Performance of simulations on a simple graph | 317 |
| K.2 | Performance of simulations on an inhibiting graph | 318 |
| K.3 | Performance of simulations on a constructed graph with inhibition | 319 |
| K.4 | Performance of simulations on a constructed graph with inhibition | 320 |
| K.5 | Detection of Synthetic Lethality within a Graph Structure | 321 |
| K.6 | Detection of Synthetic Lethality within an Inhibiting Graph Structure | 323 |

| | | |
|------|--|-----|
| K.7 | Detection of Synthetic Lethality within an Inhibiting Graph Structure . | 324 |
| K.8 | Performance of simulations on a branching graph | 325 |
| K.9 | Performance of simulations on a complex graph | 326 |
| K.10 | Performance of simulations on a large graph | 327 |
| K.11 | Performance of simulations on a branching graph with inhibition | 328 |
| K.12 | Performance of simulations on a branching graph with inhibition | 329 |
| K.13 | Performance of simulations on a complex graph with inhibition | 330 |
| K.14 | Performance of simulations on a complex graph with inhibition | 331 |
| K.15 | Performance of simulations on a large constructed graph with inhibition | 332 |
| K.16 | Performance of simulations on a large constructed graph with inhibition | 333 |
| K.17 | Performance of simulations on the $G_{\alpha i}$ signalling pathway | 334 |
| K.18 | Performance of simulations including the $G_{\alpha i}$ signalling pathway | 335 |

Glossary

synthetic lethal Genetic interactions where inactivation of multiple genes is inviable (or deleterious) which are viable if inactivated separately.

Acronyms

ANOVA Analysis of Variance.

GPCR G protein coupled receptor.

siRNA Short interfering ribonucleic acid.

SLIPT Synthetic lethal interaction prediction tool.

TCGA The Cancer Genome Atlas (genomics project).

References

- Aarts, M., Bajrami, I., Herrera-Abreu, M.T., Elliott, R., Brough, R., Ashworth, A., Lord, C.J., and Turner, N.C. (2015) Functional genetic screen identifies increased sensitivity to weel 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 Pædiatrica*, **96**(5): 644–647.
- American Cancer Society (2017) Genetics and cancer. <https://www.cancer.org/cancer/cancer-causes/genetics.html>. Accessed: 22/03/2017.
- Anjomshoaa, A., Lin, Y.H., Black, M.A., McCall, J.L., Humar, B., Song, S., Fukuzawa, R., Yoon, H.S., Holzmann, B., Friederichs, J., *et al.* (2008) Reduced expression of a gene proliferation signature is associated with enhanced malignancy in colon cancer. *Br J Cancer*, **99**(6): 966–973.
- Araki, H., Knapp, C., Tsai, P., and Print, C. (2012) GeneSetDB: A comprehensive meta-database, statistical and visualisation framework for gene set analysis. *FEBS Open Bio*, **2**: 76–82.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, **25**(1): 25–29.
- Ashworth, A. (2008) A synthetic lethal therapeutic approach: poly(adp) ribose polymerase inhibitors for the treatment of cancers deficient in dna double-strand break repair. *J Clin Oncol*, **26**(22): 3785–90.

- Audeh, M.W., Carmichael, J., Penson, R.T., Friedlander, M., Powell, B., Bell-McGuinn, K.M., Scott, C., Weitzel, J.N., Oaknin, A., Loman, N., *et al.* (2010) Oral poly(adp-ribose) polymerase inhibitor olaparib in patients with *BRCA1* or *BRCA2* mutations and recurrent ovarian cancer: a proof-of-concept trial. *Lancet*, **376**(9737): 245–51.
- Babyak, M.A. (2004) What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom Med*, **66**(3): 411–21.
- Bamford, S., Dawson, E., Forbes, S., Clements, J., Pettett, R., Dogan, A., Flanagan, A., Teague, J., Futreal, P.A., Stratton, M.R., *et al.* (2004) The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer*, **91**(2): 355–358.
- Barabási, A.L. and Albert, R. (1999) Emergence of scaling in random networks. *Science*, **286**(5439): 509–12.
- Barabási, A.L., Gulbahce, N., and Loscalzo, J. (2011) Network medicine: a network-based approach to human disease. *Nat Rev Genet*, **12**(1): 56–68.
- Barabási, A.L. and Oltvai, Z.N. (2004) Network biology: understanding the cell’s functional organization. *Nat Rev Genet*, **5**(2): 101–13.
- Barrat, A. and Weigt, M. (2000) On the properties of small-world network models. *The European Physical Journal B - Condensed Matter and Complex Systems*, **13**(3): 547–560.
- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehar, J., Kryukov, G.V., Sonkin, D., *et al.* (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, **483**(7391): 603–607.
- Barry, W.T. (2016) *safe: Significance Analysis of Function and Expression*. R package version 3.14.0.
- Baryshnikova, A., Costanzo, M., Dixon, S., Vizeacoumar, F.J., Myers, C.L., Andrews, B., and Boone, C. (2010a) Synthetic genetic array (sga) analysis in *saccharomyces cerevisiae* and *schizosaccharomyces pombe*. *Methods Enzymol*, **470**: 145–79.
- Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K., Youn, J.Y., Ou, J., San Luis, B.J., Bandyopadhyay, S., *et al.* (2010b) Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat Meth*, **7**(12): 1017–1024.
- Bass, A.J., Thorsson, V., Shmulevich, I., Reynolds, S.M., Miller, M., Bernard, B., Hinoue, T., Laird, P.W., Curtis, C., Shen, H., *et al.* (2014) Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*, **513**(7517): 202–209.

- Bates, D. and Maechler, M. (2016) *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version 1.2-7.1.
- Bateson, W. and Mendel, G. (1909) *Mendel's principles of heredity, by W. Bateson*. University Press, Cambridge [Eng.].
- Becker, K.F., Atkinson, M.J., Reich, U., Becker, I., Nekarda, H., Siewert, J.R., and Höfler, H. (1994) E-cadherin gene mutations provide clues to diffuse type gastric carcinomas. *Cancer Research*, **54**(14): 3845–3852.
- Bell, D., Berchuck, A., Birrer, M., Chien, J., Cramer, D., Dao, F., Dhir, R., DiSaia, P., Gabra, H., Glenn, P., *et al.* (2011) Integrated genomic analyses of ovarian carcinoma. *Nature*, **474**(7353): 609–615.
- Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)*, **57**(1): 289–300.
- Berx, G., Cleton-Jansen, A.M., Nollet, F., de Leeuw, W.J., van de Vijver, M., Cornelisse, C., and van Roy, F. (1995) E-cadherin is a tumour/invasion suppressor gene mutated in human lobular breast cancers. *EMBO J*, **14**(24): 6107–15.
- Berx, G., Cleton-Jansen, A.M., Strumane, K., de Leeuw, W.J., Nollet, F., van Roy, F., and Cornelisse, C. (1996) E-cadherin is inactivated in a majority of invasive human lobular breast cancers by truncation mutations throughout its extracellular domain. *Oncogene*, **13**(9): 1919–25.
- Berx, G. and van Roy, F. (2009) Involvement of members of the cadherin superfamily in cancer. *Cold Spring Harb Perspect Biol*, **1**: a003129.
- Bitler, B.G., Aird, K.M., Garipov, A., Li, H., Amatangelo, M., Kossenkov, A.V., Schultz, D.C., Liu, Q., Shih, Ie, M., Conejo-Garcia, J.R., *et al.* (2015) Synthetic lethality by targeting ezh2 methyltransferase activity in arid1a-mutated cancers. *Nat Med*, **21**(3): 231–8.
- Blake, J.A., Christie, K.R., Dolan, M.E., Drabkin, H.J., Hill, D.P., Ni, L., Sitnikov, D., Burgess, S., Buza, T., Gresham, C., *et al.* (2015) Gene Ontology Consortium: going forward. *Nucleic Acids Res*, **43**(Database issue): D1049–1056.
- Boettcher, M., Lawson, A., Ladenburger, V., Fredebohm, J., Wolf, J., Hoheisel, J.D., Frezza, C., and Shlomi, T. (2014) High throughput synthetic lethality screen reveals a tumorigenic role of adenylate cyclase in fumarate hydratase-deficient cancer cells. *BMC Genomics*, **15**: 158.
- Boone, C., Bussey, H., and Andrews, B.J. (2007) Exploring genetic interactions and networks with yeast. *Nat Rev Genet*, **8**(6): 437–49.
- Borgatti, S.P. (2005) Centrality and network flow. *Social Networks*, **27**(1): 55 – 71.

- Boucher, B. and Jenna, S. (2013) Genetic interaction networks: better understand to better predict. *Front Genet*, **4**: 290.
- Bozovic-Spasojevic, I., Azambuja, E., McCaskill-Stevens, W., Dinh, P., and Cardoso, F. (2012) Chemoprevention for breast cancer. *Cancer treatment reviews*, **38**(5): 329–339.
- Breiman, L. (2001) Random forests. *Machine Learning*, **45**(1): 5–32.
- Brin, S. and Page, L. (1998) The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, **30**(1): 107 – 117.
- Brouxhon, S.M., Kyrkanides, S., Teng, X., Athar, M., Ghazizadeh, S., Simon, M., O'Banion, M.K., and Ma, L. (2014) Soluble E-cadherin: a critical oncogene modulating receptor tyrosine kinases, MAPK and PI3K/Akt/mTOR signaling. *Oncogene*, **33**(2): 225–235.
- Bryant, H.E., Schultz, N., Thomas, H.D., Parker, K.M., Flower, D., Lopez, E., Kyle, S., Meuth, M., Curtin, N.J., and Helleday, T. (2005) Specific killing of *BRCA2*-deficient tumours with inhibitors of polyadprbose polymerase. *Nature*, **434**(7035): 913–7.
- Bussey, H., Andrews, B., and Boone, C. (2006) From worm genetic networks to complex human diseases. *Nat Genet*, **38**(8): 862–3.
- Butland, G., Babu, M., Diaz-Mejia, J.J., Bohdana, F., Phanse, S., Gold, B., Yang, W., Li, J., Gagarinova, A.G., Pogoutse, O., *et al.* (2008) esga: E. coli synthetic genetic array analysis. *Nat Methods*, **5**(9): 789–95.
- cBioPortal for Cancer Genomics (cBioPortal) (2017) cBioPortal for Cancer Genomics. <http://www.cbioportal.org/>. Accessed: 26/03/2017.
- Cerami, E.G., Gross, B.E., Demir, E., Rodchenkov, I., Babur, O., Anwar, N., Schultz, N., Bader, G.D., and Sander, C. (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res*, **39**(Database issue): D685–690.
- Chen, A., Beetham, H., Black, M.A., Priya, R., Telford, B.J., Guest, J., Wiggins, G.A.R., Godwin, T.D., Yap, A.S., and Guilford, P.J. (2014) E-cadherin loss alters cytoskeletal organization and adhesion in non-malignant breast cells but is insufficient to induce an epithelial-mesenchymal transition. *BMC Cancer*, **14**(1): 552.
- Chen, S. and Parmigiani, G. (2007) Meta-analysis of BRCA1 and BRCA2 penetrance. *J Clin Oncol*, **25**(11): 1329–1333.
- Chen, X. and Tompa, M. (2010) Comparative assessment of methods for aligning multiple genome sequences. *Nat Biotechnol*, **28**(6): 567–572.
- Chipman, K. and Singh, A. (2009) Predicting genetic interactions with random walks on biological networks. *BMC Bioinformatics*, **10**(1): 17.

- Christofori, G. and Semb, H. (1999) The role of the cell-adhesion molecule E-cadherin as a tumour-suppressor gene. *Trends in Biochemical Sciences*, **24**(2): 73 – 76.
- Ciriello, G., Gatz, 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 α -Dependent μ -Opioid Agonist-Mediated Adenylyl Cyclase Supersensitization. *Journal of Pharmacology and Experimental Therapeutics*, **310**(1): 215–222.
- Clough, E. and Barrett, T. (2016) The Gene Expression Omnibus Database. *Methods Mol Biol*, **1418**: 93–110.
- Collingridge, D.S. (2013) A primer on quantitized data analysis and permutation testing. *Journal of Mixed Methods Research*, **7**(1): 81–97.
- Collins, F.S. and Barker, A.D. (2007) Mapping the cancer genome. Pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies. *Sci Am*, **296**(3): 50–57.
- Collisson, E., Campbell, J., Brooks, A., Berger, A., Lee, W., Chmielecki, J., Beer, D., Cope, L., Creighton, C., Danilova, L., *et al.* (2014) Comprehensive molecular profiling of lung adenocarcinoma. *Nature*, **511**(7511): 543–550.
- Corcoran, R.B., Ebi, H., Turke, A.B., Coffee, E.M., Nishino, M., Cogdill, A.P., Brown, R.D., Della Pelle, P., Dias-Santagata, D., Hung, K.E., *et al.* (2012) Egfr-mediated re-activation of mapk signaling contributes to insensitivity of *BRAF*-mutant colorectal cancers to raf inhibition with vemurafenib. *Cancer Discovery*, **2**(3): 227–235.
- Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L., Toufighi, K., Mostafavi, S., *et al.* (2010) The genetic landscape of a cell. *Science*, **327**(5964): 425–31.
- Costanzo, M., Baryshnikova, A., Myers, C.L., Andrews, B., and Boone, C. (2011) Charting the genetic interaction map of a cell. *Curr Opin Biotechnol*, **22**(1): 66–74.
- Courtney, K.D., Corcoran, R.B., and Engelman, J.A. (2010) The PI3K pathway as drug target in human cancer. *J Clin Oncol*, **28**(6): 1075–1083.
- Creighton, C.J., Morgan, M., Gunaratne, P.H., Wheeler, D.A., Gibbs, R.A., Robertson, A., Chu, A., Beroukhi, 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 knowledge-base. *Nucleic Acids Res*, **42**(database issue): D472–D477.

- 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 Pax-tools. *PLoS Comput Biol*, **9**(9): e1003194.
- Deshpande, R., Asiedu, M.K., Klebig, M., Sutor, S., Kuzmin, E., Nelson, J., Piotrowski, J., Shin, S.H., Yoshida, M., Costanzo, M., *et al.* (2013) A comparative genomic approach for identifying synthetic lethal interactions in human cancer. *Cancer Res*, **73**(20): 6128–36.
- Dickson, D. (1999) Wellcome funds cancer database. *Nature*, **401**(6755): 729.
- Dienstmann, R. and Tabernero, J. (2011) *BRAF* as a target for cancer therapy. *Anti-cancer Agents Med Chem*, **11**(3): 285–95.
- Dijkstra, E.W. (1959) A note on two problems in connexion with graphs. *Numerische Mathematik*, **1**(1): 269–271.
- Dixon, S.J., Andrews, B.J., and Boone, C. (2009) Exploring the conservation of synthetic lethal genetic interaction networks. *Commun Integr Biol*, **2**(2): 78–81.
- Dixon, S.J., Fedyszyn, Y., Koh, J.L., Prasad, T.S., Chahwan, C., Chua, G., Toufighi, K., Baryshnikova, A., Hayles, J., Hoe, K.L., *et al.* (2008) Significant conservation of synthetic lethal genetic interaction networks between distantly related eukaryotes. *Proc Natl Acad Sci U S A*, **105**(43): 16653–8.

- Dong, L.L., Liu, L., Ma, C.H., Li, J.S., Du, C., Xu, S., Han, L.H., Li, L., and Wang, X.W. (2012) E-cadherin promotes proliferation of human ovarian cancer cells in vitro via activating MEK/ERK pathway. *Acta Pharmacol Sin*, **33**(6): 817–822.
- Dorogovtsev, S.N. and Mendes, J.F. (2003) *Evolution of networks: From biological nets to the Internet and WWW*. Oxford University Press, USA.
- Dorsam, R.T. and Gutkind, J.S. (2007) G-protein-coupled receptors and cancer. *Nat Rev Cancer*, **7**(2): 79–94.
- Erdős, P. and Rényi, A. (1959) On random graphs I. *Publ Math Debrecen*, **6**: 290–297.
- Erdős, P. and Rényi, A. (1960) On the evolution of random graphs. In *Publ. Math. Inst. Hung. Acad. Sci*, volume 5, 17–61.
- Eroles, P., Bosch, A., Perez-Fidalgo, J.A., and Lluch, A. (2012) Molecular biology in breast cancer: intrinsic subtypes and signaling pathways. *Cancer Treat Rev*, **38**(6): 698–707.
- Farmer, H., McCabe, N., Lord, C.J., Tutt, A.N., Johnson, D.A., Richardson, T.B., Santarosa, M., Dillon, K.J., Hickson, I., Knights, C., *et al.* (2005) Targeting the dna repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*, **434**(7035): 917–21.
- Fawcett, T. (2006) An introduction to ROC analysis. *Pattern Recognition Letters*, **27**(8): 861 – 874. {ROC} Analysis in Pattern Recognition.
- Fece de la Cruz, F., Gapp, B.V., and Nijman, S.M. (2015) Synthetic lethal vulnerabilities of cancer. *Annu Rev Pharmacol Toxicol*, **55**: 513–531.
- Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F. (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*, **136**(5): E359–386.
- Fisher, R.A. (1919) Xv.—the correlation between relatives on the supposition of mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, **52**(02): 399–433.
- Fong, P.C., Boss, D.S., Yap, T.A., Tutt, A., Wu, P., Mergui-Roelvink, M., Mortimer, P., Swaisland, H., Lau, A., O’Connor, M.J., *et al.* (2009) Inhibition of poly(adp-ribose) polymerase in tumors from BRCA mutation carriers. *N Engl J Med*, **361**(2): 123–34.
- Fong, P.C., Yap, T.A., Boss, D.S., Carden, C.P., Mergui-Roelvink, M., Gourley, C., De Greve, J., Lubinski, J., Shanley, S., Messiou, C., *et al.* (2010) Poly(adp)-ribose polymerase inhibition: frequent durable responses in BRCA carrier ovarian cancer correlating with platinum-free interval. *J Clin Oncol*, **28**(15): 2512–9.

- Forbes, S.A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S., *et al.* (2015) COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res*, **43**(Database issue): D805–811.
- Fraser, A. (2004) Towards full employment: using RNAi to find roles for the redundant. *Oncogene*, **23**(51): 8346–52.
- Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004) A census of human cancer genes. *Nat Rev Cancer*, **4**(3): 177–183.
- Futreal, P.A., Kasprzyk, A., Birney, E., Mullikin, J.C., Wooster, R., and Stratton, M.R. (2001) Cancer and genomics. *Nature*, **409**(6822): 850–852.
- Gao, B. and Roux, P.P. (2015) Translational control by oncogenic signaling pathways. *Biochimica et Biophysica Acta*, **1849**(7): 753–65.
- Gatza, M.L., Kung, H.N., Blackwell, K.L., Dewhirst, M.W., Marks, J.R., and Chi, J.T. (2011) Analysis of tumor environmental response and oncogenic pathway activation identifies distinct basal and luminal features in HER2-related breast tumor subtypes. *Breast Cancer Res*, **13**(3): R62.
- Gatza, M.L., Lucas, J.E., Barry, W.T., Kim, J.W., Wang, Q., Crawford, M.D., Datto, M.B., Kelley, M., Mathey-Prevot, B., Potti, A., *et al.* (2010) A pathway-based classification of human breast cancer. *Proc Natl Acad Sci USA*, **107**(15): 6994–6999.
- Gatza, M.L., Silva, G.O., Parker, J.S., Fan, C., and Perou, C.M. (2014) An integrated genomics approach identifies drivers of proliferation in luminal-subtype human breast cancer. *Nat Genet*, **46**(10): 1051–1059.
- Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*, **5**(10): R80.
- Genz, A. and Bretz, F. (2009) Computation of multivariate normal and t probabilities. In *Lecture Notes in Statistics*, volume 195. Springer-Verlag, Heidelberg.
- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2016) *mvtnorm: Multivariate Normal and t Distributions*. R package version 1.0-5. URL.
- 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.
- Güell, O., Sagués, F., and Serrano, M. (2014) Essential plasticity and redundancy of metabolism unveiled by synthetic lethality analysis. *PLoS Comput Biol*, **10**(5): e1003637.
- Guilford, P. (1999) E-cadherin downregulation in cancer: fuel on the fire? *Molecular Medicine Today*, **5**(4): 172 – 177.
- Guilford, P., Hopkins, J., Harraway, J., McLeod, M., McLeod, N., Harawira, P., Taite, H., Scoular, R., Miller, A., and Reeve, A.E. (1998) E-cadherin germline mutations in familial gastric cancer. *Nature*, **392**(6674): 402–5.
- Guilford, P., Humar, B., and Blair, V. (2010) Hereditary diffuse gastric cancer: translation of *CDH1* germline mutations into clinical practice. *Gastric Cancer*, **13**(1): 1–10.
- Guilford, P.J., Hopkins, J.B., Grady, W.M., Markowitz, S.D., Willis, J., Lynch, H., Rajput, A., Wiesner, G.L., Lindor, N.M., Burgart, L.J., *et al.* (1999) E-cadherin germline mutations define an inherited cancer syndrome dominated by diffuse gastric cancer. *Hum Mutat*, **14**(3): 249–55.
- Guo, J., Liu, H., and Zheng, J. (2016) SynLethDB: synthetic lethality database toward discovery of selective and sensitive anticancer drug targets. *Nucleic Acids Res*, **44**(D1): D1011–1017.
- Hajian-Tilaki, K. (2013) Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian J Intern Med*, **4**(2): 627–635.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I.H. (2009) The weka data mining software: an update. *SIGKDD Explor Newsl*, **11**(1): 10–18.
- Hammerman, P.S., Lawrence, M.S., Voet, D., Jing, R., Cibulskis, K., Sivachenko, A., Stojanov, P., McKenna, A., Lander, E.S., Gabriel, S., *et al.* (2012) Comprehensive genomic characterization of squamous cell lung cancers. *Nature*, **489**(7417): 519–525.
- Hanahan, D. and Weinberg, R.A. (2000) The hallmarks of cancer. *Cell*, **100**(1): 57–70.
- Hanahan, D. and Weinberg, R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**(5): 646–674.

- Hanna, S. (2003) Cancer incidence in new zealand (2003-2007). In D. Forman, D. Bray F Brewster, C. Gombe Mbalawa, B. Kohler, M. Piñeros, E. Steliarova-Foucher, R. Swaminathan, and J. Ferlay (editors), *Cancer Incidence in Five Continents*, volume X, 902–907. International Agency for Research on Cancer, Lyon, France. Electronic version <http://ci5.iarc.fr> Accessed 22/03/2017.
- Hansford, S., Kaurah, P., Li-Chang, H., Woo, M., Senz, J., Pinheiro, H., Schrader, K.A., Schaeffer, D.F., Shumansky, K., Zogopoulos, G., *et al.* (2015) Hereditary Diffuse Gastric Cancer Syndrome: CDH1 Mutations and Beyond. *JAMA Oncol*, **1**(1): 23–32.
- Heiskanen, M., Bian, X., Swan, D., and Basu, A. (2014) caArray microarray database in the cancer biomedical informatics gridTM (caBIGTM). *Cancer Research*, **67**(9 Supplement): 3712–3712.
- Heiskanen, M.A. and Aittokallio, T. (2012) Mining high-throughput screens for cancer drug targets-lessons from yeast chemical-genomic profiling and synthetic lethality. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, **2**(3): 263–272.
- Hell, P. (1976) Graphs with given neighbourhoods i. problèmes combinatoires at theorie des graphes. *Proc Coil Int CNRS, Orsay*, **260**: 219–223.
- Hillenmeyer, M.E. (2008) The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science*, **320**: 362–365.
- Hoadley, K.A., Yau, C., Wolf, D.M., Cherniack, A.D., Tamborero, D., Ng, S., Leiserson, M.D., Niu, B., McLellan, M.D., Uzunangelov, V., *et al.* (2014) Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*, **158**(4): 929–944.
- Hoehndorf, R., Hardy, N.W., Osumi-Sutherland, D., Tweedie, S., Schofield, P.N., and Gkoutos, G.V. (2013) Systematic analysis of experimental phenotype data reveals gene functions. *PLoS ONE*, **8**(4): e60847.
- Holm, S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**(2): 65–70.
- Holme, P. and Kim, B.J. (2002) Growing scale-free networks with tunable clustering. *Physical Review E*, **65**(2): 026107.
- Hopkins, A.L. (2008) Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol*, **4**(11): 682–690.
- Hu, Z., Fan, C., Oh, D.S., Marron, J.S., He, X., Qaqish, B.F., Livasy, C., Carey, L.A., Reynolds, E., Dressler, L., *et al.* (2006) The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics*, **7**: 96.

- Huang, E., Cheng, S., Dressman, H., Pittman, J., Tsou, M., Horng, C., Bild, A., Iversen, E., Liao, M., Chen, C., *et al.* (2003) Gene expression predictors of breast cancer outcomes. *Lancet*, **361**: 1590–1596.
- International HapMap 3 Consortium (HapMap) (2003) The International HapMap Project. *Nature*, **426**(6968): 789–796.
- Jeanes, A., Gottardi, C.J., and Yap, A.S. (2008) Cadherins and cancer: how does cadherin dysfunction promote tumor progression? *Oncogene*, **27**(55): 6920–6929.
- Jerby-Arnon, L., Pfetzer, N., Waldman, Y., McGarry, L., James, D., Shanks, E., Seashore-Ludlow, B., Weinstock, A., Geiger, T., Clemons, P., *et al.* (2014) Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. *Cell*, **158**(5): 1199–1209.
- Joachims, T. (1999) Making large-scale support vector machine learning practical. In S. Bernhard, I. Kropf, J.C.B. Christopher, and J.S. Alexander (editors), *Advances in kernel methods*, 169–184. MIT Press.
- Ju, Z., Liu, W., Roebuck, P.L., Siwak, D.R., Zhang, N., Lu, Y., Davies, M.A., Akbani, R., Weinstein, J.N., Mills, G.B., *et al.* (2015) Development of a robust classifier for quality control of reverse-phase protein arrays. *Bioinformatics*, **31**(6): 912.
- Kaelin, Jr, W. (2005) The concept of synthetic lethality in the context of anticancer therapy. *Nat Rev Cancer*, **5**(9): 689–98.
- Kaelin, Jr, W. (2009) Synthetic lethality: a framework for the development of wiser cancer therapeutics. *Genome Med*, **1**: 99.
- Kamada, T. and Kawai, S. (1989) An algorithm for drawing general undirected graphs. *Information Processing Letters*, **31**(1): 7–15.
- Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., *et al.* (2001) Functional annotation of a full-length mouse cDNA collection. *Nature*, **409**(6821): 685–690.
- Kelley, R. and Ideker, T. (2005) Systematic interpretation of genetic interactions using protein networks. *Nat Biotech*, **23**(5): 561–566.
- Kelly, S.T. (2013) *Statistical Predictions of Synthetic Lethal Interactions in Cancer*. Dissertation, University of Otago.
- Kelly, S.T., Single, A.B., Telford, B.J., Beetham, H.G., Godwin, T.D., Chen, A., Black, M.A., and Guilford, P.J. (unpublished) Towards HDGC chemoprevention: vulnerabilities in E-cadherin-negative cells identified by genome-wide interrogation of isogenic cell lines and whole tumors. Submitted to *Cancer Prev Res*.
- Kim, N.G., Koh, E., Chen, X., and Gumbiner, B.M. (2011) E-cadherin mediates contact inhibition of proliferation through Hippo signaling-pathway components. *Proc Natl Acad Sci USA*, **108**(29): 11930–11935.

- Kozlov, K.N., Gursky, V.V., Kulakovskiy, I.V., and Samsonova, M.G. (2015) Sequence-based model of gap gene regulation network. *BMC Genomics*, **15**(Suppl 12): S6.
- Kranthi, S., Rao, S., and Manimaran, P. (2013) Identification of synthetic lethal pairs in biological systems through network information centrality. *Mol BioSyst*, **9**(8): 2163–2167.
- Kroepil, F., Fluegen, G., Totikov, Z., Baldus, S.E., Vay, C., Schauer, M., Topp, S.A., Esch, J.S., Knoefel, W.T., and Stoecklein, N.H. (2012) Down-regulation of CDH1 is associated with expression of SNAIL in colorectal adenomas. *PLoS ONE*, **7**(9): e46665.
- Lander, E.S. (2011) Initial impact of the sequencing of the human genome. *Nature*, **470**(7333): 187–197.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**(6822): 860–921.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, **10**(3): R25.
- Latora, V. and Marchiori, M. (2001) Efficient behavior of small-world networks. *Phys Rev Lett*, **87**: 198701.
- Laufer, C., Fischer, B., Billmann, M., Huber, W., and Boutros, M. (2013) Mapping genetic interactions in human cancer cells with RNAi and multiparametric phenotyping. *Nat Methods*, **10**(5): 427–31.
- Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014) voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*, **15**(2): R29.
- Le Meur, N. and Gentleman, R. (2008) Modeling synthetic lethality. *Genome Biol*, **9**(9): R135.
- Le Meur, N., Jiang, Z., Liu, T., Mar, J., and Gentleman, R.C. (2014) Slgi: Synthetic lethal genetic interaction. r package version 1.26.0.
- Lee, A.Y., Perreault, R., Harel, S., Boulier, E.L., Suderman, M., Hallett, M., and Jenna, S. (2010a) Searching for signaling balance through the identification of genetic interactors of the rab guanine-nucleotide dissociation inhibitor gdi-1. *PLoS ONE*, **5**(5): e10624.
- Lee, I., Lehner, B., Vavouri, T., Shin, J., Fraser, A.G., and Marcotte, E.M. (2010b) Predicting genetic modifier loci using functional gene networks. *Genome Research*, **20**(8): 1143–1153.
- Lee, I. and Marcotte, E.M. (2009) Effects of functional bias on supervised learning of a gene network model. *Methods Mol Biol*, **541**: 463–75.

- Lee, M.J., Ye, A.S., Gardino, A.K., Heijink, A.M., Sorger, P.K., MacBeath, G., and Yaffe, M.B. (2012) Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell*, **149**(4): 780–94.
- Lehner, B., Crombie, C., Tischler, J., Fortunato, A., and Fraser, A.G. (2006) Systematic mapping of genetic interactions in *caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. *Nat Genet*, **38**(8): 896–903.
- Li, X.J., Mishra, S.K., Wu, M., Zhang, F., and Zheng, J. (2014) Syn-lethality: An integrative knowledge base of synthetic lethality towards discovery of selective anticancer therapies. *Biomed Res Int*, **2014**: 196034.
- Linehan, W.M., Spellman, P.T., Ricketts, C.J., Creighton, C.J., Fei, S.S., Davis, C., Wheeler, D.A., Murray, B.A., Schmidt, L., Vocke, C.D., *et al.* (2016) Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. *N Engl J Med*, **374**(2): 135–145.
- Lokody, I. (2014) Computational modelling: A computational crystal ball. *Nature Reviews Cancer*, **14**(10): 649–649.
- Lord, C.J., Tutt, A.N., and Ashworth, A. (2015) Synthetic lethality and cancer therapy: lessons learned from the development of PARP inhibitors. *Annu Rev Med*, **66**: 455–470.
- Lu, X., Kensche, P.R., Huynen, M.A., and Notebaart, R.A. (2013) Genome evolution predicts genetic interactions in protein complexes and reveals cancer drug targets. *Nat Commun*, **4**: 2124.
- Lu, X., Megchelenbrink, W., Notebaart, R.A., and Huynen, M.A. (2015) Predicting human genetic interactions from cancer genome evolution. *PLoS One*, **10**(5): e0125795.
- Lum, P.Y., Armour, C.D., Stepaniants, S.B., Cavet, G., Wolf, M.K., Butler, J.S., Hinshaw, J.C., Garnier, P., Prestwich, G.D., Leonardson, A., *et al.* (2004) Discovering modes of action for therapeutic compounds using a genome-wide screen of yeast heterozygotes. *Cell*, **116**(1): 121–137.
- Luo, J., Solimini, N.L., and Elledge, S.J. (2009) Principles of Cancer Therapy: Oncogene and Non-oncogene Addiction. *Cell*, **136**(5): 823–837.
- Machado, J., Olivera, C., Carvalh, R., Soares, P., Berx, G., Caldas, C., Sercuca, R., Carneiro, F., and Sorbrinho-Simoes, M. (2001) E-cadherin gene (*CDH1*) promoter methylation as the second hit in sporadic diffuse gastric carcinoma. *Oncogene*, **20**: 1525–1528.
- 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., Mastrogiannis, G.M., Olson, J.J., Mikkelsen, T., Lehman, N., Aldape, K., *et al.* (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, **455**(7216): 1061–1068.
- Miles, D.W. (2001) Update on HER-2 as a target for cancer therapy: herceptin in the clinical setting. *Breast Cancer Res*, **3**(6): 380–384.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*, **5**(7): 621–628.
- Muzny, D.M., Bainbridge, M.N., Chang, K., Dinh, H.H., Drummond, J.A., Fowler, G., Kovar, C.L., Lewis, L.R., Morgan, M.B., Newsham, I.F., *et al.* (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, **487**(7407): 330–337.
- Nagalla, S., Chou, J.W., Willingham, M.C., Ruiz, J., Vaughn, J.P., Dubey, P., Lash, T.L., Hamilton-Dutoit, S.J., Bergh, J., Sotiriou, C., *et al.* (2013) Interactions between immunity, proliferation and molecular subtype in breast cancer prognosis. *Genome Biol*, **14**(4): R34.
- Neeley, E.S., Kornblau, S.M., Coombes, K.R., and Baggerly, K.A. (2009) Variable slope normalization of reverse phase protein arrays. *Bioinformatics*, **25**(11): 1384.
- Novomestky, F. (2012) *matrixcalc: Collection of functions for matrix calculations*. R package version 1.0-3.
- Oliveira, C., Senz, J., Kaurah, P., Pinheiro, H., Sanges, R., Haegert, A., Corso, G., Schouten, J., Fitzgerald, R., Vogelsang, H., *et al.* (2009) Germline *CDH1* deletions in hereditary diffuse gastric cancer families. *Human Molecular Genetics*, **18**(9): 1545–1555.
- Oliveira, C., Seruca, R., Hoogerbrugge, N., Ligtenberg, M., and Carneiro, F. (2013) Clinical utility gene card for: Hereditary diffuse gastric cancer (HDGC). *Eur J Hum Genet*, **21**(8).
- Pandey, G., Zhang, B., Chang, A.N., Myers, C.L., Zhu, J., Kumar, V., and Schadt, E.E. (2010) An integrative multi-network and multi-classifier approach to predict genetic interactions. *PLoS Comput Biol*, **6**(9).

- Parker, J., Mullins, M., Cheung, M., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., *et al.* (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology*, **27**(8): 1160–1167.
- Pereira, B., Chin, S.F., Rueda, O.M., Vollan, H.K., Provenzano, E., Bardwell, H.A., Pugh, M., Jones, L., Russell, R., Sammut, S.J., *et al.* (2016) Erratum: The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. *Nat Commun*, **7**: 11908.
- Perou, C.M., Sørlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., *et al.* (2000) Molecular portraits of human breast tumours. *Nature*, **406**(6797): 747–752.
- Polyak, K. and Weinberg, R.A. (2009) Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nat Rev Cancer*, **9**(4): 265–73.
- Prahalad, A., Sun, C., Huang, S., Di Nicolantonio, F., Salazar, R., Zecchin, D., Beijersbergen, R.L., Bardelli, A., and Bernards, R. (2012) Unresponsiveness of colon cancer to *BRAF*(v600e) inhibition through feedback activation of egfr. *Nature*, **483**(7387): 100–3.
- R Core Team (2016) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. R version 3.3.2.
- Ravnan, M.C. and Matalaka, M.S. (2012) Vemurafenib in patients with *BRAF* v600e mutation-positive advanced melanoma. *Clin Ther*, **34**(7): 1474–86.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, **43**(7): e47.
- Robinson, M.D. and Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*, **11**(3): R25.
- Roguev, A., Bandyopadhyay, S., Zofall, M., Zhang, K., Fischer, T., Collins, S.R., Qu, H., Shales, M., Park, H.O., Hayles, J., *et al.* (2008) Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science*, **322**(5900): 405–10.
- 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.
- Rustici, G., Kolesnikov, N., Brandizi, M., Burdett, T., Dylag, M., Emam, I., Farne, A., Hastings, E., Ison, J., Keays, M., *et al.* (2013) ArrayExpress update—trends in database growth and links to data analysis tools. *Nucleic Acids Res*, **41**(Database issue): D987–990.

- Ryan, C., Lord, C., and Ashworth, A. (2014) Daisy: Picking synthetic lethals from cancer genomes. *Cancer Cell*, **26**(3): 306–308.
- 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.
- Sørli, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., *et al.* (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA*, **98**(19): 10869–10874.
- Stajich, J.E. and Lapp, H. (2006) Open source tools and toolkits for bioinformatics: significance, and where are we? *Brief Bioinformatics*, **7**(3): 287–296.
- Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009) The cancer genome. *Nature*, **458**(7239): 719–724.
- Ström, C. and Helleday, T. (2012) Strategies for the use of poly(adenosine diphosphate ribose) polymerase (parp) inhibitors in cancer therapy. *Biomolecules*, **2**(4): 635–649.
- Sun, C., Wang, L., Huang, S., Heynen, G.J.J.E., Prahallad, A., Robert, C., Haanen, J., Blank, C., Wesseling, J., Willems, S.M., *et al.* (2014) Reversible and adaptive resistance to *BRAF*(v600e) inhibition in melanoma. *Nature*, **508**(7494): 118–122.
- Telford, B.J., Chen, A., Beetham, H., Frick, J., Brew, T.P., Gould, C.M., Single, A., Godwin, T., Simpson, K.J., and Guilford, P. (2015) Synthetic lethal screens identify vulnerabilities in gpcr signalling and cytoskeletal organization in E-cadherin-deficient cells. *Mol Cancer Ther*, **14**(5): 1213–1223.
- The 1000 Genomes Project Consortium (1000 Genomes) (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**(7319): 1061–1073.
- The Cancer Genome Atlas Research Network (TCGA) (2012) Comprehensive molecular portraits of human breast tumours. *Nature*, **490**(7418): 61–70.
- The Cancer Genome Atlas Research Network (TCGA) (2017) The Cancer Genome Atlas Project. <https://cancergenome.nih.gov/>. Accessed: 26/03/2017.

- The Catalogue Of Somatic Mutations In Cancer (COSMIC) (2016) Cosmic: The catalogue of somatic mutations in cancer. <http://cancer.sanger.ac.uk/cosmic>. Release 79 (23/08/2016), Accessed: 05/02/2017.
- The Comprehensive R Archive Network (CRAN) (2017) Cran. <https://cran.r-project.org/>. Accessed: 24/03/2017.
- The ENCODE Project Consortium (ENCODE) (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*, **306**(5696): 636–640.
- The National Cancer Institute (NCI) (2015) The genetics of cancer. <https://www.cancer.gov/about-cancer/causes-prevention/genetics>. Published: 22/04/2015, Accessed: 22/03/2017.
- The New Zealand eScience Infrastructure (NeSI) (2017) NeSI. <https://www.nesi.org.nz/>. Accessed: 25/03/2017.
- Tierney, L., Rossini, A.J., Li, N., and Sevcikova, H. (2015) *snow: Simple Network of Workstations*. R package version 0.4-2.
- Tiong, K.L., Chang, K.C., Yeh, K.T., Liu, T.Y., Wu, J.H., Hsieh, P.H., Lin, S.H., Lai, W.Y., Hsu, Y.C., Chen, J.Y., *et al.* (2014) Csnk1e/ctnnb1 are synthetic lethal to tp53 in colorectal cancer and are markers for prognosis. *Neoplasia*, **16**(5): 441–50.
- Tischler, J., Lehner, B., and Fraser, A.G. (2008) Evolutionary plasticity of genetic interaction networks. *Nat Genet*, **40**(4): 390–391.
- Tomasetti, C. and Vogelstein, B. (2015) Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science*, **347**(6217): 78–81.
- Tong, A.H., Evangelista, M., Parsons, A.B., Xu, H., Bader, G.D., Page, N., Robinson, M., Raghibizadeh, S., Hogue, C.W., Bussey, H., *et al.* (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science*, **294**(5550): 2364–8.
- Tong, A.H., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G.F., Brost, R.L., Chang, M., *et al.* (2004) Global mapping of the yeast genetic interaction network. *Science*, **303**(5659): 808–13.
- 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.
- Tsai, H.C., Li, H., Van Neste, L., Cai, Y., Robert, C., Rassool, F.V., Shin, J.J., Harbom, K.M., Beaty, R., Pappou, E., *et al.* (2012) Transient low doses of dna-demethylating agents exert durable antitumor effects on hematological and epithelial tumor cells. *Cancer Cell*, **21**(3): 430–46.

- Tunggal, J.A., Helfrich, I., Schmitz, A., Schwarz, H., Gunzel, D., Fromm, M., Kemler, R., Krieg, T., and Niessen, C.M. (2005) E-cadherin is essential for in vivo epidermal barrier function by regulating tight junctions. *EMBO J*, **24**(6): 1146–1156.
- Tutt, A., Robson, M., Garber, J.E., Domchek, S.M., Audeh, M.W., Weitzel, J.N., Friedlander, M., Arun, B., Loman, N., Schmutzler, R.K., *et al.* (2010) Oral poly(adenosine diphosphate) polymerase inhibitor olaparib in patients with *BRCA1* or *BRCA2* mutations and advanced breast cancer: a proof-of-concept trial. *Lancet*, **376**(9737): 235–44.
- van der Meer, R., Song, H.Y., Park, S.H., Abdulkadir, S.A., and Roh, M. (2014) RNAi screen identifies a synthetic lethal interaction between PIM1 overexpression and PLK1 inhibition. *Clinical Cancer Research*, **20**(12): 3211–3221.
- van der Post, R.S., Vogelaar, I.P., Carneiro, F., Guilford, P., Huntsman, D., Hoogerbrugge, N., Caldas, C., Schreiber, K.E., Hardwick, R.H., Ausems, M.G., *et al.* (2015) Hereditary diffuse gastric cancer: updated clinical guidelines with an emphasis on germline CDH1 mutation carriers. *J Med Genet*, **52**(6): 361–374.
- van Steen, K. (2012) Travelling the world of gene–gene 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.
- 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 portal—a 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.