

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Quantifying and Accommodating Heterogeneity in High Throughput Drug Response Studies

Permalink

<https://escholarship.org/uc/item/6w65k84f>

Author

Ding, Kuan-Fu

Publication Date

2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

**Quantifying and Accommodating Heterogeneity in High Throughput
Drug Response Studies**

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Bioinformatics and Systems Biology

by

Kuan-Fu Ding

Committee in charge:

Professor Nicholas J. Schork, Chair
Professor Bing Ren, Co-Chair
Professor Hannah Carter
Professor Terry Gaasterland
Professor Olivier Harismendy

2017

Copyright
Kuan-Fu Ding, 2017
All rights reserved.

The dissertation of Kuan-Fu Ding is approved, and it is acceptable in quality and form for publication on micro-film and electronically:

Co-Chair

Chair

University of California, San Diego

2017

DEDICATION

To my family and friends

To my parents

To my wife and son

You made this work possible

Thank you for your unwavering support

EPIGRAPH

Statistics is the grammar of science.

— K. Pearson

...the null hypothesis is never proved or established, but is possibly disproved, in the course of experimentation. Every experiment may be said to exist only to give the facts a chance of disproving the null hypothesis.

— R.A. Fisher

No problem can be solved from the same level of consciousness that created it.

— A. Einstein

The whole is greater than the sum of its parts.

— Aristotle

TABLE OF CONTENTS

Signature Page	iii
Dedication	iv
Epigraph	v
Table of Contents	vi
List of Figures	x
List of Tables	xvi
Acknowledgements	xviii
Vita and Publications	xxi
Abstract of the Dissertation	xxii
Chapter 1	
Preface	1
1.1 Background	1
1.2 Exposing and Accommodating Issues with Heterogeneity in Drug Screens	2
1.2.1 Reproducibility	2
1.2.2 Association Analysis Involving Dose-Response Data	3
1.2.3 Network Rewiring and Target Identification and Modulation	4
1.3 Detailed Descriptions of the Data Sets Used	5
1.4 Synopsis and Discussion	7
Chapter 2	
Analysis of Variability in High Throughput Screening Data: Applications to Melanoma Cell Lines and Drug Responses . .	9
2.1 Abstract	9
2.2 Introduction	10
2.3 Results	14
2.3.1 Variation of Cell Viability Data Across Dose and Drug Replicates	14
2.3.2 Variation Between Two Laboratories	15
2.3.3 Comprehensive Analysis Considering the Entire HTS Experimental Setting via ANOVA Modeling	18
2.3.4 Exploring Interaction Effects	21
2.3.5 Re-analysis With Broader Set of Comparator Drugs and Cell Lines	22

	2.3.6	Comparing the Results of Models that Accommodate Plate-Specific Effects to Those That Do not	23
	2.3.7	Analysis of CCLE and GDSC Data	24
2.4		Discussion	24
2.5		Materials and Methods	28
	2.5.1	Data	28
	2.5.2	Data Analysis	29
2.6		Acknowledgements	31
2.7		Supplemental Figures	33
Chapter 3		An Additional Example of Quality Control Analyses for Drug Screening and Drug Target Validation Involving Longevity in Model Species	44
	3.1	Abstract	44
	3.2	Introduction	45
	3.3	Materials and Methods	46
	3.3.1	<i>Caenorhabditis elegans</i> Data Acquisition and Analysis	46
	3.4	Results	48
	3.4.1	<i>Caenorhabditis elegans</i> Screening Study	48
	3.5	Discussion	52
	3.6	Acknowledgements	53
Chapter 4		Nonlinear Mixed Effects Models for Use in High Throughput Drug Screens: Applications to Melanoma Cell Line Analysis	54
	4.1	Abstract	54
	4.2	Introduction	55
	4.3	Materials and Methods	58
	4.3.1	Cell Lines	59
	4.3.2	Drug Screening	59
	4.3.3	Gene Expression Assays	60
	4.3.4	Phosphoproteomic Assays	60
	4.3.5	CCLE Data Set	62
	4.3.6	Traditional IC_{50} DRC Analysis	62
	4.3.7	NLME-Based Analyses	63
	4.3.8	Simulation Based Tests	65
	4.4	Results	66
	4.4.1	Simulation Based Tests	66
	4.4.2	Comparing the Different Methods on the Gene Expression Data	68
	4.4.3	Comparing the Different Methods on the RPPA Data	70
	4.4.4	Assessing the Omnibus LR NLME Test	70

4.4.5	Assessing the Consistency Between the RPPA and Gene Expression Analyses	72
4.4.6	Analysis of the CCLE data with NLME-Based Models	74
4.4.7	Including Mutation Status as Random Effect in CCLE Data	74
4.5	Discussion	75
4.6	Acknowledgements	78
4.7	Supplemental Figures and Tables	79
Chapter 5	Network Rewiring in Cancer: Applications to Melanoma Cell Lines and The Cancer Genome Atlas Melanoma Patients . . .	92
5.1	Abstract	92
5.2	Introduction	93
5.3	Materials and Methods	97
5.3.1	Transcriptional Profile on Cell Lines	97
5.3.2	MAPK Pathway Information	98
5.3.3	Unsupervised Network Rewiring Analysis	99
5.3.4	Assessing Statistical Significance	99
5.3.5	<i>A Posteriori</i> Supervised Analyses	100
5.3.6	Exploring Network Architecture	100
5.3.7	Analyses of the CCLE Melanoma Cell Line Expression Data	101
5.3.8	Analyses of the TCGA Melanoma Tumor Gene Expression Data	102
5.3.9	Network Reconstruction Analyses	102
5.4	Results	103
5.4.1	Unsupervised Multivariate Mixture Model Analysis of the Transcriptomes of Melanoma Cell Lines	103
5.4.2	Does Mutation Status Explain the Clustering? . .	104
5.4.3	Degree Centrality Analysis	107
5.4.4	Are the Networks Equivalent for Mutation Status Defined Groups Against Normal Controls?	108
5.4.5	BRAF V600E+/- Differentially Expressed Genes Analysis with the Melanoma TCGA Tumor Data	112
5.4.6	Network Centrality Tests for the TGen Melanoma Cell Lines	113
5.4.7	Network Centrality Tests for the TCGA Data . .	113
5.4.8	Are Strong Edges Present in Both Mutant and Wild Type Groups?	114
5.4.9	<i>De Novo</i> Network Reconstruction Analysis	117
5.5	Discussion	117
5.6	Acknowledgements	121

5.7	Supplemental Figures	122
Chapter 6	Conclusion, Limitations and Future Research	143
6.1	Synopsis	143
6.1.1	Reproducibility and Reliability of Drug Screen and Drug Target Data	143
6.1.2	Analysis of Dose Response Data	144
6.1.3	Exploring Evidence for Re-Wiring in Cell-Lines Used in Drug Screens	144
6.2	Extensions Beyond Current Experimental Designs and Methodologies	145
6.3	Informing Clinical Trials and Patient Data Analyses . . .	146
6.4	Closing Remarks	149
	Bibliography	150

LIST OF FIGURES

Figure 2.1:	Differences in HTS plating schemes. Nine 384-well plates are used for each cell line. Left: TGen/SBP plating scheme. Each plate consists of triplicates of three doses across 120 drugs and 8 DMSOs. Right: Alternate plating scheme with each plate . . .	16
Figure 2.2:	A. Pairwise scatterplots across nine drugs comparing SBP 3-point screen against same three concentrations from SBP 9-point screen. Colors indicate concentration: 0.1 (green), 1.0 (blue), and 10.0 μ M (purple). B. Correlation coefficients from . . .	17
Figure 2.3:	A: Application of flexible linear models, using ANOVA methods, to explore the variation in CV that is explained by site, cell lines, dose, drug, and plates. Sites explained a small proportion of the variation, whereas drugs, dose, and cell lines explained	20
Figure 2.4:	ANOVA analysis on IC_{50} values from SBP, TGen, GDSC, and CCLE. Site effects were generally minimal; however, site specific effects were observed in Nilotinib.	25
Figure 2.5:	Pairwise scatterplots across all 128 drugs comparing SBP 3-point screen against same three concentrations from SBP 9-point screen. Colors indicate concentration: 0.1 (green), 1.0 (blue), and 10.0 μ M (purple).	33
Figure 2.6:	Pairwise analysis of mean Cell Viability across 46 drugs exhibiting at least one CV less than 20%. Gradient indicates concentration from 0.02 μ M (light) to 10 μ M (dark).	34
Figure 2.7:	Application of flexible linear models, using ANOVA methods, to explore the variation in CV that is explained by site, cell lines, dose, drug, and plates. Sites explained a small proportion of the variation, whereas drugs, dose, and cell lines explained a . . .	35
Figure 2.8:	Assessing the variation in CV that is explained by site, cell lines, dose, drug, and drug-site interaction. To further assess the reproducibility of HTS, we examined the drug-site interaction terms and found that 19 of the significant drugs did not have a . . .	36

Figure 2.9: Assessing the variation in CV that is explained by site, cell lines, dose, drug, and drug-dose interaction. We find that many of the significant drug-dose interaction terms were from higher concentrations, indicating that a majority of the variation signi-	37
Figure 2.10: Assessing the variation in CV that is explained by site, cell lines, dose, drug, and 3-way interaction terms. Site-by-drug-by-dose effects explained 2.49% of cell viability variance.	38
Figure 2.11: Assessing the variation in CV that is explained by site, cell lines, dose, drug, and 3-way interaction terms. Drug-by-dose-by-cell line explained 3.6% of the variance.	39
Figure 2.12: Re-assessing the variation in CV that is explained by site, cell lines, dose, and drug when using AIC-defined control drugs. Results were similar from using random baseline drug in terms of proportion of variation explained by sites, cell lines, dose, . .	40
Figure 2.13: Re-assessing the variation in CV that is explained by site, cell lines, dose, drug, and dose-drug interaction when using AIC-defined control drugs.	41
Figure 2.14: Re-assessing the variation in CV that is explained by site, cell lines, dose, drug, and site-dose-drug interaction when using AIC-defined control drugs.	42
Figure 2.15: Re-assessing the variation in CV that is explained by site, cell lines, dose, drug, and cell line-dose-drug interaction when using AIC-defined control drugs.	43
Figure 3.1: Plot of $-\log_{10}$ p-values for each experimental factor against the observed fraction of <i>C. elegans</i> alive. Vertical bars represent each group of factors analyzed with ANOVA.	49
Figure 3.2: Plot of $-\log_{10}$ p-values for each experimental factor, including column $\times$ row interaction terms, against the observed fraction of <i>C. elegans</i> alive. Vertical bars represent each group of factors analyzed with ANOVA.	50
Figure 4.1: (A) Experimental design with melanoma cell lines processed with HTS, whole genome microarray genechip, and RPPA. We performed simulations under the HTS under the null model — no influence from gene expression (GEX) (B) and under the . .	61

Figure 4.2:	Differences observed between using traditionally called IC_{50} , NLME called IC_{50} , and LLR tests. (A) $-\log_{10}$ p-values for associations between probe set expressions and traditionally called IC_{50} or NLME called IC_{50} . (B) $-\log_{10}$ p-values for associations	71
Figure 4.3:	LLR tests for assessing significant gene expression association in BRAF-related genes. (A) Heatmap of p-values across 15 drugs. (B) Dose response curves fit within KIDINS220 over-expression (Green) and under-expression (Grey). (C) Dose response curves	73
Figure 4.4:	LLR tests for assessing significant stratification of cell lines with or without mutations in CCLE and SU2C cell lines. (A) Assessment of AZD6244 in CCLE across all cell lines (Black), melanoma BRAF+ (Solid Blue), and melanoma BRAF- (Dotted	76
Figure 4.5:	GeneMANIA BRAF-related network with $K = 20$. Key: Pink = Physical interactions; Purple = Co-expression; Orange = Predicted; Light Blue = Pathway; Dark Blue = Co-localization; Green = Genetic interactions; Yellow = Shared protein domain;	79
Figure 4.6:	Power analysis for simulations. Based on the observed correlations from initial parameter simulations ($\text{cor} > .5$), 100 simulations should suffice.	80
Figure 4.7:	Simulation-based power does improve with increased sample size in both the NLME called IC_{50} associations and in the LLR associations. Statistical power in traditionally called IC_{50} associations tests yielded the worst results.	81
Figure 4.8:	Application of LLR tests to identify concordant RPPA (Phospho-PKC α/β II (Thr638)) and Gene expression (KIDINS220) that stratify Daunorubicin dose response curves.	82
Figure 4.9:	Application of LLR tests to identify concordant RPPA (anti-PKC α (Ser657)) and Gene expression (KIDINS220) that stratify Daunorubicin dose response curves.	83
Figure 4.10:	Application of LLR tests to identify concordant RPPA (pBAD) and Gene expression (MAPK3) that stratify Cladribine dose response curves.	84
Figure 4.11:	GeneMANIA pathway analysis using BRAF and BCL2 showing the network connection of MAPK3 and pBAD.	85

Figure 4.12: Application of LLR tests to identify concordant RPPA (FKHR) and Gene expression (KSR1) that stratify LDK378 dose response curves.	86
Figure 4.13: GeneMANIA pathway analysis using BRAF and BCL2 showing the network connection of FOXO1 (FKHR) and KSR1.	87
Figure 4.14: Comparing the effects of estimating IC_{50} values (uM) with NLME models and those of traditional methods. Although overall correlation is observed, there are some clear variations.	88
Figure 5.1: (A) Example of network rewiring that could occur. Under ideal conditions, Gene A is co-regulated with Gene D through gene B. When gene B is inhibited, there could be a compensatory mechanism for which gene A communicates with gene D — in .	96
Figure 5.2: Pairwise correlation between MAPK Pathway genes from the identified clusters using flexmix were used to generate heatmaps. Upper right hand half of matrix were correlation coefficients from cluster 1, whereas lower left hand half of matrix were . . .	106
Figure 5.3: Density plots of degree centrality parameters. Figures (A)-(D) are based on analysis of SU2C melanoma cell lines and figures (E)-(H) are based on analysis of CCLE melanoma cell lines. The plots are densities of centrality for (A) and (E), closeness .	109
Figure 5.4: Differential pairwise correlation of (A) BRAF associated genes differentiated by BRAF mutation status and (B) KRAS associated genes differentiated by KRAS mutation status.	111
Figure 5.5: Network centrality statistics identifying differential central nodes between BRAF+ and BRAF- patients from TCGA data. (A) Degree Centrality, (B) Closeness Centrality, (C) Betweenness Centrality, and (D) Eigenvector Centrality can all be used to . .	115
Figure 5.6: WGCNA derived modules used for <i>a priori</i> assessment of network rewiring. (A) Beanplots of overlapping modules between melanoma cell line BRAF+/- (red) and random permutation-based (gray) overlap (B) Beanplots of overlapping modules be	118
Figure 5.7: Boxplot of normalized gene expression profiles across all melanoma cell lines and skin fibroblast samples.	122

Figure 5.8: Boxplot of normalized gene expression profiles within only melanoma cell lines.	125
Figure 5.9: Density plots for the posterior probability of the flexmix groups in (A) SU2C melanoma cell lines and (B) CCLE melanoma cell lines. These plots reveal that the flexmix called clusters were most distinct in the SU2C melanoma cell lines.	126
Figure 5.10: Circular network graph of all common nodes. Top network represents the network derived from correlation associations found in cluster one, whereas the bottom network represents the edges present in cluster two. Density of the edges highlights the dif- .	127
Figure 5.11: Pairwise correlation between MAPK Pathway genes from the identified clusters in TCGA melanoma patients using flexmix were used to generate heatmaps. Upper right hand half of matrix were correlation coefficients from cluster 1, whereas lower .	128
Figure 5.12: Pairwise correlation between MAPK Pathway genes from the identified clusters in all CCLE cell lines using flexmix were used to generate heatmaps. Upper right hand half of matrix were correlation coefficients from cluster 1, whereas lower left hand .	129
Figure 5.13: QQ-plots for p-values based on association tests for gene copy number variation observed against the groups discovered in the CCLE dataset, using either the melanoma subset (top) or against all cell lines (bottom).	130
Figure 5.14: Observed p-values based on correlation strength tests. Correlations were compared between those derived from TCGA melanoma patients with BRAFv600e mutation against those derived from BRAF wt. Blue points represent edges significant	131
Figure 5.15: Density plot of node centrality statistics for flexmix clusters called on TCGA melanoma patients. We use four different degree centrality parameters (A) Centrality, (B) Closeness, (C) Betweenness, and (D) Eigenvector Centrality.	132
Figure 5.16: Density plot of node centrality statistics for flexmix clusters called across all CCLE cell lines. We use four different degree centrality parameters (A) Centrality, (B) Closeness, (C) Betweenness, and (D) Eigenvector Centrality.	133

Figure 5.17: Comparing node centrality statistics for BRAF+ and BRAF- from the SU2C cell lines. We use four different degree centrality parameters (A) Centrality, (B) Closeness, (C) Betweenness, and (D) Eigenvector Centrality.	134
Figure 5.18: Comparing node centrality statistics for KRAS+ and KRAS- from the SU2C cell lines. We use four different degree centrality parameters (A) Centrality, (B) Closeness, (C) Betweenness, and (D) Eigenvector Centrality.	135
Figure 5.19: Comparing node centrality statistics for BRAFv600e+ and BRAF- from the TCGA melanoma patients. We use four different degree centrality parameters (A) Centrality, (B) Closeness, (C) Betweenness, and (D) Eigenvector Centrality.	136
Figure 5.20: Network derived from edges present in TCGA BRAF+ patients.	137
Figure 5.21: Network derived from edges present in TCGA BRAFv600e+ patients.	138
Figure 5.22: Network derived from edges present in TCGA BRAF- patients.	139
Figure 5.23: WGCNA derived modules (with $k = 20$) used for <i>a priori</i> assessment of network rewiring. (A) Beanplots of overlapping modules between melanoma cell line BRAF+/- (red) and random permutation-based (gray) overlap (B) Beanplots of overlapping	140
Figure 5.24: Density plots of the observed pairwise correlation coefficients when comparing the BRAF+/- subgroup using the BRAF-associated gene network and the KRAS+/- subgroup using the KRAS-associated gene network.	141
Figure 5.25: GeneMANIA circular view of all genes in Supplemental Table 2, i.e., genes with strong interactions identified via co-expression analysis.	142
Figure 6.1: Flowchart of clinical trial study pursued through SU2C. The overarching goal is to incorporate algorithms developed through analysis of <i>in vitro</i> and other organisms.	148

LIST OF TABLES

Table 2.1:	Key: CLs = Cell Lines; MCLs = Melanoma Cell Lines; OCLs = Other Cell Lines; CCLs = Melanoma cell lines in common with the SBP nine-point data set; Common drugs = drugs in common with the SBP nine-point data set; SBPMDI = Sanford	15
Table 2.2:	Percentage of variance explained by the experimental factors. . .	19
Table 3.1:	Assessing the proportion of variation in experimental data explained by each of the experimental factors, including column x row interaction terms. Significance codes for F statistics: p-val < .05 ('*'), p-val < .01 ('**'), p-val < .001 ('***')	49
Table 3.2:	Assessing the proportion of variation in experimental data explained by each of the experimental factors, drugs, and days (including drug x day interaction terms). Significance codes for F statistics: p-val < .05 ('*'), p-val < .01 ('**'), p-val < .001 . .	51
Table 3.3:	Assessing the proportion of variation in experimental data explained by each of the experimental factors, pharmacological classes, and days (including pharmacological class x day interaction terms). Significance codes for F statistics: p-val < .05 . . .	52
Table 4.1:	The Mean and SD of power observed when a single parameter is fixed between two groups and all other parameters are varied in one of the groups (see Methods section).	67
Table 4.2:	Simulation-based power to detect differences between DRCs between two groups assuming different features about the groups DRCs.	69
Table 4.3:	Table of cell lines available in the SU2C study. HTS represents availability of high throughput drug screen, GEX represents availability of whole genome microarray genechips, and RPPA sets 1 and 2 represent the availability of Reverse Phase Protein .	89
Table 4.4:	Total number of significant associations based on Traditionally called IC_{50} and NLME called IC_{50} vales.	90

Table 4.5:	P-values for LLR tests stratifying gene expression at the median value. Tests were performed across 15 drugs and BRAF-related genes.	91
Table 5.1:	Comparing <i>a priori</i> called clusters within SU2C melanoma cell lines (MCL), CCLE MCL, CCLE all cell lines (ACL), and TCGA melanoma patients. Performed mean vector and covariance equality tests within groups and used log-likelihood ratio permutation	105
Table 5.2:	Table of association test results between <i>a priori</i> called clusters and common factors in melanoma. We test against available mutation status TCGA melanoma patients, SU2C melanoma cell lines, and CCLE melanoma cell lines.	107
Table 5.3:	Network node centrality statistics identified central nodes for each parameter. Consensus-based approach was used to genes to match to DGIdb to identify candidate therapeutic agents. . .	108
Table 5.4:	Table of correlation equality tests for the genes in the specified networks. Tests are between mutant and wild type samples. Asterisks represent at least one significant test within comparison (Stieger or Jennrich).	112
Table 5.5:	Assessing significant nodes by degree centrality parameters. A consensus-based approach was used to genes to match to DGIdb to identify candidate therapeutic agents.	114
Table 5.6:	A comparison between degree centrality parameters through correlations. DC = Degree Centrality, BW = Betweenness, CL = Closeness, EC = Eigenvector Centrality.	116
Table 5.7:	Correlation equality tests between WGCNA defined modules in BRAF+/- SU2C cell lines, BRAF+/- TCGA melanoma patients, and BRAFv600e+/- TCGA melanoma patients.	119
Table 5.8:	List of skin fibroblast samples obtained from GEO. Sample repository, sample name, and associated meta data is provided.	123
Table 5.9:	Identifying edges within BRAS-associated gene networks and KRAS-associated gene networks using co-expression observed within each subgroup. Edges reported had a correlation coefficient less than -.7 or greater than .7.	124

ACKNOWLEDGEMENTS

First and foremost, I must thank my advisor, Dr. Nicholas Schork. His guidance and tutelage have enabled me to become the researcher that I am. His unparalleled multi-disciplinary knowledge and unbelievable ability to distill any complex matter into a coherent thought never cease to amaze me. Yet, what I am most grateful for is his patience, kindness, thoughtfulness, and unceasing desire to impart knowledge. I cannot thank him enough for his support throughout my academic studies.

Thank you to the Schork lab for helping me work through my challenges, letting me rant and joining me in this adventure. As a team, I believe that we have made a lasting impact.

I owe a great deal of gratitude towards Professors Jeffrey M. Trent and Patricia LoRusso for allowing me to analyze data from your studies. Thank you Translational Genomics Research Institute and Sanford Burnham Prebys Medical Discovery Institute for your dedication to research. Thank you to Richard Miller and the Longevity Consortium for including me and Professor Schork's team in the analysis of your data. Without your resources and data, my work would not have been possible.

My introduction to bioinformatics is in great part due to Timothy Gallagher, Professor Jae K. Lee, and Professor Dan Theoderescu. Thanks for allowing me the opportunity to learn from you. Professor Lee, I cannot thank you enough for your support, for showing me how powerful statistics can be, and for revealing to me the potential impact mathematics and computational approaches have in healthcare. Tim, thank you for your leadership and exceptional vision. Dan, your devotion and zeal towards learning in medicine are inspiring. Thank you three for being patterns of excellence.

Thank you Professor Brett Blackman, Professor Brian Wamhoff, and James Powers for allowing me to work with you and contribute to your company. Thank you for showing me the impact of biomedical sciences in the healthcare and pharmaceutical industries. I am amazed by your work and fully expect continued success. May you continue to revolutionize the way the pharmaceutical industry

views *in vitro* and *ex vivo* models.

I also would like to thank my Doctoral Committee of Professors Bing Ren, Hannah Carter, Terry Gaasterland, and Olivier Harismendy. Thank you for your incredible support and attentiveness to the details in my research. I am very grateful to each one of you — especially due to the multi-disciplinary nature of my research. Without your expertise, the results of my research would have been much harder to achieve. Special thanks to Professors Ren and Gaasterland for your invaluable advice. I truly appreciate your guidance throughout my academic career and even in discussion regarding the next steps in my career.

To my parents, Zan and Phoebe, thank you for your love and care. Thank you for always encouraging me to pursue my interests, no matter how challenging they seemed. Your support has been unwavering and necessary. To my sisters, Jan and Yen, thank you for constant laughter and silliness. To my mother-in-law and sister-in-law, thank you for being great cheerleaders and for all that you are to our family.

To my wife and son: you are my love, happiness, and joy. I do what I do for you and I am inspired to make a difference because of you.

Thank you to all who have prayed for me — your prayers have sustained me on this journey. Praise God for His mercy and grace that have brought me to this day. To Him be the glory forever and ever. Amen.

Chapter 2 is a full reprint of the publication: **Ding, K.**, Finlay, D., Yin, H., Hendricks, W.P.D., Sereduk, C., Kiefer, J., Sekulic, A., LoRusso, P.M., Vuori, K., Trent, J.M., and Schork, N.J. Analysis of Variability in High Throughput Screening Data: Applications to Melanoma Cell Lines and Drug Responses. *Oncotarget*. 2017; 8:27786-27799, with permission from all coauthors. The dissertation author was the primary researcher and author of this material.

Chapter 4 is a full reprint of the publication: **Ding, K.**, Petricoin, E.F., Finlay, D., Yin, H., Hendricks, W.P.D., Sereduk, C., Kiefer, J., Sekulic, A., LoRusso, P.M., Vuori, K., Trent, J.M., and Schork, N.J. Applications of Nonlinear Mixed Effects Models for High Throughput Drug Screen Association Tests. (In Preparation), with permission from all coauthors. The dissertation author was the primary

researcher and author of this material.

Chapter 5 is a full reprint of the publication: **Ding, K.**, Finlay, D., Yin, H., Hendricks, W.P.D., Sereduk, C., Kiefer, J., Sekulic, A., LoRusso, P.M., Vuori, K., Trent, J.M., and Schork, N.J. Network Rewiring in Cancer: Applications to Melanoma Cell Lines and the Cancer Genome Atlas Melanoma Patients. (In Preparation), with permission from all coauthors. The dissertation author was the primary researcher and author of this material.

VITA

2016-Present	Senior Data Scientist, Intrexon
2017	Ph.D. in Bioinformatics and Systems Biology, University of California, San Diego
2009-2012	Bioinformatics Specialist, HemoShear
2011	M.S. in Statistics, University of Virginia
2008-2010	Biostatistics Manager, Key Genomics
2008	B.S. in Mathematics, University of Texas at Austin

PUBLICATIONS

Jamison McCorrison*, **Kuan-Fu Ding***, Andrew Pickering, Dan Evans, Trina Norton, Pui Kwok, Ludmila Pawlikowska, Richard Miller, and Nicholas J. Schork. Multi-Reference Transcriptomic Analysis of Bird Species: Interpreting Reference-specific Signatures of Longevity. (In Preparation) *McCorrison and Ding contributed equally to the work.

Kuan-Fu Ding, Emanuel F. Petricoin, Darren Finlay, Hongwei Yin, William P.D. Hendricks, Chris Sereduk, Jeffrey Kiefer, Aleksandar Sekulic, Patricia M LoRusso, Kristiina Vuori, Jeffrey M. Trent, and Nicholas J. Schork. Applications of Nonlinear Mixed Effects Models for High Throughput Drug Screen Association Tests. (In Preparation)

Kuan-Fu Ding, Hongwei Yin, William P.D. Hendricks, Chris Sereduk, Jeffrey Kiefer, Aleksandar Sekulic, Patricia M LoRusso, Jeffrey M. Trent, and Nicholas J. Schork. Network Rewiring in Cancer: Applications to Melanoma Cell Lines and the Cancer Genome Atlas Melanoma Patients. (In Preparation)

Kuan-Fu Ding, Darren Finlay, Hongwei Yin, William P.D. Hendricks, Chris Sereduk, Jeffrey Kiefer, Aleksandar Sekulic, Patricia M LoRusso, Kristiina Vuori, Jeffrey M. Trent, and Nicholas J. Schork. Analysis of Variability in High Throughput Screening Data: Applications to Melanoma Cell Lines and Drug Responses. *Oncotarget*, Volume 8, February 2017, Pages 27786-27799.

Smith, Steven, Alexander Baras, Garrett Dancik, Yuanbin Ru, **Kuan-Fu Ding**, Christopher Moskaluk, Yves Fradet, Jan Lehmann, Michael Stickle, Arndt Hartmann, Jae Lee, and Dan Theodorescu. A 20-Gene Model for Molecular Nodal Staging of Bladder Cancer: Development and Prospective Assessment. *The Lancet Oncology*, Volume 12, No 2, January 2011, Pages 137-143.

ABSTRACT OF THE DISSERTATION

**Quantifying and Accommodating Heterogeneity in High Throughput
Drug Response Studies**

by

Kuan-Fu Ding

Doctor of Philosophy in Bioinformatics and Systems Biology

University of California, San Diego, 2017

Professor Nicholas J. Schork, Chair

Professor Bing Ren, Co-Chair

Over the past decade, there has been steady increase of studies leveraging genomic and related molecular profiling technologies to identify drug targets and characterize drug effects for particular diseases. This is particularly true for cancer, in which researchers can potentially identify genetic and other factors in tumors to indicate appropriate treatments. Many relevant studies have leveraged high throughput drug screen (HTS) strategies using, e.g., cancer cell lines that have been profiled at the genetic and other levels. These strategies allow researchers to assess the effects of many different drugs and compounds on cell lines exhibiting wide variation in their genomic profiles, so that associations can be identified that

relate elements of those profiles with drug response. Unfortunately, given the manner in which HTS studies are pursued and the amount of data they generate, problems such as false positives or a loss of statistical power must be addressed. These problems fall into three general categories: (i) the reliability of the screening data and the procedures used to generate the data; (ii) the statistical analysis methods used to identify associations between drug responses and other factors collected on the cells used in the screening; and (iii) the differences between the cell lines used in terms of their genetic architectures. I have taken a data-driven approach to address each of these concerns.

First, I assessed the reliability and reproducibility of HTS data leveraging two different collaborations. One set of analyses involved melanoma cancer cell lines subjected to two independent laboratory drug screens. I ultimately assessed the proportion of variation that could be explained by laboratory and technical effects associated with the design of the experiments and found that when sources of variation are quantified and controlled for, signals beyond “noise” can be detected that reflect true drug response. A second set of analyses involved HTS to identify drugs that influence lifespan in *Caenorhabditis elegans*. As with the study of the melanoma cell lines, I examined the variability in the screening outcome data that could be attributed to plate and plate-specific effects.

Second, I considered different ways of statistically analyzing dose-response data arising from HTS experiments. I ultimately evaluated the performance of nonlinear mixed effects (NLME) models relative to traditional models based on an analysis of IC_{50} values derived from individual cell line drug response profiles. Through simulation studies as well as applications to actual data, I found that testing for differences in dose-response curves using the NLME models has greater statistical power to detect gene associations with drug responses than tests involving traditional IC_{50} -based values.

Third, I assessed differences in genetic co-expression among cell lines used in drug screening studies. Such differences can dramatically affect identification of gene/drug relationships. I find evidence for differences in the way genes are related to each other between cell lines used in screening experiments and show

how this “re-wiring” of genes can affect interpretation of resulting drug screen data and identification of drug targets. Since more and more emphasis will be placed on choosing the right treatment for an individual based on his or her genetic and related profile in the future — as this is the goal of “personalized,” “individualized” and “precision” medicine — I believe my analyses and approaches will motivate future studies and lead to more reliable drug screening strategies and results.

Chapter 1

Preface

1.1 Background

There is tremendous interest in using genomic and related molecular profiling technologies with cell lines and other cellular systems to identify drug targets and to characterize drug effects in pharmaceutical studies [Lamb et al., 2006]. This is because genes, proteins, metabolites, and other molecular entities can be evaluated as drug targets using a number of strategies involving cell lines [Krogan et al., 2015]. In addition, the impact of specific drugs on those molecular entities can be studied in a high throughput fashion using state-of-the-field infrastructure, making it possible to efficiently screen thousands of experimental compounds and drugs. Such studies can provide insights into the molecular mechanisms associated with disease; the molecular and physiologic processes that are modulated via pharmacologic manipulation; and the factors that may mitigate response to a drug and ultimately shed light on how the research community might specifically target or “personalize” medicines [Barretina et al., 2012] [Yang et al., 2013]. But, there are a number of complications associated with conducting and interpreting the results of cell line-based drug targeting and screening studies. Though numerous, these problems are generally associated with sources of variation and heterogeneity, and fall into three broad categories: (i) the reliability and reproducibility of the data generated by screening studies given the large number of potential sources of both technical and biological variation inherent to their execution; (ii) the manner in

which claims about relationships among targets, drugs and modifying factors are drawn via statistical methods, especially in light of the substantial heterogeneity exhibited by individual cell lines at the genetic and molecular physiologic levels; and (iii) whether or not the molecular systems to be interrogated exhibit variation across cell lines so extreme that the study of any particular set of cell lines precludes easy generalization. The following expands on these issues as well as the need for strategies that might be used to assess and overcome them.

1.2 Exposing and Accommodating Issues with Heterogeneity in Drug Screens

As noted, drug screening studies require a great sensitivity to the inherent heterogeneity across cell lines used and the different sources of technical variation that complicate inferences from data analyses of those cell lines and their drug responses. The specific sources of heterogeneity and how they complicate relevant data analyses have been discussed in the literature to some degree, which we review in the following.

1.2.1 Reproducibility

The reliability and reproducibility of cell line-based drug target identification and screening initiatives have become a significant contemporary concern [The CCLE and The GDSC Consortia, 2015] [Haibe-Kains et al., 2013] [Safikhani et al., 2016a]. The massive amount of data generated in drug screening experiments raises questions about not only the reliability of the data, but also the rate of potential false positive results if important variation sources are not accounted for. For example, there is a need to quantify and control for purely experimental sources of variation (e.g., different plating schemes, assays done in different laboratories, different drug dosages, etc.). To address this, I studied various sources of technical and biological variability in high throughput drug screen (HTS) data generated by Sanford Burnham Prebys Medical Discovery Institute (SBP) and

Translational Genomics Research Institute (TGen) as part of a project sponsored by the Stand-Up-To-Cancer (SU2C) Melanoma Dream Team [Ding et al., 2017]. I also assessed variation sources in publicly available databases, such as the Cancer Cell Line Encyclopedia (CCLE) [Barretina et al., 2012] and Genomics of Drug Sensitivity in Cancer (GDSC) [Yang et al., 2013]. Lastly, I considered experimental sources of variation (e.g., plate position, drug dosage, drug class, etc.) in an independent organism-specific large-scale studies. In this study, I assessed variation sources in a HTS screen of drugs meant to modulate *Caenorhabditis elegans* life-span [Ye et al., 2014].

1.2.2 Association Analysis Involving Dose-Response Data

Current strategies for identifying factors that influence drug response using cell lines rely on analytical approaches that may not be optimal for detecting, e.g., gene/drug relationships. Previous attempts in community-driven drug discovery challenges have demonstrated the utility of HTS studies but raised questions about the approaches to analyzing the generated data [Bansal et al., 2014]. Typical analysis methods rely heavily on extracting single parameters meant to capture dose-response relationships, like the IC_{50} value, that may generate issues for associating those parameters with covariates and variation in drug response patterns. Comparing existing methodologies with novel methodologies could shed light on how to reliably and powerfully identify factors influencing drug response. I applied both traditional and novel approaches, such as nonlinear mixed effects (NLME) models, for relating covariates to drug response in an effort to identify factors that mitigate drug response using the aforementioned SU2C, CCLE and GDSC data sets. Based on the SU2C cell lines, I also performed simulation studies to investigate the utility of NLME models relative to traditional analysis models, in terms of their level-accuracy (sensitivity) and power (specificity) to detect the modifying influence of factors on drug response. NLME model-based methods advantageously balance the use of individual curves with group averaging, but are also more computationally demanding than traditional analysis models. These simulations show that NLME models accommodate statistical noise and are more powerful than

the standard models that only exploit, e.g., IC_{50} values obtained from individual dose-response curves.

1.2.3 Network Rewiring and Target Identification and Modulation

The manner in which genes, proteins and metabolites interact and function as a whole in normal as opposed to disease states is likely to be very different for obvious reasons. However, many drug target identification and target modulation studies focus on the behavior of a drug or modifiable target in wild-type, normal, or available and conveniently obtained cell lines. If differences of gene interactions between normal and disease states are pronounced, then generalization of drug activity assessed in normal cells to cells representing the disease state is compromised. I explored differences among cell lines in terms of connections and correlations among genes to characterize how much “re-wiring” among transcriptional networks might exist between cell lines in normal and diseased states.

In my analysis, I incorporated both “supervised” and “unsupervised” methods of machine learning to identify evidence of rewiring. “Supervised” approaches such as Evaluation of Dependency DifferentialitY (EDDY) [Jung and Kim, 2014]; Gene Set Enrichment Analysis (GSEA) [Subramanian et al., 2005] [Mootha et al., 2003]; and Gene Set Co-expression Analysis (GSCA) [Choi and Kendzierski, 2009] are limited to the settings in which the grouping of the cell lines is set in advance and hence requires prior knowledge of, e.g., disease and normal conditions. Ultimately, “supervised” methods require that each sample be identified as representing a disease or normal state. GSEA further assumes that each set of potentially differentially expressed genes is independent and the probability of observing a set of genes differentially expressed between conditions is equal. Similarly, GSCA depends on identifying differences in the correlations between genes in previously determined disease and normal states. Finally, although EDDY does aim to identify differential networks among groups of cell lines by incorporating probability distributions defining dependency networks across many genes, it still requires the prior specification of disease and normal conditions. “Unsupervised” methods do

not require, e.g., disease and normal states to be specified in advance but generate results that might be difficult to interpret or reconcile with the results of other “unsupervised” methods. This is because they must search for groups in a data set in addition to characterizing the differences between those groups. In addition, although there are plenty of unsupervised methods used almost routinely in many data analysis setting, most, if not all, of these methods *only* rely on exploring differences in the means or mean vectors of the factors (e.g., gene expression values, protein abundances, etc.) among groups. Moreover, they do not consider the correlations, connections or relationships — i.e., the “wiring” — of those factors by exploring differences in, e.g., covariance matrices that reflect the connections among the factors between identified groups. We assess evidence for network rewiring using both “supervised” approaches as well as “unsupervised” methods in drug screen data and consider the implications such rewiring might have for making claims about cell line responses to drugs.

1.3 Detailed Descriptions of the Data Sets Used

The data sets briefly mentioned in the previous section contain many nuances that were important in my analyses. I was able to collaborate with investigators from various institutions to obtain complex proprietary data sets as well as leverage public databases and resources that are also quite complex. In fact, these complexities motivated many aspects of the quality control (QC) analyses I pursued. The data sets I used to conduct my analyses are described in some detail below.

1. SU2C data: The SU2C Melanoma Dream Team high throughput drug screen data [Ding et al., 2017] was generated at the SBP in La Jolla, California and TGen in Phoenix, Arizona. This data includes information about the experimental factors and technical constructs used in executing the screens (e.g., plate, laboratory, drug, dosage, etc.), which allows for the quantification and accommodation of sources of variation in drug response attributable to these factors in the

HTS experiments. In addition to this information, I obtained information about different sources of biological variation that could impact the drug responses in the screen. This included: Reverse Phase Protein Array (RPPA) protein abundance, microarray whole genome gene expression, and melanoma-related cell line mutational status information for each cell line used in the screens. The SU2C cell line data was also used to define parameters for simulating HTS. We used these simulations to assess different methods for identifying drug x gene interactions (e.g., by defining drug sensitivity and resistance with NLME models and subsequently associating response with mutation status, gene expression and protein abundance information). Lastly, this data set was also used to assess evidence of network rewiring.

2. CCLE and GDSC data: The CCLE [Barretina et al., 2012] and GDSC [Yang et al., 2013] data sets are publicly available HTS dose-response curve data sets that I used to perform an assessment of the sources of variation influencing drug response. The CCLE and GDSC data analyses I performed are complementary to my analysis of the SU2C data [Ding et al., 2017]. Since the CCLE and GDSC HTS do not use the same drug concentration ranges, I assessed sources of variation in the context of previously determined and publicly available IC_{50} values for each cell line and drug studied. I was able to obtain dose concentrations and cell viability for each dose and drug administered in the CCLE data, but only had access to the IC_{50} values for the GDSC data set. Additionally, plate information was not available for these studies and could not be included in my comparative analyses.

3. *Caenorhabditis elegans* life span data from Michael Petrascheck: This data set was used to assess technical and experimental variation sources (e.g., plate position, drug dosage, drug class, etc.) contributing to 1280 drugs screened for their ability to modulate *C. elegans* life span. I assessed the variability of the screening data in the context of plate effects (including specific row and column of plates), date, and the laboratory performing the HTS experiments.

4. The Cancer Genome Atlas (TCGA) [Cancer Genome Atlas Network, 2015]: This is a very large, scientific community organized data set with thousands of patient tumors all having some kind of genome profiling done on them, such as gene expression pattern characterization. I used the gene expression data to find evidence of network rewiring in melanoma patients using both “supervised” and “unsupervised” data analysis methods that consider (i) differences across groups of tumors that are reflected in mean gene expression differences, and (ii) differences in their correlations as reflected in covariance matrices. I obtained Melanoma TCGA patient RNA-sequencing samples from the TCGA BROAD GDAC Firehose.

1.4 Synopsis and Discussion

Each of the analyses I pursued was designed to address a specific question related to cell line-based drug assessment studies, but could be extended in a number of ways. For example, as there are a number of published methodologies designed to relate particular pharmacologic manipulations and drug screens to factors measured on cell lines and the cell lines’ responsiveness [Costello et al., 2014], a comparison of those methodologies would be ideal. My approach was to (i) address issues plaguing HTS and drug target identification studies that have received attention in literature and (ii) expose areas of additional immediate and future research. This is highly appropriate for three reasons: (i) as genomic and related profiling technologies become more efficient and cost-effective, they will be used to identify potential drug targets and assess drug response in unprecedented ways; (ii) the use of standard cell lines, such as available immortalized cell lines, will likely be replaced with more sophisticated cellular constructs, such as fresh cells, induced pluripotent stem cells, and organoids with similar problems as those in traditional cell lines; and (iii) it is becoming increasingly clear that many factors mitigate drug response clinically as it is well-known that most commonly used drugs only help a fraction of users. Identifying those factors in cellular models could motivate the development of assays, companion diagnostic, and prognostic

markers that could ultimately lead to “personalized medicines” [Johnson and Gallagher, 2010] [Lupski et al., 2010] [Bainbridge et al., 2011]. These three reasons for pursuing my studies also support the importance of anticipating future analytical needs for HTS studies, since they will undoubtedly involve the combination of (i) better and cheaper data generation technologies; (ii) a desire to understand drug activity as a function of the potentially unique and nuanced profiles of individuals and their cells; and (iii) more biologically relevant and sophisticated cellular systems. Thus, I believe our analyses and approaches will motivate future studies and lead to more reliable drug screening strategies and results.

Chapter 2

Analysis of Variability in High Throughput Screening Data: Applications to Melanoma Cell Lines and Drug Responses

2.1 Abstract

High-throughput screening (HTS) strategies and protocols have undergone significant development in the last decade. It is now possible to screen hundreds of thousands of compounds, each exploring multiple biological phenotypes and parameters, against various cell lines or model systems in a single setting. However, given the vast amount of data such studies generate, the fact that they use multiple reagents, and are often technician-intensive, questions have been raised about the variability, reliability and reproducibility of HTS results. Assessments of the impact of the multiple factors in HTS studies could arguably lead to more compelling insights into the robustness of the results of a particular screen, as well as the overall quality of the study. We leveraged classical, yet highly flexible, analysis of variance (ANOVA)-based linear models to explore how different factors contribute to the variation observed in a screening study of four different melanoma

cell lines and 120 drugs over nine dosages studied in two independent academic laboratories. We find that factors such as plate effects, appropriate dosing ranges, and to a lesser extent, the laboratory performing the screen, are significant predictors of variation in drug responses across the cell lines. Further, we show that when sources of variation are quantified and controlled for, they contextualize claims of inconsistencies and reveal the overall quality of the HTS studies performed at each participating laboratory. In the context of the broader screening study, we show that our analysis can also elucidate the robust effects of drugs, even those within specific cell lines.

2.2 Introduction

High-throughput screening (HTS) strategies allow researchers to assess the effects of thousands of compounds on drug responses in one large experimental setting. Over the past decade, applications of HTS for drug discovery and drug effect characterization studies have steadily increased, ranging from studies focusing on the assessment of multiple phenotypic endpoints in high-content screening [Taylor, 2010], the evaluation of drugs on traits such as lifespan through the sophisticated use of model species such as *Caenorhabditis elegans* [Ye et al., 2014], as well as drug response pattern identification using massive amounts of genomic information made available for crowd-sourcing efforts and community driven challenges [Eduati et al., 2015]. HTS studies have also been pursued to advance personalized medicine, especially in oncology settings, since tumor-derived cell lines can be used in the screening studies to identify compounds that are active against them or some subset of them. For example, large HTS databases — such as the NCI-60 [Shoemaker, 2006], the Cancer Cell Line Encyclopedia [Barretina et al., 2012] (CCLE) and the Genomics of Drug Sensitivity in Cancer [Yang et al., 2013] (GDSC) have been made available to researchers for the express purpose of uncovering drugs that exhibit unique effects against tumor-generated cell lines with specific genomic profiles. In addition, very recent work involving the Connectivity Map [Lamb et al., 2006] has exploited genetic network and pathway reconstruction

methods to identify sets of genes that mediate specific drug responses in subsets of cancers. In this light, the Cancer Cell Map Initiative [Krogan et al., 2015] (CCMI) and related initiatives have not only drastically reduced drug discovery costs, but have also guided efforts to identify genomically-informed, patient-specific cancer treatment strategies. Unfortunately, as timely and as sophisticated as these efforts have been, very recent studies comparing the quality of different HTS studies meant to advance insights into personalized cancer care have raised questions and concerns about their reliability and reproducibility as well as the interpretation of the data they generated [Haibe-Kains et al., 2013] [The CCLE and The GDSC Consortiums, 2015].

Assessing the reliability of HTS studies is not trivial given the number of compounds typically considered, the number of reagents used, the way in which constructs such as plating schemes and distributed robotic handlers are set up, the manner in which dose response curves are constructed, and the fact that different labs likely follow slightly, if not overtly, different protocols; i.e., the sources of HTS cells, tissues or organisms, cell culture and assay conditions, reagents, consumables and instrumentation are not standardized within the community. This is particularly true for, e.g., tumor-derived cell line-based screening studies where the nature of the source cell lines, their procurement and sustenance as well as the responses measured on them may vary widely between different laboratories. In addition, although diverse in execution, many cancer-oriented HTS studies focus on cell counts upon stimulation with a drug that reflect that drug’s ability to kill cells derived from a specific cell lines across differing drug concentrations. These concentrations often range from inducing no response (i.e., no cells are killed) to a very strong response (e.g., all the cells are killed). The dose ranges necessary to achieve variation in the number of cells killed and establish a dose response curve are very hard to anticipate often leading to different labs using different concentrations and numbers of concentrations.

After having established the drug concentrations or doses to be used and applying them to cells derived from a single cancer cell line, sigmoidal curves are often fit to the cell counts associated with each drug dose to generate drug-specific

dose response curves (DRCs). This is repeated for each cell line. The half minimal inhibitory concentration (IC_{50}) is then extracted from these curves [Shoemaker, 2006] [Barretina et al., 2012]. These IC_{50} values, which are often couple with related response measures such as the area under each dose response curve (AUC), are then used to determine the sensitivity or resistance of each cell line to the different drugs. Recently, the IC_{50} and AUC results from two large cancer cell line HTS studies, the Cancer Cell Line Encyclopedia (CCLE) and the Genomics of Drug Sensitivity in Cancer (GDSC) studies, were used to assess the variability of HTS assays pursued in this manner [Barretina et al., 2012] [The CCLE and The GDSC Consortiums, 2015] [Yang et al., 2013]. The results of an assessment of the comparability and reproducibility of CCLE and GDSC data sets by two different research teams yielded opposing interpretations, which underscores the complexity of HTS studies and their interpretation [Haibe-Kains et al., 2013] [The CCLE and The GDSC Consortiums, 2015] [Safikhani et al., 2016b]. A third research team recently reevaluated the CCLE and GDSC data and came to yet a different a conclusion [Haverty et al., 2016].

In order to assess the reliability of HTS data, we conducted a study of intra- and inter-site experimental variability across melanoma cell lines treated with 120 different drugs that are either in use in clinical trials or have been FDA-approved for use in treating cancers. Our study was motivated by not only the controversies surrounding the reliability of the CCLE and GDSC data sets, but also by our engagement in a large clinical trial exploring the utility of personalized treatment for late-stage BRAF wild-type melanoma [LoRusso et al., 2015]. We first measured variability across replicated dose and drug applications to 29 melanoma cell lines pursued within a single institution, the Sanford Burnham Prebys (SBP) Medical Discovery Institute in La Jolla, California. The SBP studies were pursued using two independent HTS formats and screens: a nine-concentration and a three-concentration dose-response screens. Using the same 120 drugs and four of the 29 cell lines, we performed an independent nine-concentration dose response screen at the Translational Genomics Research Institute (TGen) in Phoenix, Arizona. To enable analysis of inter-site experimental variability, two copies of the master drug

plates were generated at SBP. One was then ultimately used onsite at SBP and the other was provided to TGen for their respective screens. Furthermore, the two sites used the same final dosing concentrations and the same cell lines. All other aspects of the screen were independent, resulting in variation in the environments in the which the screens were pursued, personnel, compound freeze/thaw cycles, cell passages, culture conditions, plating density, actual plates and other consumables, and instrumentation.

To analyze the data produced from the two nine-concentration HTS studies, and to assess the variability of the results, we used flexible linear models and analysis of variance (ANOVA) techniques. These traditional techniques allowed us to examine how variation in drug responses (i.e., variation in the fraction of cells killed for a particular cell line, drug and dose) is impacted by different factors, such as the laboratory, the drug used, the plate on which specific assays were conducted. We also considered interaction terms in the relevant (e.g., dose x drug interactions). We chose not to generate dose-response curves and extract IC_{50} values for use in our analyses, since our interests were in quantifying as many sources of variation as possible and not condensing or obscuring any of them in dose-response relationships reflected in single derived value. Thus, we modeled the dosage effects of each drug as a separate independent or explanatory variable for drug response variation. It is well-known that models that are “saturated” in that they exhaustively model the effects of independent variables and their interactions are inherently linear. This fact is exploited in many contexts, most notably econometrics, to help draw causal inferences between independent and dependent variables [Angrist and Pischke, 2009]. Although we did not consider all possible interaction terms in our models, we did consider most of them. Ultimately, our linear modeling and ANOVA analyses allowed us to make comprehensive claims about that effects of particular drugs and dosages on specific cell lines while accounting for factors built into the design of the HTS, such a plate effects, that could induce variation in drug responses. Thus, we believe our analyses can help identify signals of truly statistically-significant drug effects over-and-above the “noise” created by various factors, including individual laboratories and/or the individual plates upon which

cells were placed for drug effect characterization. We firmly believe that more sophisticated analytical methods, careful analyses, and interpretations of drug effect claims in HTS experiments are necessary and will likely lead to the identification and characterization of correctable sources of variation that may obscure HTS results and shed light on claims about their lack of reproducibility.

2.3 Results

2.3.1 Variation of Cell Viability Data Across Dose and Drug Replicates

Table 2.1 provides a brief summary of the data sets we considered in our analyses. As an initial assessment of the consistency of drug-by-dose effects across the 29 cell lines in common between the SBP nine-concentration and three-concentration HTS studies (Figure 2.1), which comprise all HTS data obtained at matching doses within the SBP HTS data, we considered the use of simple correlation analyses. We found that the data cell viability data are not normally distributed (Shapiro-Wilks test p -value $< 2.2\text{e-}16$ for drug responses for each drug) and that the non-parametric Spearman correlation coefficient, rather than the standard Pearson correlation coefficient would be more appropriate for use in assessing the consistency between the two data sets. We computed pairwise Spearman correlations for the replicates at 0.1, 1.0, and 10.0 μM concentrations across each drug and cell line (Figure 2.2A, Figure 2.5). Additionally, we calculated the correlation coefficients using all available concentrations. As expected, the pairwise correlations at the three concentration points suggested that a subset of the cell lines showed greater evidence for reproducibility. These cell lines were identified as those most likely to be sensitive to the drugs. This makes sense since the cell lines exhibiting no response (i.e., does response curve) contributed only noise to the correlations. In addition, within each dose, the distribution of the correlation coefficients for each drug was skewed left (Figure 2.2B), with a long tail towards negative correlations. This was the same when considering all dosages together,

Table 2.1: Key: CLs = Cell Lines; MCLs = Melanoma Cell Lines; OCLs = Other Cell Lines; CCLs = Melanoma cell lines in common with the SBP nine-point data set; Common drugs = drugs in common with the SBP nine-point data set; SBPMDI = Sanford Burnham Prebys Medical Discovery Institute; TGen = Translational Genomics Research Institute; CCLE = Cancer Cell Line Encyclopedia; GDSC = Genomics of Drug Sensitivity in Cancer

Data Set	Doses (μm)	MCLs	OCLs	# Drugs	CCLs	Common Drugs
SBPMDI 1 dose	10	49	0	747	37	89
SBPMDI 3 dose	.1, 1, 10	30	0	120	29	120
SBPMDI 9 dose	.02, .04, .1, .2, .4, 1, 2, 4, 10	40	0	120	40	120
TGen 9 doses	.02, .04, .1, .2, .4, 1, 2, 4, 10	4	0	120	4	120
CCLE 8 doses	.0025, .008, .025, .08, .25, .8, 2.53, 8	59	888	24	4	9
GDSC 8 doses	varied	45	1209	139	4	6

although there was an observed improvement in the correlation coefficient, which may reflect a larger sample size (Figure 2.2C). This highlights the advantage of considering the overall pattern of consistency for drug sensitivity profiles, as opposed to considering each dose individually. Thus, studies exploring the influence of different factors on HTS results should pursue analyses reflecting variation in the entire experiment, instead of just focusing on each individual drug, cell line or dose in isolation.

2.3.2 Variation Between Two Laboratories

For drugs with at least one cell line exhibiting a 20% cell viability at higher doses, which is consistent with a drug response ($n=46$), we also calculated the Spearman correlation coefficient between the SBP and TGen response data for each of the nine concentrations and also across all nine concentrations (Figure 2.2D, Figure 2.2E, Figure 2.6). The correlation coefficients across all available

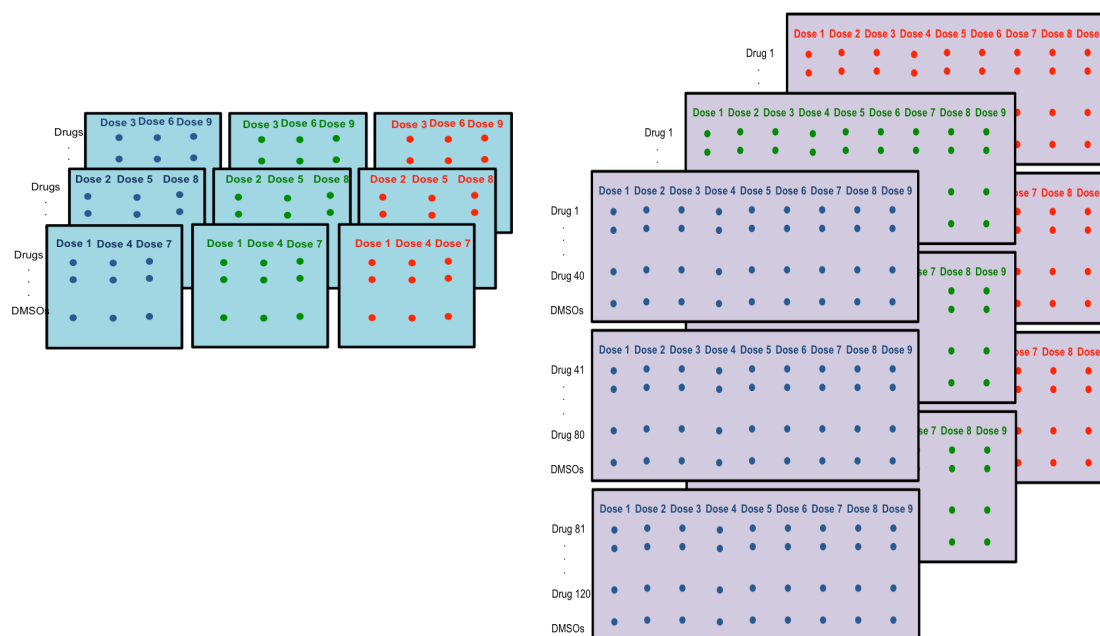


Figure 2.1: Differences in HTS plating schemes. Nine 384-well plates are used for each cell line. Left: TGen/SBP plating scheme. Each plate consists of triplicates of three doses across 120 drugs and 8 DMSOs. Right: Alternate plating scheme with each plate including all 9 doses.

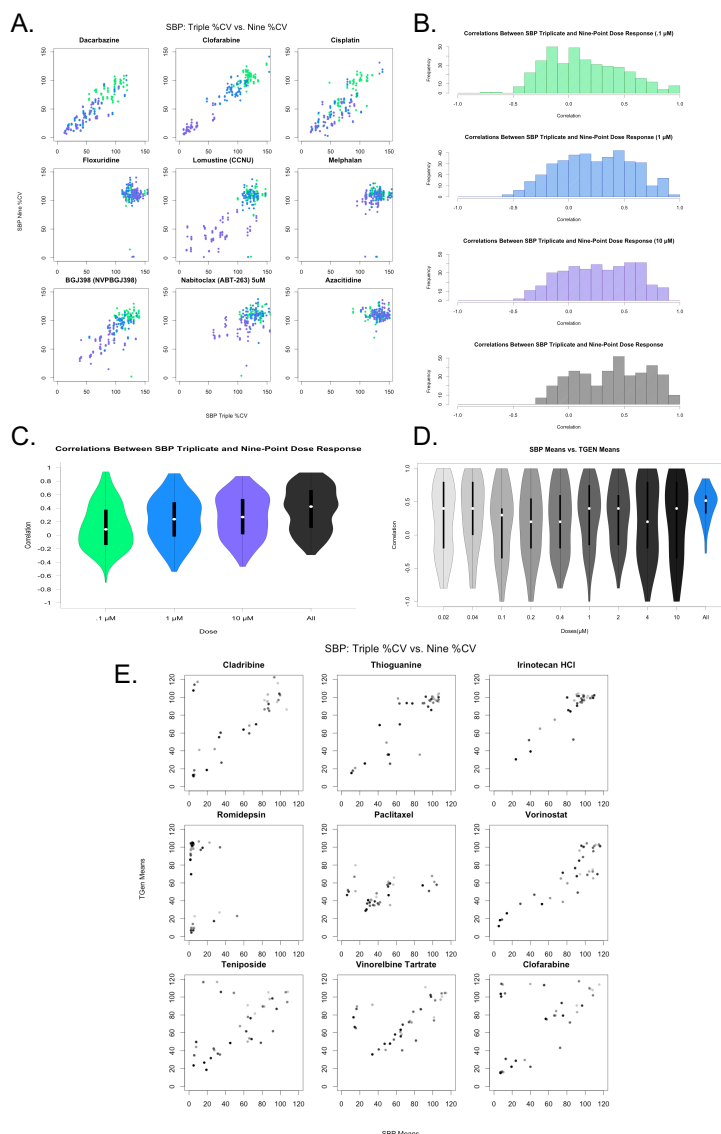


Figure 2.2: A. Pairwise scatterplots across nine drugs comparing SBP 3-point screen against same three concentrations from SBP 9-point screen. Colors indicate concentration: 0.1 (green), 1.0 (blue), and 10.0 μ M (purple). B. Correlation coefficients from pairwise spearman correlation across same three concentrations. Colors indicate concentration: 0.1 (green), 1.0 (blue), 10.0 μ M (purple), and across all three concentrations. C. Violin plot of correlation coefficients. D: Violin plot of correlation coefficients. E. Pairwise analysis of mean Cell Viability across nine drugs exhibiting at least one CV less than 20%. Gradient indicates concentration from 0.02 M (light) to 10 μ M (dark). Pairwise scatterplots comparing SBP 9-point HTS against TGen 9-point HTS.

dosages were greater than 0.0 in 44 out of the 46 drugs (Bonferroni-adjusted p -value $< .05$ in 22 of the 44 drugs). Individual pairwise scatterplots for each dose across the 46 drugs indicate that for a large number of them, there is a high degree of between-laboratory consistency. As in the three-concentration drug response analysis performed, the drug responses were more consistent at higher doses and across all dosages when considered together (Figure 2.2E, Figure 2.6). The stronger correlation at higher dosages, especially in the context of lower doses that do not induce an effect or response, reveal technical variation and “noise” that should be considered in analyses seeking to identify bona fide drug-induced effects producing signals that rise above this experiment-wise noise. Likewise, the improved correlation coefficients observed when comparing the nine-concentration dose-response curves (DRCs) against the three-concentration DRCs suggest that a full range of DRCs may be better at revealing true biological variation in the HTS data.

2.3.3 Comprehensive Analysis Considering the Entire HTS Experimental Setting via ANOVA Modeling

For a more comprehensive assessment of the factors contributing to the variation in drug response associated with our HTS studies, we applied flexible linear regression modeling within an ANOVA context. We ultimately wanted to partition the variation in cell viability data arising from the entire HTS study into factors representing different experimental and biological conditions (i.e., across all plates on which the samples were arranged, drugs, drug concentrations, cell lines, and laboratories; Supplemental Table 2.1). We limited this analysis to the four cell lines in common at the two independent laboratories. In order to conduct the modeling appropriately, we initially had to choose a comparator to be contrasted with all the other subgroupings for each of the different factors we studied. This was achieved by creating simple zero (absence) or one (presence) dummy variables for each factor subgroup. We randomly selected SBP as the comparator laboratory, Cladribine as the comparator drug, plate 36 (from the SBP experiment) as the comparator plate, cell line A375 as the comparator cell line, and the lowest dose ($0.02 \mu\text{M}$) as the comparator dose. Our analysis found that the laboratory used

Table 2.2: Percentage of variance explained by the experimental factors.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	% Variance explained
CellLines	3	1409075	469692	1064.446	< 2.2e-16	4.94
SBP	1	7975	7975	18.0742	2.13E-05	0.03
Ldose	8	1493445	186681	423.0677	< 2.2e-16	5.24
Drug	119	12966053	108958	246.9287	< 2.2e-16	45.46
Plates	65	921052	14170	34.236	< 2.2e-16	3.23
Residuals	25723	11721774	441			

explained 0.028% of the variation in drug response, whereas plate (3.23%) and other biological factors such as drugs (45.5%), concentration (5.24%), and cell lines (4.94%) explained approximately 60% of the variation (Table 2.2). To identify the individual factors that were most statistically significant sources of variation, we carried out simple t-tests on each factors regression coefficient. We used a conservative Bonferroni-correction to accommodate the multiple tests. The results suggested that laboratory was only marginally significant factor relative to the others (Figure 2.3A, Figure 2.7). Analysis of the cell lines indicated that two of the four cell lines, MeWo and SK-MEL-2, had an effect on drug responses that were statistically significantly different from the comparator cell line A375. This could be due to the BRAF mutation status in the cell lines: A375 (the comparator cell line) and UACC-0257 are cell lines with the BRAF V600E mutation, whereas MeWo and SK-Mel-2 are BRAF wild type cell lines. Obviously, more work on this hypothesis is needed before attributing differences to the presence of the BRAF V600E mutation. Additionally, the higher concentrations (i.e., 2.0, 4.0, and 10.0 μ M) were the most statistically significant, consistent with the existence of overall dose-response relationships in the experiment.

Interestingly, we also found that a large number of SBP plates were moderately statistically different from other plates both used at SBP and TGen; however, a single TGen plate (plate 25) was highly significant. Furthermore, only one SBP plate yielded a more significant t-statistic (plate 27) than the TGen plate 25 (Supplemental Table 2.2). As expected, the outlying plates produced DRCs with greater variability (e.g., Mitoxantrone DRCs for UACC-0257 at 0.1, 1.0, and 10.0

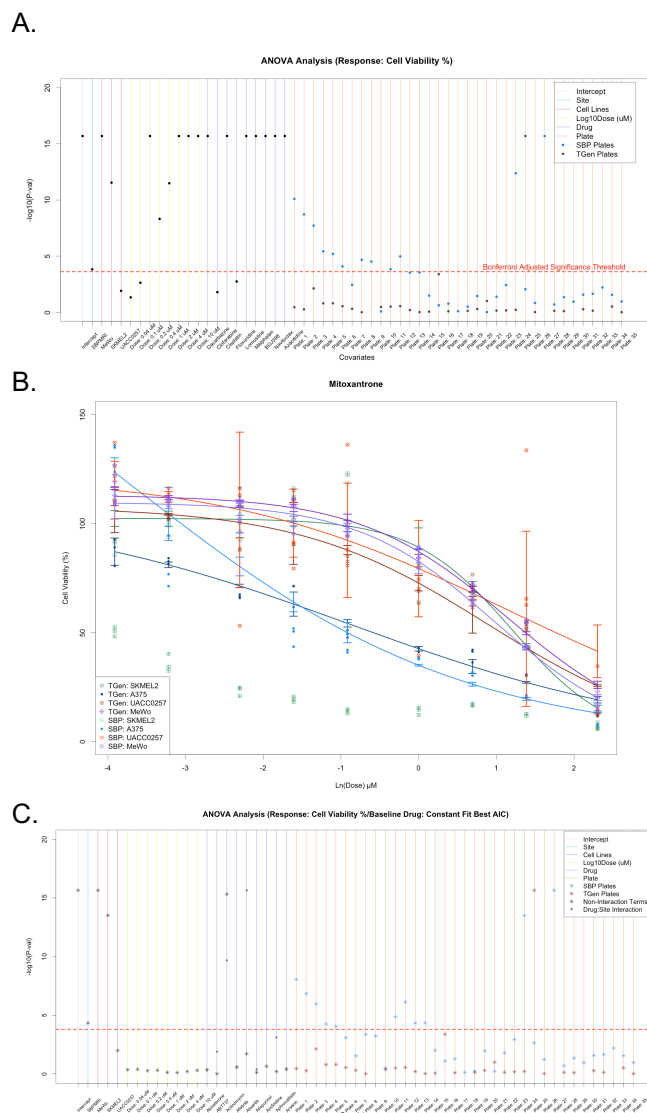


Figure 2.3: A: Application of flexible linear models, using ANOVA methods, to explore the variation in CV that is explained by site, cell lines, dose, drug, and plates. Sites explained a small proportion of the variation, whereas drugs, dose, and cell lines explained a majority of the variation observed in CV. Nine of the 120 drugs are shown. B. Plots of dose response curves for each of the cell line and site combinations. For example, greater variance is observed within Mitoxantrone in concentration from outlying plates detected by ANOVA-like methods. C. Assessing the variation in CV that is explained by site, cell lines, dose, drug, and drug-site interaction. To further assess the reproducibility of HTS, we examined the drug-site interaction terms and found that 19 of the significant drugs did not have a significant drug-site interaction term.

μM for SBP plate 27 and at 0.02, 0.2, and 2.0 μM for TGen plate 25, Figure 2.3B). Although plate-effects explained 3.23% of the variation in cell viability, the small number of highly significant plates suggests that plate analysis and ways of accommodating plate effects in HTS data analyses, or subsequent removal of outlying plates, should be performed to assess the overall quality of the HTS data and potentially lead to explanations for why some drugs don't replicate across site-specific studies (e.g., because some plates were outlying and should be removed as opposed to a more global analysis and rationale). The reasons for plate effects should be explored, but could reflect contamination, technician, or robot error when setting up the plate or experiment in question. Importantly, if the drugs, concentrations and cell lines used on the plate led to biologically meaningful effects, then one would not be able to separate the biological significance from a potential technical artifact, which suggests the use of controls and designed plating schemes are necessary. Finally, we found that most drugs in the study exhibited strong, statistically significant p-values, far beyond what would be expected by chance alone.

2.3.4 Exploring Interaction Effects

To identify whether specific drugs, cell lines, doses, and sites were significant predictors of drug response (i.e., cell viability) while considering other factors, we added interaction terms to the linear models. When we accounted for laboratory x drug interactions, we found that many of the drugs exhibited significant interaction term p-values; in fact many more than would be expected by chance alone. These results indicate that although drugs may influence variation in cell viability across the experiment as a whole, some of the cell viability variation may be laboratory-specific (Supplemental Table 2.3). 19 drugs yielded significant drug p-values, yet non-significant drug-laboratory interaction effects (Figure 2.3C, Figure 2.8), indicating that these drugs exhibited overtly reproducible effects (Supplemental Table 2.4). As expected, when we incorporated the drug x dose interaction terms into our analysis models, we found that a majority of the variation significantly accounted for by dose was limited to higher concentrations, which of course makes sense (Figure 2.9). Importantly, drug x laboratory interaction effects only

explained an additional 3.94% of the cell viability variation, whereas drug x dose interaction effects explained an additional 11.02% of the variation in cell viability.

The consideration of three-way interaction effects led to additional insights into the factors contributing to cell viability in a pronounced enough way to rise above the “noise” in cell viability across the HTS experiments as a whole. Laboratory x drug x dose effects explained 2.5% of cell viability variance and drug x dose x cell line explained 3.6% of the variance (Supplemental Table 2.5). These findings provided further evidence that laboratory effects were not as influential on the experiment as a whole relative to other factors. While nine drugs appeared to have laboratory-specific effects (i.e., a non-significant effect when considering the drug alone, but a significant drug x laboratory interaction effect), drug x dose interactions had the largest impact on the drug-associated cell viability, which it to be expected (Figure 2.10). Again, we identified a subset of drug x dose x cell line interaction terms that were significant (Supplemental Table 2.6), suggesting that there may be some heterogeneity among the cell lines (likely due to genetic differences) affecting the drug response. This is important in that it suggests unique cell-line features may indicate efficacy of certain drugs, which is consistent with the goal of personalized medicine. The significant differences in cell viability across the cell lines generally appear in the intermediate concentration ranges (Figure 2.11).

2.3.5 Re-analysis With Broader Set of Comparator Drugs and Cell Lines

To investigate whether the use of an arbitrary drug as the comparator drug impacted our initial analyses in any way, we set out to identify a group of cell line and drug combinations that did not exhibit any dose-response or general drug effect. These combinations would then effectively act as a group of controls whose cell viabilities represent simple noise and technical variation when compared to the other cell line-drug combinations. These analyses would allow us to see if compelling drug or drug x cell line effects can be identified that rise above the noise and variation exhibited in the HTS experiment as a whole. As noted in the Methods section, we fit a three-parameter sigmoidal, a four-parameter sigmoidal, a

linear, and a constant model to each of the cell lines for each drugs and dosages to characterize dose response curves. We then selected the drug-cell line pairs where the constant model fit the data the best based on the Akaike Information Criterion (AIC), since this would be indicative of no evidence of a drug or dose-dependent response. The set of drug x cell line combinations that suggested no evidence for a drug or dose-dependent response was then used as the comparator group in a re-analysis of the cell viability data. The re-analysis suggested that the laboratory in which the assays were done explained a minimal proportion of the variance (0.03%). However, drug effects explained 41.2%, dose 5.24%, cell line 4.94%, and plates 3.23% of the variation in cell viability across the experiment (Supplemental Table 7). Interestingly, this re-analysis greatly reduced the number of statistically significant drug effects, suggesting that only a small subset of drugs studied may be exhibiting actual effects that rise above the noise in cell viability across the experimental setting as a whole (Figures 12-15 and Supplemental Tables 2.7-2.10).

2.3.6 Comparing the Results of Models that Accommodate Plate-Specific Effects to Those That Do not

To assess the impact that technically-deficient or outlying plates have on the interpretation of HTS experiments, we assessed the statistical significance of the effects on cell viability of each drug in the context of the entire experiment using our linear model. We used the drugs that did not exhibit evidence for a dose response relationship as a comparator group, as discussed in the previous section. When not accommodating any site, plate, and dose (since doses were plated in triplicates on each plate) effects in the model, we found 54 significant drugs with drug-effect p-values surpassed Bonferroni-adjusted significance thresholds (Supplemental Table 2.11). When the site, plate and dose effects were considered in the model, we found 44 drugs with drug effect p-values that surpassed Bonferroni-adjusted significance thresholds (Supplemental Table 11). Strikingly, only 28 of the 44 drugs (64%) resulting from these analyses were in common, further confirming that experimental factors ultimately affect downstream analyses as well as the interpretation of analyses meant to identify drugs with pronounced effects

(Supplemental Table 2.11).

2.3.7 Analysis of CCLE and GDSC Data

We used ANOVA analyses on IC_{50} values provided in the CCLE and GDSC data sets along with our SBP and TGen data on the four cell lines and six common drugs they had in common. In these models, we randomly selected CCLE as the comparator laboratory, the drug Crizotinib as the comparator drug, and cell line A375 as the comparator cell line. Similar to the analysis comparing the raw SBP and TGen data, we found that laboratories were not significant predictors of IC_{50} values when we also accommodated drug, cell line, and laboratory effects in the model. We note, however, that we did observe some laboratory-specific effects for one of the six drugs, Nilotinib (Figure 2.4).

2.4 Discussion

Although there has been a considerable amount of interest and development in HTS technologies, there have been few attempts at standardizing protocols and ensuring the reliability and reproducibility of the resulting data. Many sources of variation implicated in HTS studies, such as the laboratory used to pursue the experiment, technicians, reagents and version of reagents used, plating schemes, cell culture conditions and dosing ranges, will inevitably have an effect on the resulting data. However, properly recording and accommodating these factors in the analysis of HTS data can aid in not only an assessment of the quality of the data, but also the interpretation of the results.

We sought to quantify the sources of variability in an HTS setting in order to gauge the, reliability and reproducibility of the resulting data. We did this using HTS data from a study on melanoma cell lines and simple correlation analyses coupled with often-used ANOVA modeling methods. The methods we discussed and applied can be used with any HTS data set including many of those in the public domain, although many studies have either used methods like we have proposed or different methods for assessing the concordance between datasets⁹

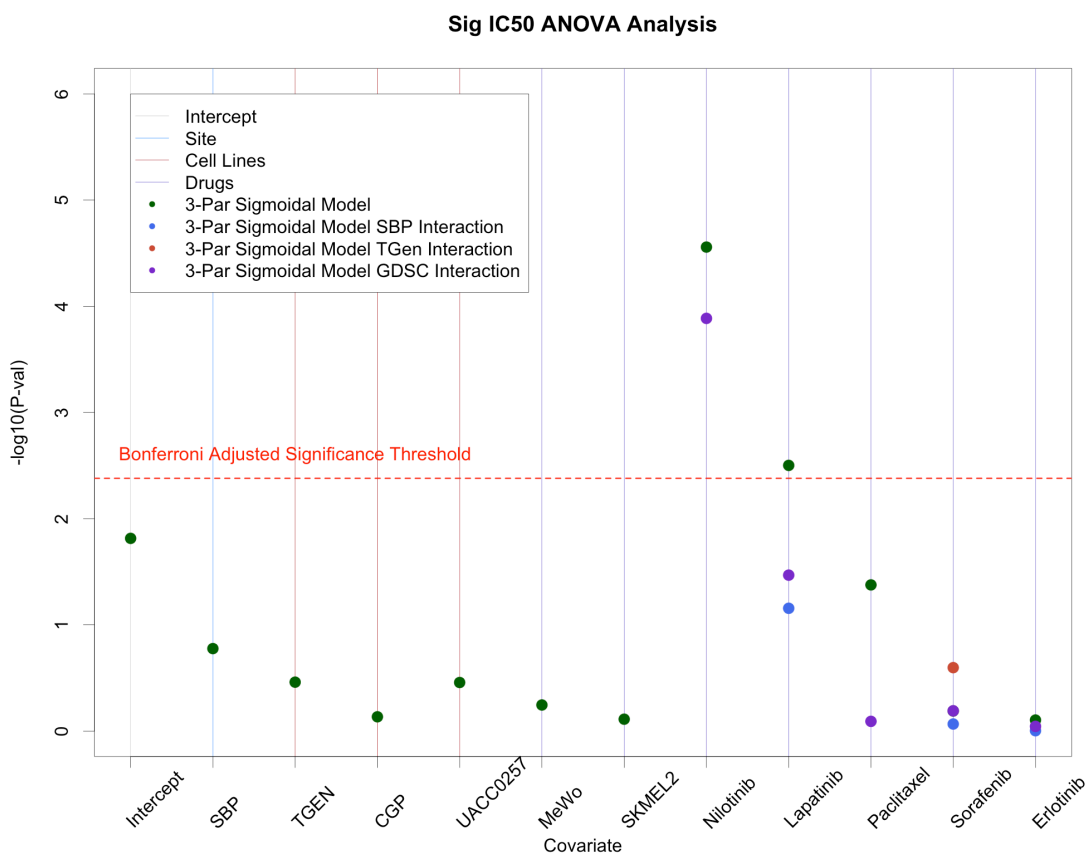


Figure 2.4: ANOVA analysis on IC_{50} values from SBP, TGen, GDSC, and CCLE. Site effects were generally minimal; however, site specific effects were observed in Nilotinib.

of all sorts (e.g., gene expression microarray, next generation sequencing, reverse phase protein arrays, etc.). Using the ANOVA setting, we, for example, identified evidence that certain plates used in the experiment were problematic and thus could undermine the reliability of the HTS experiment as whole if not accounted for properly.

We chose not to construct dose-response curves (DRCs) for each cell line for each drug and then compare, e.g., IC_{50} values extracted from those curves across sites or among different conditions, since this would ultimately ignore variation introduced by different technical factors impacting dose effects on cell viability (e.g., plate effects) and hence the DRCs. In this context, once outlying plates have been identified, it is possible to reassess the reliability of the HTS data using the traditional methods while excluding those outlying plates (alternatively, one could obtain coefficients associated with significant experimental effects and weight them accordingly). Additionally, ANOVA methodology allows for the simultaneous analysis of all cell viability data across all drugs to identify true and very compelling signals that have been obtained in a HTS experiment. Although the specific plating scheme in our experiments (Figure 2.1) allowed for the simultaneous analysis of various factors, it could also lead to confounding effects between plate and dose because triplicate concentrations were constrained to individual plates (i.e., one of three concentrations sets were used for each plate: concentrations 1, 4, and 7; concentrations 2, 5, and 8; and concentrations 3, 6, and 9).

In terms of specific findings, we observed that laboratory effects only explained a small fraction (0.03%) of the cell viability variation in our HTS setting. This contrasts with the results of studies described in recent publications about the reliability of the data associated with two very large cancer cell line screens pursued by different groups at different sites [Barretina et al., 2012] [Yang et al., 2013] [Krogan et al., 2015] [The CCLE and The GDSC Consortiums, 2015]. We also observed that the main source of variation in our study could be attributed to the drugs used further indicating that in certain settings, HTS data may indeed be reliable. By considering site x drug interactions in our analyses, we identified drugs with reproducible effects. Our consideration of three-way interaction

terms allowed us to identify some heterogeneity among the cell line response to drugs, suggesting that genetic differences among cell lines may explain why some cell lines are responsive to certain drugs. This is consistent with the notion that in vitro drug screening can shed light on the potential for personalized medicine [Crystal et al., 2014]. We also find that the downstream analyses of drug association tests are greatly impacted by whether one accommodates or does not accommodate experimental and technical factors, such a plating scheme, in the analysis models. In this light, we found that many of the drugs identified without experimental factors were false positives. Similarly, we find that many drugs that were significantly associated with drug response were missed when the models did not accommodate experimental factors. Our experience suggests that interaction terms in ANOVA-like analysis settings should be considered in order to tease out important and compelling drug associations that would otherwise go undetected due to the masking of particular drug effects by “noise.”

Ultimately, assessing the variability, reliability and reproducibility of HTS data by designing the study to contrast different experimental conditions could add to the overall experimental costs. As a result, alternative methods that can account for sources of variation are needed. Our analysis at the very least highlights the importance of using consistent dosing and plating schemes that include specific controls, which would allow researchers to not only measure and test for the impact of experimental factors on the outcomes, but also adjust for these specific experimental effects when making claims about, e.g., drug and cell-line specific effects. Thus, ANOVA-like methods can accommodate different experimental factors that may influence the HTS assays in pronounced ways and also reveal compelling signals attributable to drugs and experimental compounds. Although our analysis was performed only within the context of melanoma, and, importantly, only within the context of four melanoma cell lines and 120 drugs, our overall ANOVA-based approach can be used to reassess the quality of publicly available data as well as additional HTS data.

2.5 Materials and Methods

2.5.1 Data

We initially performed a nine-concentration (i.e., dose) HTS study (drug concentrations: 0.02, 0.04, 0.1, 0.2, 0.4, 1.0, 2.0, 4.0, and 10.0 μM) on 40 melanoma cell lines across 120 drugs (Table 2.1) at Sanford Burnham Prebys Medical Discovery Institute (SBP). We used 384-well plates with three concentrations of a drug assigned to each plate across all 120 drugs in triplicate (Figure 2.1). Drugs were spotted on 384-well clear bottom tissue culture treated plates (Greiner Bio-One, #781098) using an Echo Liquid Handler (Labcyte Inc., Sunnyvale, CA) such that addition of 25Ls of cells (100k cells/mL in RPMI +10% FBS +Pen./ Strep./ Glut., Omega Scientific, Tarzana, CA) resulted in the above described final drug concentrations and 2.5k cells/well. Upon plating, the cells were gently spun down at 1k rpm for one minute and incubated with drugs for 96 hours at 37 °C in a standard tissue culture incubator. After this time course, plates were allowed to equilibrate to room temperature for 30 minutes before 10 μL s per well of freshly prepared CellTiterGlo reagent (G7571, Promega Corp., Madison WI) were added. Samples were incubated for ten minutes with gentle agitation (100 rpm) before luminescence was read on a BioTek Synergy2 plate reader using Gen5 software (BioTek, Winooski, VT).

Each plate was assayed in triplicate and included 24 vehicle only DMSO controls. In total, nine plates were used for each cell line. For each cell line, drug, and drug concentration combination, cell viability (CV) measures were obtained post drug administration and were normalized to the plate-specific average DMSO cell:

$$CV_{norm} = 100 * CellCount_{Drug} / CellCount_{DMSO}$$

We also pursued an independent three-concentration HTS study (drug concentrations: 0.1, 1.0, and 10.0 μM) at SBP across 30 melanoma cell lines (29 in common with the nine-point screen) on all 120 drugs as described above. Finally,

we performed the primary nine-concentration HTS study across all 120 drugs for four melanoma cell lines in common with the initial nine-point SBP screen (four screened cell lines: UACC-0257, MeWo, SK-Mel-2, and A375) at the Translational Genomics Research Institute (TGen). For the TGen HTS study, the experimental protocols were similar to the SBP screen, with minor differences. Specifically, the compounds were pre-spotted to white, solid-bottom 384-well assay plates (Greiner Bio-One) using ATS (Biosero, San Diego, CA). Additionally, prior to measuring the luminescence using an Analyst GT plate reader (Molecular Devices, Sunnyvale, CA), 25 μ L CellTiterGlo reagent (G7571, Promega Corp., Madison WI) was added to assay plates and incubated at room temperature for one hour (as opposed to ten minutes at SBP).

MeWo and A-375 cell lines were obtained directly from American Type Culture Collection, all SK- cell lines were received directly from Memorial Sloan Kettering Cancer Center, TGen generated all UACC- cell lines, and all were of low passage number. Cells were maintained according to the manufacturer’s or collaborators instructions: A-375 cells are grown using DMEM medium, MeWo and Sk-Mel-2 cells are grown using EMEM medium, and UACC-0257 cells are grown using RPMI1640 medium. All media have 10% FBS and 1%AA added to final growth media. All cell lines were banked at low passages in multiple aliquots as liquid nitrogen stocks to reduce risk of phenotypic drift. All cells were cultured for less than three months before reinitiating culture from the frozen stock. All cells were routinely inspected for identity by morphology and growth curve analysis and validated to be mycoplasma free. All cell lines were free of contaminants.

2.5.2 Data Analysis

We assessed the variability of the HTS data by first examining the correlations between measures obtained at matching doses using the two SBP screens (i.e., nine-concentration and three-concentration screens) and then across the SBP and TGen screens (i.e., the two nine-concentration screens). We calculated the Spearman correlation of the CV_{norm} between the data obtained from each experimental pair setting at each dose (i.e., single concentration pairs) and the correlation

coefficient across all available doses (i.e., all three-concentration doses when comparing the two SBP screens or all nine-concentration doses when comparing SBP and TGen screens). Since the Spearman correlation is robust to outliers, is non-parametric, and does not necessarily assume linear relationships, we chose use it as implemented in base stats package in R. We assessed the normality of the data using Shapiro-Wilk test in the base stat package for R.

For a more comprehensive assessment of the assays variability and reliability on the four common cell lines across institutes, we used flexible linear models and ANOVA (also within the base stats11 package in R) to simultaneously assess drug effects (across all drugs) as well as all other experimental (e.g., plate and lab) and biological (e.g., drug, dose, and cell line) factor effects by creating (0, 1) dummy variables for each factor (e.g., we had $120-1=119$ dummy variables for the drugs, $9-1=8$ for doses, $4-1=3$ for cell lines, $72-1=71$ for plates, and one for site). After combining the lab, plate, drug, dose, and cell line factors, there were a total of 202 factor effects. Notably, we performed a multi-factorial analysis in an unbalanced experimental design. Under this approach, ANOVA in the context of linear regression models was used to assess the proportion of variation that can be attributed to each factor (e.g., laboratory, plate, drug, dose, and cell line). Interaction terms between the factors were also used to assess the non-additive influence and combinations of different factors. We then fit dose-response curves (DRCs) to each of the SBP and TGen nine-concentration HTS experiments using four separate models: a four-parameter sigmoidal model, a three-parameter sigmoidal model, a linear model, and a constant model (i.e., no dose-response effect). We fit these DRCs using the nplr package in R. We used the Akaike Information Criterion (AIC) from the base stats package in R to identify the “best” model for each cell line and drug combination. The AIC provides an assessment of model performance while considering the number of parameters used for the model. Thus, this approach allowed us to assess the fit of each model while adjusting for the number of parameters assumed within those models. We did this to identify individual drugs that exhibited no evidence of a dose-response relationship. The drugs and cell lines that exhibited no evidence of a dose response effect were then treated as

a set of controls for the other drugs because the variation they exhibited across the different doses reflect noise-induced and technical variation. We used this set of control drug-cell line combinations in re-analysis with the ANOVA model, coding the controls as the baseline drug (i.e. for the “controls”, binary variables for each drug were set to 0 in the model). To evaluate the impact of controlling for plate-specific effects, we performed drug association tests using linear models with interaction terms. We compared models that included and excluded plate effects. The significance of each drugs effects in the context of the entire experiment was determined based on a Bonferroni-adjusted significance threshold.

Because we were unable to access the raw experimental data for the Cancer Cell Line Encyclopedia (CCLE) data, we explored the consistency of the IC_{50} values provided from the CCLE5 repository along with IC_{50} values from the Genomics of Drug Sensitivity in Cancer6 (GDSC) data repository. We also computed the IC_{50} values from our SBP and TGen data using three-parameter sigmoidal curve fits. The four cell lines in common between the SBP and TGen data sets were also available for the CCLE and GDSC data, but only six drugs were in common. Further, replicates were only available in the GDSC dataset, plating schemes used between the sites were inconsistent, and there were differences in dose-concentrations between same cell line and drug combinations from the various datasets. Nevertheless, we explored the variation in the HTS data using information on the different sites ($n = 4$), cell lines ($n = 4$), and drugs ($n = 6$). All analyses were performed in R and all figures were generated using the graphics package in R.

2.6 Acknowledgements

Research supported by a Stand Up To Cancer Melanoma Research Alliance Melanoma Dream Team Translational Cancer Research Grant. Stand Up To Cancer is a program of the Entertainment Industry Foundation administered by the American Association for Cancer Research. Dr. Schork and Mr. Ding are also supported by NIH grants: U19 AG023122-09; R01 DA030976-05; R01 MH094483-03; R01 MH100351-02; R21 AG045789-01A1; UL1TR001442-01; and

U24AG051129-01, in addition to grants from Janssen, Human Longevity, Inc., and the Tanner Project Foundation.

Chapter 2 is a full reprint of the publication: Ding, K., Finlay, D., Yin, H., Hendricks, W.P.D., Sereduk, C., Kiefer, J., Sekulic, A., LoRusso, P.M., Vuori, K., Trent, J.M., and Schork, N.J. Analysis of Variability in High Throughput Screening Data: Applications to Melanoma Cell Lines and Drug Responses. *Oncotarget*. 2017; 8:27786-27799, with permission from all coauthors. The dissertation author was the primary researcher and author of this material.

2.7 Supplemental Figures

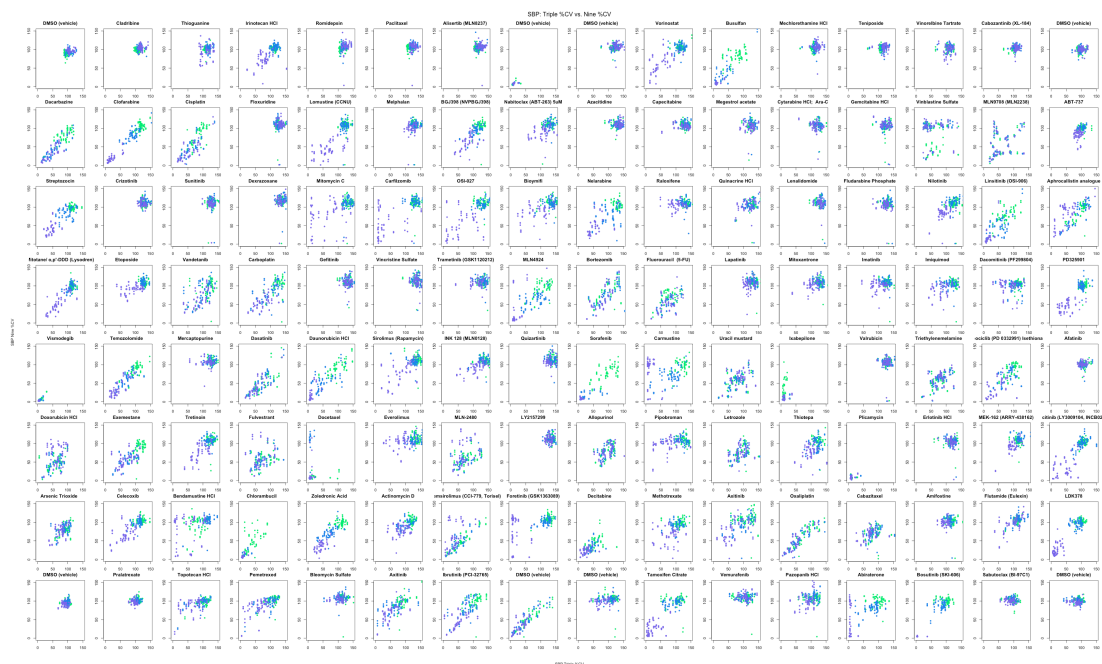


Figure 2.5: Pairwise scatterplots across all 128 drugs comparing SBP 3-point screen against same three concentrations from SBP 9-point screen. Colors indicate concentration: 0.1 (green), 1.0 (blue), and 10.0 μM (purple).

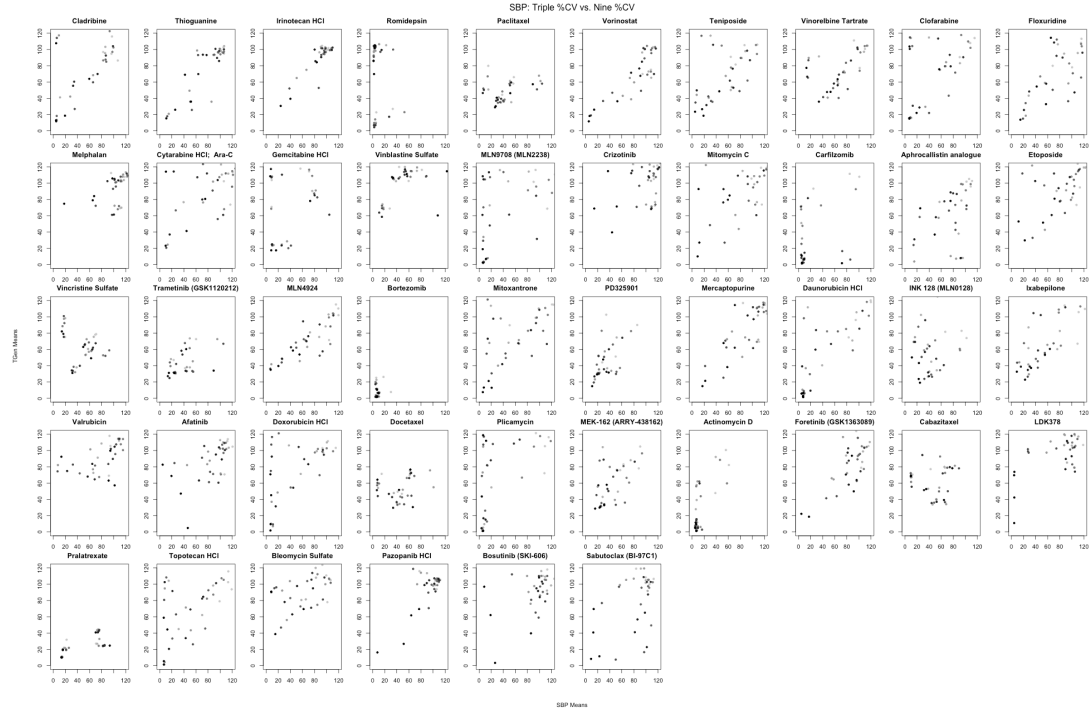


Figure 2.6: Pairwise analysis of mean Cell Viability across 46 drugs exhibiting at least one CV less than 20%. Gradient indicates concentration from 0.02 μM (light) to 10 μM (dark).

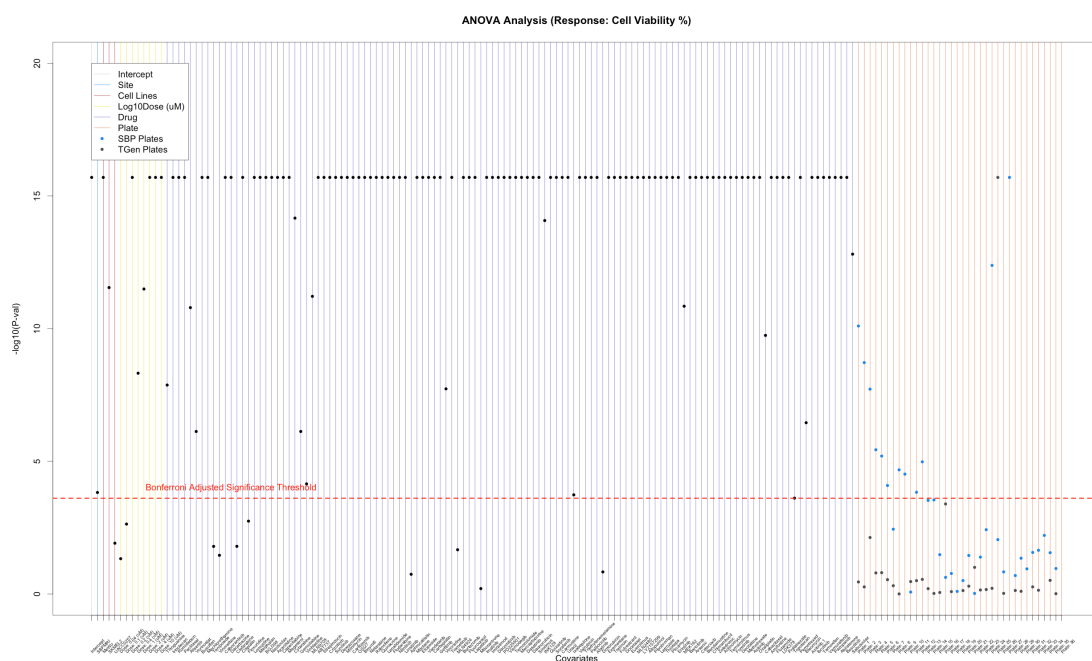


Figure 2.7: Application of flexible linear models, using ANOVA methods, to explore the variation in CV that is explained by site, cell lines, dose, drug, and plates. Sites explained a small proportion of the variation, whereas drugs, dose, and cell lines explained a majority of the variation observed in CV.

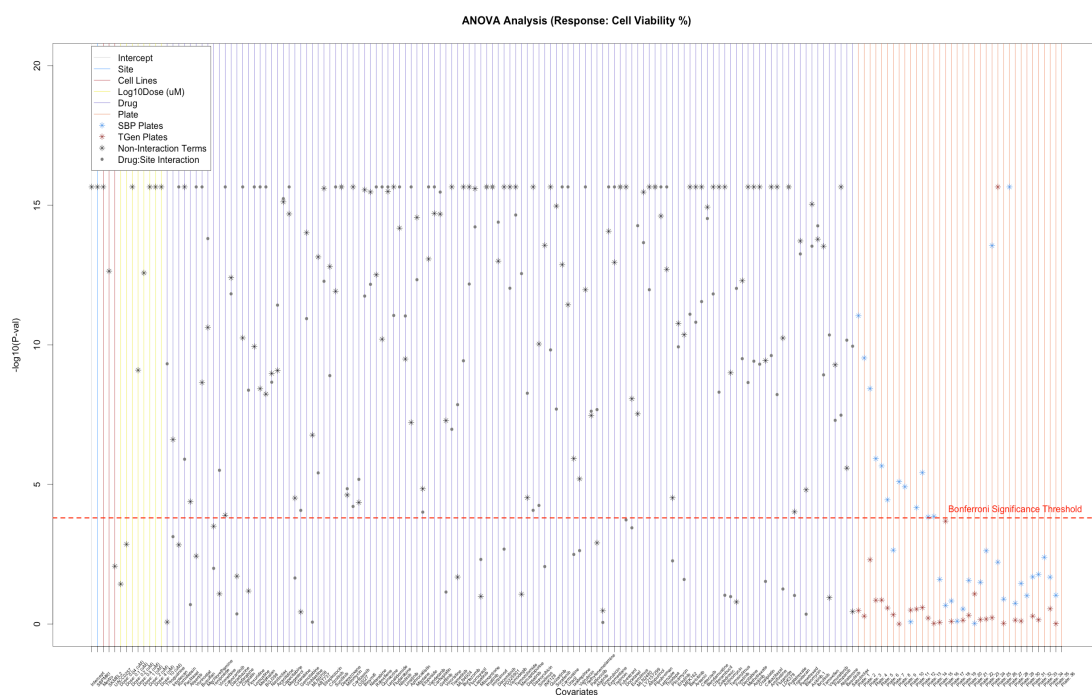


Figure 2.8: Assessing the variation in CV that is explained by site, cell lines, dose, drug, and drug-site interaction. To further assess the reproducibility of HTS, we examined the drug-site interaction terms and found that 19 of the significant drugs did not have a significant drug-site interaction term.

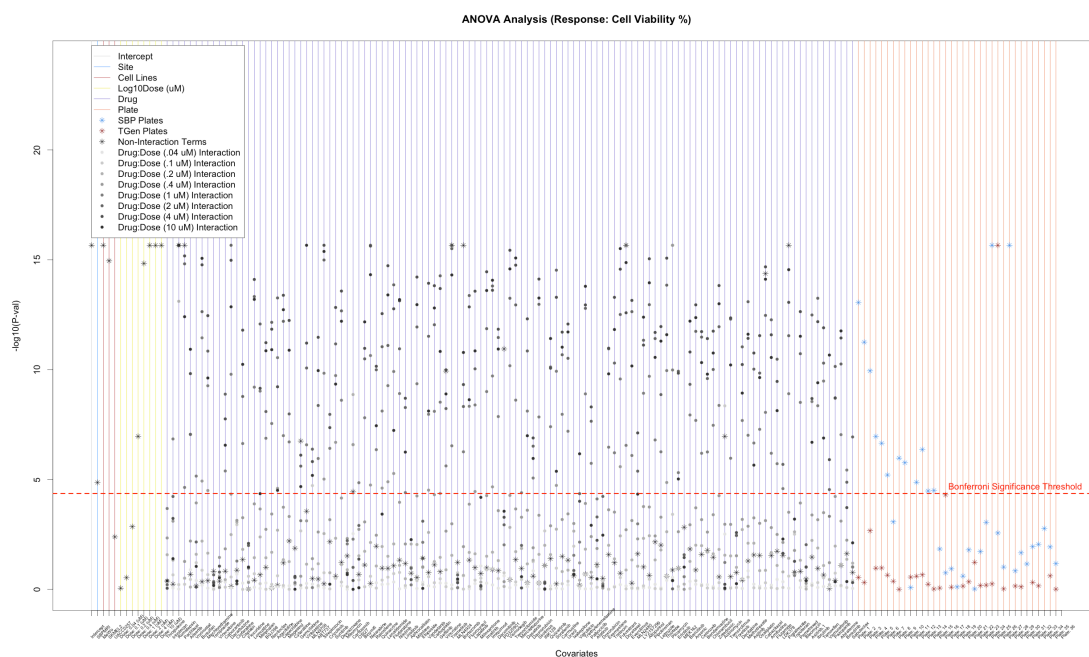


Figure 2.9: Assessing the variation in CV that is explained by site, cell lines, dose, drug, and drug-dose interaction. We find that many of the significant drug-dose interaction terms were from higher concentrations, indicating that a majority of the variation significantly accounted for by dose was limited to higher concentrations.

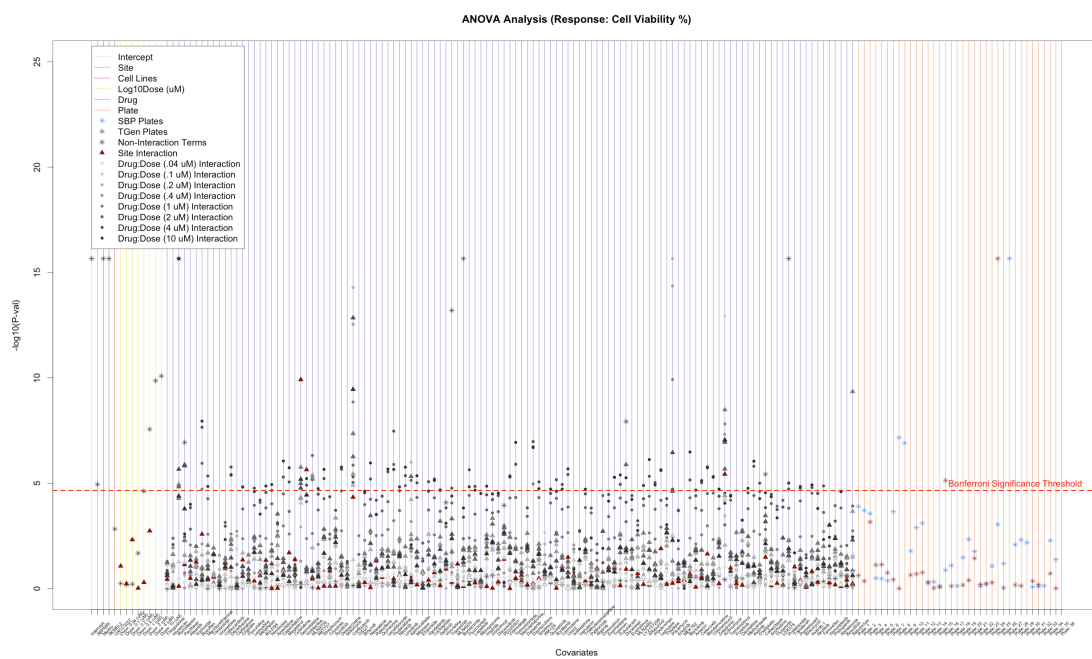


Figure 2.10: Assessing the variation in CV that is explained by site, cell lines, dose, drug, and 3-way interaction terms. Site-by-drug-by-dose effects explained 2.49% of cell viability variance.

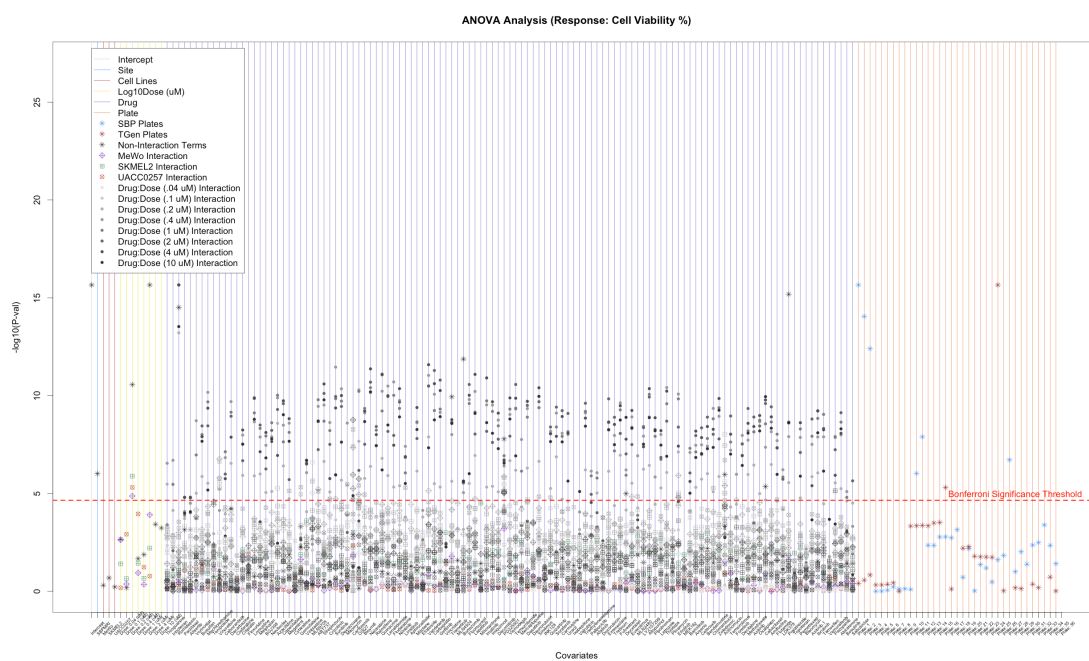


Figure 2.11: Assessing the variation in CV that is explained by site, cell lines, dose, drug, and 3-way interaction terms. Drug-by-dose-by-cell line explained 3.6% of the variance.

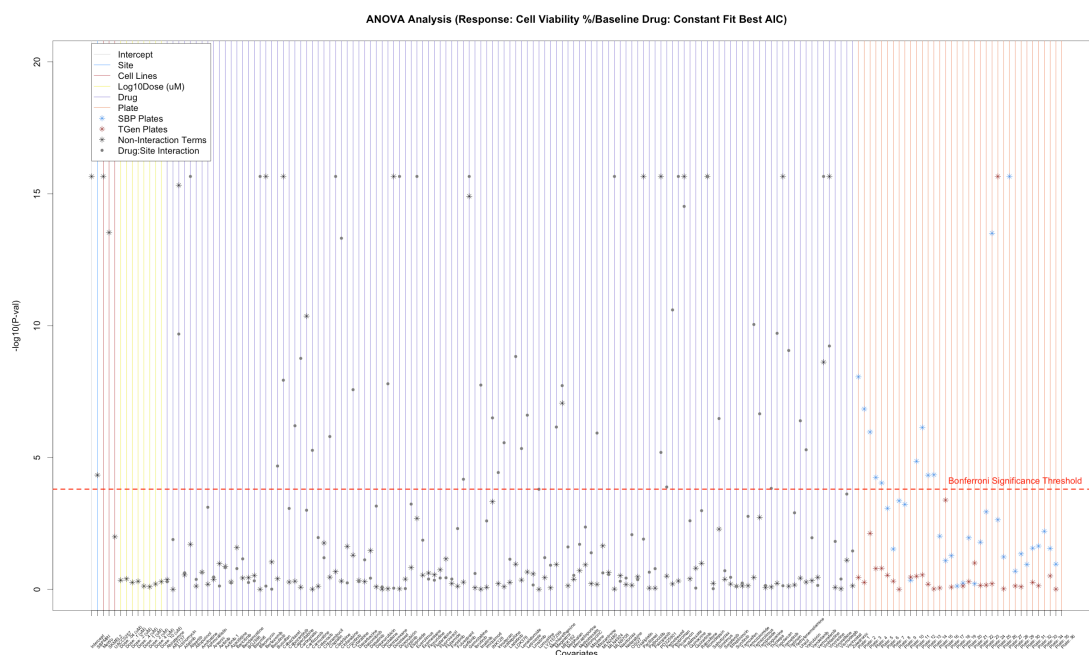


Figure 2.12: Re-assessing the variation in CV that is explained by site, cell lines, dose, and drug when using AIC-defined control drugs. Results were similar from using random baseline drug in terms of proportion of variation explained by sites, cell lines, dose, and drug. However, a smaller number of statistically significant drug effects were observed.

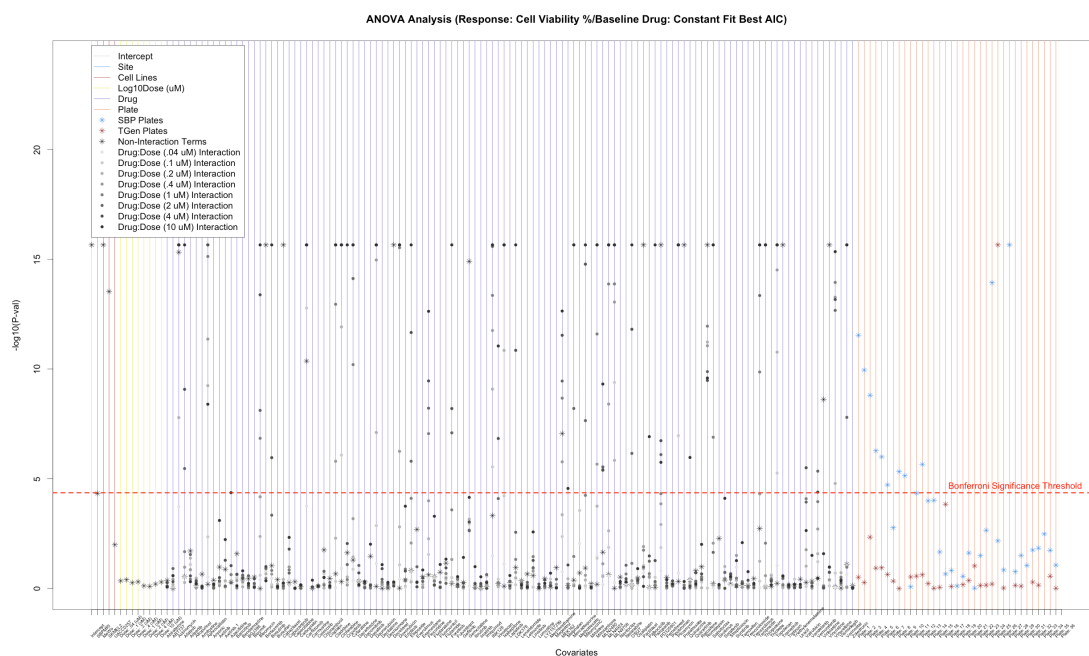


Figure 2.13: Re-assessing the variation in CV that is explained by site, cell lines, dose, drug, and dose-drug interaction when using AIC-defined control drugs.



Figure 2.14: Re-assessing the variation in CV that is explained by site, cell lines, dose, drug, and site-dose-drug interaction when using AIC-defined control drugs.

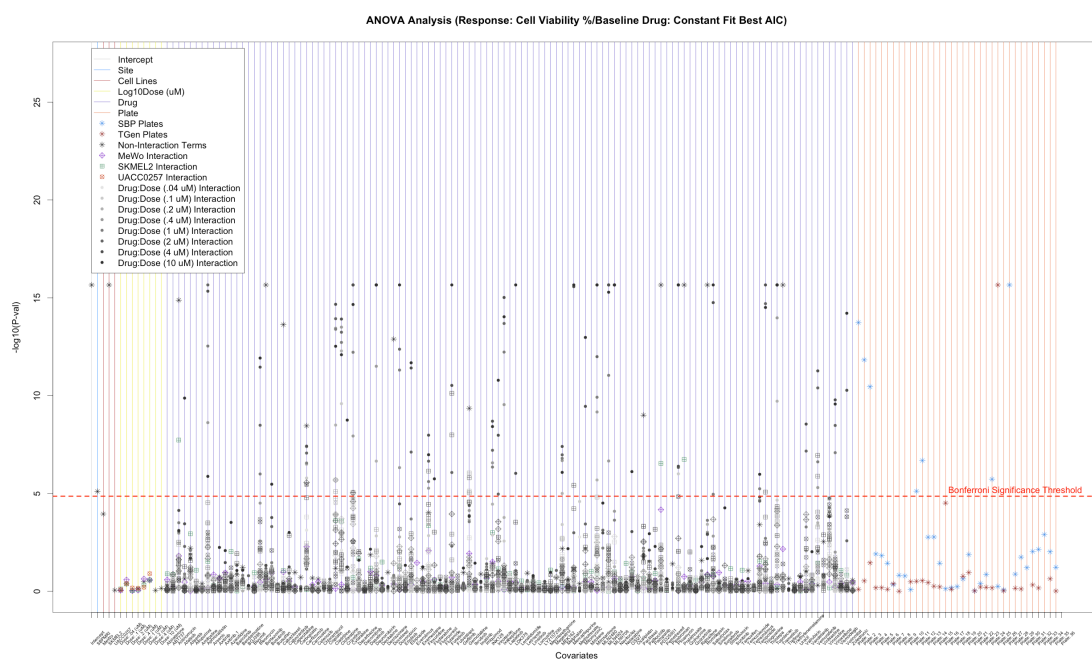


Figure 2.15: Re-assessing the variation in CV that is explained by site, cell lines, dose, drug, and cell line-dose-drug interaction when using AIC-defined control drugs.

Chapter 3

An Additional Example of Quality Control Analyses for Drug Screening and Drug Target Validation Involving Longevity in Model Species

3.1 Abstract

Model species have been used to identify drugs and drug targets that delay the onset of age-associated diseases and promote longevity. “Forward pharmacology” approaches are used routinely when compounds screened on a large scale impact lifespan in a model species amenable to high-throughput characterization such as yeast or *Caenorhabditis elegans* [Powers et al., 2006] [Park et al., 2011] [Torres et al., 2013] [Ye et al., 2014] [Miller et al., 2007]. The genes that emerge from these studies are then subjected to various manipulations or drug targeting studies to see if those manipulations impact longevity [Miller et al., 2007] [Strong et al., 2008] [Harrison et al., 2009] [Miller et al., 2011].

Under this approach, quality control issues arise and must be dealt with

overtly to avoid false positive and false negative results. Similar to high throughput studies in cancer, screens involving large contrasts among responses to different drugs potentially affecting longevity of a single species are likely to include experimental “noise” (e.g., inter- and intra- lab variability — plate effects, etc.). If not carefully modeled and quantified, these factors could potentially affect and bias the downstream association tests. In the following, methods are discussed that address the aforementioned concerns in large-scale drug screening studies involving *C. elegans*.

3.2 Introduction

Model organism and *in vitro* studies are used in various settings to understand human biology, especially in the contexts of drug response and drug targeting studies. Given ethical restraints and limited resources for human clinical trials, *in vitro* models are often desirable and effective substitutes. However, *in vitro* studies, even in ideal laboratory conditions, may not reveal drug response patterns translatable to humans *in vivo*. In light of this fact, animal models have been subject to increasing scrutiny [Aitman et al., 2011] [Warren et al., 2016]. Animal models that facilitate the identification of human biological factors and drugs include studies of disease mechanisms [Jubb et al., 2016], drug candidate screens [Ye et al., 2014] and associations with longevity [Warner, 2006]. For example, in the context of cancer, xenograft models are now routinely used to study disease mechanisms and to characterize DNA levels, RNA levels, and even differential protein expression patterns — all to shed light on mechanisms that influence anti-tumor drug response. In addition, model organism-based high throughput drug screens are often used to simulate human disease conditions so that sensitivity and resistance to drugs designed to treat those conditions can be studied. Ultimately, animal models have the potential to uncover mechanisms impacting drug response that would be otherwise unattainable in human studies. Further, any biological and technical limitations in such studies are probably less significant than those in purely *in vitro* studies.

The use of model organisms in studies of longevity has a long and somewhat complicated history. To evaluate drugs that might affect lifespan, model organisms with relatively short life expectancies can be used to both minimize the duration of drug investigative studies and maximize the number of drugs tested. But any results suggesting strong associations between increased lifespan and particular drugs should account for and accommodate “noise” before expending additional valuable resources and especially before involving human subjects. As an example, in a recent study involving *Caenorhabditis elegans*, 57 compounds were actually found to increase longevity [Ye et al., 2014] [Carretero et al., 2015]. However, the analyses pursued to identify these compounds did not quantify or account for experimental variability which potentially impacted the results. As a result, we obtained the raw data from this study in order to assess the proportion of variation in the drug response outcomes that could be attributed to technical variation, such as the specific plating of the experiments. Based on our findings, we then re-evaluated the drugs and compounds considered in the study while accommodating and quantifying the experimental and biological sources of variation.

In the following, I describe our experiences and results trying to accommodate sources of variation in the large-scale drug screening study involving *C. elegans*. Based on our experience, care must be taken in studies like these going forward if compelling insights into drugs and drug targets affecting lifespan are to be obtained.

3.3 Materials and Methods

3.3.1 *Caenorhabditis elegans* Data Acquisition and Analysis

In collaboration with The Scripps Research Institute (TSRI), we obtained data on *Caenorhabditis elegans* from a previous study conducted by Michael Petrascheck [Ye et al., 2014] [Rangaraju et al., 2015b] [Carretero et al., 2015] [Rangaraju et al., 2015a]. The data consists of 1280 different compounds that are

grouped into pharmacological classes according to their mammalian targets. This commercial collection of pharmacological agents with known or suspected targets in humans is known as the Library of Pharmacologically Active Compounds (LOPAC). As described in Petrascheck’s published results, animals were grown in liquid medium at 20 °C in 96-well plates with 5-15 animals per well [Ye et al., 2014] [Rangaraju et al., 2015b] [Carretero et al., 2015] [Rangaraju et al., 2015a]. Then, the animals were treated with either 33 μ M concentration of a drug or vehicle control (.33% DMSO). From the 1280 LOPAC drug screen, we obtained plate IDs; row and column specifications; drugs treated; pharmacological classes; and number of animals alive, dead, and days alive.

For each drug, we converted the number of animals alive and dead into a fraction of *C. elegans* alive. Then, using this fraction of *C. elegans* as the dependent variable, we applied multiple linear models to assess the proportion of variation explained by creating (0, 1) dummy variables for each experimental and biological factor (e.g., we had 1281-1 = 1280 dummy variables for the drugs; 56-1 = 55 for the pharmacological class; 16-1 = 15 for the plates, 12-1 = 11 for plate column; 8-1 = 7 for plate rows; 4-1 = 3 for replicates; and 2-1 = 1 for DMSO). For each set of factors, we also needed to specify the reference factor — we chose the control drug (DMSO only), the Vanilloid pharmacological class, plate 1, column 1, row 1, replicate 1, and DMSO 1. This approach allowed us to measure the influence of experimental factors and potential plate-specific outliers (e.g., entire plates, rows, columns, and row $\times$ column interaction terms). Additionally, these linear models allowed us to identify candidate drugs after accommodating these sources of variation. For example, we used interaction terms between days and drugs in order to identify drugs that potentially modulate lifespan. Significance was reassessed under these models and p-values were adjusted using Benjamini Hochberg FDR procedures [Benjamini and Hochberg, 1995]. We then compared significant drug candidates and pharmacological classes against the results from the previously published manuscript [Powers et al., 2006].

3.4 Results

3.4.1 *Caenorhabditis elegans* Screening Study

3.4.1.1 Assessing Sources of Variation in *Caenorhabditis elegans* Longevity Study

To ascertain the impact of experimental factors on the sources of variation in the study, we fitted linear models with only experimental factors and associated interaction terms. We found that the combined covariates and interaction terms explained about .35% of the variation (Table 3.1). Although none of the groups significantly affected the variation in the data, plots of the individual factors did reveal that there might have been an outlying column (Figure 3.1), suggesting that shifts within specific columns should be accounted for prior to making claims about observations. We then re-assessed the significance for the plate effects while including column $\times$ row interaction terms (Figure 3.2). Although none of the factors nor interaction terms produced p-values beyond a bonferroni adjusted threshold, we again found that column 10 produced strong p-values associated with the fraction of *C. elegans* alive. Moreover, the plots of the p-values indicated that these plate-effects were mainly attributable to three rows (3, 5, and 8) within the plate. In light of the interaction terms, we also noticed other potential experimental outliers (e.g., column 9 $\times$ row 3). These results provided further arguments that experimental factors — such as plate, column, row, and experimental site — should be included in the analyses of HTS studies.

3.4.1.2 Reassessing Candidate Longevity Drugs and Pharmacological Classes

To assess the potential impact of experimental factors on association tests, we reanalyzed the LOPAC drugs and pharmacological classes used in the study. Unlike previous methods, we performed association tests in the context of the overall study. However, when we analyzed the drugs, we could not use row $\times$ column interaction terms due to singularity issues (Table 3.2). To identify can-

Table 3.1: Assessing the proportion of variation in experimental data explained by each of the experimental factors, including column $\times$ row interaction terms. Significance codes for F statistics: p-val < .05 ('*'), p-val < .01 ('**'), p-val < .001 ('***')

	DF	Sum Sq	Mean Sq	F Value	Pr(> F)	Sig	% Var Explained
Plate	15	2.8	0.18571	1.9608	0.01424	*	0.05%
Column	11	3.5	0.3226	3.4063	9.65E-05	***	0.06%
Row	7	4.6	0.65047	6.8682	3.47E-08	***	0.08%
Replicate	3	0.2	0.06608	0.6978	0.55328		0.00%
DMSO	1	0.3	0.2546	2.6883	0.10109		0.01%
Column:Row Interaction	77	9	0.11663	1.2314	0.08229	.	0.16%
Residuals	60806	5758.8	0.09471				99.65%

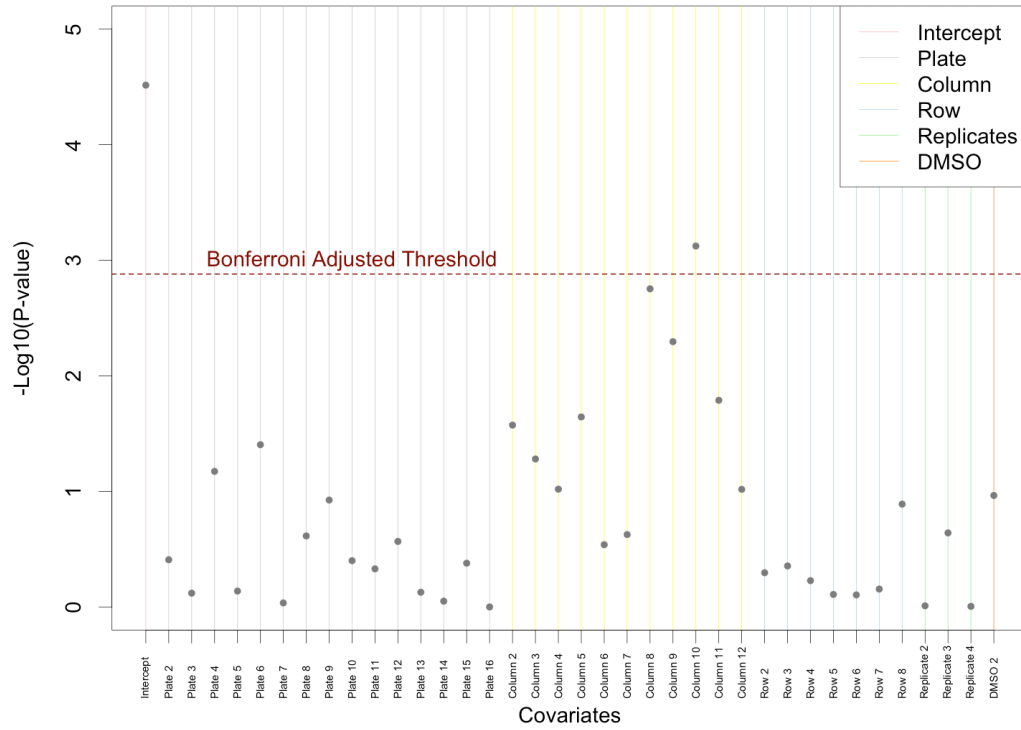


Figure 3.1: Plot of $-\log_{10}$ p-values for each experimental factor against the observed fraction of *C. elegans* alive. Vertical bars represent each group of factors analyzed with ANOVA.

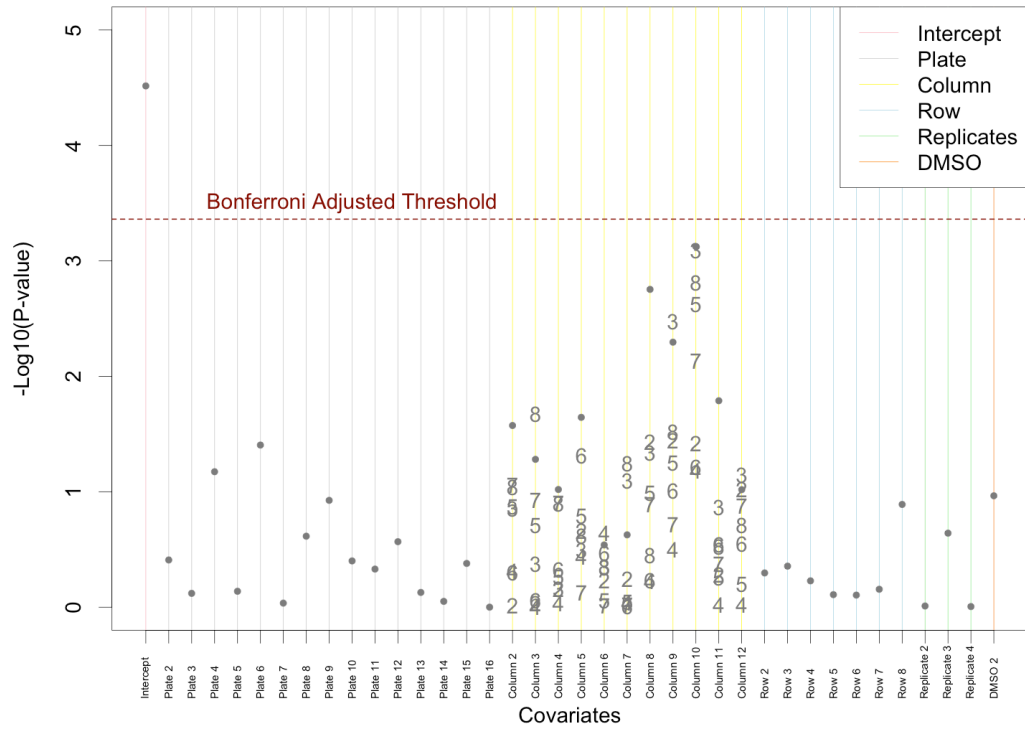


Figure 3.2: Plot of $-\log_{10}$ p-values for each experimental factor, including column $\times$ row interaction terms, against the observed fraction of *C. elegans* alive. Vertical bars represent each group of factors analyzed with ANOVA.

Table 3.2: Assessing the proportion of variation in experimental data explained by each of the experimental factors, drugs, and days (including drug x day interaction terms). Significance codes for F statistics: p-val < .05 ('*'), p-val < .01 ('**'), p-val < .001 ('***')

	DF	Sum Sq	Mean Sq	F Value	Pr(> F)	Sig	% Var Explained
Plate	15	2.79	0.19	3.8812	5.14E-07	***	0.05%
Column	11	3.55	0.32	6.7423	2.00E-11	***	0.06%
Row	7	4.55	0.65	13.5947	<2.2e-16	***	0.08%
Drug	1270	171.41	0.13	2.8209	<2.2e-16	***	2.97%
Days	1	2624.12	2624.12	54843.4665	<2.2e-16	***	45.41%
Replicate	3	0.33	0.11	2.2749	0.077713	.	0.01%
DMSO	1	0.42	0.42	8.8777	0.002888	**	0.01%
Drug:Days Interaction	1280	180.89	0.14	2.9536	<2.2e-16	***	3.13%
Residuals	58332	2791.04	0.05				48.30%

didate drugs for modulating lifespan, we also included interaction terms between days and drugs. After performing Benjamini Hochberg FDR corrections on the p-values for the drugs, we found only two drugs significant at an alpha level of .05 — Se(methyl)selenocysteine hydrochloride (adjusted p-value 0.00045) and Tyrphostin A9 (adj p-value 0.025). Strikingly, these two drugs were not reported as candidate longevity drugs in the original manuscript [Powers et al., 2006]. Upon relaxing the stringencies of the p-value threshold, we recovered two additional drugs that were significant in the original manuscript (Pergolide methanesulfonate and Phorbol 12-myristate 13-acetate; adj p-val .0996). However, when we evaluated significant drug x day interactions, we found 74 drugs significant (15 in common with previous findings) at the FDR adjusted p-value threshold of .01 and an additional 57 drugs significant (seven in common with previous findings) at the FDR adjusted p-value threshold of .05.

When analyzing the pharmacological class, we included row x plate interaction terms (Table 3.3). In these analyses, we were able to recapitulate the results from previous studies and found both Dopamine and Phosphorylation pharmacological classes to be highly significant (FDR adj p-val < 2.46e-45).

Table 3.3: Assessing the proportion of variation in experimental data explained by each of the experimental factors, pharmacological classes, and days (including pharmacological class $\times$ day interaction terms). Significance codes for F statistics: p-val < .05 (*), p-val < .01 (**), p-val < .001 (***)

	DF	Sum Sq	Mean Sq	F Value	Pr(> F)	Sig	% Var Explained
Plate	15	2.8	0.19	3.3982	8.39E-06	***	0.05%
Column	11	3.5	0.32	5.9032	1.12E-09	***	0.06%
Row	7	4.6	0.65	11.9029	2.97E-15	***	0.08%
Replicate	3	0.2	0.07	1.2092	0.3046		0.00%
DMSO	1	0.3	0.25	4.659	0.0309	*	0.01%
Pharmaco Class	53	9.1	0.17	3.1341	1.44E-13	***	0.16%
Days	1	2382.4	2382.4	43595.4136	<2.2e-16	***	41.22%
Column:Row Interaction	77	24	0.31	5.6917	<2.2e-16	***	0.42%
Pharmaco Class:Days Interaction	55	35.4	0.64	11.7623	<2.2e-16	***	0.61%
Residuals	60697	3317	0.05				57.39%

3.5 Discussion

Model organism studies have the potential to alter the pharmacological landscape, especially in the context of age association studies. However, issues exist organism-specific experiments. In species-specific HTS studies, we observed plate-specific effects that impact the downstream longevity association analyses. Upon accounting for these experimental factors, the overall pharmacological classes identified were the same, but we identified a large set of drugs associated with longevity that were not previously reported. Of the 57 previously identified drugs, our analysis identified 22 of the compounds after correcting for experimental effects and multiple hypothesis testing. Although the sets of analysis were vastly different — ours considered drug associations in the context of the overall screen, while previous methods used individual replicates and subsequent merging of results — it is difficult to reconcile the differences between these results without further experiments on candidate drugs. Still, these results suggest that experimental “noise” should be carefully considered and accounted for prior to downstream analyses and interpretation.

Our analyses highlight some of the common quality control issues faced when performing model organism studies. We accommodate these issues through the use of statistical and computational approaches. However, these methods

require careful record keeping of experimental factors so that they can be included as factors in downstream analyses.

3.6 Acknowledgements

Research supported by a Stand Up To Cancer Melanoma Research Alliance Melanoma Dream Team Translational Cancer Research Grant. Stand Up To Cancer is a program of the Entertainment Industry Foundation administered by the American Association for Cancer Research. Dr. Schork and Mr. Ding are also supported by NIH grants: U19 AG023122-09; R01 DA030976-05; R01 MH094483-03; R01 MH100351-02; R21 AG045789-01A1; UL1TR001442-01; and U24AG051129-01, in addition to grants from Janssen, Human Longevity, Inc., and the Tanner Project Foundation.

Chapter 4

Nonlinear Mixed Effects Models for Use in High Throughput Drug Screens: Applications to Melanoma Cell Line Analysis

4.1 Abstract

Cancer cell lines are used often in high throughput drug screens (HTS) to explore the relationship between characteristics of the cell lines (e.g., their genetic profiles) and responsiveness to different compounds and drugs. Many current analysis methods used to draw inferences about such relationships focus on one aspect of drug-specific dose-response curves (DRCs) obtained from the cell lines, the concentration causing 50% inhibition of a phenotypic endpoint (IC_{50}), that is often estimated in isolation from a small number of doses and responses gathered on each cell line. Such analysis methods do not consider properties of DRCs that might not be captured by IC_{50} values, nor do they leverage information about drug response patterns estimated with the different cell lines simultaneously, which could lead to increased false positive and false negative rates in drug response associations. We consider the application of two methods, each rooted in nonlinear mixed effects

(NLME) models, that test the relationship relationships between estimated cell line DRCs and factors that might mitigate response. Both methods leverage estimation and testing techniques that consider the simultaneous analysis of different cell lines to draw inferences about any one cell line. One of the methods is designed to provide an omnibus test of the differences between cell lines DRCs that is not focused on any one aspect of the DRCs (such as the IC_{50} value). We simulated different settings and compared the different methods on the simulated data. We also compared the proposed methods against traditional IC_{50} -based methods using 40 melanoma cell lines whose transcriptomes, proteomes, and, importantly, BRAF and related mutation profiles were available. Ultimately, we find that the NLME-based methods are more robust, powerful and, for the omnibus test, more flexible, than traditional methods. Their application to the melanoma cell lines reveals insights into factors that may be clinically useful.

4.2 Introduction

Cancers have proven notoriously difficult to treat because of their cellular complexity and heterogeneity, their ability to coopt a number of naturally-occurring molecular processes and factors to sustain their growth, their capacity to evade and adapt to treatments through a rapid evolution, and a general lack of available drugs that can combat different mechanisms contributing to their initiation and growth. The identification of specific factors and processes that reveal points of vulnerability in cancers should reveal therapeutic targets, but relevant strategies for doing so are far from trivial. In addition, there is now growing consensus that the factors contributing to the initiation and growth of a tumor are nuanced and often very individual-specific, so that the best way to treat cancer is to identify the potentially unique determinants of an individuals tumor and therapeutically target those determinants. Such “personalization” efforts for cancer treatments are receiving a great deal of attention [Garraway et al., 2013] [Shrager and Tenenbaum, 2014].

Strategies for identifying the very heterogeneous determinants of cancer

that might be used in the tailoring therapies to an individual are varied. Many large-scale epidemiological and tumor characterization efforts, such as the The Cancer Genome Atlas (TCGA) [The Cancer Genome Atlas Research Network et al., 2013] initiative, have proven to be useful but suffer from quality control and inadequate sample size issues [Buckley et al., 2016]. Strategies that consider testing different treatment regimens on individual patients until an effective therapy is found — which is often what happens spontaneously in the clinical care of patients refractory to initial treatments — is problematic from both biological and ethical standpoints, since purposely using a drug that might be known to be ineffective simply to test another drug is unethical. In addition, drugs have effects on tumors that change their composition, as noted, which could thwart the logic behind attempts to switch a patient to a pre-specified list of drugs [Schork, 2015].

Alternatives to large-scale epidemiological and focused clinical studies on individual patients to find ways of identifying connections between drugs and tumor characteristics include tumorgraft (or xenograft) models and tumor cell line-based studies. Implanting tumors in mice or other organisms and studying the engrafted tumors response to drugs has been used to great effect, but depends critically on the tumor environment in mice mimicking that in humans, which is not often the case [Lum et al., 2012] [Monsma et al., 2012]. Cell line based models derived from actual patient tumors have been used, also to great effect, for some time [Barretina et al., 2012] [Yang et al., 2013]. They are relatively easy and inexpensive to perform and can be used in high-throughput drug screening protocols (HTS), making them especially attractive for gaining at least initial insights into what drugs might be most effective against cells making up a tumor. Cell line models do suffer from the problem of extrapolating their *in vitro* responses to actual *in vivo* settings [Gillet et al., 2013], but their aforementioned ease of use and amenability to HTS makes them particularly attractive [Wang et al., 2016]. In addition, emerging strategies for creating cell lines using, e.g., induced pluripotent stem (iPS) cells, have added to their sophistication [Mercola et al., 2013].

Unfortunately, the reproducibility of cell line-based HTS experiments have been called into question [Haibe-Kains et al., 2013]. Differences in protocols used,

the media within which the cells are suspended, the precise formulation of the drugs used in the screening, the origins of the cell lines used, among other items, can create differences in outcome of studies involving what are thought to be the same cell lines. It is arguable that many of these technical sources of variation in cell line-based HTS studies can be identified and possibly controlled for in well-designed experiments [Ding et al., 2017]. However, the analytical methods used to draw inferences about the relationship between a subset of cell lines observed dose-dependent responses to a drug and characteristics of those cell lines (e.g., their genomic, transcriptomic, or proteomic profiles) also play an important (and often overlooked) role in the identification of factors that might mitigate a tumors response to a drug.

Traditional statistical methods for analyzing cell line-based HTS data involve fitting (often sigmoidal) dose response curves (DRCs) to each cell lines dose response data in isolation and then extracting a single estimated parameter, the concentration causing 50% inhibition of a phenotypic endpoint, like the cell growth, (i.e., the “ IC_{50} ”) value, from those curves [Haibe-Kains et al., 2013] [The CCLE and The GDSC Consortiums, 2015] [Shoemaker, 2006]. These IC_{50} values are then tested for association with other factors, like the expression levels of genes measured on those cell lines, to identify markers of response. Analyzing each cell line and drug combination in isolation is highly problematic since it ignores variation exhibited across the cell lines that might inform the response profile of any single cell line. Leveraging variation across all the cell lines when making claims about drug responsive in any single cell line can reduce noise and lead to more reliable inferences.

In this light, we considered the use of non-linear mixed effects (NLME) models in analyzing HTS dose-response data, which considers variation across the data on all or sets of the cell lines in order to estimate and assign a metric of drug responsiveness to each cell line. We considered two different NLME-based tests to compare to traditional methods. Both tests considered the evaluation of the cell lines collectively when estimating IC_{50} values and other DRC-related parameters. One test focuses exclusively on the IC_{50} values estimated from the

cell lines simultaneously and hence is similar to traditional methods that focus on IC_{50} values, whereas the other test leverages a Likelihood Ratio (LR) formulation of an omnibus test of differences in DRC profiles among groups of cell lines and is thus much more flexible in theory than IC_{50} -based tests since it can capture differences between the DRCs among subgroups of cell lines beyond IC_{50} values.

We applied the proposed NLME-based models and tests to 40 available melanoma cells lines with genomic, transcriptomic, and proteomic profiles available that had been screened against 120 drugs. We compared the results of analyses obtained with traditional, IC_{50} -based data analytic methods to the proposed NLME-based methods. We also considered analyses with the available Cancer Cell Line Encyclopedia [Garraway et al., 2013] [Barretina et al., 2012] (CCLE) data sets, and performed simulation studies to explore the relative advantages of each method in different situations that go beyond those focusing on IC_{50} values. We find that the proposed NLME-based tests are more powerful, sensitive and, for the omnibus test, flexible than traditional methods. In fact, the LR-based NLME test can be used to uncover associations among various factors collected on the cell lines, such as gene expression levels, and drug responsiveness in a very robust and compelling way, and that are simply beyond the reach of analyses focusing exclusively on IC_{50} values. Both proposed methods may reveal factors mitigating cancer drug response that could inform personalized or targeted therapeutic approaches to cancer in actual patient care.

4.3 Materials and Methods

We made use of a number of resources in our evaluation of different methods for analyzing cell line-based DRC data generated from HTS data. We briefly describe these resources before providing greater detail about the construction and execution of the two proposed NLME-based models and tests.

4.3.1 Cell Lines

Thirty-three melanoma cell lines from the University of Arizona Cancer Center (UACC) repository were used for our analyses. Additionally, the SK-MEL-2 cell line was received directly from Memorial Sloan Kettering Cancer Center through the Cancer Cell Line Encyclopedia (CCLE) repository and MeWo cell line from the Developmental Therapeutics Program’s NCI-60 repository. The Translational Genomics Research Institute (TGen) had access to all other UACC-cell lines through the UACC [Ding et al., 2017]. All cell lines were of low passage number. Cells were maintained according to the manufacturer’s or collaborators instructions. All media used to grow and harvest the cell lines had 10% FBS and 1% AA added to final growth media, and all cell lines were banked at low passages in multiple aliquots and liquid nitrogen stocks to reduce risk of phenotypic and genetic drift. All cells were cultured for less than three months before reinitiating culture from the frozen stock, were routinely inspected for identity by morphology and growth curve analysis and validated to be free of mycoplasma and contaminants.

4.3.2 Drug Screening

Each of the 40 cell lines was used in a nine-point (i.e., concentration/dose) HTS study (drug concentrations: 0.02, 0.04, 0.1, 0.2, 0.4, 1.0, 2.0, 4.0, and 10.0 μ M) across 120 drugs (Figure 4.1A, Table 4.3) at the Sanford Burnham Prebys Medical Discovery Institute (SBPMDI or SBP in the sequel). Drugs were spotted on 384-well clear bottom tissue culture treated plates (Greiner Bio-One, #781098) using an Echo Liquid Handler (Labcyte Inc., Sunnyvale, CA) such that addition of 25 μ Ls of cells (100k cells/mL in RPMI +10% FBS +Pen./ Strep./ Glut., Omega Scientific, Tarzana, CA) resulted in the above-described final drug concentrations and 2.5k cells/well. Upon plating, the cells were gently spun down at 1k rpm for one minute and incubated with drugs for 96 hours at 37 °C in a standard tissue culture incubator. After this time course, plates were allowed to equilibrate to room temperature for 30 minutes before μ Ls per well of freshly prepared CellTiterGlo reagent (G7571, Promega Corp., Madison WI) were added. Samples were incu-

bated for ten minutes with gentle agitation (100 rpm) before luminescence was read on a BioTek Synergy2 plate reader using Gen5 software (BioTek, Winooski, VT). Each plate was assayed in triplicate and included 24 vehicle-only DMSO controls.

4.3.3 Gene Expression Assays

The cell line samples were subjected to nucleic acid extraction, verification, amplification, and hybridization per the protocol regarding the use of the Affymetrix HG-U133 plus 2.0 arrays (54,675 probesets, Affymetrix, Santa Clara, CA). Affymetrix HG-U133 Plus 2.0 microarrays were normalized (background adjustment, interquartile normalization, and median polish) using robust multichip averaging [Irizarry et al., 2003] in R.

4.3.4 Phosphoproteomic Assays

Cell lysates from the cell lines were printed in triplicate onto nitrocellulose-coated slides (Grace Bio-Lab, Bend, OR) using a 2470 Aushon arrayer equipped with 185 μm pins (Aushon BioSystems, Burlington, MA) and subjected to Reverse Phase Protein Array (RPPA) analysis at George Washington University. For quality control purposes, standard curves were printed on each array along with the samples. To quantify the amount of total protein in each sample, selected arrays were stained with Sypro Ruby staining solution (Molecular Probes, Eugene, OR) pursuant to manufacturer recommendation. Before proceeding with immunostaining, arrays were first incubated in Reblot Stripping solution (Chemicon, Temecula, CA) for 15 minutes, washed twice with PBS (Life Technologies, Carlsbad, CA) and incubated in I-Block (Tropix, Bedord, MA) for one hour. Using an automatic system (Dako Cytomation, Carpinteria, CA), arrays were then incubated with commercially available 3% hydrogen peroxidase solution, avidin-biotin blocking system, and protein block (Dako Cytomation, Carpinteria, CA). The expression/activation level of a panel of FDA-approved and/or under investigation drug targets and their downstream effectors was measured using a single primary antibody targeting the protein and phosphorylation site of interest. Each

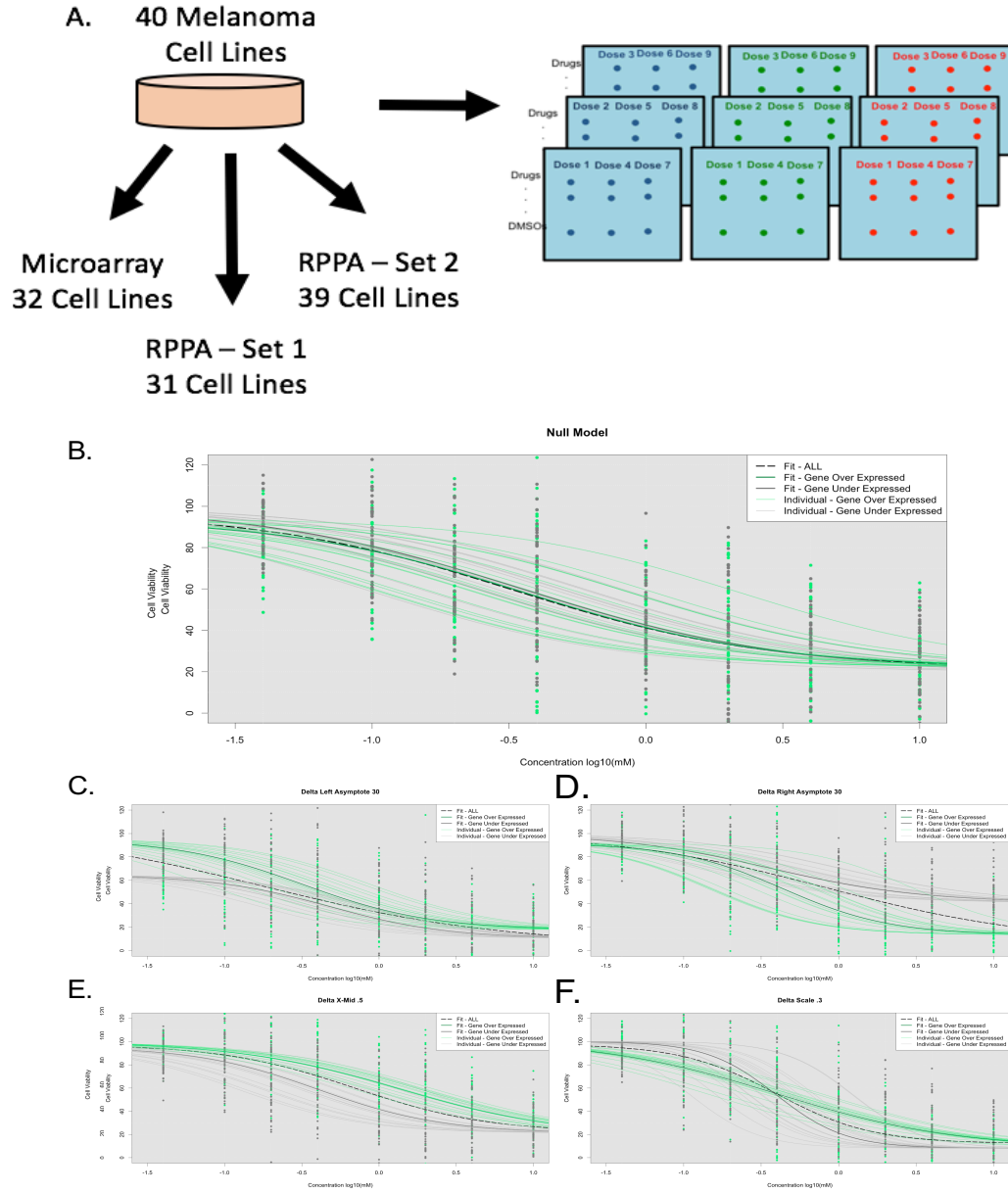


Figure 4.1: (A) Experimental design with melanoma cell lines processed with HTS, whole genome microarray genechip, and RPPA. We performed simulations under the HTS under the null model — no influence from GEX (B) and under the alternative model for left asymptote differences (C), right asymptote differences (D), X-Mid differences (E), Scale differences (F), and a combination of the parameters.

antibody used on the array was previously validated using western blot to confirm its specificity. A commercially available tyramide-based avidin/biotin amplification system (CSA; Dako Cytomation, Carpinteria, CA) and fluorescent detection (LI-COR Biosciences, Lincoln, NE) were used to quantify the amount of protein present in each sample. Antibody stained slides were scanned using the Tecan laser scanner (Tecan PowerScanner Tecan group Ltd., Mnnedorf, Switzerland). Images were analyzed with MicroVigene 5.1.0.0 (Vigenetech, Carlisle, MA) as previously described [Kaushik et al., 2014]. Because the output of the RPPA platform is quantitative, the intensity values obtained from the analysis were reported on a continuous variable.

4.3.5 CCLE Data Set

To provide another comparison and evaluation of the performance of different analysis methods, we downloaded and processed raw CEL files from the CCLE. We obtained gene expression profiles and genome-wide mutation calls for melanoma cell lines from the CCLE web portal.

4.3.6 Traditional IC_{50} DRC Analysis

For each of the 40 cell lines 9-point dose response data generated for each of 120 drugs from the melanoma drug screen, we fit DRCs using a traditional four parameter sigmoidal curve. The parameters governing these curves were: left asymptote of the curve, right asymptote of the curve, midpoint of the curve (referred to as the “x-mid” value), and a scale parameter [Commo and Bot, 2015]. All analyses, including those involving the proposed NLME models and simulation studies described below were performed on the San Diego Supercomputer Center (TSCC). To fit DRCs to each cell of the melanoma cell lines 9-concentration DRC data, we used the `nplr` [Commo and Bot, 2015] package in R. We used a similar DRC analyses for the CCLE data and simulated data described below. To determine whether the IC_{50} values were significantly associated with the gene expression, we used Pearson correlation tests with the paired gene expression and

estimated IC_{50} values across the cell lines.

4.3.7 NLME-Based Analyses

As noted, we considered two different NLME-based models and tests for the DRC data, each assuming the same four parameter sigmoidal curve as in the traditional IC_{50} analyses discussed in the previous section. For the first NLME-based model, which we will refer to as the “ IC_{50} NLME” model, we fit four parameter sigmoidal curves to all the cell lines simultaneously allowing for variation in IC_{50} values across the cell lines. We leveraged the lme4 [Bates et al., 2015] package in R with fixed effects for each of the four parameters in the sigmoidal curve and an additional random effect for the “x-mid” (i.e., an estimate of the IC_{50}) parameter while using the cell line as a grouping factor to do so. Note that the IC_{50} NLME analysis model produces IC_{50} values for each cell line that can be used in Pearson correlation tests to determine their association with other factors collected on those cell lines, such as gene expression levels, but just estimated simultaneously with the other cell lines.

For the second test, which we will refer to as the “Omnibus LR NLME” model and test, we fit four parameter sigmoidal curves to stratified subsets of the cell lines based on grouping factors, such as the cell lines expressing a gene above and below the median gene expression value (real or simulated, depending on the data set) or cell lines with and without a mutation. The likelihoods obtained from each groups curve fits were then used to calculate a log-likelihood ratio statistic comparing the DRCs between the two groups as:

$$LLR = -2 * LL_{all} + 2 * (LL_{Gene+} + LL_{Gene-})$$

Note that since the Omnibus LR NLME test does not focus on a single parameter, like the x-mid or IC_{50} value, it can in theory identify differences between DRCs obtained from two (or more) groups that could involve the left or right asymptote, the scale parameter, the x-mid parameter or any combination of these. To determine the significance of the observed LLR statistic, 100 permutations were performed for, e.g., the over and under expressed status for each cell line. For each

permuted sample, we calculated LLRs and for each drug i and each permutation j , we calculated C_{ij} as:

$$C_{ij} = \begin{cases} 1, & \text{if } LLR_{ij} > LLR_i \\ 0, & \text{otherwise} \end{cases}$$

Then, based on these permutations, for n total permutations, the p-value for the observed stratification was:

$$p_i = \frac{1 + \sum_{j=1}^n C_{ij}}{n + 1}$$

To save computational resources, when association test p-values were going to exceed 0.05 based on the number of permutations we set (i.e., they were not significant and could not achieve significance given the number of permutations), we stopped the permutations for that setting and proceeded to the next test setting. Additionally, we employed adaptive permutations [Che et al., 2014]; i.e., for the associations that were significant at a 0.05 p-value, we increased the number of permutations to meet multiple hypothesis testing criterion (e.g., we increased permutations to 10,000 when testing for associations between 23 genes and 15 drug tests).

We explored the utility of the proposed Omnibus LR NLME tests relative to the traditional tests in a number of scenarios with the melanoma cell lines. We computed statistics and p-values using the RPPA values stratifying the cell lines at the median intensity value. We also ran tests on all BRAF associated genes for which we had gene expression data and the top five most significant phosphoproteins using Pearson correlation tests for the IC50 NLME and traditional IC50 test. As a check on the robustness of the analyses, we also assessed significance on five random analytes for each drug. Bonferroni corrections were used to account for multiple comparisons given the number of genes, proteins and mutations we tested for association with drug responses across the cell lines. For the analyses involving the CCLE data set, we compared BRAF mutant and wild type cell lines, and compared the results against analyses using the IC50 values provided the Broad Institute for the CCLE data set [Barretina et al., 2012].

4.3.8 Simulation Based Tests

We simulated High Throughput Screen (HTS) nine-point Dose Response Curves (DRCs) data assuming a four parameter sigmoidal model. We note that each simulation required heavy computational resources (30-120+ hours and 8GB of RAM per simulated parameter set). We assumed that the standard deviation of the response for each concentration followed that of the observed drugs' standard deviations calculated from the actual melanoma cell line data. We simulated null models in which there was no assumed relationship between a hypothetical factor and drug response.^c In addition, we simulated cell line DRCs assuming sample sizes of 34, 50, and 100 in triplicate with a left asymptote of 100, right asymptote of 10, scale of -0.5, and x-mid values following a normal distribution with mean -0.5 and standard deviation of 0.3 (i.e., $\sim N(-0.5, 0.3)$) (Figure 4.1B). These values were determined based on averages from analyses involving all the actual melanoma samples. We then considered simulations in which half of the cell lines were randomly assigned as over-expressing a gene with a distribution of $N(8,1)$ and the other half were assigned as expressing that gene to a lesser degree with a distribution of $N(3,1)$. Based on these simulated parameters and standard deviation at each concentration, DRCs were simulated for each cell line. For each set of parameters, 100 simulations were performed and subsequent association tests were applied as follows: t-test of over-expressers vs. under-expressers with traditionally called IC_{50} , t-test of over-expressers vs. under-expressers with NLME model-derived IC_{50} values, and the Omnibus LR test with the NLME-fitted DRCs. Keep in mind that differences in the test results between the IC_{50} values from the traditional analysis and the IC_{50} NLME called IC_{50} values reflect the influence of analyzing all the cell lines together rather than individually.

To evaluate the power of the association tests by exploring the impact of changes in DRCs not associated with IC_{50} values, we simulated dose response curves under different alternative models. In these simulations, under-expressed genes were simulated with different sigmoidal curve parameters described in Figure 4.1C-E. Note that in some settings, we averaged the power of the simulations in which one parameter was fixed but the others were allowed to vary to determine

how easily the three tests could detect differences in DRCs between two groups that were not exclusively due to differences in IC_{50} values. To create realistic simulations, we explored various combinations (153 of the possible 192 combinations, approximately 80%) of the parameters and performed 100 simulations for each set of parameters.

4.4 Results

4.4.1 Simulation Based Tests

As noted, we simulated DRCs assuming triplicate cell line evaluations (to mimic are actual data) for sample sizes of 34, 50, and 100 triplicate cell lines. Based on the results of the analyses of this simulated data using each method (i.e., traditional tests using individual cell line-based IC_{50} values, the IC_{50} NLME test and Omnibus LR NLME tests) we calculated the false positive rate or type I error rate (i.e., using the null model simulation settings) and the false negative rate, power or type II error rate (using the alternative model simulation settings). Power analysis revealed that 100 simulations for these analyses would suffice (Figure 4.6). For the type I error rate, we found that each of the three association tests produced results close to the expected value of 5% false positives, regardless of the number of cell line samples. However, the simulation results for the Omnibus LR NLME tests suggested it performed better in nearly every scenario (with the exception of the largest right asymptote changes, where the IC_{50} NLME association tests performed better (Table 4.1). We note that for Table 1, the three tests can be seen as not being equally sensitive to differences in DRCs that exhibit differences in parameters beyond the IC_{50} , as we fixed one of the parameters and averaged the estimates of the power estimates when the other parameters were allowed to vary. This was to be expected, since the traditional IC_{50} test that estimates IC_{50} values for each cell line independently and then explores the relationship between those estimated IC_{50} values and some other factor on those cell lines and essentially ignores other properties about the cell lines. The NMLE IC_{50} test, although focusing on the IC_{50} values, is influenced by the other DRC parameter differences between the

Table 4.1: The Mean and SD of power observed when a single parameter is fixed between two groups and all other parameters are varied in one of the groups (see Methods section).

Fixed Parameter	Varied Parameters	Trad IC50	NLME IC50	LLR
Left Asym = 100	Right Asym, X-Mid, Scale	0.12 +/- 0.11	0.73 +/- 0.35	0.77 +/- 0.24
Left Asym = 90	Right Asym, X-Mid, Scale	0.13 +/- 0.10	0.61 +/- 0.38	0.73 +/- 0.24
Left Asym = 80	Right Asym, X-Mid, Scale	0.11 +/- 0.08	0.42 +/- 0.32	0.85 +/- 0.07
Left Asym = 70	Right Asym, X-Mid, Scale	0.09 +/- 0.08	0.49 +/- 0.33	0.83 +/- 0.09
Right Asym = 10	Left Asym, X-Mid, Scale	0.12 +/- 0.10	0.41 +/- 0.33	0.69 +/- 0.29
Right Asym = 20	Left Asym, X-Mid, Scale	0.12 +/- 0.10	0.46 +/- 0.33	0.78 +/- 0.14
Right Asym = 30	Left Asym, X-Mid, Scale	0.12 +/- 0.10	0.61 +/- 0.36	0.87 +/- 0.08
Right Asym = 40	Left Asym, X-Mid, Scale	0.11 +/- 0.08	0.93 +/- 0.13	0.86 +/- 0.08
X-Mid = -0.5	Right Asym, Left Asym, Scale	0.05 +/- 0.02	0.48 +/- 0.34	0.79 +/- 0.21
X-Mid = -0.4	Right Asym, Left Asym, Scale	0.05 +/- 0.02	0.47 +/- 0.35	0.76 +/- 0.23
X-Mid = -0.25	Right Asym, Left Asym, Scale	0.12 +/- 0.03	0.59 +/- 0.37	0.80 +/- 0.20
X-Mid = 0	Right Asym, Left Asym, Scale	0.28 +/- 0.05	0.79 +/- 0.32	0.82 +/- 0.10
Scale = -0.5	Right Asym, Left Asym, X-Mid	0.12 +/- 0.09	0.62 +/- 0.36	0.74 +/- 0.20
Scale = -0.4	Right Asym, Left Asym, X-Mid	0.12 +/- 0.11	0.58 +/- 0.35	0.76 +/- 0.24
Scale = -0.2	Right Asym, Left Asym, X-Mid	0.12 +/- 0.09	0.53 +/- 0.38	0.88 +/- 0.07

two groups since it estimates parameters for the groups using all the data and can therefore detect differences in DRCs beyond differences in IC_{50} values.

When exploring specific settings, the association tests using traditionally derived IC_{50} values (i.e., that are estimated independently from each cell line) exhibited a type 1 error rate consistent with the null model (e.g., 0.05) in cases where differences in DRCs were simulated that involved changes in left asymptote, right asymptote, and scale (Table 4.2). Note that in Table two, the values highlighted in red denote differences in assumed parameter values between the two groups. Moreover, we do find that the traditional IC_{50} association tests performed significantly poorer when the x-mid parameters were fixed (i.e., x-mid were the same in both groups; first two rows of Table 4.2). This was expected since, as noted, the IC_{50} association tests are designed to detect differences between changes in the X-Mid (i.e., estimated IC_{50} values) parameters (Table 4.2). However, in both the IC_{50} NLME test and Omnibus LR NLME tests, there is a marked improvement in the power to detect the DRC differences (Table 4.2). Although there was a slight

improvement in the simulated settings where differences in x-mid values were assumed, interestingly the traditionally-derived IC_{50} association tests (i.e., where each cell lines DRC x-mid or estimated IC_{50} value is estimated independently) exhibited much less power.

We compared the statistical power of each of the tests using paired two-sample t-tests and found that the traditionally derived IC_{50} association tests performed worse than the NLME (paired 2-sample t-tests statistics = 11, pval < 2.2e-16) and LLR tests (paired 2-sample t-tests statistics = 27, pval < 2.2e-16). Even upon increasing the number of sample sizes to 50 and 100 cell lines, the Omnibus LR NLME tests had greater statistical power than IC_{50} NLME and traditional IC_{50} -based association tests (Table 4.5, Figure 4.7). As noted, when the x-mid parameter is held constant, i.e. the x-mid is the same in both group, then the traditional IC_{50} association tests are close to the assumed Type I error rate, as expected (Table 4.2). Ultimately, we observed that the Omnibus LR NLME tests exhibited the greatest power to detect differences in the dose response curves, without sacrificing type I errors or generating many false positive associations. However, it also proved to be the most computationally demanding method for association tests.

4.4.2 Comparing the Different Methods on the Gene Expression Data

To assess the properties of association tests based on traditionally estimated IC_{50} (x-mid) values against the IC_{50} NLME test, we performed association tests using drugs with NLME curve fits to cell lines that had at least one cell line with cell viability less than 20%. Since a subset of the drugs in our study target the BRAF-mediated pathway, we suspect that differential drug response may be observed based on the BRAF associated genes' expression patterns. However, based on analyses of these 15 drugs, no BRAF associated genes were significantly associated when using IC_{50} NLME test, and only four significant associations were found with traditional IC_{50} calls: HSPA9 (adjusted p-value < 0.027) and KSR1 (adjusted p-value < 0.028) with Vorinostat; RAF1 (adjusted p-value < 0.007) and

Table 4.2: Simulation-based power to detect differences between DRCs between two groups assuming different features about the groups DRCs.

Group 1 Parameters					Group 2 Parameters					Power		
Simulated CLs	Left	Right	XMid	Scale	Simulated CLs	Left	Right	XMid	Scale	Trad IC50	NLME IC50	LLR Test
50	100	10	-0.5	-0.5	50	100	10	-0.5	-0.5	0.06	0.11	0.01
100	100	10	-0.5	-0.5	100	100	10	-0.5	-0.5	0.04	0.06	0.04
50	100	10	-0.5	-0.5	50	100	10	-0.4	-0.5	0.07	0.18	0.02
50	100	10	-0.5	-0.5	50	100	10	-0.25	-0.5	0.16	0.66	0.24
50	100	10	-0.5	-0.5	50	100	10	0	-0.5	0.39	1.00	0.90
50	100	10	-0.5	-0.5	50	100	10	-0.5	-0.4	0.00	0.06	0.49
50	100	10	-0.5	-0.5	50	100	10	-0.5	-0.2	0.06	0.08	0.86
50	100	10	-0.5	-0.5	50	100	20	-0.5	-0.5	0.02	0.34	0.81
50	100	10	-0.5	-0.5	50	100	30	-0.5	-0.5	0.07	0.90	0.99
50	100	10	-0.5	-0.5	50	100	40	-0.5	-0.5	0.04	1.00	0.88
50	100	10	-0.5	-0.5	50	90	10	-0.5	-0.5	0.05	0.34	0.69
50	100	10	-0.5	-0.5	50	80	10	-0.5	-0.5	0.04	0.86	0.98
50	100	10	-0.5	-0.5	50	70	10	-0.5	-0.5	0.02	1.00	0.98
100	100	10	-0.5	-0.5	100	100	10	-0.4	-0.5	0.09	0.20	0.07
100	100	10	-0.5	-0.5	100	100	10	-0.25	-0.5	0.29	0.96	0.63
100	100	10	-0.5	-0.5	100	100	10	-0.5	-0.4	0.05	0.07	0.84
100	100	10	-0.5	-0.5	100	100	40	-0.5	-0.5	0.09	1.00	1.00
100	100	10	-0.5	-0.5	100	90	10	-0.5	-0.5	0.05	0.63	0.99
100	100	10	-0.5	-0.5	100	80	10	-0.5	-0.5	0.09	1.00	1.00
100	100	10	-0.5	-0.5	100	70	10	-0.5	-0.5	0.04	1.00	1.00

Key: Values shown present the number of cell lines (CLs), the left asymptote, right asymptote, X-Mid, and Scale for each simulation. Power results were calculated using tests based on traditionally estimated IC_{50} values, NLME estimated IC_{50} , and the Omnibus LLR tests. In each simulation, one group was held constant with the following parameters. The values highlighted in red denote differences in the assumed parameter values for the two groups. Note that the first two rows are simulations under the null hypothesis (i.e., no difference in parameter settings between the two groups).

HSPA9 (adjusted p-value < 0.03) with Crizotinib. Scatterplots comparing these correlations highlighted the differences in the nature of the associations between the traditional IC_{50} and IC_{50} NLME-based methods (Figure 4.2A). Although there was a general positive correlation, we observed a great deal of variation observed when using the two different curve fit methods.

4.4.3 Comparing the Different Methods on the RPPA Data

Similar to the gene expression analysis, we compared the associations between proteins and drug responses defined by the traditional IC_{50} and the IC_{50} NLME-based tests across the cell lines (Table 4.5). As with the results observed for the gene expression correlation analysis, we found that using different methods for obtaining the IC_{50} parameters yielded different sets of genes associated with drug response (Figure 4.2B-C). Across all drugs, we found only significant RPPA associations when using IC_{50} NLME tests (Caspase 6 (cleaved Asp162) with the drug Thioguanine Bonferroni adjusted p-value = 0.026). However, using traditional methods, we found seven significant associations in the first RPPA set and eleven significant associations in the second RPPA set.

4.4.4 Assessing the Omnibus LR NLME Test

To test whether gene expression was associated with overall drug response profiles and not just IC_{50} values, we stratified cell lines based on their gene expression levels for those genes in the BRAF pathway. For each gene, the samples were either labeled as over-expressed or under-expressed based on the median intensity level across all cell lines. LLRs were calculated for each gene and drug combination, and permutation-based tests were performed to assess statistical significance (Figure 4.3A-B, Table 4.5). Notably, only a small fraction of the differentially expressed genes were significantly associated with drug response. This result is particularly noticeable in the heatmaps representing the differential effects (Figure 4.3A). Specifically, the more significant gene associations appear to be limited to Daunorubicin, Doxorubicin, Vorinostat, and Cladribine (Figure 4.3A). To ac-

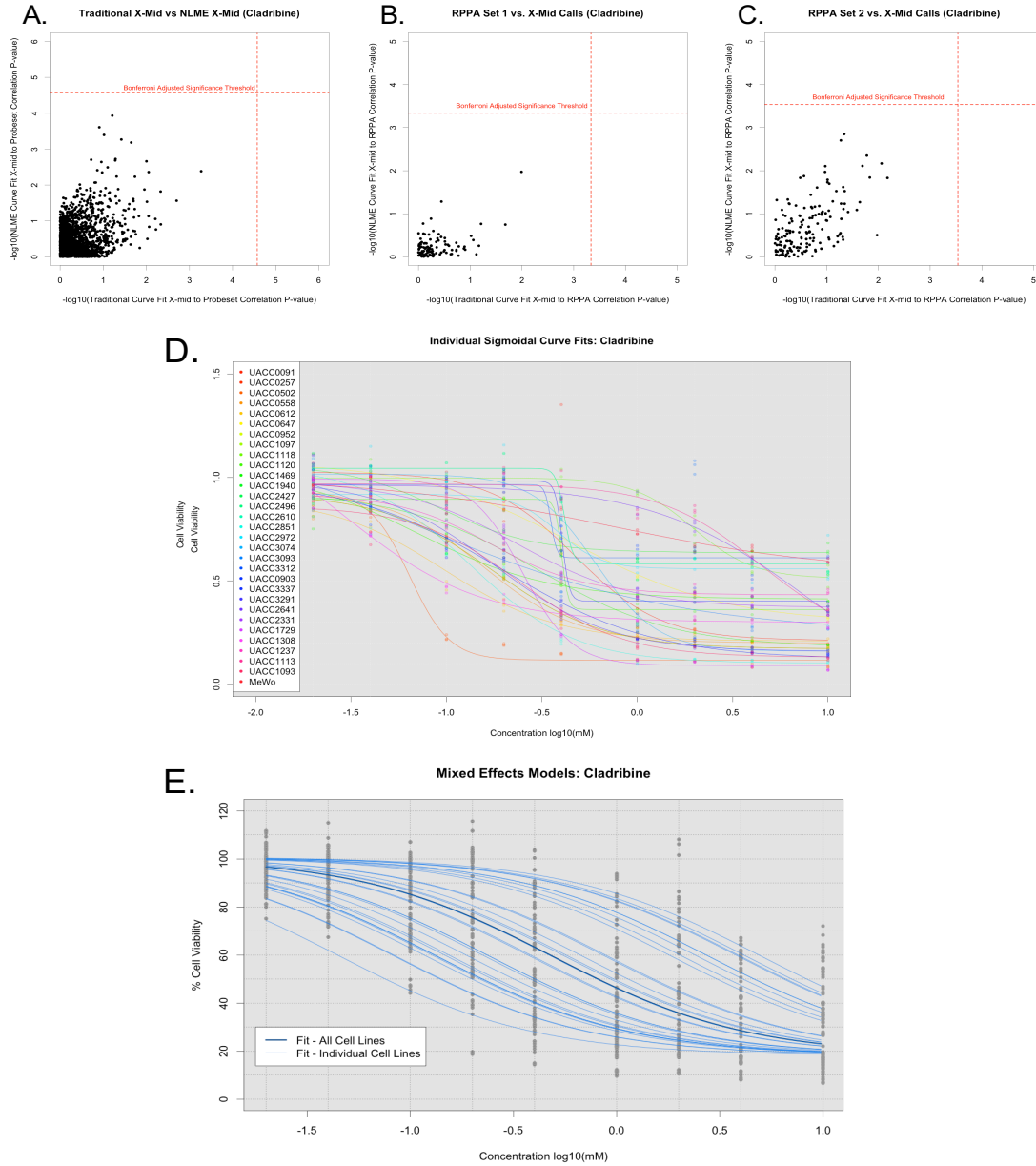
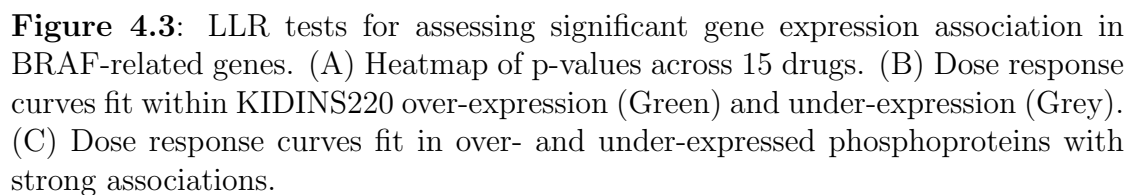


Figure 4.2: Differences observed between using traditionally called IC_{50} , NLME called IC_{50} , and LLR tests. (A) $-\log_{10}p$ -values for associations between probe set expressions and traditionally called IC_{50} or NLME called IC_{50} . (B) $-\log_{10}p$ -values for associations between RPPA set 1 expression and traditionally called IC_{50} or NLME called IC_{50} . (C) $-\log_{10}p$ -values for associations between RPPA set 2 expression and traditionally called IC_{50} or NLME called IC_{50} . (D) Dose response curves fit individually for Cladribine. (E) Dose response curves leveraging information across cell lines, which balances inter- and intra- cell line variability.

count for multiple hypothesis testing, we increased the permutations to a total of 9,999 for those with potentially significant p-values based on the initial Omnibus LR NLME tests. We found a single gene association significant beyond multiple hypothesis testing corrections (KIDDINS220 with Daunorubicin, p-value $< 1e-4$). However, there were several other cases that yielded strong (but not beyond multiple hypothesis testing correction thresholds) association between gene expression and drug response (Figure 4.3C).

4.4.5 Assessing the Consistency Between the RPPA and Gene Expression Analyses

To investigate potential relationships between gene and protein expression, we used un-adjusted NLME association p-values associated with each drug to identify proteins encoded by specific genes that were also seemingly associated with drug response. Using this strategy, many genes exhibited gene expression levels and their corresponding protein levels that were significantly associated with drug response. After consistencies between the GEX and RPPA analyses were found, we mined molecule-interaction databases to identify a biological basis for the observed results. Admittedly, this was not an exhaustive search, but we found several compelling cases to support efforts to combine insights from transcriptomics and proteomics using the NLME models and LLR tests. For Daunorubicin, we found the increased expression of phosphorylated MSK1 (anti-MSK1: phosphor S360) conferred resistance, whereas the increased expression of phosphorylated PKC (Phospho-PKC α/β II: Thr638) and PKC α (anti-PKC: Ser657) were associated with sensitivity to the drug (Figure 4.3C, Figures 4.8-4.9). In another compelling example, over-expression of MAPK3 was a significant predictor of response to Cladibrine. Furthermore, we observed that cell lines responsive to Cladibrine also had lower expression in the protein pBAD S112, which is known to be associated with the BCL pathway (Figures 4.10-4.11). Similarly, we found that over-expression of the KSR1 gene was associated with response to the drug LDK378 (Figure 4.12). KSR1 and FKHR have many physical interactions or are co-expressed with many of the same genes (Figure 4.13).



4.4.6 Analysis of the CCLE data with NLME-Based Models

To explore whether or not the manner in which the IC_{50} values were estimated had an impact on the identification of relationships between drug response and gene expression values, we fit NLME models across all the CCLE cell lines for each available drug and compared the results of these NLME-assigned x-mid (IC_{50}) values against the available IC_{50} values computed by the CCLE research team (Figure 4.14). Although we observed a general concordance between the two IC_{50} estimates, we found that the methods yielded different values across the cell lines. Additionally, fitting NLME curves across cancer types allowed us to assess each drug’s response profile within cancer types. For example, we identified AZD-6244 as a candidate drug in melanoma since it had a more significant NLME curve fit relative to other cancer types.

4.4.7 Including Mutation Status as Random Effect in CCLE Data

Fitting NLME curves within groups of cell lines stratified by mutation status could provide more accurate IC_{50} assessments if mutation status impacts drug responsiveness. Based on our NLME curve fits to the CCLE cell lines, we found that, overall, melanoma cell lines were more sensitive to the drug AZD-6244 than other cancer types. To assess the impact of incorporating mutation status into NLME curve fits, we stratified the cell lines into BRAF mutant versus BRAF wild type cell lines. Although the mutation status did not impact curve fits for most CCLE drugs, we found that the BRAF wild type samples were much more resistant to AZD-6244 (Figure 4.4A). NRAS mutations were also associated with response to AZD-6244. However, melanoma cell lines with NRAS mutations appeared to have a greater sensitivity than those with BRAF mutations (Figure 4.4A).

To assess concordance between our UACC melanoma cell lines and the CCLE melanoma cell lines, we compared NLME curve fits by subdividing the cell lines by BRAF and NRAS mutation status in both data sets. We found cases where

mutation status had an opposite impact on curve fits between the two data sets – though this is likely due to differences in cell lines used. For example, for the drug Topotecan, we found that NRAS mutations led to greater sensitivity and BRAF mutations led to greater resistance in the CCLE cell lines (Figure 4.4B) compared to the UACC cell lines. However, in our UACC cell lines, both NRAS and BRAF mutations led to greater sensitivity to the Topotecan (Figure 4.4C). This suggests that while the mutation status could help improve NLME curve fits, careful data analyses should be considered. In this particular case, we suggest incorporating the NRAS mutation status, but omitting the BRAF mutation status, in fitting NLME curves for Topotecan in melanoma cell lines.

4.5 Discussion

HTS studies using tumor cells lines can illuminate the relationship between drug responsiveness and various factors collected on those cell lines (like gene expression levels, mutation status, etc.). Such relationships can be exploited clinically to “personalize” cancer treatments. However, there have been recent concerns surrounding the reliability HTS studies using tumor cell lines [Garraway et al., 2013] [Shrager and Tenenbaum, 2014] [Barretina et al., 2012] [Yang et al., 2013] [Gillet et al., 2013] [Wang et al., 2016]. One question about such screens concerns the analysis methods used to draw inferences about the relationships between cell line drug responses and characteristics of the cell lines. Traditional methods of measuring drug response in HTS consider cell lines in isolation and do not leverage information across all available cell lines used in the screen. To overcome this limitation, we considered NLME models to analyze multiple dose response curves simultaneously to account for the variability across the cell lines. Thus, instead of using single parameters extracted from individual dose response curves to assess associations, we use overall curve fits to a group of cell lines that accommodates association tests. The two tests we introduced, the IC_{50} NLME-based test and the Omnibus LR NLME test, each have advantages and disadvantages. The clear advantage of the Omnibus LR test is that it does not consider a single parame-

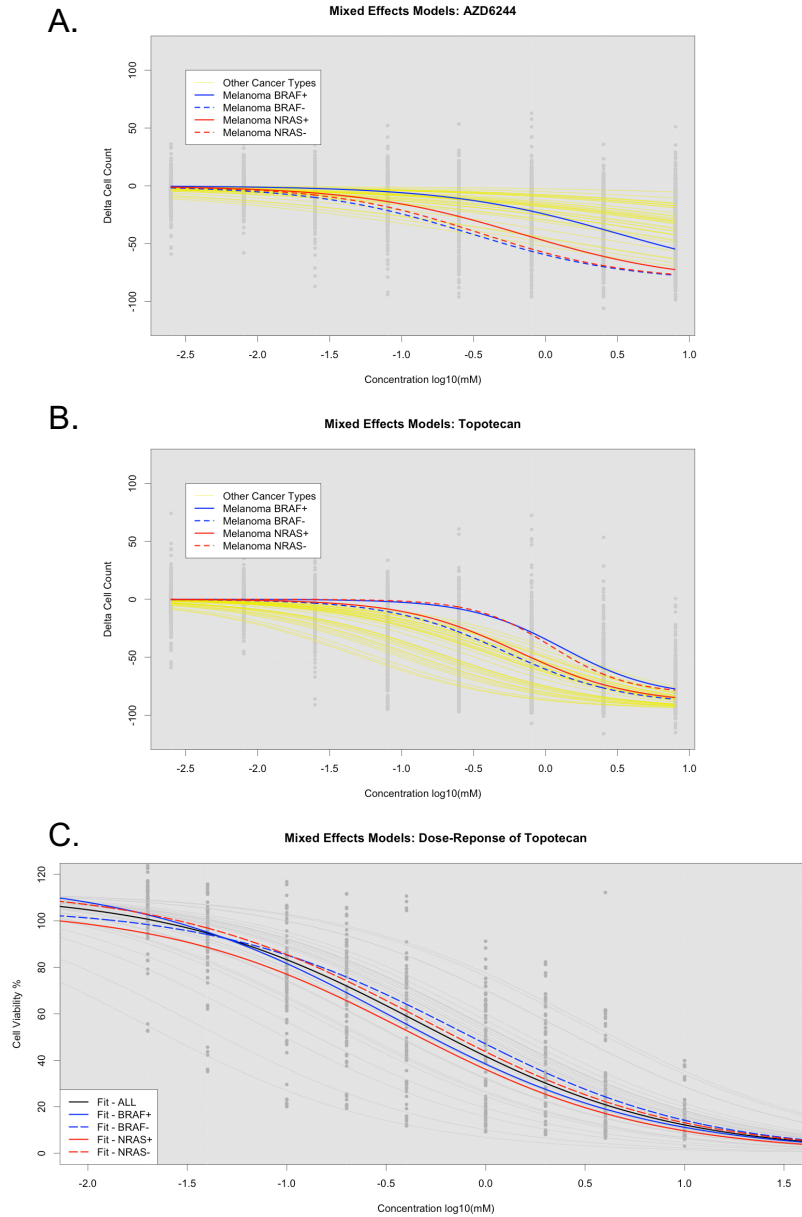


Figure 4.4: LLR tests for assessing significant stratification of cell lines with or without mutations in CCLE and SU2C cell lines. (A) Assessment of AZD6244 in CCLE across all cell lines (Black), melanoma BRAF+ (Solid Blue), and melanoma BRAF- (Dotted Blue). (B) Same as A, but with Topotecan. (C) Assessment of Topotecan in SU2C across all cell lines (Black), BRAF+ (Solid Blue), BRAF- (Dotted Blue), NRAS+ (Solid Red), and NRAS- (Dotted Red)

ter in the analysis of DRCs, but rather assumes the overall shapes of the curves between two groups (we note that it would be easy to generalize the test to more than two groups) defined on an assumed associated cell line characteristic like gene expression level, are different.

We pursued simulation studies to assess false positive rates and the statistical power of NLME methods relative to traditional methods. Our simulations showed that NLME-based tests, in particular the proposed Omnibus LR NLME test, are more powerful than standard tests. However, these simulations are very computationally expensive, especially for the Omnibus LR NLME test, and as a result we suggest that if adequate computational resources are not available, then the use of more efficient tests that still leverage NLME models, such as the IC_{50} NLME test, can be used.

We applied the NLME-based tests in order to see if they could identify differential gene expression and protein level associations with drug response in more powerful and compelling ways than traditional individual DRC IC_{50} based analytical methods. We found evidence for many associations that could not be attributed to chance. Upon mining gene and protein interaction databases, we observed that many of the associations we observed several were consistent with previously published data. For example, we found that the differential expression of the MAPK3 gene stratified melanoma cell lines into Cladribine responders and non-responders, which is consistent with what is known about the mechanism of action of Cladribine. Based on this finding, and in conjunction with the interrogation of the DGIdb database, we found that kinase inhibitors 5-Iodotubercidin and Purvalanol are likely drug candidates for these cell lines.

Our analyses and observations suggest that NLME-based models for fitting and analyzing DRCs could complement current HTS strategies and infrastructure. However, these NLME methods require, as noted, a fairly heavy computational burden, especially if one wants to rely on permutation-based methods for obtaining p-values for the Omnibus LR NLME test, and may not be optimal in every context; e.g., when one wants to search the entire HTS space for gene by drug interactions which could be huge. In an ideal world, every association among genes, proteins,

and drugs could be explored. Given the computational burden, we opted for a biologically-guided search (e.g., searching for BRAF-related genetic associations in melanoma) that ultimately led to compelling associations between genes and drug responses. The use of NLME-based analysis approaches that we document here can be extended to test associations between other important factors (e.g., Copy Number Variations, DNA methylation status, Single Nucleotide Polymorphisms, etc.) and drug response. Ultimately, we believe that the NLME approach to analysis of HTS dose-response data is flexible, robust and powerful enough to be used routinely.

4.6 Acknowledgements

Research supported by a Stand-Up-To-Cancer/Melanoma Research Alliance Melanoma Dream Team Translational Cancer Research Grant. Stand-Up-To-Cancer is a program of the Entertainment Industry Foundation administered by the American Association for Cancer Research. Dr. Schork and Mr. Ding have also been supported in part by NIH grants: U19 AG023122-09; R01 DA030976-05; R01 MH094483-03; R01 MH100351-02; R21 AG045789-01A1; UL1TR001442-01 and U24AG051129-01, in addition to support from the Tanner Project Foundation.

Chapter 4 is a full reprint of the publication: **Ding, K.**, Petricoin, E.F., Finlay, D., Yin, H., Hendricks, W.P.D., Sereduk, C., Kiefer, J., Sekulic, A., LoRusso, P.M., Vuori, K., Trent, J.M., and Schork, N.J. Applications of Nonlinear Mixed Effects Models for High Throughput Drug Screen Association Tests. (In Preparation), with permission from all coauthors. The dissertation author was the primary researcher and author of this material.

4.7 Supplemental Figures and Tables

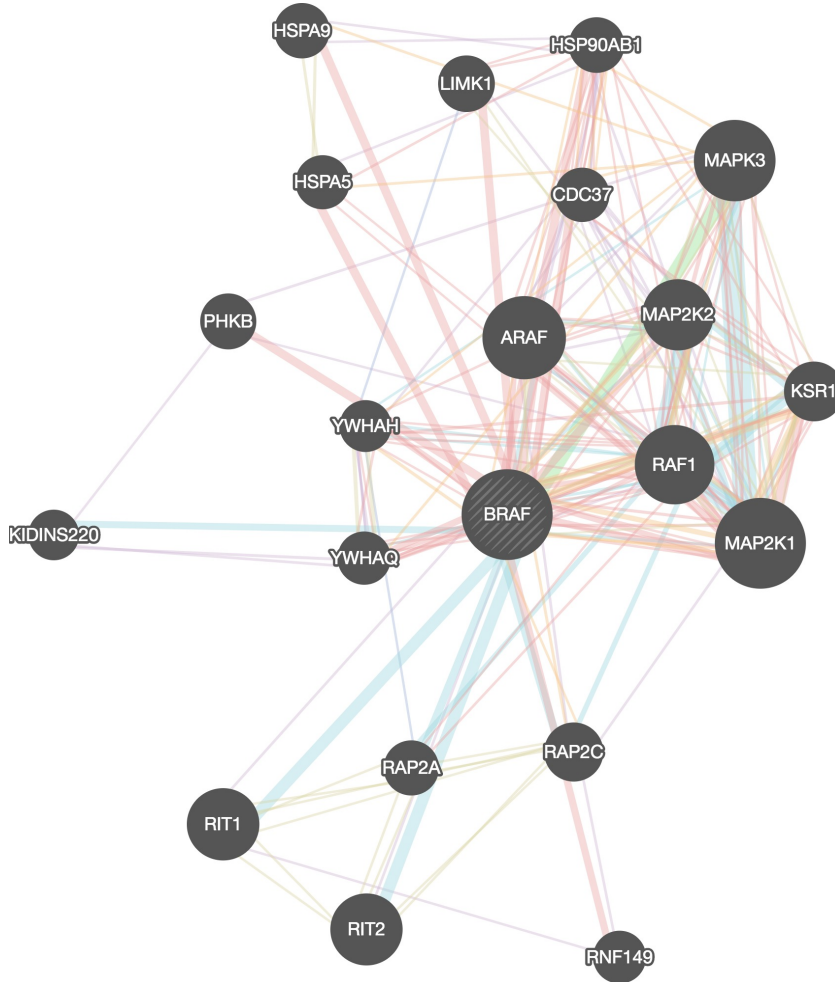


Figure 4.5: GeneMANIA BRAF-related network with $K = 20$. Key: Pink = Physical interactions; Purple = Co-expression; Orange = Predicted; Light Blue = Pathway; Dark Blue = Co-localization; Green = Genetic interactions; Yellow = Shared protein domain; and size of circle is proportional to the number of connections.

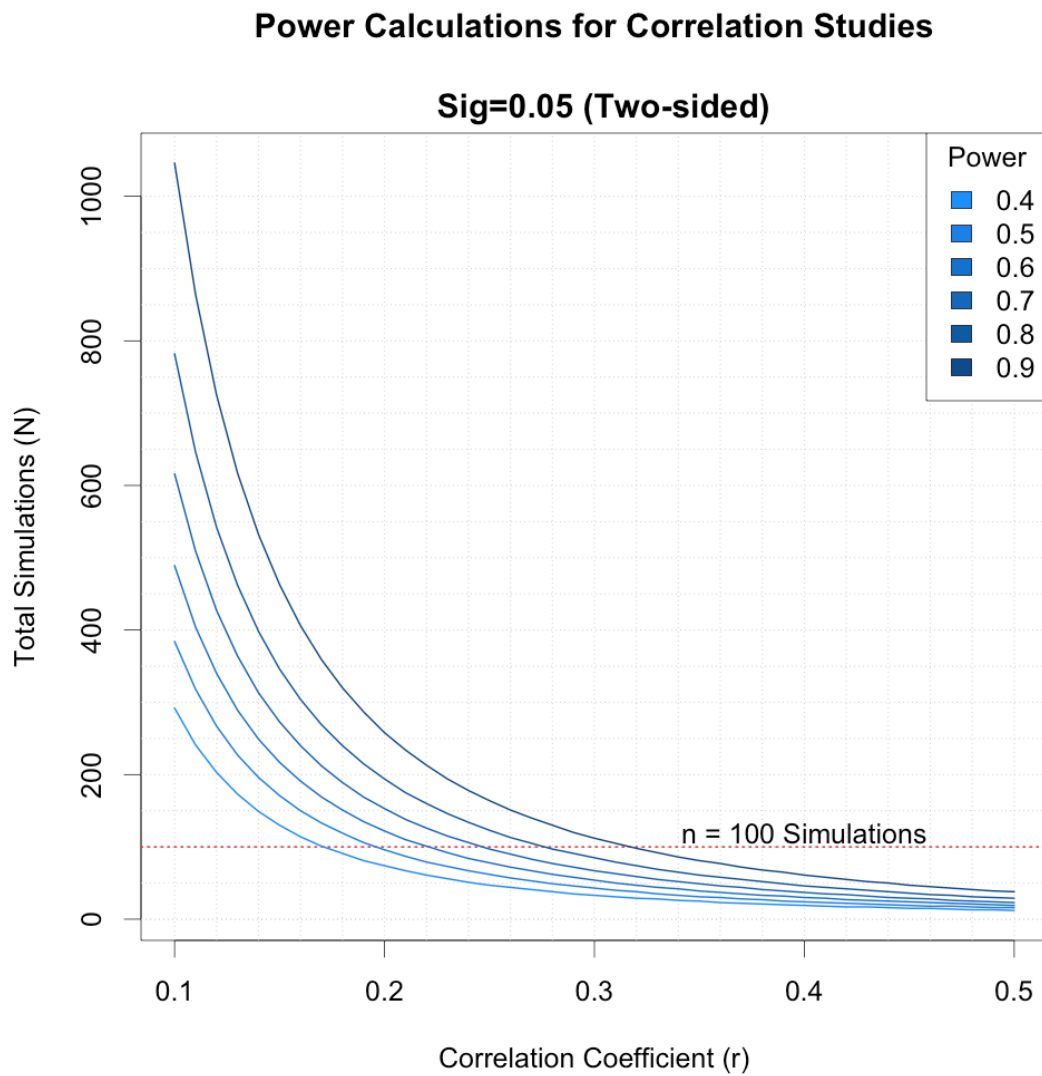


Figure 4.6: Power analysis for simulations. Based on the observed correlations from initial parameter simulations ($\text{cor} > .5$), 100 simulations should suffice.

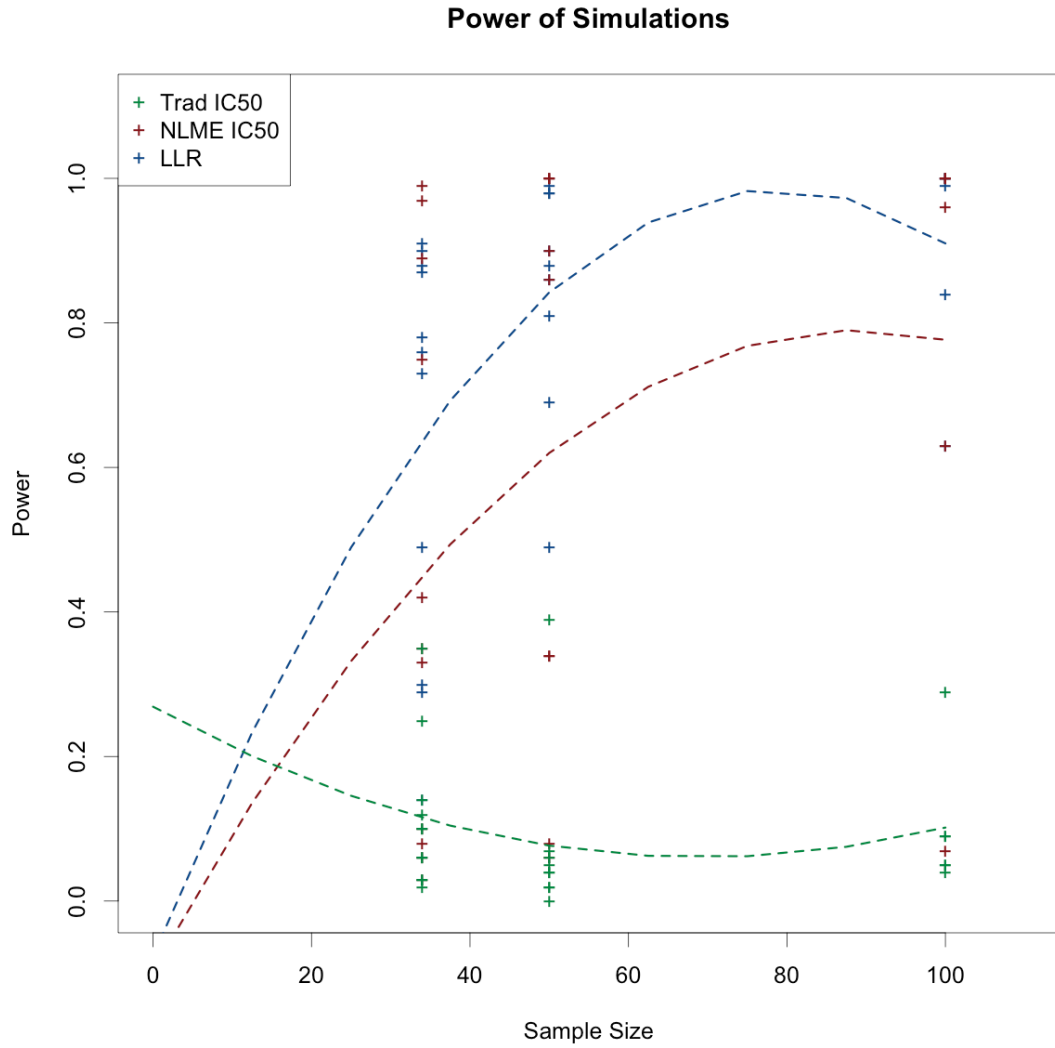


Figure 4.7: Simulation-based power does improve with increased sample size in both the NLME called IC_{50} associations and in the LLR associations. Statistical power in traditionally called IC_{50} associations tests yielded the worst results.

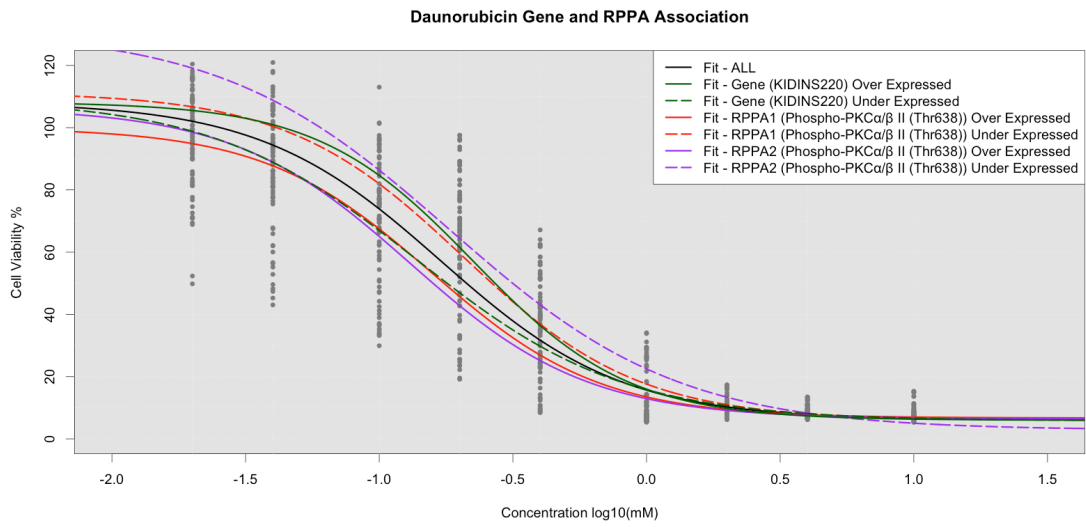


Figure 4.8: Application of LLR tests to identify concordant RPPA (Phospho-PKC α / β II (Thr638)) and Gene expression (KIDINS220) that stratify Daunorubicin dose response curves.

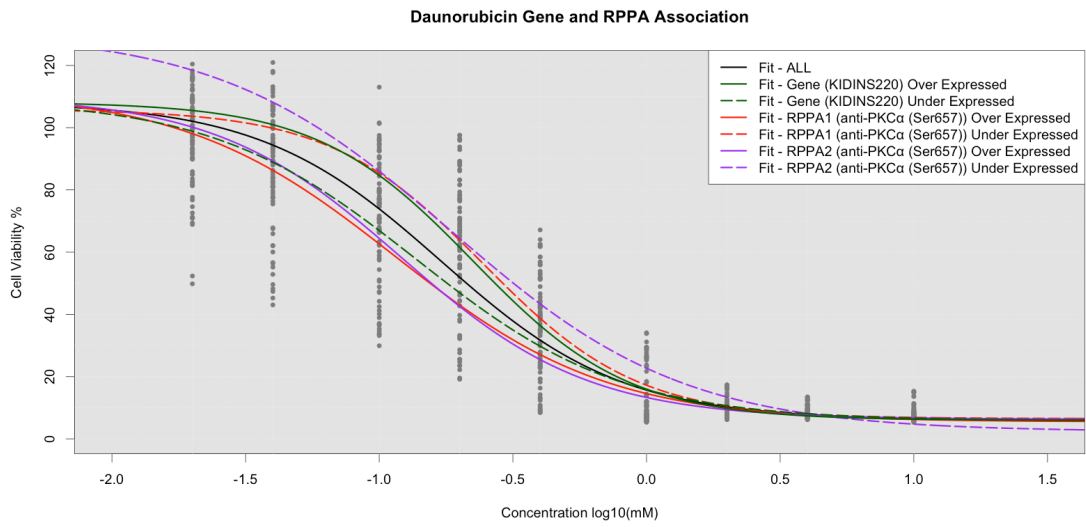


Figure 4.9: Application of LLR tests to identify concordant RPPA (anti-PKC α (Ser657)) and Gene expression (KIDINS220) that stratify Daunorubicin dose response curves.

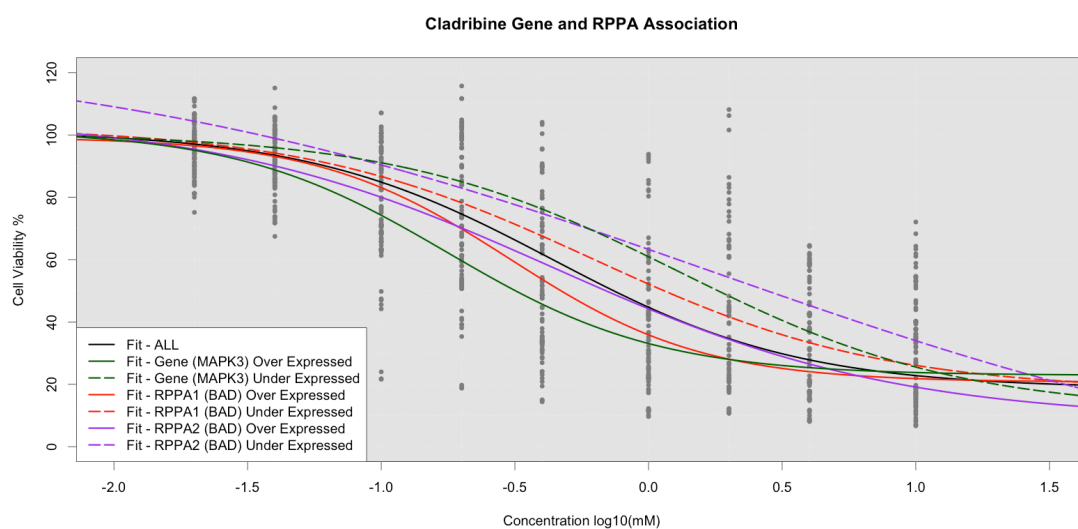


Figure 4.10: Application of LLR tests to identify concordant RPPA (pBAD) and Gene expression (MAPK3) that stratify Cladribine dose response curves.

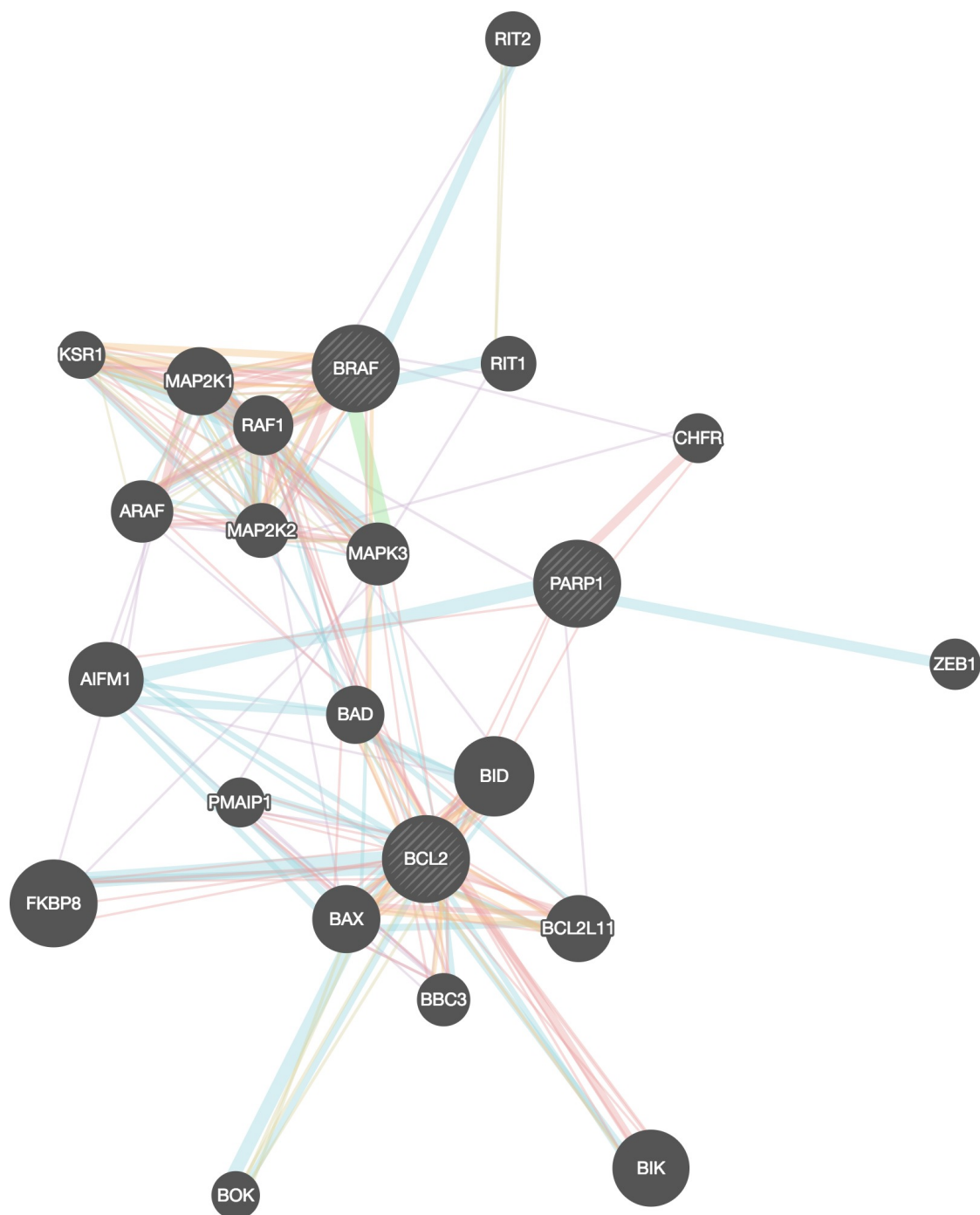


Figure 4.11: GeneMANIA pathway analysis using BRAF and BCL2 showing the network connection of MAPK3 and pBAD.

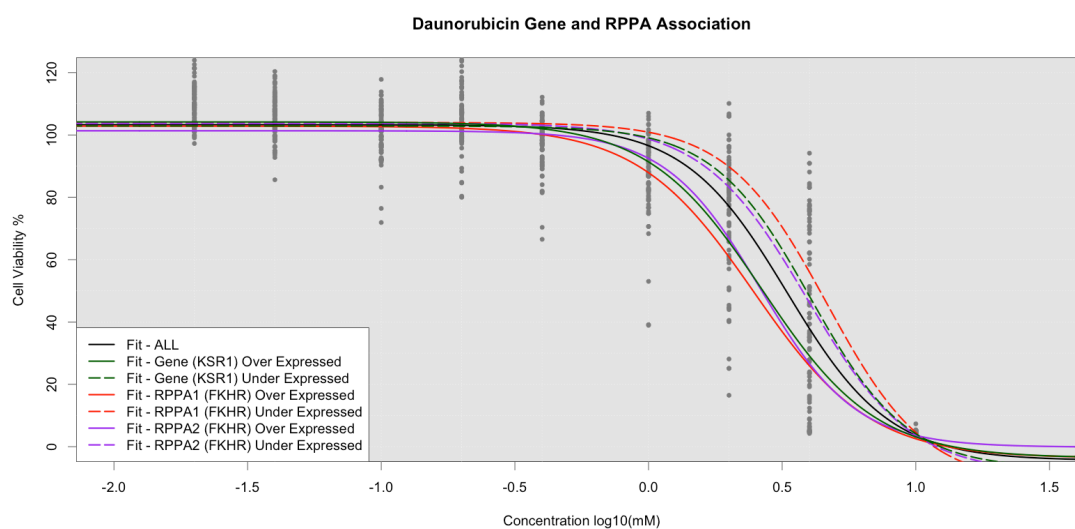


Figure 4.12: Application of LLR tests to identify concordant RPPA (FKHR) and Gene expression (KSR1) that stratify LDK378 dose response curves.

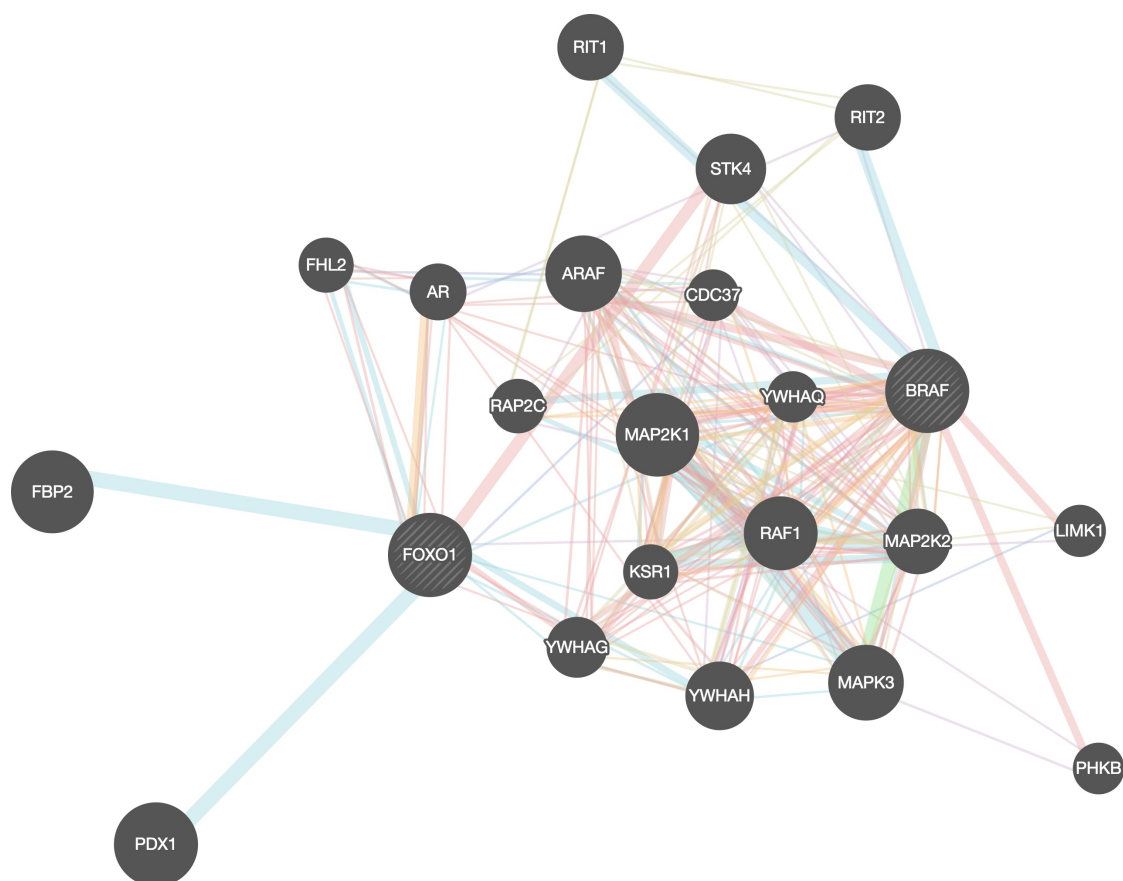


Figure 4.13: GeneMANIA pathway analysis using BRAF and BCL2 showing the network connection of FOXO1 (FKHR) and KSR1.

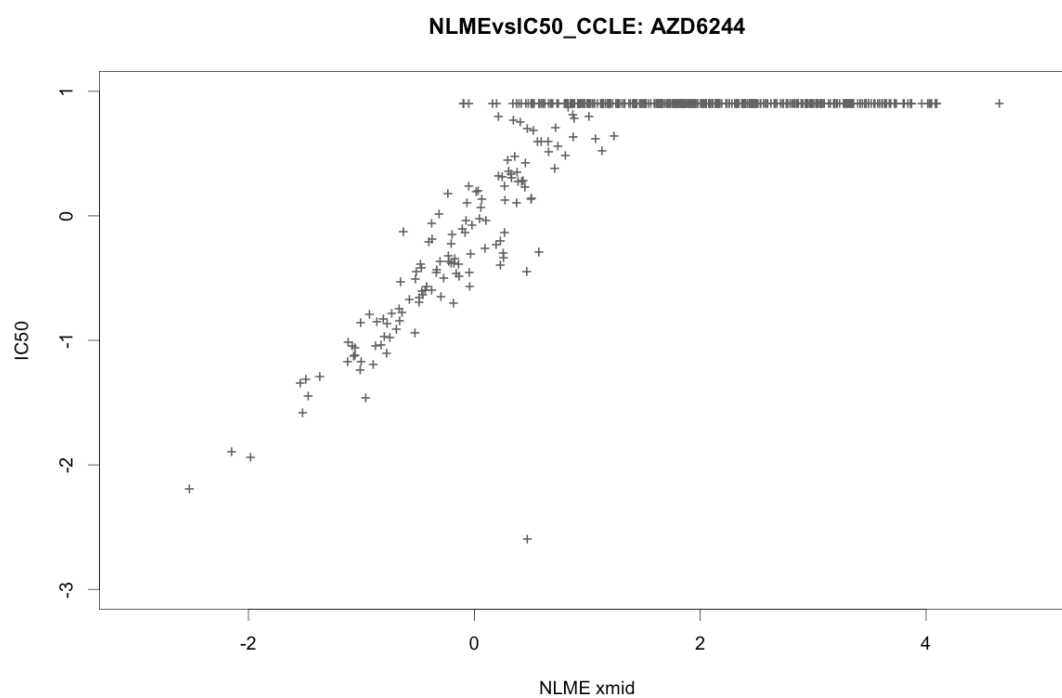


Figure 4.14: Comparing the effects of estimating IC_{50} values (uM) with NLME models and those of traditional methods. Although overall correlation is observed, there are some clear variations.

Table 4.3: Table of cell lines available in the SU2C study. HTS represents availability of high throughput drug screen, GEX represents availability of whole genome microarray genechips, and RPPA sets 1 and 2 represent the availability of Reverse Phase Protein Arrays.

SU2C Cell Lines	Nine Point HTS	GEX	RPPA Set 1	RPPA Set 2
A375	Y	N	N	Y
HEMnLP	Y	Y	N	N
MeWo	Y	Y	N	Y
SKMEL113	Y	N	N	Y
SKMEL119	Y	N	N	Y
SKMEL2	Y	Y	N	Y
SKMEL21	Y	N	N	Y
SKMEL217	Y	N	N	Y
UACC0091	Y	Y	Y	Y
UACC0257	Y	Y	Y	Y
UACC0502	Y	Y	Y	Y
UACC0558	Y	Y	Y	Y
UACC0612	Y	Y	Y	Y
UACC0647	Y	Y	Y	Y
UACC0903	Y	Y	Y	Y
UACC0952	Y	Y	Y	Y
UACC1093	Y	Y	N	Y
UACC1097	Y	Y	Y	Y
UACC1113	Y	Y	Y	Y
UACC1118	Y	Y	Y	Y
UACC1120	Y	Y	Y	Y
UACC1237	Y	Y	Y	Y
UACC1265	N	Y	N	N
UACC1308	Y	Y	Y	Y
UACC1469	Y	Y	Y	Y
UACC1649	Y	N	N	Y
UACC1729	Y	Y	Y	Y
UACC1940	Y	Y	Y	Y
UACC2331	Y	Y	Y	Y
UACC2427	Y	Y	Y	Y
UACC2496	Y	Y	Y	Y
UACC2610	Y	Y	Y	Y
UACC2641	Y	Y	Y	Y
UACC2851	Y	Y	Y	Y
UACC2972	Y	Y	Y	Y
UACC2994	Y	N	N	Y
UACC3074	Y	Y	Y	Y
UACC3093	Y	Y	Y	Y
UACC3291	Y	Y	Y	Y
UACC3312	Y	Y	Y	Y
UACC3337	Y	Y	Y	Y

Table 4.4: Total number of significant associations based on Traditionally called IC_{50} and NLME called IC_{50} vales.

	Azacitidine	Doxorubicin	Cladribine	Vorinostat	Crizotinib	Thioguanine	Daunorubicin	OSI027
RPPA Set 1 Trad Sig	0	0	1	0	0	3	3	1
RPPA Set 1 NLME Sig	0	0	0	0	0	1	0	0
RPPA Set 2 Trad Sig	1	1	0	2	1	4	1	1
RPPA Set 2 NLME Sig	NA	0	0	0	0	NA	0	0

Table 4.5: P-values for LLR tests stratifying gene expression at the median value. Tests were performed across 15 drugs and BRAF-related genes.

Gene Name	Azacitidine	Doxorubicin	Cladribine	Vorinostat	Clofarabine	Etoposide	Thioguanine
RAP2A	NA	1.000	0.010	0.250	0.866	1.000	1.000
KIDINS220	NA	0.010	0.667	1.000	0.904	1.000	0.944
RAF1	NA	0.091	0.667	0.667	1.000	0.286	0.500
HSP90AB1	NA	0.500	1.000	0.500	0.500	1.000	1.000
HSPA9	NA	0.167	1.000	1.000	0.500	0.941	1.000
YWHAQ	NA	0.200	0.667	1.000	1.000	0.667	1.000
YWHAH	1.000	0.667	0.051	1.000	0.222	1.000	NA
ARAF	1.000	0.080	0.667	1.000	1.000	0.667	1.000
MAP2K2	1.000	1.000	1.000	0.667	1.000	1.000	0.300
MAP2K1	1.000	1.000	0.111	0.010	0.400	0.500	1.000
PHKB	1.000	1.000	0.125	0.400	1.000	1.000	1.000
LIMK1	NA	1.000	0.059	0.010	0.074	0.500	NA
BRAF /// KIAA1549	NA	0.286	0.100	0.333	0.222	0.500	0.500
RIT2	NA	1.000	1.000	0.500	1.000	1.000	1.000
RIT1	NA	1.000	1.000	1.000	0.667	0.300	0.400
CDC37	NA	0.143	0.667	0.667	1.000	1.000	1.000
HSPA5	0.750	0.500	0.010	0.400	0.800	1.000	0.667
MAPK3	NA	0.286	1.000	1.000	0.667	0.667	1.000
KSR1	1.000	0.020	0.200	0.333	0.222	NA	1.000
RAP2A /// RAP2B	0.667	0.154	1.000	1.000	0.111	0.852	0.500
RAP2C	1.000	0.333	0.667	0.667	0.667	NA	1.000
RNF149	NA	1.000	0.250	0.333	1.000	1.000	1.000
BRAF	NA	0.667	0.667	1.000	1.000	NA	1.000

Gene Name	Irinotecan	Mitoxantrone	Mitomycin	Daunorubicin	OSI	MLN4924	LDK378
RAP2A	0.125	0.667	0.600	0.500	0.125	0.667	0.222
KIDINS220	1.000	1.000	1.000	0.010	1.000	0.667	0.200
RAF1	0.111	0.667	0.500	0.010	1.000	1.000	1.000
HSP90AB1	0.667	1.000	1.000	0.067	1.000	0.667	0.667
HSPA9	0.222	NA	NA	0.059	0.400	1.000	0.051
YWHAQ	1.000	1.000	0.667	1.000	1.000	0.667	1.000
YWHAH	1.000	NA	1.000	1.000	0.400	1.000	0.500
ARAF	1.000	NA	NA	0.200	1.000	0.400	1.000
MAP2K2	1.000	1.000	NA	1.000	0.333	1.000	1.000
MAP2K1	0.182	1.000	1.000	0.111	1.000	0.667	1.000
PHKB	0.667	1.000	1.000	1.000	1.000	1.000	1.000
LIMK1	0.667	1.000	1.000	0.500	0.167	NA	1.000
BRAF /// KIAA1549	1.000	0.286	1.000	1.000	1.000	0.667	1.000
RIT2	1.000	0.500	1.000	1.000	1.000	0.333	0.182
RIT1	0.286	1.000	0.400	0.333	0.200	0.500	0.333
CDC37	1.000	NA	1.000	0.667	1.000	1.000	0.667
HSPA5	0.667	0.800	1.000	1.000	1.000	1.000	0.182
MAPK3	0.200	NA	0.667	0.167	0.500	1.000	1.000
KSR1	0.400	NA	1.000	0.010	1.000	NA	0.400
RAP2A /// RAP2B	0.154	1.000	0.267	0.667	0.010	0.667	0.020
RAP2C	0.667	0.167	0.172	0.167	0.667	1.000	1.000
RNF149	1.000	1.000	1.000	0.125	1.000	1.000	0.500
BRAF	1.000	NA	0.333	0.333	0.667	0.667	1.000

Chapter 5

Network Rewiring in Cancer: Applications to Melanoma Cell Lines and The Cancer Genome Atlas Melanoma Patients

5.1 Abstract

Genes do not work in isolation, but rather as part of networks that have many feedback and redundancy mechanisms. Studying the properties of genetic networks and how individual genes contribute to the overall functioning of a network can provide insight into genetically-mediated disease processes. This is particularly true in the context of cancer, where there are known differences between cancerous and non-cancerous tissues with respect to how genes function together as networks. Most analytical techniques for identifying genes contributing to dysfunctional genetic networks assume a network topology (i.e., connections between genes) that is based on how the network functions in normal or non-diseased states. This is problematic since perturbations in genes often have “ripple” effects that lead to the rewiring of relevant networks and impact relationships among other genes. We apply a suite of analysis methodologies to assess the degree of tran-

scriptional network rewiring observed in different sets of melanoma cell lines using whole genome gene expression microarray profiles, skin fibroblasts obtained from Gene Expression Omnibus (GEO), and cancer cell lines from the Cancer Cell Line Encyclopedia (CCLE). We also assess evidence for network rewiring attributable to mutations in melanoma patient tumor samples using RNA-sequence data available from The Cancer Genome Atlas (TCGA). We make a distinction between “unsupervised” and “supervised” network-based methods and contrast their use in the identification of consistent differences in networks between subsets of cell lines and tumor samples. Ultimately, we argue that our results have important implications for understanding the molecular pathology of melanoma as well as the choice of treatments to combat that pathology.

5.2 Introduction

Many studies leveraging genomic assays explore associations between genes and diseases. If any associations are found then strategies to make broader claims about genetically-mediated processes, networks, and pathways that mediate those diseases are pursued; for example, by studying the expression levels or protein function of disease-associated genes. Identifying genetically-mediated molecular physiologic processes underlying a disease could lead to insights into how to combat or treat the disease. One strategy for identifying molecular physiologic processes responsible for a disease is to measure gene expression of diseased and non-diseased tissues. Then, differentially expressed genes could be extracted to determine if some subset of regulated genes participate in a coherent network or contribute to particular processes. Although intuitive, there are some complications to such analyses. For example, in making claims about whether or not the differentially expressed genes (DEGs) all participate in a particular network, the list of DEGs is compared to databases, such as Kyoto Encyclopedia of Genes and Genomes [Kanehisa and Goto, 2000] (KEGG) or WikiPathways [Pico et al., 2008] [Kutmon et al., 2016] [Kelder et al., 2012], which provide network, function, or pathways associated with each gene. If the DEGs match lists of genes in known networks,

then it can be inferred that those networks are likely to be contributing to the pathogenesis of the disease [Costanzo et al., 2010] [Blomen et al., 2015].

Such strategies do not address crucial questions about the relationships between the genes, i.e., their correlations, which might be necessary to determine the potential effect of a drug on those genes as a network. For example, it might be necessary to know if the inhibition of one gene in the network will lead to the inhibition or excitation of another. In order to assess the relationships between the genes, and whether those relationships are themselves different in the diseased or normal state, one would need to have some basal “wiring” diagram or “topology” associated with the network that provides insight into which genes might influence other genes. Unfortunately, most network topologies or wiring diagrams have been devised from studying genes in the normal (i.e., non-diseased) state [Snider et al., 2015] [Califano, 2011] [Ideker and Krogan, 2012]. This is problematic if the networks have become “rewired” (i.e., the topology or connections between the genes is fundamentally different) in the disease state, since then the simple assessment of the differences in the relationships between pairs of genes in the normal and diseased tissues would not reveal differences at the more holistic or network level, and the choice of which pairs of genes to assess guided by the network topology obtained in the normal state may not point out crucial relationships unique to the diseased state.

As an example of this problem, consider the fact that many drug target identification and modulation studies focus on how drugs impact levels of gene expression or protein in wild-type, normal, or conveniently available cell lines [St Onge et al., 2007]. If the communication between genes is fundamentally different in the diseased state, then extrapolating a drug’s effects on genes or protein levels in the normal state to that of the diseased state may be problematic (see Figure 5.1A for simple hypothetical example). Another issue with the characterization of the relationships between genes is that there may be many genes in a given network, creating a large number of relationships to consider [Hofree et al., 2013]. This creates computational and inferential challenges. In addition, such analyses require that one pre-specify the groups whose gene-gene relationships are to be contrasted

(e.g., BRAF mutant positive tumors vs. BRAF wild-type tumors in melanoma).

To identify differences in network architecture within a given data set, one could pursue a “unsupervised” or “supervised” approach. Supervised approaches such as Evaluation of Dependency Differentiality [Jung and Kim, 2014] (EDDY), Gene Set Enrichment Analysis [Mootha et al., 2003] [Subramanian et al., 2005], (GSEA), and Gene Set Co-expression Analysis [Choi and Kendzierski, 2009] (GSCA) are limited to the settings in which the grouping of the cells or tissues is set in advance, and hence requires prior knowledge of, e.g., disease and normal conditions. Unsupervised methods do not require that the groups be specified *a priori*, but often generate results that might be difficult to interpret or reconcile with the results of other unsupervised methods. In addition, most available unsupervised genetic analysis methods focus on exploring differences in the means or mean vectors of the factors among groups (e.g., gene expression values, protein abundances, etc.) as a way of determining which factors may be contributing to differences between them. However, if, as noted, the rewiring of the genetic networks among the groups upsets the relationships or correlations between the genes (i.e., not the mean differences between them), then focusing on mean gene expression levels will not identify the group differences.

We consider the value of both supervised and unsupervised approaches in the analysis of network data [Ideker and Krogan, 2012] [Yang et al., 2014] [Oros Klein et al., 2016]. We assessed evidence for network rewiring in a variety of melanoma drug screen data and The Cancer Genome Atlas [The Cancer Genome Atlas Research Network et al., 2013] (TCGA) tumor profile data sets. We further consider the implications such rewiring might have for making claims about cell line responses to drugs. Specifically, we pursued supervised analyses of genetic networks by stratifying melanoma cell lines into groups defined by BRAF and/or KRAS mutation status and then evaluating differences in the overall network structure (i.e., relationships between genes). Using this strategy, we sought to identify genes that appeared to have a more central role in those networks. We then performed network analyses with patient samples from TCGA melanoma data. We observed that the cell lines exhibited a greater degree of network rewiring or dif-

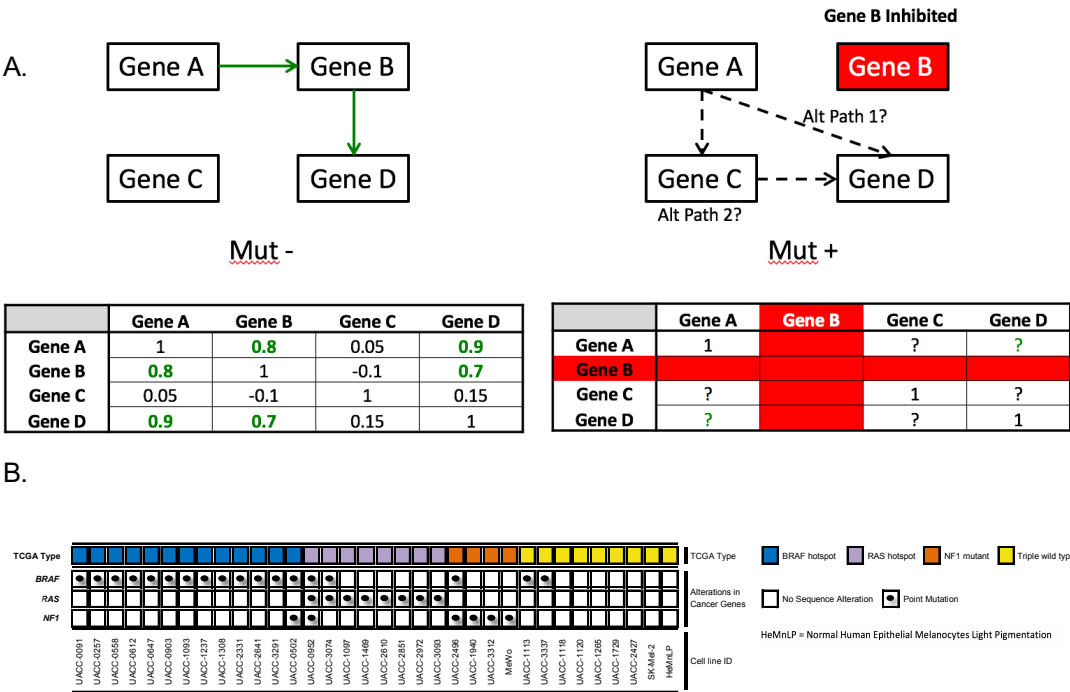


Figure 5.1: (A) Figure 1: (A) Example of network rewiring that could occur. Under ideal conditions, Gene A is co-regulated with Gene D through gene B. When gene B is inhibited, there could be a compensatory mechanism for which gene A communicates with gene D in this example, that is potentially gene C. (B) Mutation status of 33 melanoma cell lines and HeMnLP processed with whole genome microarray gene chips at the Translational Genomics Research Institute. TCGA type calls are based on known pathogenic hotspots for BRAF and RAS (e.g. BRAF V600 mutations), but are more open-ended for NF1. Point mutations are provided for BRAF, RAS, and NF1. Cell lines are ordered by TCGA mutation type, point mutation, and then cell line number.

ferences than the patient samples and therefore leveraged unsupervised methods to explore the variation in network behavior. These analyses allowed us to identify potential drug targets and could shed light on how to match patient genetic profiles to specific treatments by exploiting the Drug Gene Interaction database [Wagner et al., 2016] (DGIdb). Ultimately, our proposed methods provide a strategy for identifying unique network features among tumors and subsequently exploiting those unique features for drug targeting purposes.

5.3 Materials and Methods

5.3.1 Transcriptional Profile on Cell Lines

We leveraged a number of melanoma cell lines. Including the UACC cell lines maintained at the Translational Genomics Research Institute (TGen) in Phoenix, AZ. All the TGen UACC cell lines were of low passage number (Figure 5.1B) and were maintained according to the manufacturer’s or collaborator’s instructions. All cell lines were stored and evaluated in media with 10% FBS and 1% AA added. The low passage cell lines were maintained in multiple aliquots as liquid nitrogen stocks to reduce risk of phenotypic drift. All cells were cultured for less than three months before reinitiating culture from the frozen stock and routinely inspected for identity by morphology and growth curve analysis and validated to be mycoplasma free. All cell lines were free of contaminants.

The cell line samples were used for nucleic acid extraction, verification, amplification, and hybridisation to Affymetrix HG-U133 plus 2.0 arrays (54,675 probesets, Affymetrix, Santa Clara, CA) using standard protocols for Affymetrix GeneChip microarrays. Affymetrix HG-U133 Plus 2.0 microarrays were normalized (background adjustment, interquartile normalization, and median polish) using robust multichip averaging [Irizarry et al., 2003] in R. To assess network rewiring that occurs within melanoma cell lines against a control group, we normalized expression profiles across melanoma cell line and skin fibroblast samples (Table 5.8, Figure 5.7). All subsequent analyses were based on normalization only within melanoma cell lines (Figure 5.8).

We also obtained the SK-MEL-2 melanoma cell line from the Cancer Cell Line Encyclopedia (CCLE), and the MeWo melanoma cell line from the Developmental Therapeutics Program’s NCI-60. In addition, we leveraged microarray gene expression data on skin fibroblast samples ($n = 29$) from Gene Expression Omnibus, (GEO) (Table 5.8). The skin fibroblast samples varied in age and ethnicity (Caucasian and Puerto Rican). Finally, we obtained gene expression data from the TCGA melanoma samples from the Broad GDAC Firehose ($n = 474$). Raw CEL files were downloaded from the CCLE web portal. All samples from the CCLE, NCI-60, and GEO were performed on Affymetrix HG-U133 plus 2.0 arrays. All microarray samples were re-normalized together to ensure minimal batch effects. Normalization of all microarray data (background adjustment, interquartile normalization, and median polish) was completed using robust multichip averaging [Irizarry et al., 2003] [Gautier et al., 2004] in R. We obtained RAF and RAS mutation status on the CCLE cell lines from the Oncomap mutation provided in the CCLE web portal and additional binary calls for copy number and mutation data from the CCLE web portal.

5.3.2 MAPK Pathway Information

We took advantage of a wide variety of data analysis techniques, as described in detail below but focused much of our analyses on sets of genes in the MAPK pathway, which were extracted from the information in the WikiPathways [Kutmon et al., 2016] [Kelder et al., 2012] resource. We ultimately matched MAPK pathway genes to probe sets in the Affymetrix arrays used on for transcriptional profiling on the cell lines. We also matched MAPK genes to those evaluated in TCGA RNAseq dataset. For genes with multiple probe set matches, the probe set with the maximum average intensity was used. We fit multivariate mixture models using flexmix [Leisch, 2004] [Grn and Leisch, 2008] [Grn and Leisch, 2007], assuming that there were two (BRAF+ vs BRAF-) or three (possible sub-grouping for BRAF-) overarching different groups with gene expression values as the dependent variables.

5.3.3 Unsupervised Network Rewiring Analysis

In order to conduct analyses exploring evidence for network heterogeneity without specifying a grouping factor *a priori* (i.e., in an unsupervised manner) we used multivariate mixture models as implemented in the R module flexmix [Leisch, 2004] [Grn and Leisch, 2008] [Grn and Leisch, 2007]. For the multivariate mixture models, we hypothesized that groups identified in the data had different covariance (correlation) structures reflecting the strength of the associations of the expression values of genes within networks. We tested for the existence of two groups in the data, but our approach could have easily tested for three or more groups; however, sample size limitations would have resulted in poor power for such analyses. Based on the patterns of the correlations between expression levels of the genes within each group, we sought to identify blocks of highly correlated genes by using simple hierarchical clustering of the pairwise correlations. Dendrograms or trees reflecting this clustering could then be cut at varying heights to assess the extent of heterogeneity within the flexmix analysis-derived groups.

5.3.4 Assessing Statistical Significance

In order to determine whether there was a significant statistical evidence for subgroups within a data set regulated exhibiting different transcriptional networks, we leveraged simulation-based tests. For a given analysis, we simulated 1000 samples with an n -dimensional mean vector and an $n \times n$ covariance matrix consistent with the observed mean vectors and covariance matrices for the full data set (i.e., under the assumption of homogeneity or non-subgroups). We then used flexmix to identify two subgroups with these simulated data sets. From these flexmix analyses, we calculated t-statistics comparing the estimated mean vectors, χ^2 statistics comparing correlation matrices, as well as log-likelihood ratios, delta AICs, and delta BICs comparing the two group vs. single group fits. We calculated p-values for the statistics obtained from the non-simulated data as the number of simulated sample statistics greater than the actual test statistics divided by 1001.

5.3.5 *A Posteriori* Supervised Analyses

To determine whether or not the two groups identified in an analysis could be explained by the existence of gene mutations present or not in the cell lines, we performed logistic regression analysis with the flexmix group identifier for a cell line as the dependent variable and BRAF, NRAS, and KRAS mutation information on those cell lines as explanatory variables. For the CCLE cell lines, additional association tests were pursued using gene mutation, gene amplification, insertion and deletion information as independent variables. To account for the multiple hypothesis tests, we used Q:Q-plots and Benjamini-Hochberg [Benjamini and Hochberg, 1995] FDR-adjusted p-values.

5.3.6 Exploring Network Architecture

To determine whether particular genes exhibited stronger correlations with other genes and were acting as potential “hubs” within the transcriptional networks, we calculated network centrality metrics for each gene. To do this, individual “edges” between each pair of nodes (i, j ; where i and j denote genes or nodes in the transcriptional network) were set equal to:

$$E_{ij} = \begin{cases} 0, & \text{if } \rho(r_{ij}^2) > .05 \\ r_{ij}^2, & \text{otherwise} \end{cases}$$

where r_{ij}^2 reflects the correlation between genes i and j and $\rho(r_{ij}^2)$ is the p-value associated with that correlation, to create graphs of the network. These calculations were performed for each group identified from the mixture analysis using flexmix. Based on the resulting graphs, we then calculated the metrics degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality for each gene within each groups network. With these metrics for each gene, we examined the degree of similarity of genes across groups (i.e., we compared the metrics for each gene across the mixture model-derived groups). For those genes that emerged as the most central for each network, we identified drugs that target those genes using the Drug Gene Interaction Database (DGIdb).

5.3.7 Analyses of the CCLE Melanoma Cell Line Expression Data

To assess evidence for network rewiring associated with groups of tumors defined by mutational status (i.e., a supervised analysis), we tested the equality of correlation matrices computed for each group using correlation matrix equivalence tests. Specifically, we extracted gene networks centered on BRAF and KRAS genes using geneMANIA. For this analyses, we identified the 20 most inter-connected genes associated with the BRAF and KRAS pathways. Although NRAS is a more common mutation in melanomas, we pursued more focused analysis on the KRAS mutation because the cell lines and patient splits had a greater proportion of KRAS mutant samples (i.e., there were enough KRAS samples to pursue an analysis, but only a limited number of NRAS mutant samples). We then calculated the pair-wise spearman correlations for the gene expression values for those genes within mutant melanoma cell lines and wild type melanoma cell lines. We compared the resulting correlation matrices through permutation-based tests (see above). We calculated χ^2 statistics using the Steiger and Jennrich correlation equivalence test methods with the psych [Revelle, 2015] package in R. As noted in the package manual, the Steiger method uses the sum of the squared correlations, whereas the Jennrich method compares the differences between two matrices to the averages of the two matrices. Thus, the Steiger method emphasizes non-zero values to a greater degree, whereas the Jennrich method places a greater emphasis the change in correlations. We generated null distributions of χ^2 statistics through permutations of the mutation status. For each test, we randomly permuted mutation status 10,000 times, repeated the analyses, and derived the distribution of test statistics from these permutations to get a null distribution as well as p-values for χ^2 statistics.

5.3.8 Analyses of the TCGA Melanoma Tumor Gene Expression Data

We performed similar correlation and permutation-based tests within the TCGA melanoma gene expression data. For each melanoma patient’s tumor gene expression data in the TCGA, we obtained expression profiles from normal and tumor from Cancer Genomics Hub. We used BAM slicer to identify BRAF and KRAS gene regions and we assigned somatic mutation status using the default settings in VarScan. We stratified patient samples into three groups: BRAF mutant (BRAF+), BRAF V600E mutant (BRAF V600E+), and BRAF wildtype (BRAF-). We then performed correlation equivalence tests to compare BRAF+ against BRAF- derived spearman correlation matrices and BRAF V600E+ against BRAF- derived spearman correlation matrices. As before, we generated null distributions of relevant chi2 statistics by permuting the mutation status. To determine whether evidence for rewiring was limited to the extracted BRAF network, we explored evidence for broader network differences by identifying genes that exhibited differential expression between mutant and wild type groups using the DESeq and edgeR packages in R. We used these genes as inputs to identify candidate networks by assessing them with geneMANIA and GO. Resulting networks were then assessed for differences across the mutation-defined groups, using correlation equivalence tests and degree centrality statistics on these additional DEG-derived networks.

5.3.9 Network Reconstruction Analyses

As a complement to analyses comparing correlation matrices, we used the Weighted Gene Correlation Network Analysis [Langfelder and Horvath, 2008] [Langfelder and Horvath, 2012] (WGCNA) package in R to identify significant transcriptional “modules” within mutation bearing and non-mutation-bearing samples. We created modules (setting WGCNA parameter k to $k = 10$ and $k = 20$) within BRAF+/- and KRAS+/- groups for the SU2C melanoma cell lines; and BRAF+/- and BRAF V600E+/BRAF- for the TCGA patient tumor samples. Correspond-

ing modules were defined as the modules with the maximum overlapping gene matches between BRAF⁻ or KRAS⁻ modules and BRAF⁺, BRAF V600E⁺, or KRAS⁺ modules. For the intersections of genes identified, we assessed network rewiring using the previously mentioned node centrality statistics and correlation equivalence tests.

5.4 Results

5.4.1 Unsupervised Multivariate Mixture Model Analysis of the Transcriptomes of Melanoma Cell Lines

As noted in the Methods section, we extracted genes from the WikiPathway version of the MAPK pathway ($n = 224$). These 224 genes were then mapped to the probesets used in the Affymetrix gene expression chips applied to the CCLE and TGen cells lines based on the maximum average intensity for each gene (Supplemental Table 2). We also obtained gene expression levels for these genes from the RNA-seq data available on the TCGA melanoma samples. However, only 216 of these 224 genes could be confidently mapped to the TCGA RNAseq data. We applied flexmix to identify gene expression-based clusters in the SU2C melanoma cell lines. Across different datasets two groups consistently emerged. For the SU2C cell lines the two groups emerging from this analysis were made up of 27 and 7 cell lines. In applying it to the CCLE melanoma cell line group sizes of 39 and 19 were identified (Table 5.1). Moreover, for the TCGA data, the two groups were made up of 401 and 72 tumors. When all the CCLE cell lines were considered together the groups emerging had 375 and 542 cell lines. Density plots for the posterior probability of the flexmix groups revealed that the clusters were most distinct in the SU2C melanoma cell lines (Figures 5.9A-B). This is also evident when viewing the number of edges present when comparing graphs with identical nodes (Figure 5.10). Interestingly, when we used permutation tests to assess the statistical significance of the mean gene expression levels between the two groups we found that mean equality tests were not significant. However, tests of the

equality of the correlation matrices between the two groups yielded statistically significant p-values for the TCGA melanoma patients and the CCLE all cell lines cohort, moderately significant p-values for the CCLE melanoma cell lines, and a non-significant p-value for the SU2C melanoma cell lines. However, we found that in each cohort, the mixture groups were significant in at least one statistic (LLR, delta AIC, or delta BIC, Table 5.1).

In order to visualize the relationships between the genes based on their expression levels and assess their consistency across the two groups identified in each data set, we extracted pairwise spearman correlation coefficients from the correlation matrices and then generated heatmaps on the matrices in which the entries above the diagonal were from one group set and those below the diagonal were from the other (Figures 5.2A-D, Figure 5.11, Figure 5.12). If the correlations were the same the matrices would be completely symmetric. Clearly, the heatmaps suggest differences exist between the flexmix clusters in each of the data sets, although the differences were more pronounced in the SU2C melanoma cell lines and the CCLE melanoma cell lines.

5.4.2 Does Mutation Status Explain the Clustering?

To assess whether common factors in melanoma cell lines and tumors explain the mixture model clustering, we performed logistic regression using BRAF and KRAS mutation status as predictors of cluster membership; BRAF and NRAS mutation status for the CCLE melanoma cell line cluster membership; and the BRAF V600E mutation status for the TCGA patient samples cluster membership. In each analysis, we found that the mutation status was not a significant predictor of the clusters (Table 5.2). For the CCLE cell lines, we also performed logistic regression with copy number variation, mutation, insertion, and deletion status as predictors of cluster membership. Interestingly, for the subset of melanoma cell lines in the CCLE data, the QQ-plots and FDR-adjusted p-values did not indicate any significant predictors of the cluster (Figure 5.9C), although over the entire set of CCLE cell lines, the QQ-plots and FDR-adjusted p-values did suggest that genetic variations could predict or explain the two groups that were identified (Figure

Table 5.1: Comparing *a priori* called clusters within SU2C melanoma cell lines (MCL), CCLE MCL, CCLE all cell lines (ACL), and TCGA melanoma patients. Performed mean vector and covariance equality tests within groups and used log-likelihood ratio permutation-based tests to evaluate strength of one, two, or three groups.

Mean Equality Test	SU2C MCL	CCLE MCL	CCLE ACL	TCGA Mel
T-Test Statistic	4.9867	6.144	6.5957	4.314
df	223	223	223	216
p-value	1.23E-06	3.66E-09	3.03E-10	2.44E-05
Simulation p-value	0.309327	0.4325581	0.3452381	0.3583333
Correlation Equality Test	SU2C MCL	CCLE MCL	CCLE ACL	TCGA Mel
Chi-sq Statistic	1320.269	1915.88	66408.1	76275.79
df	561	1653	22791	22366
p-value	4.30E-63	6.50E-06	0	0
Simulation p-value	0.08891109	0.04595405	0.000999001	0.000999001
Flexmix 2 groups vs. 1 group	SU2C MCL	CCLE MCL	CCLE ACL	TCGA Mel
LLR Statistic	9.349944	0.444172	9.479712	34.39548
LLR Permuted P-Value	0.008991009	0.3786214	0.01298701	0.000999001
Delta AIC Statistic	-3.349944	5.555828	-3.479712	-28.39548
Delta AIC Permuted P-Value	0.008991009	0.3796204	0.01298701	0.000999001
Delta BIC Statistic	1.229137	11.73716	10.98361	-15.9182
Delta BIC Permuted P-Value	0.000999001	0.02897103	0.03196803	0.000999001
Flexmix 3 groups vs. 1 group	SU2C MCL	CCLE MCL	CCLE ACL	TCGA Mel
LLR Statistic	9.354378	0.5115797	9.557176	39.45247
LLR Permuted P-Value	0.02797203	0.4825175	0.02997003	0.000999001
Delta AIC Statistic	2.645622	11.48842	2.442824	-27.45247
Delta AIC Permuted P-Value	0.02797203	0.4825175	0.02997003	0.000999001
Delta BIC Statistic	11.80379	23.85108	31.36947	-2.497894
Delta BIC Permuted P-Value	0.000999001	0.00999001	0.1048951	0.000999001

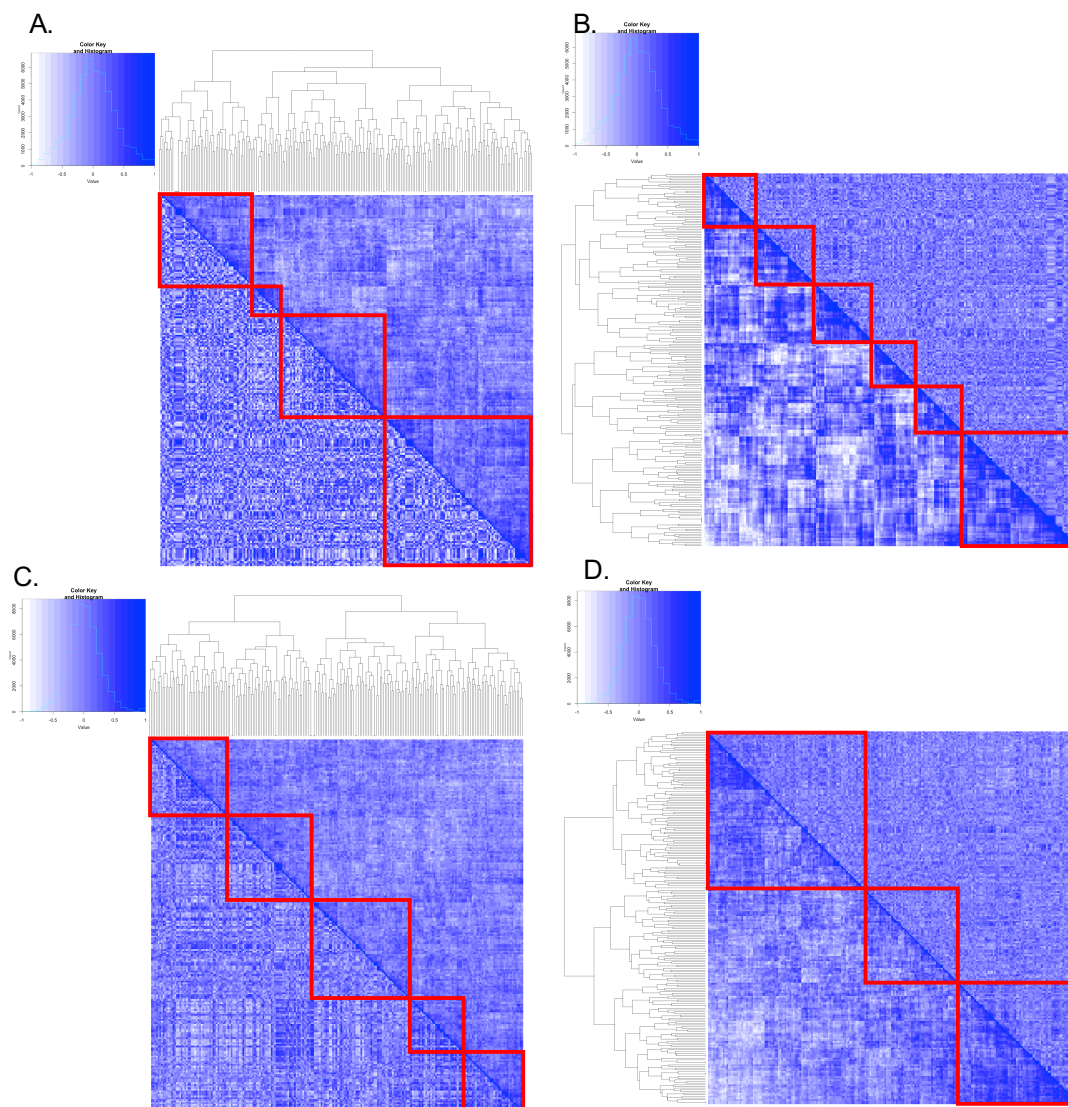


Figure 5.2: Pairwise correlation between MAPK Pathway genes from the identified clusters using flexmix were used to generate heatmaps. Upper right hand half of matrix were correlation coefficients from cluster 1, whereas lower left hand half of matrix were correlation coefficients from cluster 2. Hierarchical clustering was subsequently performed using either distances from cluster 1 coefficients (panels A, C) or cluster 2 coefficients (panels B, D). (A) and (B) are from flexmix derived clusters on SU2C melanoma cell lines (C) and (D) are flexmix derived clusters on CCLE melanoma cell lines.

Table 5.2: Table of association test results between *a priori* called clusters and common factors in melanoma. We test against available mutation status TCGA melanoma patients, SU2C melanoma cell lines, and CCLE melanoma cell lines.

SU2C MAPK Pathway Cluster				
Mutation	Estimate	Std. Error	z value	Pr(>—z—)
BRAF +/-	0.2136	0.8554	0.25	0.8028
KRAS +/-	0.3365	0.9562	0.352	0.72492
TCGA MAPK Pathway Cluster				
Mutation	Estimate	Std. Error	z value	Pr(>—z—)
BRAF v600e +/-	-0.1234	0.2786	-0.443	0.658
CCLE Melanoma Cell Lines MAPK Pathway Cluster				
Mutation	Estimate	Std. Error	z value	Pr(>—z—)
BRAF +/-	-0.2231	0.8878	-0.251	0.8016
NRAS +/-	-0.5522	0.9289	-0.594	0.552

5.9D). These results were similarly observed when regressing on the clusters with copy number variation (Figure 5.13), suggesting that the gene expression associations identified across the entire CCLE cohort potentially arose from copy number variation.

5.4.3 Degree Centrality Analysis

For each group or cluster identified from the multivariate mixture model analysis (for $k = 2$ groups), we tested the statistical significance of the pairwise gene correlations (where edges were kept if $\rho(r_{ij}^2) \leq .05$). We then removed all edges (i.e., gene pairs) without a significant correlation. We calculated the degree centrality statistics on the genes based on the remaining correlation strengths. The more connected or central genes were then extracted based on each degree centrality statistic and mapped to the Drug Gene Interaction Database (Table 5.3). This analysis suggested Bortezomib and Carfilzomib as potential drug targets in one of the clusters identified for each of the SU2C, CCLE melanoma, and all CCLE all cell lines. However, this same analysis applied to the second cluster for each of these data sets suggested that different drugs (Tofacitinib, AT9283, AZD1480,

Table 5.3: Network node centrality statistics identified central nodes for each parameter. Consensus-based approach was used to genes to match to DGIdb to identify candidate therapeutic agents.

Dataset	Degree	Closeness	Betweenness	Eigenvector
SU2C K1	201274_at (PSMA5)	201274_at (PSMA5)	201274_at (PSMA5)	201532_at (PSMA3)
SU2C K2	227677_at (JAK3)	227677_at (JAK3)	201648_at (JAK1)	227677_at (JAK3)
CCLE Mel K1	201388_at (PSMD3)	201979_s_at (PPP5C)	201388_at (PSMD3)	201388_at (PSMD3)
CCLE Mel K2	226163_at (SYNGAP1, ZBTB9)	226163_at (SYNGAP1, ZBTB9)	207243_s_at (CALM2)	226163_at (SYNGAP1, ZBTB9)
CCLE All K1	201274_at (PSMA5)	201274_at (PSMA5)	201274_at (PSMA5)	201274_at (PSMA5)
CCLE All k2	208799_at (PSMB5)	208799_at (PSMB5)	208799_at (PSMB5)	201400_at (PSMB3)
TCGA Mel K1	x819 (APBB1IP)	x14671 (RASGRF1)	x14671 (RASGRF1)	x819 (APBB1IP)
TCGA Mel K2	x819 (APBB1IP)	x14632 (RAP1A)	x14632 (RAP1A)	x819 (APBB1IP)
Dataset	Consensus	Predicted Drug Targets		
SU2C K1	PSMA5	BORTEZOMIB, CARFILZOMIB		
SU2C K2	JAK3	TOFACITINIB, AT9283, AZD1480, PACRITINIB, TOFACITINIB CITRATE		
CCLE Mel K1	PSMD3	BORTEZOMIB, CARFILZOMIB		
CCLE Mel K2	SYNGAP1, ZBTB9	NA		
CCLE All K1	PSMA5	BORTEZOMIB, CARFILZOMIB		
CCLE All k2	PSMB5	BORTEZOMIB, CARFILZOMIB, OPROZOMIB		
TCGA Mel K1	APBB1IP, RASGRF1	NA		
TCGA Mel K2	APBB1IP, RAP1A	NA		

Pacritinib, Tofacitinib, and Citrate) would be better candidates, as least based on the connectedness or centrality of the genes given the identified transcriptional correlation strengths or networks. For the CCLE melanoma cell lines and the TCGA tumors, no drugs were identified based on this analysis for the second group. Finally, we compared the overall degree centrality statistic, across all the data sets to assess their consistency (Figure 5.3). These density plots revealed that network rewiring was more pronounced in the SU2C melanoma cell lines and the CCLE melanoma cell lines, as there was a more obvious mixture of densities in these cohorts in the plots. The density plots for the TCGA samples and the complete set of cell lines for the CCLE data set were very similar and indicated that the differences in identified networks were again not as pronounced (Figures 5.15-5.16). This is likely due to the heterogeneity between TCGA patient samples and cross tissue pan-CCLE analysis (i.e., the cell lines are from different tissues, thus likely have very different underlying baseline networks).

5.4.4 Are the Networks Equivalent for Mutation Status Defined Groups Against Normal Controls?

We tested the possibility that network rewiring occurs as the consequence of mutation (i.e., the subgroups identified with supervised clustering were associated

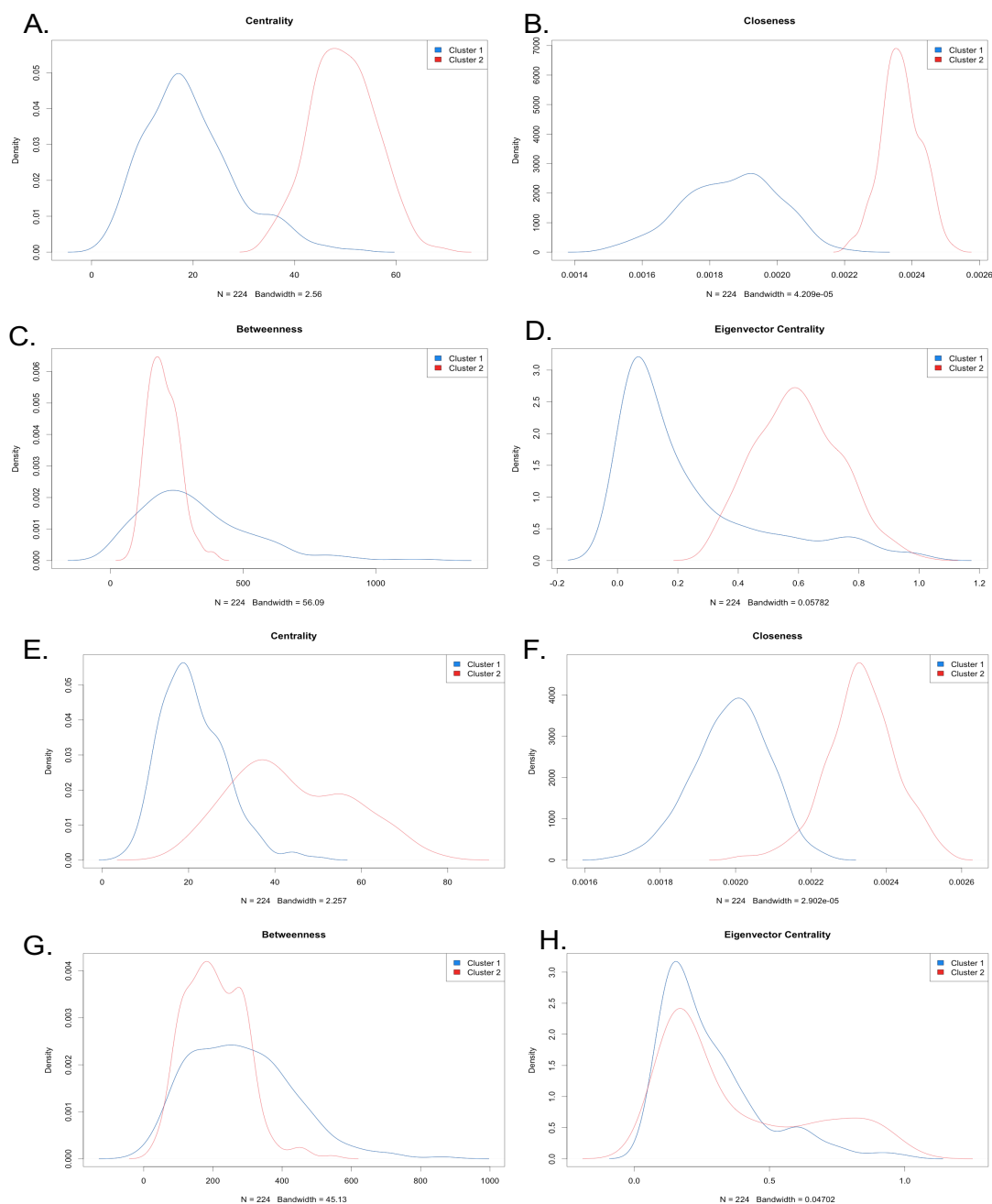


Figure 5.3: Density plots of degree centrality parameters. Figures (A)-(D) are based on analysis of SU2C melanoma cell lines and figures (E)-(H) are based on analysis of CCLE melanoma cell lines. The plots are densities of centrality for (A) and (E), closeness for (B) and (F), betweenness for (C) and (G), and Eigenvector centrality for (D) and (H).

with the presence of a specific mutation). We first tested the equality of the correlation matrices derived from groups defined by mutation status for each data set defined by: BRAF+, BRAF-, KRAS+, and KRAS- against correlation matrices derived from skin fibroblast control gene expression data (as previously described in the methods section). Based on these matrix equivalency tests, we found evidence for statistically significant rewiring when comparing BRAF+, BRAF-, and KRAS- defined cell lines against skin fibroblast data (Table 5.4). However, when testing the KRAS+ against the control groups, the correlation matrix differences was not statistically significant from that derived from the skin fibroblast data (this may have been a result of the small sample size for the KRAS+ cell lines; $n = 8$).

When we tested the equivalence of correlation matrices across BRAF and KRAS mutation status defined groups within each data set, we found evidence for statistically significant differences based on tests that involved permuting mutation status (Table 5.4). We used differential correlations between mutant-defined and wild type-defined groups to assess the overall degree of rewiring. For the BRAF-associated genes, we found strong correlation differences between BRAF+ and BRAF- groups within a small set of gene pairs. However, for the KRAS-associated genes, roughly half of the genes pairs exhibited statistically significant correlation changes when assessing differences between KRAS+ and KRAS- groups (Figures 5.4A-B). Interestingly, when we considered the genes defined in the BRAF network for the TCGA melanoma samples, we also observed differences when comparing BRAF+/BRAF- derived correlation matrices as well as BRAF V600E+/BRAF- derived correlation matrices (Jennrich method significant, $p\text{-val} < .027$ and $p\text{-val} < .032$, respectively; Table 5.4). However, when we used VarScan to call the somatic mutation status and assign mutation status to the patient tumor samples, the signal from the matrix equivalence test was lost, suggesting that the network equivalence changes were only observable in the filtered samples.

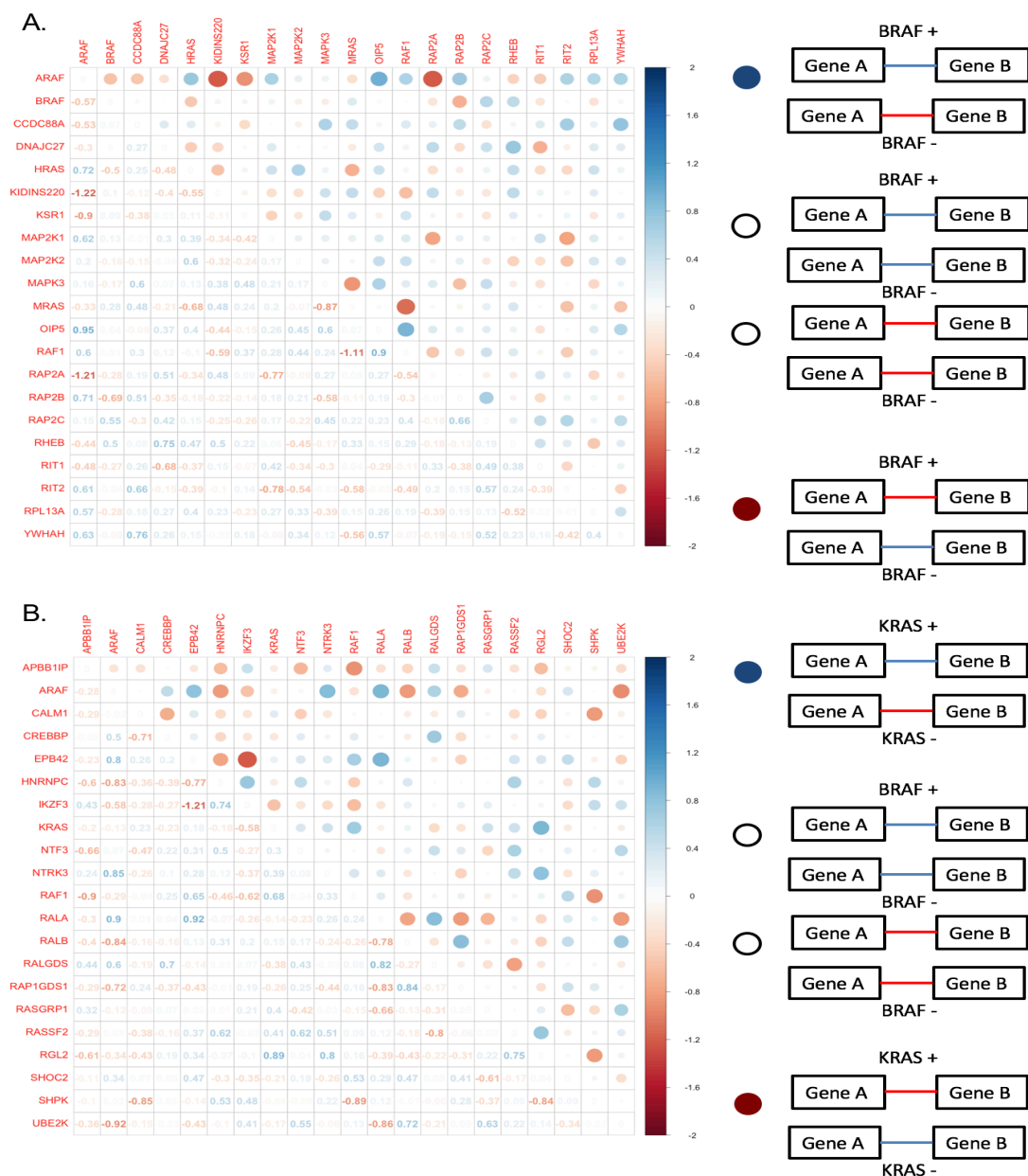


Table 5.4: Table of correlation equality tests for the genes in the specified networks. Tests are between mutant and wild type samples. Asterisks represent at least one significant test within comparison (Stieger or Jennrich).

Dataset	Samples	Network	St Chi-sq	Sim Mut St pval	JR Chi-sq	Sim Mut JR pval	Sig in 1 model?
TGen Cell Line	BRAF+/-	BRAF GM	209	0.019	802	0.754	*
	BRAF+/control	BRAF GM	316.9404	0.00909909	972.9073	0.2678732	*
	BRAF-/Control	BRAF GM	333.675	0.00309969	670.7589	0.8981102	*
	KRAS+/Control	BRAF GM	228.3091	0.1333195	1086.123	0.6022398	
	KRAS-/Control	BRAF GM	373.2865	0.00049995	792.552	0.3406659	*
	KRAS+/-	KRAS GM	158	0.437	678	0.042	*
	BRAF+/control	KRAS GM	310.6259	0.00629937	722.1048	0.820018	*
	BRAF-/Control	KRAS GM	285.9183	0.01479852	706.6058	0.910309	*
	KRAS+/Control	KRAS GM	207.9291	0.2406963	1077.38	0.7448255	
	KRAS-/Control	KRAS GM	338.746	0.00079992	535.796	0.9529047	*
TCGA Samples	BRAF+/-	BRAF GM	673	0.1812	1325	0.0269	*
	BRAF V600E +/BRAFWt	BRAF GM	658	0.2263	1332	0.0319	*
TCGA (WT somatic calls)	BRAF+/- (add somatic called)	BRAF GM	636	0.3107	1195	0.3592	
	BRAF V600E +/BRAFWt	BRAF GM	625	0.3643	1236	0.2996	
	BRAF V600E +/BRAFWt	26 DEGs GM	19513.81	0.01887179	92424.32	0.9985938	*

5.4.5 BRAF V600E+/- Differentially Expressed Genes Analysis with the Melanoma TCGA Tumor Data

We identified genes that exhibited average gene expression level differences between BRAF V600E+ and BRAF- tumor samples in the entire TCGA data using DESeq (FDR adjusted $p < .05$) and edgeR (FDR $< .01$) between and identified 26 differentially expressed genes (DEGs) (Table 5.4). These genes were then used to extract a new gene network from geneMANIA. As anticipated, when we compared the correlation matrices based on the genes in this new network between BRAF+ and BRAF- groups, we observed statistically significant network changes. However, an analysis of the GO networks derived from these genes produced mostly non-significant differences between the BRAF V600E+ and BRAF-patient samples. These results indicate that the patient-specific network rewiring may have been isolated to specific nodes and were less pronounced than comparisons to the melanoma cell lines. To further investigate the overall differences in the correlations among gene expression levels between the TCGA BRAF V600E+/- samples, we calculated the spearman correlation between 348 DEGs (edgeR FDR $< .01$ and DESeq FDR-adjusted p -value $< .05$) and all other genes we had expression level data on. Using a strict Bonferroni-adjusted significance threshold, we observed 63,118 (0.88%) significant gene correlations in BRAF V600E+ patients,

81,765 (1.14%) significant gene correlations in BRAF- samples, and 21,844 (0.31%) significant gene pairs in common. Overall, we found that the majority of the significant gene pair correlations were significant in only one of the groups (Figure 5.16), further indicating that the correlation strengths and overall transcriptional networks between TCGA BRAF V600E+ and BRAF- tumor samples are unique.

5.4.6 Network Centrality Tests for the TGen Melanoma Cell Lines

To investigate the importance of individual genes or pairs of genes in the transcriptional correlation networks we observed in our analyses of the TGen melanoma cell line data, we calculated the node centrality statistics for each network. These analyses suggested that the most important genes associated with BRAF+ cell lines for all genes in the BRAF network were ARAF, RAF1, and OIPS; whereas the most central or connected genes in the BRAF- cell lines were RAF1 and HRAS (Figure 5.17, Figure 5.18, and Table 5.5). For the genes in the BRAF geneMANIA-derived network, the average correlation across all node centrality statistics was 0.1121 for BRAF+/- and 0.0001 for KRAS+/- indicating the existence of different central nodes (Table 5.5) for the networks. For the KRAS geneMANIA network, we observed negative correlations when comparing them across KRAS+/- cell lines for all node centrality statistics, which indicates that extensive transcriptional rewiring is occurring in the networks. Additionally, based on the genes that appeared as most central to the identified networks, DGIdb analysis suggested different drug targets. For the BRAF+ cell lines, the most central or connected genes in the networks indicated that MEK inhibitors would be likely candidate drugs (Table 5.5). However, for the BRAF- cell lines, mTOR inhibitors were predicted as best targets.

5.4.7 Network Centrality Tests for the TCGA Data

Centrality and connection analysis of the TCGA patient samples based on BRAF mutation status also identified distinct central genes in transcriptional

Table 5.5: Assessing significant nodes by degree centrality parameters. A consensus-based approach was used to genes to match to DGIdb to identify candidate therapeutic agents.

Dataset	Samples	Network	DC	BW	CL	EC	Consensus	Predicted Drug Targets (Top Genes Only)
TGen Cell Line	BRAF+	BRAF GM	ARAF	ARAF	RAF1, ARAF, OHS	ARAF	ARAF	Sorafenib, Trametinib, Regorafenib, BMS-908662, Adenosine Triphosphate
TGen Cell Line	BRAF-	BRAF GM	HRAS	RAF1, HRAS	RAF1, HRAS	HRAS	HRAS	AZD8055, Everolimus, DB08751
TGen Cell Line	KRAS+	KRAS GM	RASSF2, RGL2	RASSF2, RGL2	RASSF2, RGL2	RASSF2, RGL2	RASSF2, RGL2	
TGen Cell Line	KRAS-	KRAS GM	NTF3	NTF3	NTF3	NTF3	NTF3	
TCGA Samples	BRAF+	BRAF GM	CCDC88A	CNKSRI	CNKSRI	CCDC88A	CCDC88A, CNKSRI	
TCGA Samples	BRAF V600E+	BRAF GM	CCDC88A	CNKSRI	RHEBL1	CCDC88A	CCDC88A	
TCGA Samples	BRAF-	BRAF GM	RAP2C	RHEBL1	RAP2C	RAP2C	RAP2C	
TCGA Samples	BRAF V600E+	26 DEGs GM	ATF1, CNF3, GTG2E1, GPT2E2, GTF3A, PHTF2, TFEC, TFF1, TF, BCL2L15, FAM3C, PMCHL1	BCAP29	CSTF3	BCAP29	BCAP29	
TCGA Samples	BRAF-	26 DEGs GM	TFF1	BCAP29	TFF1	BCAP29	TFF1, BCAP29	

networks. Although there was a strong positive correlation for each node centrality statistic ($r > .48$) across BRAF+ and BRAF V600E+ when compared against BRAF-, the networks derived from the gene expression-based correlation coefficients for BRAF+ patients had CCDC88A and CNKSRI as most central or connected genes, whereas networks derived from BRAF V600E+ samples had only CCDC88A as the most important node (Figure 5.5, Figure 5.19, and Table 5.6). CCDC88A is a key modulator of the AKT-mTOR signaling pathway [UniProt Consortium, 2015] and CNKSRI promotes invasion of cancer cells through NFB dependent signaling [Fritz and Radziwill, 2010] (Table 5.5). For the BRAF- derived networks, the central node was found to be RAP2C, which is a RAS-related protein in the MAPK/ERK signaling pathway. The consensus central nodes for melanoma cell lines with and without BRAF mutations were mapped to candidate drugs using DGIdb. Interestingly, RAF and MEK inhibitors (Sorafenib, Trametinib, Regorafenib, BMS-908662, and Adenosine Triphosphate) were predicted to be the best drugs give the observed BRAF+ transcriptional network, whereas the mTOR inhibitors (AZD8055, Everolimus, and DB08751) were predicted to be the best drugs in the observed BRAF- transcriptional network (Figures 5.19-5.21, Table 5.5).

5.4.8 Are Strong Edges Present in Both Mutant and Wild Type Groups?

To evaluate the strength of the correlation coefficients obtained, we assessed the overall distribution of coefficients (Figure 5.24). We found that the distribution of pairwise correlation for BRAF-associated genes were approximately the same in both BRAF+ and BRAF- subgroups. However, when evaluating the KRAS-

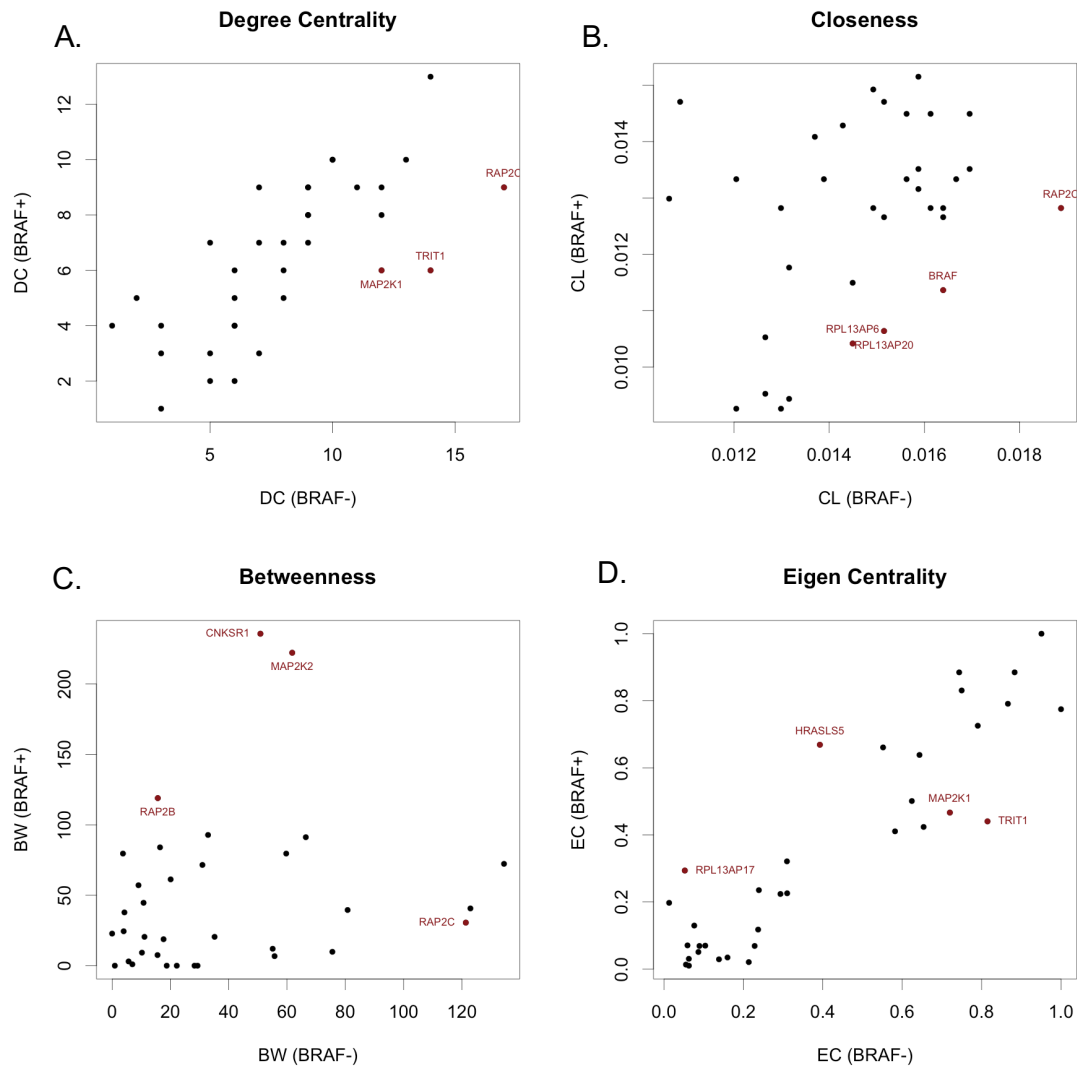


Figure 5.5: Network centrality statistics identifying differential central nodes between BRAF+ and BRAF- patients from TCGA data. (A) Degree Centrality, (B) Closeness Centrality, (C) Betweenness Centrality, and (D) Eigenvector Centrality can all be used to assess different properties of the network.

Table 5.6: A comparison between degree centrality parameters through correlations. DC = Degree Centrality, BW = Betweenness, CL = Closeness, EC = Eigenvector Centrality.

Dataset	Samples	Network	DC Cor	BW Cor	CL Cor	EC Cor	Avg Cor
TGen Cell Line	BRAF+/-	BRAF GM	0.05340418	0.1522432	0.1027221	0.1401849	0.112138595
TGen Cell Line	KRAS+/-	BRAF GM	0.103506	0.03017858	0.003084305	-0.1362763	0.000123146
TGen Cell Line	BRAF+/-	KRAS GM	-0.06479926	-0.09663954	-0.3566459	-0.4211979	-0.23482065
TGen Cell Line	KRAS+/-	KRAS GM	-0.5021395	-0.1873172	-0.5892824	-0.3236824	-0.400605375
TCGA Samples	BRAF+/-	BRAF GM	0.7678829	0.2597188	0.28202	0.8450726	0.538673575
TCGA Samples	BRAF V600E +/-BRAFwt	BRAF GM	0.777453	0.363136	0.4762343	0.8624035	0.6198067
TCGA (WT somatic calls)	BRAF+/- (add somatic called)	BRAF GM	0.7786251	0.5123316	0.4362009	0.8747135	0.650467775
TCGA (WT somatic calls)	BRAF V600E +/-BRAFwt	BRAF GM	0.7920405	0.6776392	0.6469417	0.8510963	0.741929425
TCGA (WT somatic calls)	BRAF V600E +/-BRAFwt	26 DEGs GM	0.2356217	0.8135823	0.3493597	0.4647389	0.46582565

associated genes, there was an increased proportion of correlation coefficients near the tail for KRAS+ subgroup when compared against the KRAS- subgroup, which suggests a stronger network in the KRAS+ subgroup exists. This is further confirmed when assessing the upper 5% tail in correlation coefficients (for BRAF-associated genes, $r = .54$ in both BRAF+/-BRAF- subgroups; for KRAS-associated genes, $r = .46$ in KRAS- subgroup and $r = .76$ for KRAS+ subgroup).

To assess whether the strongest co-expression patterns were observed in normal laboratory conditions, we extracted all correlation coefficients less than $-.7$ or greater than $.7$ (Table 5.9). We then used the resulting gene list as an input into geneMANIA in order to calculate the proportion of edges that were previously reported. Surprisingly, we found that a majority of the strong co-expression edges were not previously identified (Figure 5.25). Specifically, for the BRAF- subgroup, RAP2B and RAP2C are co-localized and share protein domains; for the BRAF+ subgroup, BRAF and RHEB are in the same pathway and have physical interactions; and for the KRAS+ subgroup, RALA and UBE2K are co-expressed, KRAS and RASSF2 have physical interactions, RGL2 and KRAS are in the same pathway and have physical interactions, and RASSF2 and APBB1IP have shared protein domains. Thus, only six of the 24 identified strong edges were previously reported.

5.4.9 *De Novo* Network Reconstruction Analysis

We also considered reconstructing transcriptional networks *de novo* using WCGNA [Langfelder and Horvath, 2008] [Langfelder and Horvath, 2012] across mutation-status (supervised) groups on the TGen melanoma cell lines and TCGA melanoma patient samples. For the TGen melanoma cell lines, WCGNA modules were created using BRAF and KRAS mutation groupings (Figure 5.6A-B, 5.22). For each of the modules identified from the WCGNA assessment of the wild-type cell lines, we identified the modules identified in the mutation-bearing cell lines with the highest intersection of genes. We then tested these intersecting genes for network rewiring using network equivalence and permutation tests (although we were unable to test for KRAS because of the limited number of KRAS+ melanoma cell lines). Our results confirmed the network rewiring on the *de novo* reconstructed networks (Table 5.7). When creating ten modules for the melanoma cell lines, six out of the ten networks had strong p-values; and when creating twenty modules for the melanoma cell lines, seven out of the 20 networks had strong p-values. Although the network rewiring in the TCGA patients was not as extensive, there were still some modules that had statistically significant p-values (Figure 5.6C, Figure 5.23 and Table 5.7). We found that there was network rewiring in a small number of *a priori* created modules in both BRAF+ vs. BRAF- and BRAFV600E+ vs. BRAF- melanoma cohorts.

5.5 Discussion

Although there has been a great deal of attention given to the development and use of pathway and genetic network analysis tools in understanding disease pathogenesis and drug targeting, these tools often rely upon the use of pathway and network information derived from analyses of genes in normal and non-diseased cells, cell lines and tissues [Hofree et al., 2013] [Jung and Kim, 2014] [Mootha et al., 2003] [Subramanian et al., 2005]. This compromises relevant analyses since pathways and genetic networks are not static and are often perturbed by various diseases of interest in those analyses. This is especially true in the

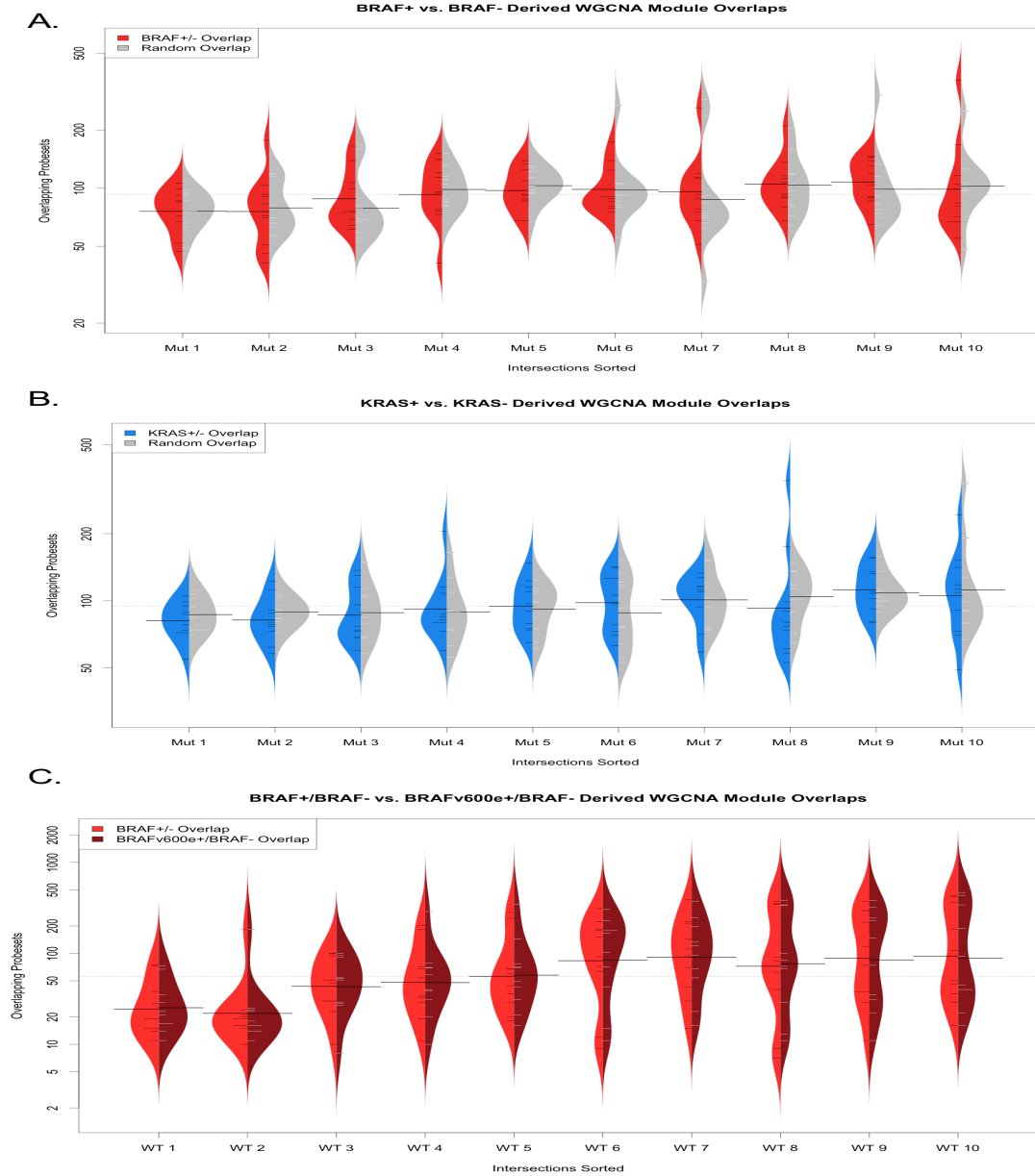


Figure 5.6: WGCNA derived modules used for *a priori* assessment of network rewiring. (A) Beanplots of overlapping modules between melanoma cell line BRAF+/- (red) and random permutation-based (gray) overlap (B) Beanplots of overlapping modules between melanoma cell line KRAS+/- (blue) and random permutation-based (gray) and (C) Beanplots of overlapping modules between TCGA melanoma patient BRAF+/- (red) and BRAFv600e+/- (dark red) overlap

Table 5.7: Correlation equality tests between WGCNA defined modules in BRAF+/- SU2C cell lines, BRAF+/- TCGA melanoma patients, and BRAFv600e+/- TCGA melanoma patients.

SU2C Cell Lines: BRAF+ vs BRAF-				TCGA Melanoma Patients: BRAF+ vs BRAF-				TCGA Melanoma Patients: BRAFv600e+ vs BRAF-			
WGCNA Module	Network Size	St Chi-sq	Sim Mut St pval	WGCNA Module	Network Size	St Chi-sq	Sim Mut St pval	WGCNA Module	Network Size	St Chi-sq	Sim Mut St pval
WGCNA 1 (k = 10)	177	21137	0.003	WGCNA 1 (k = 10)	388	104850.9	0.05494505	WGCNA 1 (k = 10)	375	90564.64	0.1248751
WGCNA 2 (k = 10)	210	22813	0.152	WGCNA 2 (k = 10)	374	80097.34	0.2507493	WGCNA 2 (k = 10)	379	81201.04	0.2617383
WGCNA 3 (k = 10)	363	NA		WGCNA 3 (k = 10)	311	46890.96	0.5984016	WGCNA 3 (k = 10)	304	46140.55	0.5124875
WGCNA 4 (k = 10)	173	21100	0.001	WGCNA 4 (k = 10)	203	26916.43	0.1318681	WGCNA 4 (k = 10)	289	59361.26	0.1248751
WGCNA 5 (k = 10)	146	9770	0.135	WGCNA 5 (k = 10)	375	103328.7	0.1298701	WGCNA 5 (k = 10)	380	106081.3	0.1278721
WGCNA 6 (k = 10)	151	12584	0.042	WGCNA 6 (k = 10)	100	5847.691	0.08191808	WGCNA 6 (k = 10)	100	7580.433	0.001998002
WGCNA 7 (k = 10)	139	10240	0.044	WGCNA 7 (k = 10)	432	110979.6	0.2277722	WGCNA 7 (k = 10)	462	133053.9	0.1668332
WGCNA 8 (k = 10)	106	6058	0.002	WGCNA 8 (k = 10)	75	4024.792	0.04195804	WGCNA 8 (k = 10)	73	4256.042	0.02397602
WGCNA 9 (k = 10)	261	27640	0.343	WGCNA 9 (k = 10)	282	54268.77	0.04195804	WGCNA 9 (k = 10)	346	79351.82	0.06293706
WGCNA 10 (k = 10)	165	16857	0.069	WGCNA 10 (k = 10)	185	18780.28	0.3806194	WGCNA 10 (k = 10)	184	18863.78	0.3366933
WGCNA 1 (k = 20)	115	6637	0.183	WGCNA 1 (k = 20)	127	9636.389	0.1228771	WGCNA 1 (k = 20)	128	10609.87	0.06693307
WGCNA 2 (k = 20)	86	2174	0.987	WGCNA 2 (k = 20)	161	13476.35	0.5004995	WGCNA 2 (k = 20)	159	12823.37	0.5134865
WGCNA 3 (k = 20)	41	868	0.121	WGCNA 3 (k = 20)	139	8823.917	0.7122877	WGCNA 3 (k = 20)	136	8576.759	0.7122877
WGCNA 4 (k = 20)	50	1093	0.007	WGCNA 4 (k = 20)	160	12930.49	0.4565435	WGCNA 4 (k = 20)	140	10142.85	0.185814
WGCNA 5 (k = 20)	38	731	0.089	WGCNA 5 (k = 20)	138	10933.09	0.2257742	WGCNA 5 (k = 20)	118	8564.582	0.2657942
WGCNA 6 (k = 20)	58	1269	0.563	WGCNA 6 (k = 20)	405	112287.8	0.1418581	WGCNA 6 (k = 20)	370	89285.78	0.1838162
WGCNA 7 (k = 20)	172	11040	0.481	WGCNA 7 (k = 20)	201	35418.36	0.08191808	WGCNA 7 (k = 20)	181	28131.66	0.09496509
WGCNA 8 (k = 20)	38	834	0.053	WGCNA 8 (k = 20)	105	7231.11	0.1248751	WGCNA 8 (k = 20)	159	14801.29	0.2457542
WGCNA 9 (k = 20)	58	2249	0.017	WGCNA 9 (k = 20)	171	17254.12	0.2557443	WGCNA 9 (k = 20)	173	17306.31	0.2737263
WGCNA 10 (k = 20)	48	881	0.400	WGCNA 10 (k = 20)	187	20291.88	0.2167832	WGCNA 10 (k = 20)	189	21524.95	0.1848152
WGCNA 11 (k = 20)	88	4670	0.150	WGCNA 11 (k = 20)	69	3021.248	0.1108891	WGCNA 11 (k = 20)	64	2667.571	0.0797921
WGCNA 12 (k = 20)	124	NA		WGCNA 12 (k = 20)	246	34916.67	0.2657343	WGCNA 12 (k = 20)	222	28239.58	0.2717283
WGCNA 13 (k = 20)	35	971	0.003	WGCNA 13 (k = 20)	55	1602.315	0.3716284	WGCNA 13 (k = 20)	54	1688.05	0.2007992
WGCNA 14 (k = 20)	88	3814	0.215	WGCNA 14 (k = 20)	218	26417.98	0.3296703	WGCNA 14 (k = 20)	230	27211.54	0.0685315
WGCNA 15 (k = 20)	50	2179	0.016	WGCNA 15 (k = 20)	21	293.7024	0.02397602	WGCNA 15 (k = 20)	15	174.5572	0.01798202
WGCNA 16 (k = 20)	46	1193	0.049	WGCNA 16 (k = 20)	134	12282.96	0.1608392	WGCNA 16 (k = 20)	84	4154.571	0.2887113
WGCNA 17 (k = 20)	45	1638	0.023	WGCNA 17 (k = 20)	128	10300.94	0.2267732	WGCNA 17 (k = 20)	153	15411.85	0.1918082
WGCNA 18 (k = 20)	72	3108	0.077	WGCNA 18 (k = 20)	170	15247.5	0.4195804	WGCNA 18 (k = 20)	169	15297.47	0.4155844
WGCNA 19 (k = 20)	41	1211	0.016	WGCNA 19 (k = 20)	117	8769.188	0.2357642	WGCNA 19 (k = 20)	115	9200.047	0.1338661
WGCNA 20 (k = 20)	42	1030	0.056	WGCNA 20 (k = 20)	12	89.48155	0.09190809	WGCNA 20 (k = 20)	13	77.14802	0.5094905

study of cancer, where there are many stresses including gene perturbations, drugs and environmental influences, that all impact the way genes interact and enter into, e.g., signaling or regulatory relationships. This is not to say that researchers cannot use information about the relationships among genes derived from studies of normal cells and tissues as an initial starting point for detecting differences in gene relationships in a disease state, but it does suggest that a great deal of caution should be made in making claims about the relationships between genes in a disease state, i.e., their “wiring” as reflected in the networks they form.

We assessed evidence for network rewiring *a posteriori* among a set of melanoma cell lines and tumor profiles by leveraging multivariate mixture models in the analysis of the relationships among gene expression values. The mixture models allowed us to identify subsets of cell lines and tumors that exhibited the most evidence for rewiring (i.e., differences in the relationships among genes based on their expression patterns) without having to specify *a priori* which cell lines or tumors we wanted to contrast for network differences. We focused on the MAPK pathway in particular and found that, based on very robust simulation-based tests, heterogeneity in networks reflective of rewiring did indeed exist in our cell lines and tumor profiles. Furthermore, we explored various factors that could explain the differences between the identified cell lines, tumor profiles identified from the

mixture models, but found that, in the melanoma cell lines, common melanoma gene alterations could not explain the group differences we identified, suggesting that some other factor is responsible for the rewiring in these groups of cell lines and tumors. We also applied degree centrality and other network node importance statistics to the networks in the subgroups of cell lines and tumors identified. These analyses suggested that different genes played more central roles in the different subgroup networks and hence were likely better drug targets. These results suggest that the use of pathway and genetic network information on a *biologically relevant* subset of cell lines or tumors may reveal greater insights into drug candidates.

As a complement to using mixture models to identify novel subgroups of melanoma cell lines and tumors exhibiting evidence for genetic network rewiring, we also compared the genetic networks based on correlations between expression levels of genes in melanoma cell lines and patient samples across groups defined by the presence of specific mutations in supervised analyses. We found that there was evidence of network rewiring when stratifying the cell lines based on mutation status. Although network rewiring was also observed in the patient samples, it was less pronounced. The cancer cell lines may have gone through numerous passages, contributing to changes in genetic network and gene expression patterns. Another potential explanation is that cell lines represent a fairly homogeneous population with chaotic genomes, while patient tumor samples are heterogeneous with respect to the mutations they carry and are often comprised of tumor, stroma, infiltrating immune cells, and other non-tumor cells. It is likely that this heterogeneity makes it much more difficult to clearly identify changes in correlation in gene expression. Tumor heterogeneity also makes identifying consensus between-patient drug targets more difficult since patient-specific “rewiring” could affect the downstream drug target identification.

Our analyses also involved the use of various network centrality statistics to identify candidate genes (i.e., nodes in the networks) that might be ideal drug targets among both BRAF mutant cell lines and BRAF wild-type cell lines, especially for groups defined (in a supervised manner) by mutation status. As expected, we found that MEK inhibitors were the most likely drugs to affect BRAF mutant cell

lines, whereas the mTOR inhibitors were the most likely drugs to affect BRAF wild-type cell lines. Ultimately, we feel that our analyses strongly suggest that an understanding of how the relationships among genes or other factors influencing, or associated with, disease pathogenesis can facilitate treatment decisions. However, the identification of these relationships requires a clear sensitivity to the fact that those relationships change in the disease state in ways that cannot be easily anticipated from, e.g., mutation status or average gene expression levels alone.

5.6 Acknowledgements

Research supported by a Stand Up To Cancer Melanoma Research Alliance Melanoma Dream Team Translational Cancer Research Grant. Stand Up To Cancer is a program of the Entertainment Industry Foundation administered by the American Association for Cancer Research. Dr. Schork and Mr. Ding are also supported by NIH grants: U19 AG023122-09; R01 DA030976-05; R01 MH094483-03; R01 MH100351-02; R21 AG045789-01A1; UL1TR001442-01; and U24AG051129-01, in addition to grants from Janssen, Human Longevity, Inc., and the Tanner Project Foundation. We would like to thank Dr. Hannah Carter for her thoughtful comments and scientific discussions.

Chapter 5 is a full reprint of the publication: **Ding, K.**, Finlay, D., Yin, H., Hendricks, W.P.D., Sereduk, C., Kiefer, J., Sekulic, A., LoRusso, P.M., Vuori, K., Trent, J.M., and Schork, N.J., Network Rewiring in Cancer: Applications to Melanoma Cell Lines and the Cancer Genome Atlas Melanoma Patients. (In Preparation), with permission from all coauthors. The dissertation author was the primary researcher and author of this material.

5.7 Supplemental Figures

Table 5.8: List of skin fibroblast samples obtained from GEO. Sample repository, sample name, and associated meta data is provided.

GEO Dataset	Sample ID	Sample Label
GSE70818	GSM1820029	Normal 0J
GSE70818	GSM1820030	Normal 0.5J 4hr
GSE70818	GSM1820031	Normal 0.5J 12hr
GSE70818	GSM1820032	Normal 5J 4hr
GSE70818	GSM1820033	Normal 5J 12hr
GSE41751	GSM1023557	Father_rep1
GSE41751	GSM1023558	Father_rep2
GSE41751	GSM1023559	AgeControl_rep1
GSE41751	GSM1023560	AgeControl_rep2
GSE20538	GSM516173	fibroblast C1-0
GSE20538	GSM516174	fibroblast C1-1
GSE20538	GSM516175	fibroblast C1-10
GSE20538	GSM516176	fibroblast C2-0
GSE20538	GSM516177	fibroblast C2-10
GSE20538	GSM516178	fibroblast C3-0
GSE20538	GSM516179	fibroblast C3-1
GSE20538	GSM516180	fibroblast C3-10
GSE17549	GSM437504	CONT_CD14_1
GSE17549	GSM437505	CONT_FB_18
GSE17549	GSM437506	CONT_CD14_2
GSE17549	GSM437507	CONT_CD14_3
GSE17549	GSM437508	CONT_CD14_14
GSE15537	GSM389321	CRL-2072(1)
GSE15537	GSM389322	CRL-2072(2)
GSE15537	GSM389323	CRL-2072(3)
GSE16524	GSM415037	Normal Puerto Rican control 1
GSE16524	GSM415038	Normal Puerto Rican control 2
GSE16524	GSM415039	Normal Puerto Rican control 3
GSE16524	GSM415040	Normal Caucasian control 4

Table 5.9: Identifying edges within BRAS-associated gene networks and KRAS-associated gene networks using co-expression observed within each subgroup. Edges reported had a correlation coefficient less than -.7 or greater than .7.

BRAF-	BRAF+	KRAS-	KRAS+
MAPK3 & CCDC88A	ARAF & KIDINS220		IKZF3 & EPB42
RAP2B & RAP2C	BRAF & RHEB		IKZF3 & HNRPC
	KIDINS220 & RAP2A		NTF3 & APBB1IP
			NTRK3 & RALB
			RAF1 & IKZF3
			RALA & ARAF
			RALA & UBE2K
			RALB & NTRK3
			RALGDS & CREBBP
			RAP1GDS1 & NTRK3
			RAP1GDS1 & RALB
			RASSF2 & KRAS
			RASSF2 & NTRK3
			RGL2 & KRAS
			RGL2 & NTRK3
			RGL2 & RASSF2
			UBE2k & ARAF
			RASSF2 & APBB1IP

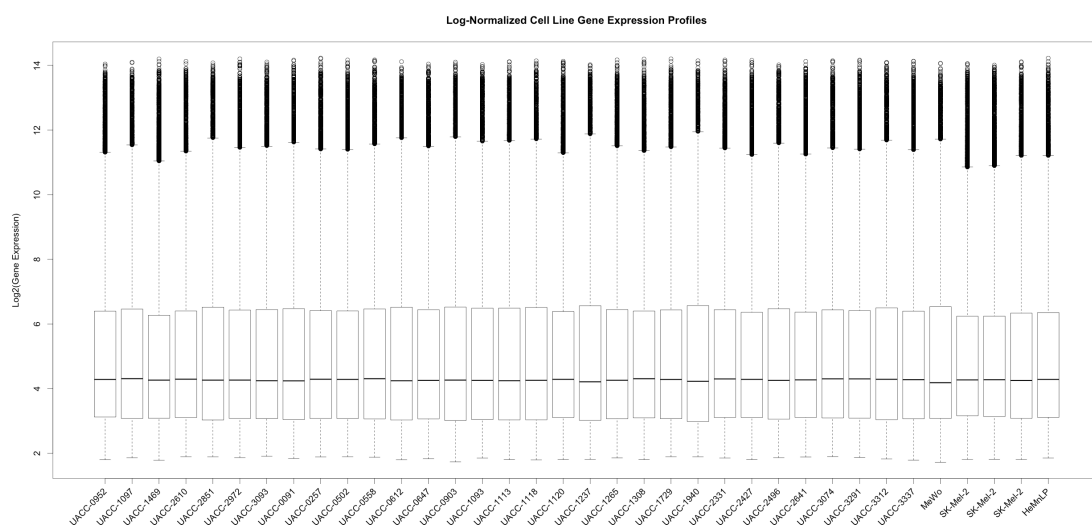


Figure 5.8: Boxplot of normalized gene expression profiles within only melanoma cell lines.

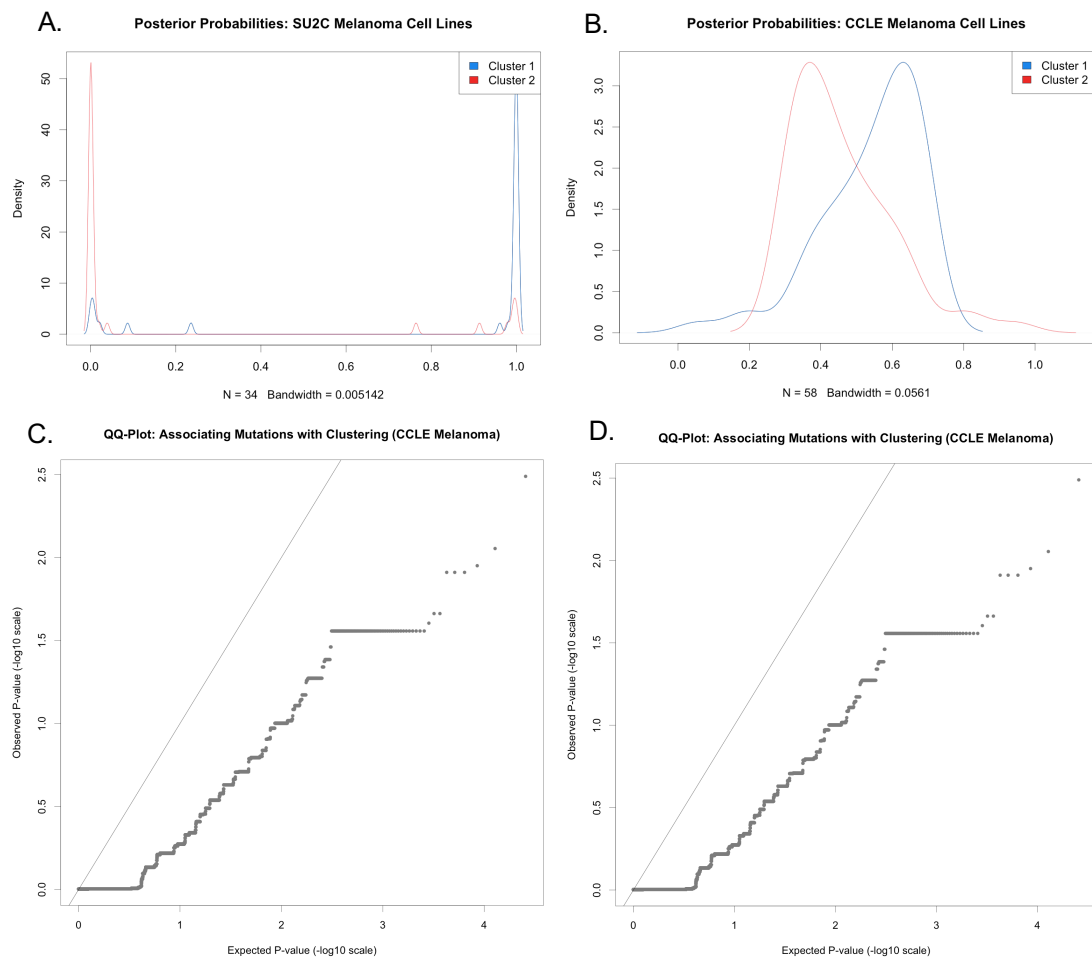


Figure 5.9: Density plots for the posterior probability of the flexmix groups in (A) SU2C melanoma cell lines and (B) CCLE melanoma cell lines. These plots reveal that the flexmix called clusters were most distinct in the SU2C melanoma cell lines.

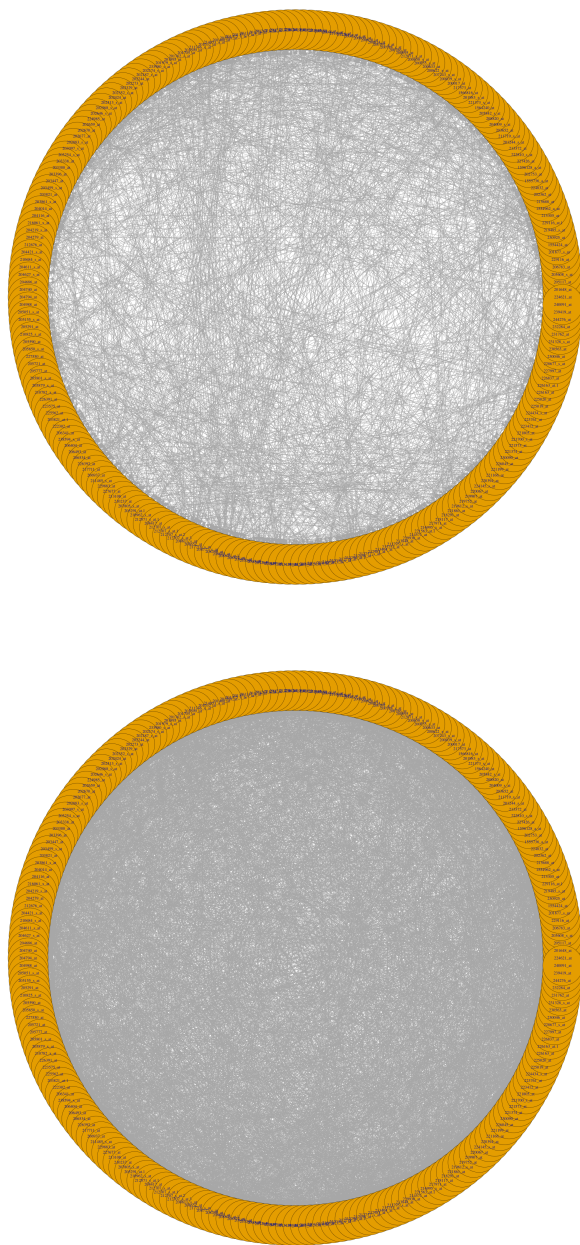


Figure 5.10: Circular network graph of all common nodes. Top network represents the network derived from correlation associations found in cluster one, whereas the bottom network represents the edges present in cluster two. Density of the edges highlights the differences in networks.

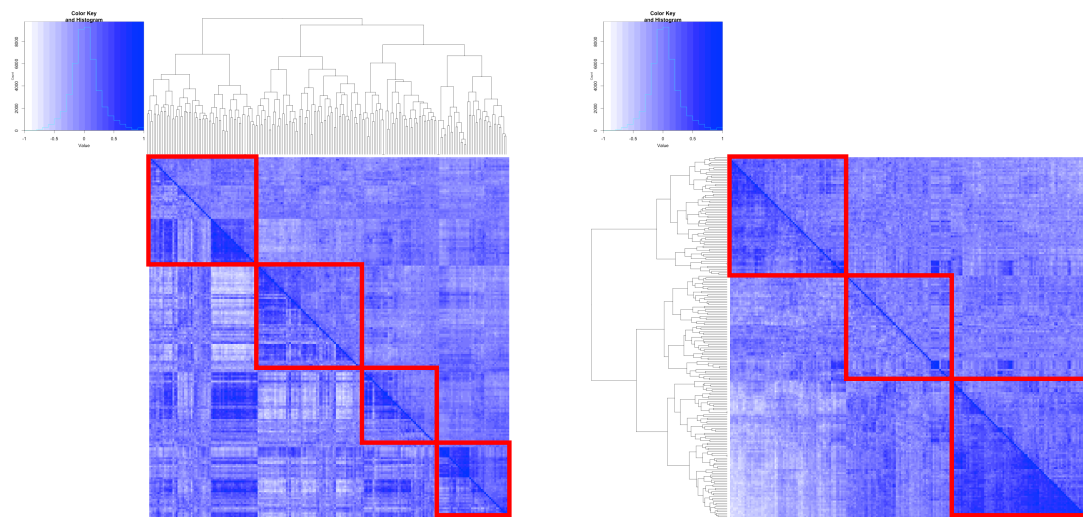


Figure 5.11: Pairwise correlation between MAPK Pathway genes from the identified clusters in TCGA melanoma patients using flexmix were used to generate heatmaps. Upper right hand half of matrix were correlation coefficients from cluster 1, whereas lower left hand half of matrix were correlation coefficients from cluster 2. Heirarchical clustering was subsequently performed using either distances from cluster 1 coefficients (left) or cluster 2 coefficients (right).

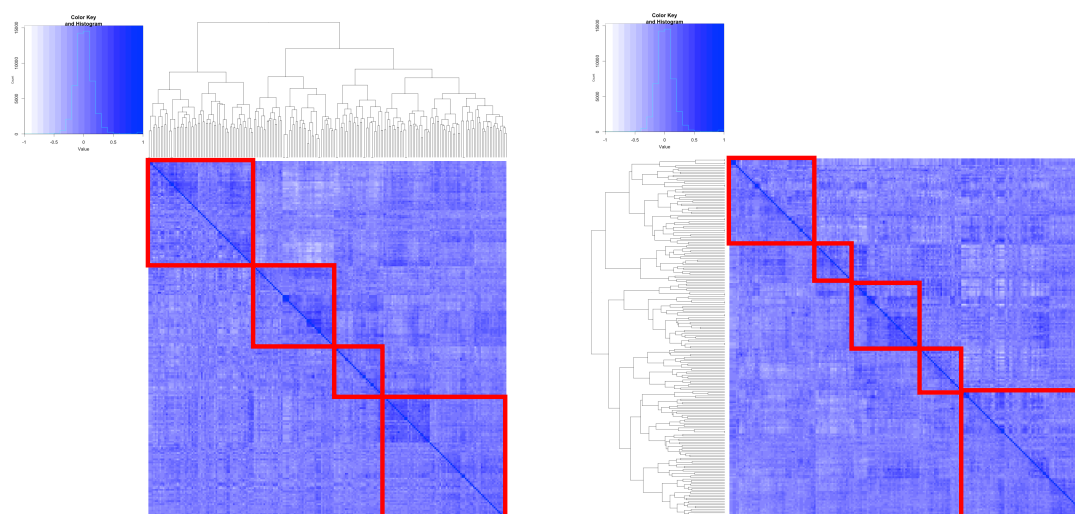


Figure 5.12: Pairwise correlation between MAPK Pathway genes from the identified clusters in all CCLE cell lines using flexmix were used to generate heatmaps. Upper right hand half of matrix were correlation coefficients from cluster 1, whereas lower left hand half of matrix were correlation coefficients from cluster 2. Hierarchical clustering was subsequently performed using either distances from cluster 1 coefficients (left) or cluster 2 coefficients (right).

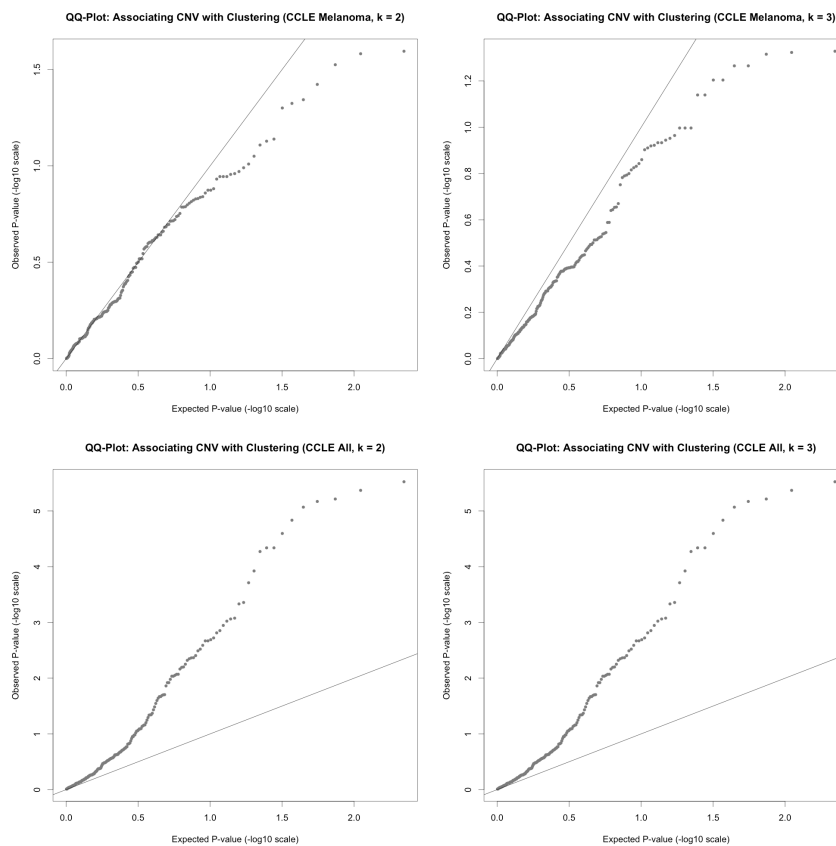


Figure 5.13: QQ-plots for p-values based on association tests for gene copy number variation observed against the groups discovered in the CCLE dataset, using either the melanoma subset (top) or against all cell lines (bottom).

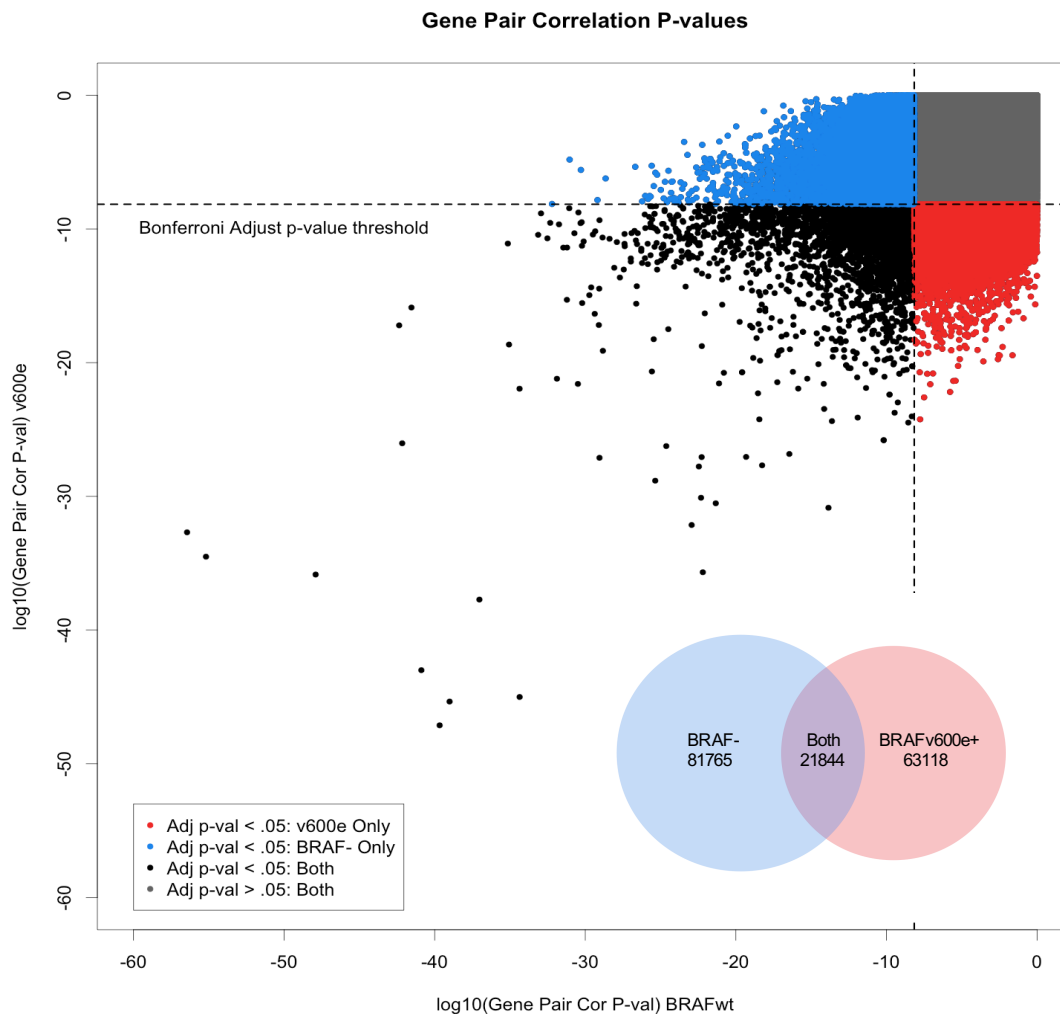


Figure 5.14: Observed p-values based on correlation strength tests. Correlations were compared between those derived from TCGA melanoma patients with BRAFv600e mutation against those derived from BRAF wt. Blue points represent edges significant in only BRAF-, red points represent points significant in only BRAFv600e+, black points represent significance in both.

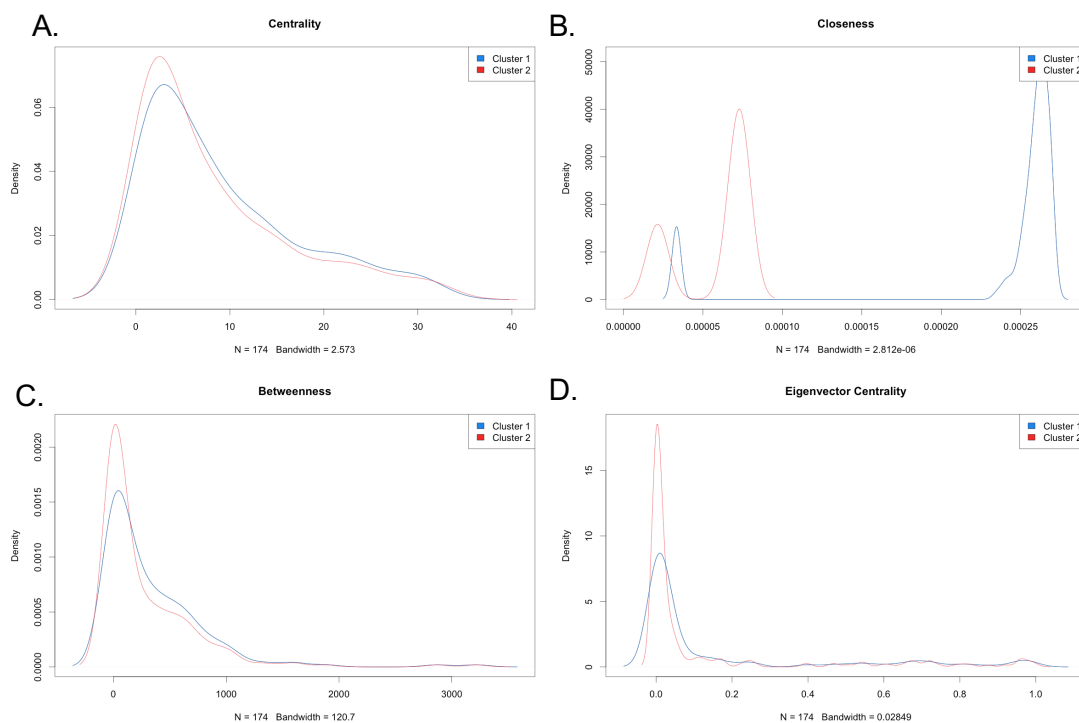


Figure 5.15: Density plot of node centrality statistics for flexmix clusters called on TCGA melanoma patients. We use four different degree centrality parameters (A) Centrality, (B) Closeness, (C) Betweenness, and (D) Eigenvector Centrality.

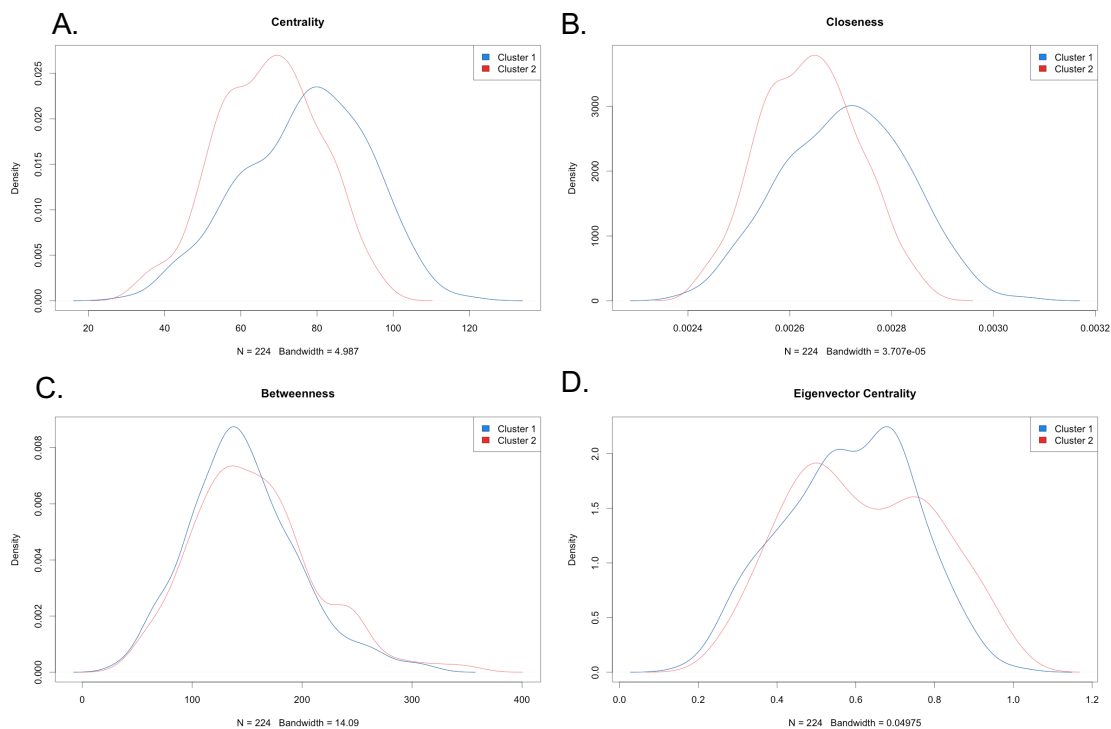


Figure 5.16: Density plot of node centrality statistics for flexmix clusters called across all CCLE cell lines. We use four different degree centrality parameters (A) Centrality, (B) Closeness, (C) Betweenness, and (D) Eigenvector Centrality.

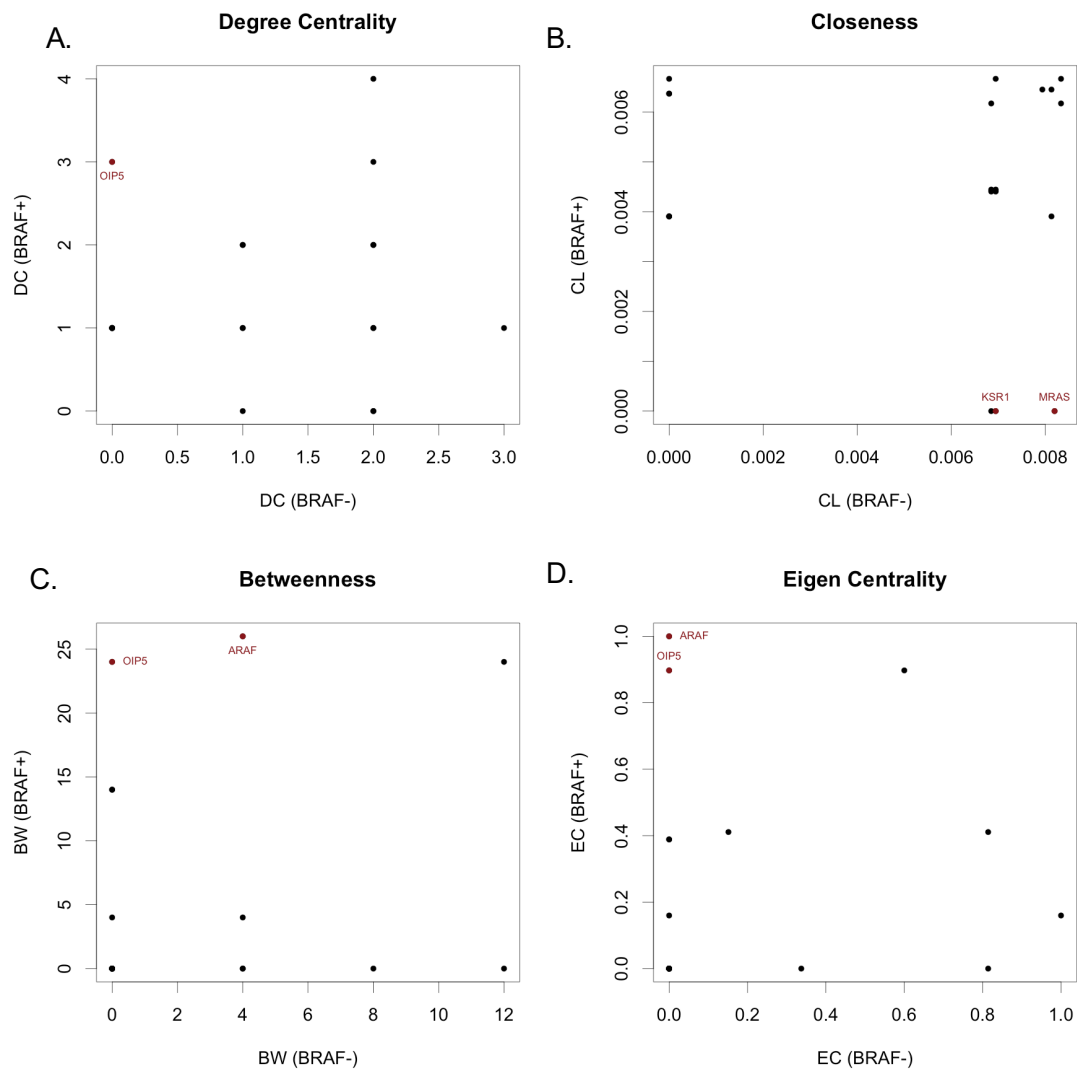


Figure 5.17: Comparing node centrality statistics for BRAF+ and BRAF- from the SU2C cell lines. We use four different degree centrality parameters (A) Centrality, (B) Closeness, (C) Betweenness, and (D) Eigenvector Centrality.

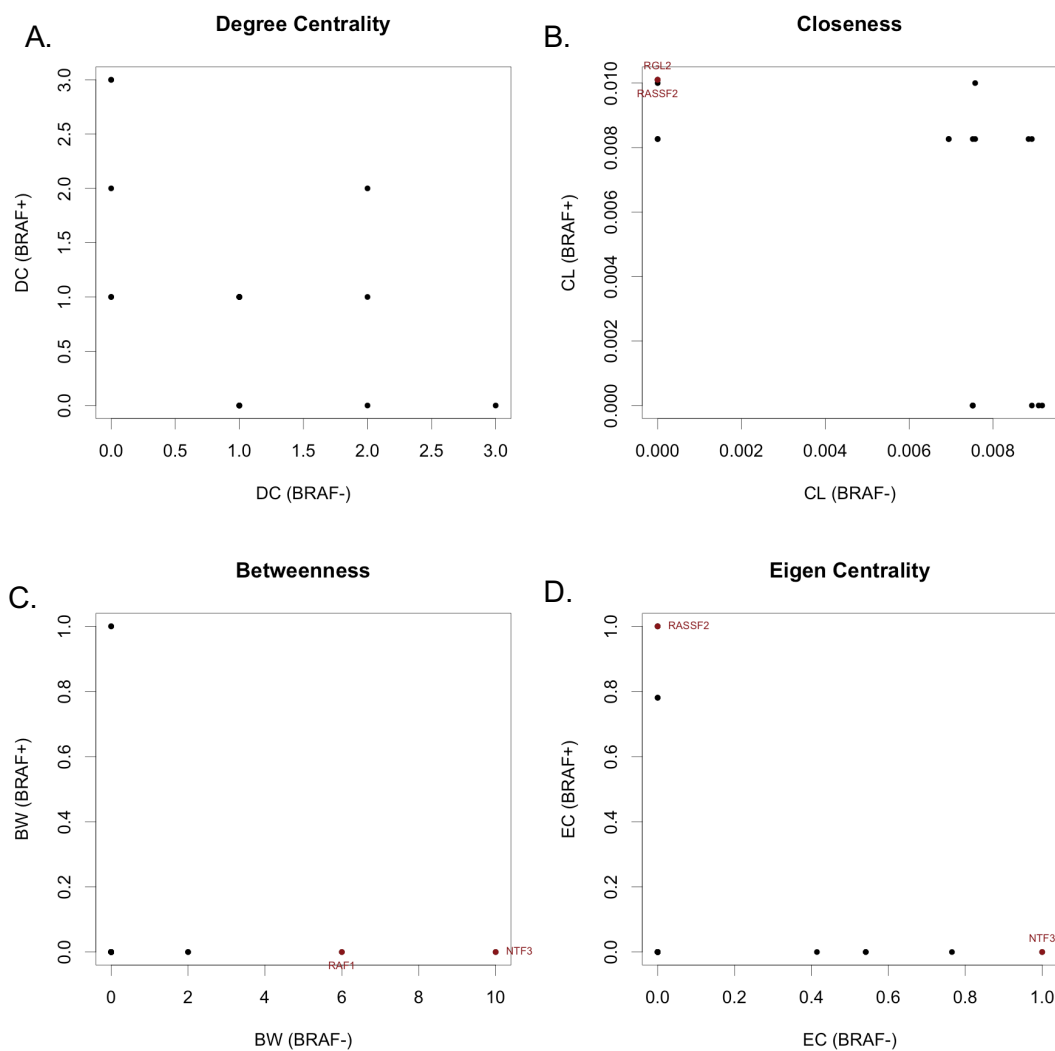


Figure 5.18: Comparing node centrality statistics for KRAS+ and KRAS- from the SU2C cell lines. We use four different degree centrality parameters (A) Centrality, (B) Closeness, (C) Betweenness, and (D) Eigenvector Centrality.

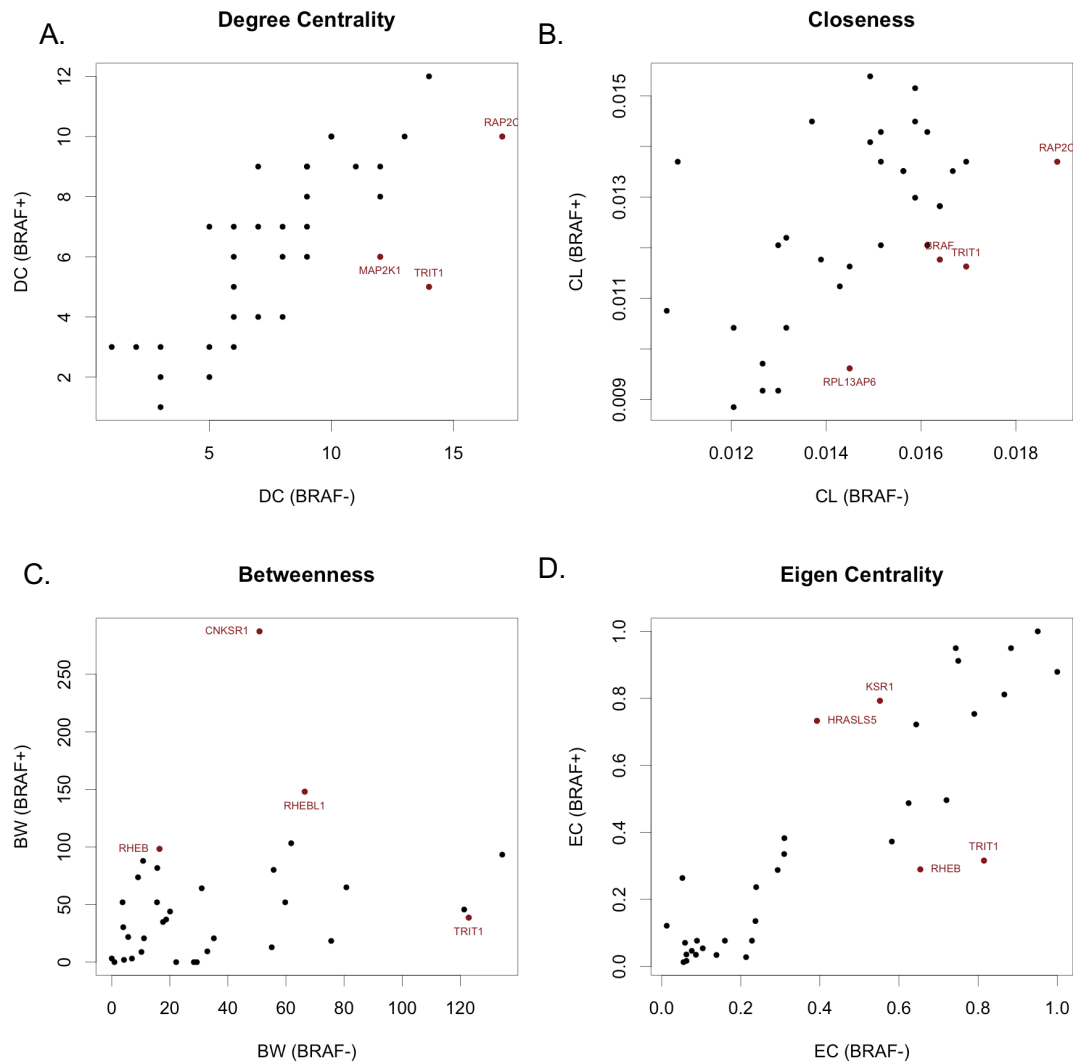


Figure 5.19: Comparing node centrality statistics for BRAFv600e+ and BRAF- from the TCGA melanoma patients. We use four different degree centrality parameters (A) Centrality, (B) Closeness, (C) Betweenness, and (D) Eigenvector Centrality.

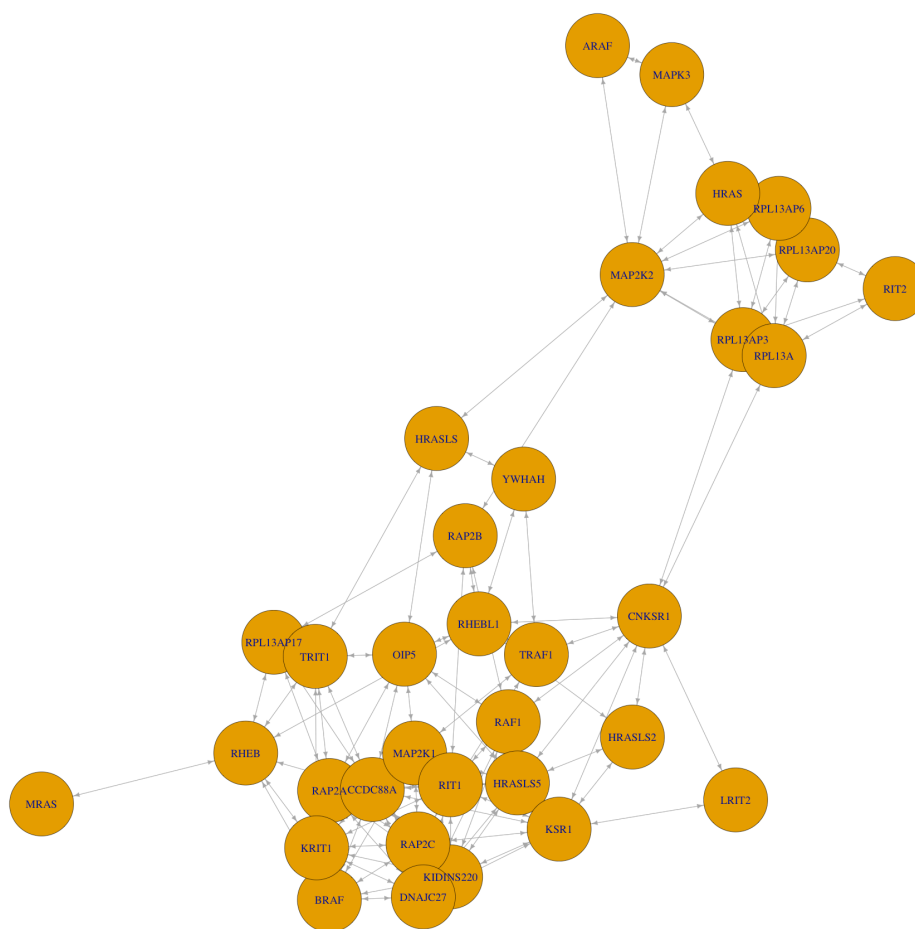


Figure 5.20: Network derived from edges present in TCGA BRAF+ patients.

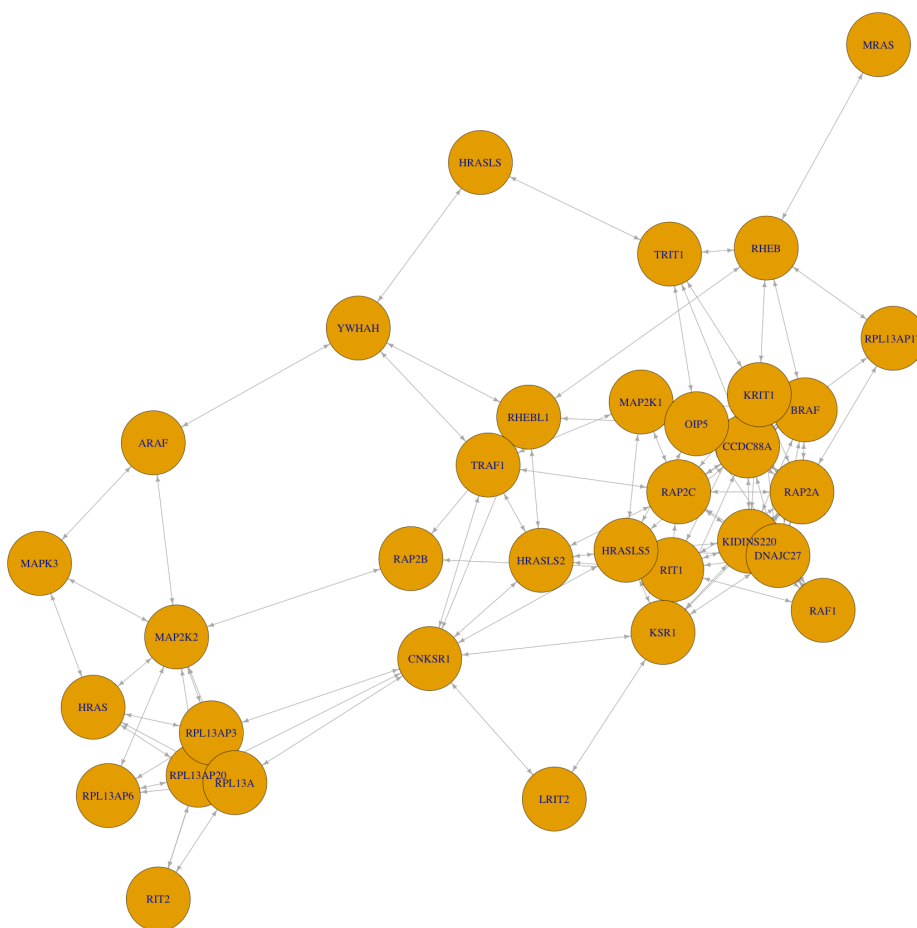


Figure 5.21: Network derived from edges present in TCGA BRAFv600e+ patients.

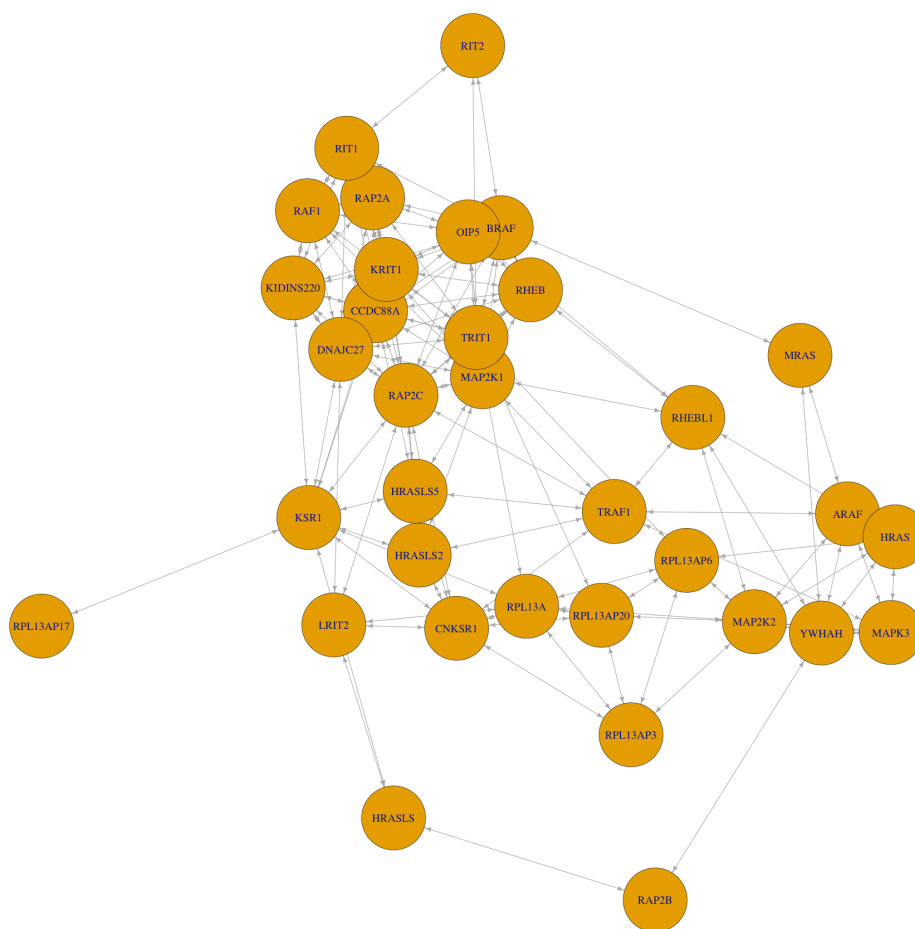


Figure 5.22: Network derived from edges present in TCGA BRAF- patients.

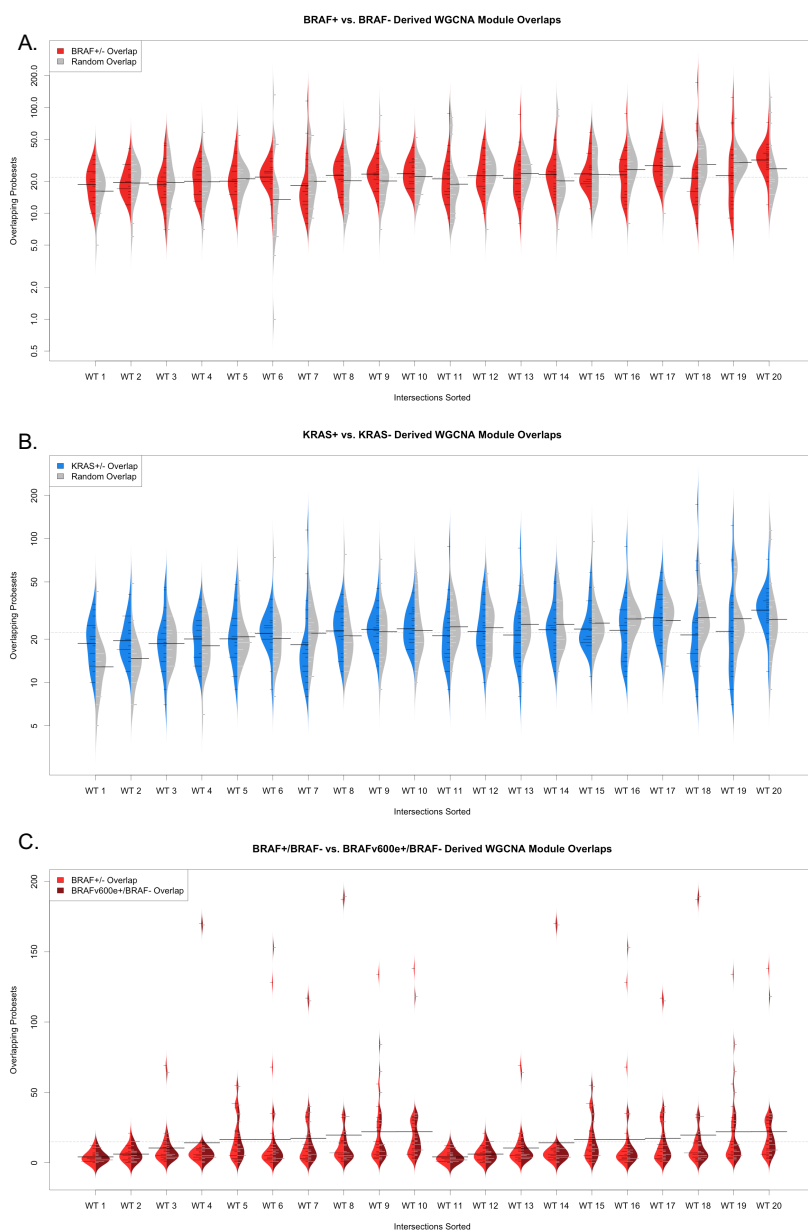


Figure 5.23: WGCNA derived modules (with $k = 20$) used for *a priori* assessment of network rewiring. (A) Beanplots of overlapping modules between melanoma cell line BRAF+/- (red) and random permutation-based (gray) overlap (B) Beanplots of overlapping modules between melanoma cell line KRAS+/- (blue) and random permutation-based (gray) and (C) Beanplots of overlapping modules between TCGA melanoma patient BRAF+/- (red) and BRAFv600e+/- (dark red) overlap

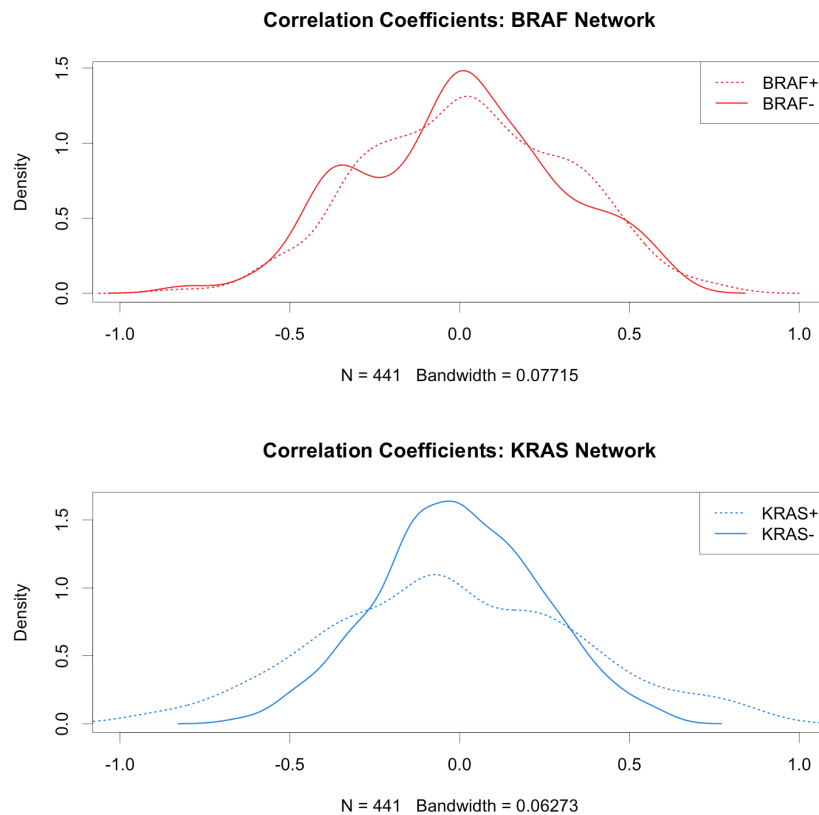


Figure 5.24: Density plots of the observed pairwise correlation coefficients when comparing the BRAF+/- subgroup using the BRAF-associated gene network and the KRAS+/- subgroup using the KRAS-associated gene network.

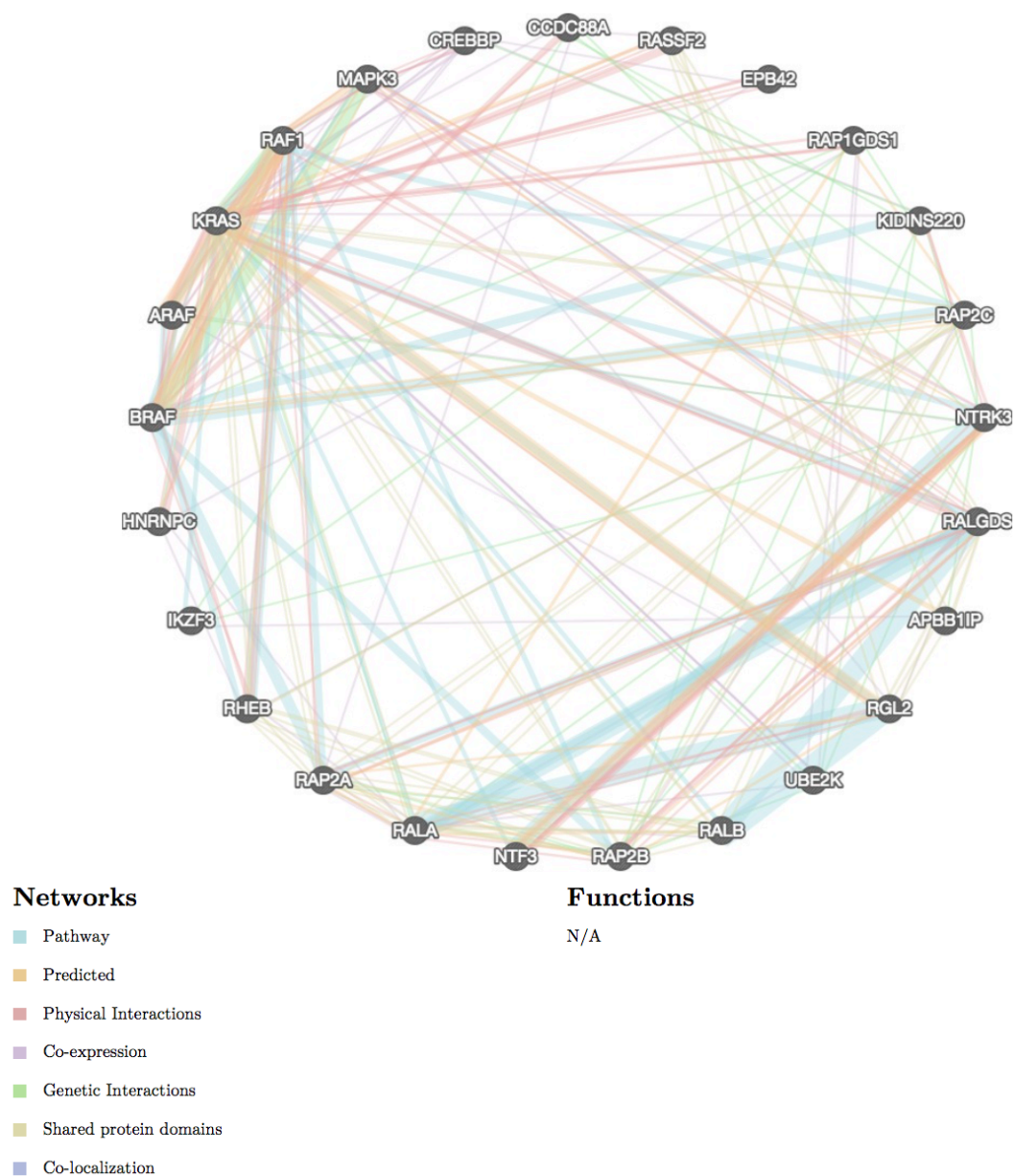


Figure 5.25: GeneMANIA circular view of all genes in Supplemental Table 2, i.e., genes with strong interactions identified via co-expression analysis.

Chapter 6

Conclusion, Limitations and Future Research

6.1 Synopsis

6.1.1 Reproducibility and Reliability of Drug Screen and Drug Target Data

There have been several attempts to assess the reliability and reproducibility of high throughput drug screen data (HTS). However, these attempts may not have accounted for important technical and biological factors (e.g., dose concentrations, efficacy, multiple cancer types, heterogeneity, etc.) or for certain experimental factors (e.g., plate effects, the laboratory in which assays were performed, etc.). I assessed the variability in HTS data attributable to both biological and experimental factors in order to ascertain the reliability of such studies. By accounting for variation attributable to, e.g., the plating scheme, the specific cell lines studied, the drugs used, and the concentration ranges, I was able to partition the sources of variation into coherent factors for more compelling inferences in those screens about the activity of the drugs, or potential targets for drugs.

6.1.2 Analysis of Dose Response Data

In HTS studies investigating chemotherapeutic agents, there are numerous analytical issues that need to be considered to reach robust and reliable results. Through simulation tests, I showed that traditional methods of assessing specific cell line response to drugs independently of other cell lines can lead to lower sensitivity to detect important factors mitigating drug responsiveness. Additionally, I found that fitting dose response curves and exploring those curves' relationships to other factors (such as mutational status or gene expression values) by leveraging all the data available in the drug screen produced greater power to detect associations. In particular, I demonstrated that using statistical analysis leveraging entire dose response curve profiles, rather than single point statistics such as the traditional IC_{50} value, was more powerful to detect associations between drug response and other factors. I also found that incorporating significantly associated covariates into an analysis directly, rather than in a traditional *ad hoc* manner, mitigated many problems with traditional analysis methods. Ultimately, the methods I described, implemented and evaluated provide more robust methods for assessing drug responses in HTS data because they directly consider different sources of variation and confounding effects.

6.1.3 Exploring Evidence for Re-Wiring in Cell-Lines Used in Drug Screens

In my analyses of melanoma cell lines and TCGA melanoma patient data, I observed significant heterogeneity that could plague traditional analyses. I found evidence of heterogeneity by using mixture models to search for clusters or groups of cell lines and tumors that appeared to exhibit evidence of different relationships among genes based on correlations in their expression values. These differences suggest that genetic networks may be “rewired” in subsets of cell lines. The identification of these network perturbations could impact our understanding of disease pathogenesis and discovery of drug targets. I did not find evidence for single alterations (e.g., mutation status) that could explain the clusters, which further

suggests that the factors mediating rewiring are complex. Subsequent analyses with different data sets validated the presence of the overall network differences and evidence of re-wiring. Based on the differences in genetic networks that I identified in subgroups of cell lines and tumor profiles I concluded that the genes acted as central network “hubs” and considered them candidate drug targets. These analyses suggest that different genes play a more pronounced role in the different genetic network architectures that the groups of cell lines exhibited and thus revealed different drug targets.

6.2 Extensions Beyond Current Experimental Designs and Methodologies

There are many issues that my research did not address that potentially limited my analyses and results. For example, one of the main challenges in cell line-based HTS and drug screen studies in general is that the dosing schemes used often do not result in noticeable effects on the cell lines for many drugs. This could be due to the drugs simply not having an effect or the dosages used not being appropriate to elicit an effect. If the higher doses considered show partial signs of an effect in cell lines, then one could expand the dose range, but this would come at a higher experimental cost. Partial dose response data causes major problems for the proposed NLME models assessing dose-response relationships. In most cases, drugs with a lack of response across all samples fail to produce reliable NLME model fits, as may be expected. Extrapolating the fitted dose response curves to dosages not considered in the actual screening could be considered, but is highly problematic because concentrations beyond the range considered in the experiment are arguably non-translatable to patient treatment due to the potential toxic effect larger doses would have.

Another important consideration in HTS and drug screening studies involves the use of drug combinations [Winter et al., 2012]. Because cancers often arise from perturbed networks that implicate many genes or even additional genetic networks, understanding combined drug effects could shed light on the utility

of “cocktail” chemotherapeutic agents provided to a patient. Applying computational approaches that consider drug synergy and antagonism could help expose why certain chemotherapies only work with other drugs. However, testing each and every pair of drugs or set of drugs that may exhibit an effect would be burdensome. Thus, computational approaches would need to guide experiments towards the most likely set of potentially synergistic and antagonistic drug combinations. Identifying multiple drug targets for drug combination studies could be facilitated through the proposed network analyses because multiple genes playing more central roles in those networks could be identified. Complementary to the drug efficacy experiments and methods, is the importance of uncovering any potential combined drug toxicity effects through experimental and computational approaches [Bulusu et al., 2016]. This could help identify potential drug combinations that would be too lethal to administer to patients.

Ideally, any identified central network “hubs” that emerge from a network analysis of the type I pursued would also be experimentally validated. Although there are many methods for pursuing this, two methods for validating the role of specific genes contributing to drug response based on re-wiring analysis I considered involve: (i) obtaining gene expression microarray and RPPA data on cell lines pre- and post-treatment of drug candidates, treating the cell lines with the candidate drugs, and then determining if changes in the networks thought to be rewired occur; and (ii) running clustered regularly interspaced short palindromic repeats (CRISPR) [Horvath and Barrangou, 2010] [Cong et al., 2013] experiments to remove or inactivate the central genes revealed from the gene centrality analyses of the identified rewired networks, in order to create an overall pathway of the type anticipated from the network centrality analyses.

6.3 Informing Clinical Trials and Patient Data Analyses

The real motivation for drug screens and the analytical methods I presented is to encourage the use of the insights obtained in actual clinical trials

and patient assessments. In this light, the melanoma cell line experiments leveraged material that was obtained in the context of the well-known but controversial Stand-Up-To-Cancer (SU2C) Melanoma Dream Team clinical trial (see Figure 5.1 for a schematic on how they were envisioned for use in the trial). In the clinical trial, patients were either randomized to a control arm (providing standard-of-care assignment of chemotherapy) or to an experimental treatment arm (providing therapies determined by an analysis of patient tumor profiles). The primary endpoint was tumor-free survival at eight weeks. The motivation for providing specific drugs to a patient, based on the analysis of that patient's tumor profiles, in the experimental therapy arm of the trial is to leverage available knowledge linking specific drugs to features in the patient's tumor. This knowledge could come from the results of other trials, knowledge of the basic mechanisms of action of a drug or drugs, physician intuitions, tumor board deliberations, or the results of empirical drug screens on the patient's tumor or a similar patient's tumor. I intended to incorporate results from my analyses on connections among drug responses, cell lines and patient tumors into prediction models that guide therapeutic choices. Toward this goal, I began developing models based on the cell line gene expression and RPPA data and applied cross-validation methods to demonstrate the feasibility of predicting drug responses, at least in a retrospective way given the drugs ultimately provided to the patients in the trial. However, due to a delay in patient accrual and subsequent data processing, I have not been able to validate these models on the data obtained from the outcomes of the patients. Nonetheless, I believe that our methods provide a robust framework for identifying genes and networks associated with drug response. These methods provide a path forward to understand the underlying mechanism in e.g., BRAF mutant and BRAF wild type patients. Furthermore, it is hoped that by matching patient tumor profiles to specific cell line profiles for which we have evaluated drug responses using the methods described herein, patients could be assigned a probability of responding to each drug based on e.g., receiver operating characteristics (ROC) curves [Hanley and McNeil, 1982], and this could provide a broad framework to measure the effectiveness of our models.

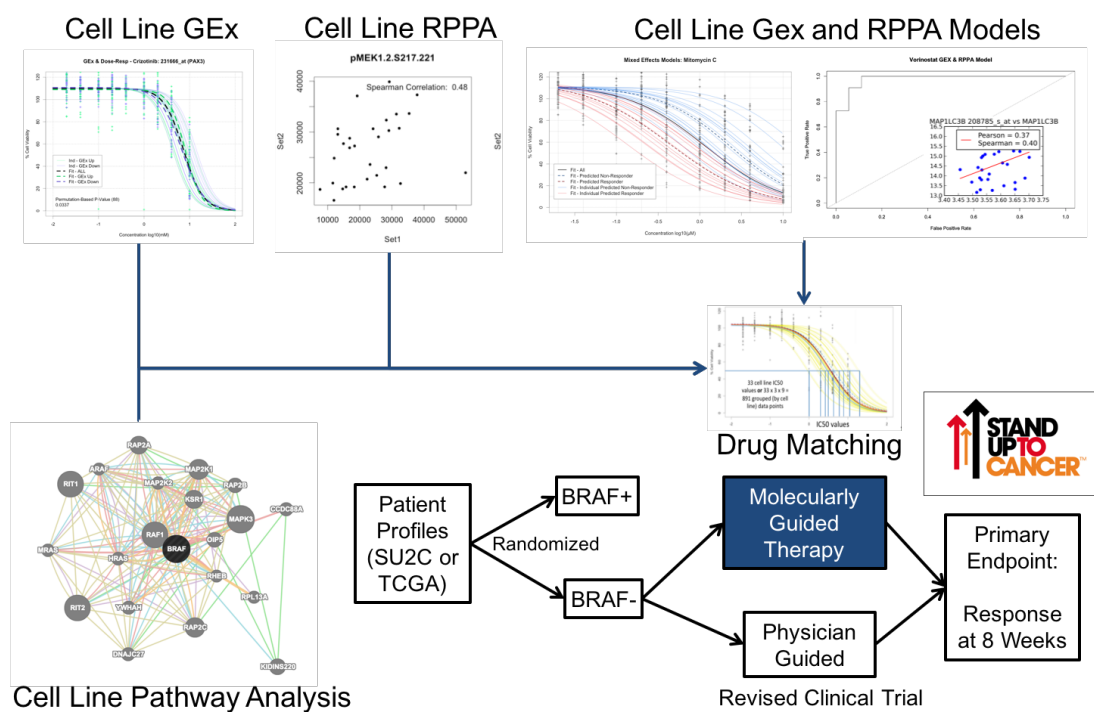


Figure 6.1: Flowchart of clinical trial study pursued through SU2C. The overarching goal is to incorporate algorithms developed through analysis of *in vitro* and other organisms.

6.4 Closing Remarks

The underlying motivation for the studies and analyses I pursued was to identify and provide solutions for problems that have received a great deal of attention in literature concerning the use of cell-line based drug targeting, HTS and general drug screening strategies. I have shown that there is a need for great sensitivity to factors likely to impact the reproducibility and reliability of drug screen and drug target data. I have also shown that existing ways of analyzing dose-response curves derived from drug screen data are flawed and could be complemented by more sophisticated analytical methods. Finally, I have shown that many cell lines and tumor profiles considered in drug target identification and screening analyses may exhibit differences in the genetic networks that could impact the identification of logical drug targets or inferences about molecular disease pathogenesis. In light of these results, I consider my efforts as opening up new avenues of research and hope to see the community improve upon them.

Bibliography

- [Aitman et al., 2011] Aitman, T. J., Boone, C., Churchill, G. A., Hengartner, M. O., Mackay, T. F. C., and Stemple, D. L. (2011). The future of model organisms in human disease research. *Nature Reviews Genetics*, 12(8):575–582.
- [Angrist and Pischke, 2009] Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton University Press, Princeton, 1 edition edition.
- [Bainbridge et al., 2011] Bainbridge, M. N., Wiszniewski, W., Murdock, D. R., Friedman, J., Gonzaga-Jauregui, C., Newsham, I., Reid, J. G., Fink, J. K., Morgan, M. B., Gingras, M.-C., Muzny, D. M., Hoang, L. D., Yousaf, S., Lupski, J. R., and Gibbs, R. A. (2011). Whole-Genome Sequencing for Optimized Patient Management. *Science Translational Medicine*, 3(87):87re3.
- [Bansal et al., 2014] Bansal, M., Yang, J., Karan, C., Menden, M. P., Costello, J. C., Tang, H., Xiao, G., Li, Y., Allen, J., Zhong, R., Chen, B., Kim, M., Wang, T., Heiser, L. M., Realubit, R., Mattioli, M., Alvarez, M. J., Shen, Y., NCI-DREAM Community, Gallahan, D., Singer, D., Saez-Rodriguez, J., Xie, Y., Stolovitzky, G., Califano, A., and NCI-DREAM Community (2014). A community computational challenge to predict the activity of pairs of compounds. *Nature Biotechnology*, 32(12):1213–1222.
- [Barretina et al., 2012] Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., Wilson, C. J., Lehr, J., Kryukov, G. V., Sonkin, D., Reddy, A., Liu, M., Murray, L., Berger, M. F., Monahan, J. E., Morais, P., Meltzer, J., Korejwa, A., Jan-Valbuena, J., Mapa, F. A., Thibault, J., Bric-Furlong, E., Raman, P., Shipway, A., Engels, I. H., Cheng, J., Yu, G. K., Yu, J., Aspesi, P., de Silva, M., Jagtap, K., Jones, M. D., Wang, L., Hatton, C., Palescandolo, E., Gupta, S., Mahan, S., Sougnez, C., Onofrio, R. C., Liefeld, T., MacConaill, L., Winckler, W., Reich, M., Li, N., Mesirov, J. P., Gabriel, S. B., Getz, G., Ardlie, K., Chan, V., Myer, V. E., Weber, B. L., Porter, J., Warmuth, M., Finan, P., Harris, J. L., Meyerson, M., Golub, T. R., Morrissey, M. P., Sellers, W. R., Schlegel, R., and Garraway, L. A. (2012). The Cancer Cell

- Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391):603–607.
- [Bates et al., 2015] Bates, D., Mchler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.
- [Benjamini and Hochberg, 1995] 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.
- [Blomen et al., 2015] Blomen, V. A., Mjek, P., Jae, L. T., Bigenzahn, J. W., Nieuwenhuis, J., Staring, J., Sacco, R., van Diemen, F. R., Olk, N., Stukalov, A., Marceau, C., Janssen, H., Carette, J. E., Bennett, K. L., Colinge, J., Superti-Furga, G., and Brummelkamp, T. R. (2015). Gene essentiality and synthetic lethality in haploid human cells. *Science (New York, N.Y.)*, 350(6264):1092–1096.
- [Buckley et al., 2016] Buckley, A. R., Standish, K. A., Bhutani, K., Ideker, T., Carter, H., Harismendy, O., and Schork, N. J. (2016). Pan-Cancer Analysis Reveals Technical Artifacts in The Cancer Genome Atlas (TCGA) Germline Variant Calls. *bioRxiv*, page 092163.
- [Bulusu et al., 2016] Bulusu, K. C., Guha, R., Mason, D. J., Lewis, R. P. I., Muratov, E., Kalantar Motamedi, Y., Cokol, M., and Bender, A. (2016). Modelling of compound combination effects and applications to efficacy and toxicity: state-of-the-art, challenges and perspectives. *Drug Discovery Today*, 21(2):225–238.
- [Califano, 2011] Califano, A. (2011). Rewiring makes the difference. *Molecular Systems Biology*, 7(1):463.
- [Cancer Genome Atlas Network, 2015] Cancer Genome Atlas Network (2015). Genomic Classification of Cutaneous Melanoma. *Cell*, 161(7):1681–1696.
- [Carretero et al., 2015] Carretero, M., Gomez-Amaro, R. L., and Petrascheck, M. (2015). Pharmacological classes that extend lifespan of *Caenorhabditis elegans*. *Frontiers in Genetics*, 6:77.
- [Che et al., 2014] Che, R., Jack, J. R., Motsinger-Reif, A. A., and Brown, C. C. (2014). An adaptive permutation approach for genome-wide association study: evaluation and recommendations for use. *BioData Mining*, 7:9.
- [Choi and Kendzierski, 2009] Choi, Y. and Kendzierski, C. (2009). Statistical methods for gene set co-expression analysis. *Bioinformatics*, 25(21):2780–2786.

- [Commo and Bot, 2015] Commo, F. and Bot, B. M. (2015). *nplr: N-Parameter Logistic Regression*. R package version 0.1-4.
- [Cong et al., 2013] Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., and Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science (New York, N.Y.)*, 339(6121):819–823.
- [Costanzo et al., 2010] Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., Sevier, C. S., Ding, H., Koh, J. L. Y., Toufighi, K., Mostafavi, S., Prinz, J., Onge, R. P. S., VanderSluis, B., Makhnevych, T., Vizeacoumar, F. J., Alizadeh, S., Bahr, S., Brost, R. L., Chen, Y., Cokol, M., Deshpande, R., Li, Z., Lin, Z.-Y., Liang, W., Marback, M., Paw, J., Luis, B.-J. S., Shuteriqi, E., Tong, A. H. Y., Dyk, N. v., Wallace, I. M., Whitney, J. A., Weirauch, M. T., Zhong, G., Zhu, H., Houry, W. A., Brudno, M., Ragibizadeh, S., Papp, B., Pl, C., Roth, F. P., Giaever, G., Nislow, C., Troyanskaya, O. G., Bussey, H., Bader, G. D., Gingras, A.-C., Morris, Q. D., Kim, P. M., Kaiser, C. A., Myers, C. L., Andrews, B. J., and Boone, C. (2010). The Genetic Landscape of a Cell. *Science*, 327(5964):425–431.
- [Costello et al., 2014] Costello, J. C., Heiser, L. M., Georgii, E., Gnen, M., Menden, M. P., Wang, N. J., Bansal, M., Ammad-ud din, M., Hintsanen, P., Khan, S. A., Mpindi, J.-P., Kallioniemi, O., Honkela, A., Aittokallio, T., Wennerberg, K., NCI DREAM Community, Collins, J. J., Gallahan, D., Singer, D., Saez-Rodriguez, J., Kaski, S., Gray, J. W., and Stolovitzky, G. (2014). A community effort to assess and improve drug sensitivity prediction algorithms. *Nature Biotechnology*, 32(12):1202–1212.
- [Crystal et al., 2014] Crystal, A. S., Shaw, A. T., Sequist, L. V., Friboulet, L., Niederst, M. J., Lockerman, E. L., Frias, R. L., Gainor, J. F., Amzallag, A., Greninger, P., Lee, D., Kalsy, A., Gomez-Caraballo, M., Elamine, L., Howe, E., Hur, W., Lifshits, E., Robinson, H. E., Katayama, R., Faber, A. C., Awad, M. M., Ramaswamy, S., Mino-Kenudson, M., Iafrate, A. J., Benes, C. H., and Engelman, J. A. (2014). Patient-derived models of acquired resistance can identify effective drug combinations for cancer. *Science*, 346(6216):1480–1486.
- [Ding et al., 2017] Ding, K.-F., Finlay, D., Yin, H., Hendricks, W. P. D., Sereduk, C., Kiefer, J., Sekulic, A., LoRusso, P. M., Vuori, K., Trent, J. M., Schork, N. J., Ding, K.-F., Finlay, D., Yin, H., Hendricks, W. P. D., Sereduk, C., Kiefer, J., Sekulic, A., LoRusso, P. M., Vuori, K., Trent, J. M., and Schork, N. J. (2017). Analysis of variability in high throughput screening data: applications to melanoma cell lines and drug responses. *Oncotarget*, 5(0).
- [Eduati et al., 2015] Eduati, F., Mangravite, L. M., Wang, T., Tang, H., Bare, J. C., Huang, R., Norman, T., Kellen, M., Menden, M. P., Yang, J., Zhan,

- X., Zhong, R., Xiao, G., Xia, M., Abdo, N., Kosyk, O., The NIEHS-NCATS-UNC DREAM Toxicogenetics Collaboration, Friend, S., Dearry, A., Simeonov, A., Tice, R. R., Rusyn, I., Wright, F. A., Stolovitzky, G., Xie, Y., and Saez-Rodriguez, J. (2015). Prediction of human population responses to toxic compounds by a collaborative competition. *Nature Biotechnology*, 33(9):933–940.
- [Fritz and Radziwill, 2010] Fritz, R. D. and Radziwill, G. (2010). CNK1 promotes invasion of cancer cells through NF-kappaB-dependent signaling. *Molecular cancer research: MCR*, 8(3):395–406.
- [Garraway et al., 2013] Garraway, L. A., Verweij, J., and Ballman, K. V. (2013). Precision oncology: an overview. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 31(15):1803–1805.
- [Gautier et al., 2004] Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A. (2004). affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics (Oxford, England)*, 20(3):307–315.
- [Gillet et al., 2013] Gillet, J.-P., Varma, S., and Gottesman, M. M. (2013). The Clinical Relevance of Cancer Cell Lines. *JNCI Journal of the National Cancer Institute*, 105(7):452–458.
- [Grn and Leisch, 2007] Grn, B. and Leisch, F. (2007). Fitting finite mixtures of generalized linear regressions in R. *Computational Statistics & Data Analysis*, 51(11):5247–5252.
- [Grn and Leisch, 2008] Grn, B. and Leisch, F. (2008). Flexmix version 2: Finite mixtures with concomitant variables and varying and constant parameters. *Journal of Statistical Software*, 28(1):1–35.
- [Haibe-Kains et al., 2013] Haibe-Kains, B., El-Hachem, N., Birkbak, N. J., Jin, A. C., Beck, A. H., Aerts, H. J. W. L., and Quackenbush, J. (2013). Inconsistency in large pharmacogenomic studies. *Nature*, 504(7480):389–393.
- [Hanley and McNeil, 1982] Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1):29–36.
- [Harrison et al., 2009] Harrison, D. E., Strong, R., Sharp, Z. D., Nelson, J. F., Astle, C. M., Flurkey, K., Nadon, N. L., Wilkinson, J. E., Frenkel, K., Carter, C. S., Pahor, M., Javors, M. A., Fernandez, E., and Miller, R. A. (2009). Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. *Nature*, 460(7253):392–395.
- [Haverty et al., 2016] Haverty, P. M., Lin, E., Tan, J., Yu, Y., Lam, B., Lianoglou, S., Neve, R. M., Martin, S., Settleman, J., Yauch, R. L., and Bourgon, R.

- (2016). Reproducible pharmacogenomic profiling of cancer cell line panels. *Nature*, 533(7603):333–337.
- [Hofree et al., 2013] Hofree, M., Shen, J. P., Carter, H., Gross, A., and Ideker, T. (2013). Network-based stratification of tumor mutations. *Nature Methods*, 10(11):1108–1115.
- [Horvath and Barrangou, 2010] Horvath, P. and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science (New York, N.Y.)*, 327(5962):167–170.
- [Ideker and Krogan, 2012] Ideker, T. and Krogan, N. J. (2012). Differential network biology. *Molecular Systems Biology*, 8:565.
- [Irizarry et al., 2003] Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M., Hobbs, B., and Speed, T. P. (2003). Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Research*, 31(4):e15.
- [Johnson and Gallagher, 2010] Johnson, M. and Gallagher, K. (2010). One in a Billion: A boy’s life, a medical mystery - A baffling illness.
- [Jubb et al., 2016] Jubb, A. W., Young, R. S., Hume, D. A., and Bickmore, W. A. (2016). Enhancer turnover is associated with a divergent transcriptional response to glucocorticoid in mouse and human macrophages. *Journal of immunology (Baltimore, Md. : 1950)*, 196(2):813–822.
- [Jung and Kim, 2014] Jung, S. and Kim, S. (2014). EDDY: a novel statistical gene set test method to detect differential genetic dependencies. *Nucleic Acids Research*, page gku099.
- [Kanehisa and Goto, 2000] Kanehisa, M. and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1):27–30.
- [Kaushik et al., 2014] Kaushik, P., Molinelli, E. J., Miller, M. L., Wang, W., Korkut, A., Liu, W., Ju, Z., Lu, Y., Mills, G., and Sander, C. (2014). Spatial Normalization of Reverse Phase Protein Array Data. *PLOS ONE*, 9(12):e97213.
- [Kelder et al., 2012] Kelder, T., van Iersel, M. P., Hanspers, K., Kutmon, M., Conklin, B. R., Evelo, C. T., and Pico, A. R. (2012). WikiPathways: building research communities on biological pathways. *Nucleic Acids Research*, 40(Database issue):D1301–D1307.
- [Krogan et al., 2015] Krogan, N. J., Lippman, S., Agard, D. A., Ashworth, A., and Ideker, T. (2015). The Cancer Cell Map Initiative: Defining the Hallmark Networks of Cancer. *Molecular Cell*, 58(4):690–698.

- [Kutmon et al., 2016] Kutmon, M., Riutta, A., Nunes, N., Hanspers, K., Wil-
lighagen, E. L., Bohler, A., Mlius, J., Waagmeester, A., Sinha, S. R., Miller,
R., Coort, S. L., Cirillo, E., Smeets, B., Evelo, C. T., and Pico, A. R. (2016).
WikiPathways: capturing the full diversity of pathway knowledge. *Nucleic Acids
Research*, 44(D1):D488–494.
- [Lamb et al., 2006] Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C.,
Wrobel, M. J., Lerner, J., Brunet, J.-P., Subramanian, A., Ross, K. N., Reich,
M., Hieronymus, H., Wei, G., Armstrong, S. A., Haggarty, S. J., Clemons, P. A.,
Wei, R., Carr, S. A., Lander, E. S., and Golub, T. R. (2006). The Connectivity
Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes,
and Disease. *Science*, 313(5795):1929–1935.
- [Langfelder and Horvath, 2008] Langfelder, P. and Horvath, S. (2008). WGCNA:
an R package for weighted correlation network analysis. *BMC Bioinformatics*,
9:559.
- [Langfelder and Horvath, 2012] Langfelder, P. and Horvath, S. (2012). Fast R
Functions for Robust Correlations and Hierarchical Clustering. *Journal of Sta-
tistical Software*, 46(11).
- [Leisch, 2004] Leisch, F. (2004). Flexmix: A general framework for finite mixture
models and latent class regression in r. *Journal of Statistical Software*, 11(1):1–
18.
- [LoRusso et al., 2015] LoRusso, P. M., Boerner, S. A., Pilat, M. J., Forman, K. M.,
Zuccaro, C. Y., Kiefer, J. A., Liang, W. S., Hunsberger, S., Redman, B. G.,
Markovic, S. N., Sekulic, A., Bryce, A. H., Joseph, R. W., Cowey, C. L., Fecher,
L. A., Sosman, J. A., Chapman, P. B., Schwartz, G. K., Craig, D. W., Carpten,
J. D., and Trent, J. M. (2015). Pilot Trial of Selecting Molecularly Guided
Therapy for Patients with NonV600 BRAF-Mutant Metastatic Melanoma: Ex-
perience of the SU2c/MRA Melanoma Dream Team. *Molecular Cancer Thera-
peutics*, 14(8):1962–1971.
- [Lum et al., 2012] Lum, D. H., Matsen, C., Welm, A. L., and Welm, B. E. (2012).
Human Primary Tumorgraft Models: Comparisons with Traditional Oncology
Pre-Clinical Models and The Clinical Relevance and Utility of Primary Tu-
morgrafts in Basic and Translational Oncology Research. *Current protocols in
pharmacology / editorial board, S.J. Enna (editor-in-chief) ... [et al.]*, CHAP-
TER:Unit14.22.
- [Lupski et al., 2010] Lupski, J. R., Reid, J. G., Gonzaga-Jauregui, C., Rio Deiros,
D., Chen, D. C., Nazareth, L., Bainbridge, M., Dinh, H., Jing, C., Wheeler,
D. A., McGuire, A. L., Zhang, F., Stankiewicz, P., Halperin, J. J., Yang, C.,
Gehman, C., Guo, D., Irikat, R. K., Tom, W., Fantin, N. J., Muzny, D. M., and

- Gibbs, R. A. (2010). Whole-Genome Sequencing in a Patient with Charcot-Marie-Tooth Neuropathy. *New England Journal of Medicine*, 362(13):1181–1191.
- [Mercola et al., 2013] Mercola, M., Colas, A., and Willems, E. (2013). Induced pluripotent stem cells in cardiovascular drug discovery. *Circulation Research*, 112(3):534–548.
- [Miller et al., 2011] Miller, R. A., Harrison, D. E., Astle, C. M., Baur, J. A., Boyd, A. R., de Cabo, R., Fernandez, E., Flurkey, K., Javors, M. A., Nelson, J. F., Orihuela, C. J., Pletcher, S., Sharp, Z. D., Sinclair, D., Starnes, J. W., Wilkinson, J. E., Nadon, N. L., and Strong, R. (2011). Rapamycin, but not resveratrol or simvastatin, extends life span of genetically heterogeneous mice. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, 66(2):191–201.
- [Miller et al., 2007] Miller, R. A., Harrison, D. E., Astle, C. M., Floyd, R. A., Flurkey, K., Hensley, K. L., Javors, M. A., Leeuwenburgh, C., Nelson, J. F., Ongini, E., Nadon, N. L., Warner, H. R., and Strong, R. (2007). An Aging Interventions Testing Program: study design and interim report. *Aging Cell*, 6(4):565–575.
- [Monsma et al., 2012] Monsma, D. J., Monks, N. R., Cherba, D. M., Dylewski, D., Eugster, E., Jahn, H., Srikanth, S., Scott, S. B., Richardson, P. J., Everts, R. E., Ishkin, A., Nikolsky, Y., Resau, J. H., Sigler, R., Nickoloff, B. J., and Webb, C. P. (2012). Genomic characterization of explant tumorgraft models derived from fresh patient tumor tissue. *Journal of Translational Medicine*, 10:125.
- [Mootha et al., 2003] Mootha, V. K., Lindgren, C. M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstr le, M., Laurila, E., Houstis, N., Daly, M. J., Patterson, N., Mesirov, J. P., Golub, T. R., Tamayo, P., Spiegelman, B., Lander, E. S., Hirschhorn, J. N., Altshuler, D., and Groop, L. C. (2003). PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics*, 34(3):267–273.
- [Oros Klein et al., 2016] Oros Klein, K., Oualkacha, K., Lafond, M.-H., Bhatnagar, S., Tonin, P. N., and Greenwood, C. M. T. (2016). Gene Coexpression Analyses Differentiate Networks Associated with Diverse Cancers Harboring TP53 Missense or Null Mutations. *Frontiers in Genetics*, 7:137.
- [Park et al., 2011] Park, S.-K., Pegan, S. D., Mesecar, A. D., Jungbauer, L. M., LaDu, M. J., and Liebman, S. W. (2011). Development and validation of a yeast high-throughput screen for inhibitors of A β 21 oligomerization. *Disease Models & Mechanisms*, 4(6):822–831.

- [Pico et al., 2008] Pico, A. R., Kelder, T., Iersel, M. P. v., Hanspers, K., Conklin, B. R., and Evelo, C. (2008). WikiPathways: Pathway Editing for the People. *PLOS Biology*, 6(7):e184.
- [Powers et al., 2006] Powers, R. W., Kaerberlein, M., Caldwell, S. D., Kennedy, B. K., and Fields, S. (2006). Extension of chronological life span in yeast by decreased TOR pathway signaling. *Genes & Development*, 20(2):174–184.
- [Rangaraju et al., 2015a] Rangaraju, S., Solis, G. M., Andersson, S. I., Gomez-Amaro, R. L., Kardakaris, R., Broaddus, C. D., Niculescu, A. B., and Petrascheck, M. (2015a). Atypical antidepressants extend lifespan of *Caenorhabditis elegans* by activation of a non-cell-autonomous stress response. *Aging Cell*, 14(6):971–981.
- [Rangaraju et al., 2015b] Rangaraju, S., Solis, G. M., and Petrascheck, M. (2015b). High-throughput small-molecule screening in *Caenorhabditis elegans*. *Methods in Molecular Biology (Clifton, N.J.)*, 1263:139–155.
- [Revelle, 2015] Revelle, W. (2015). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 1.5.8.
- [Safikhani et al., 2016a] Safikhani, Z., El-Hachem, N., Quevedo, R., Smirnov, P., Goldenberg, A., Juul Birkbak, N., Mason, C., Hatzis, C., Shi, L., Aerts, H. J., Quackenbush, J., and Haibe-Kains, B. (2016a). Assessment of pharmacogenomic agreement. *F1000Research*, 5:825.
- [Safikhani et al., 2016b] Safikhani, Z., Smirnov, P., Freeman, M., El-Hachem, N., She, A., Rene, Q., Goldenberg, A., Birkbak, N. J., Hatzis, C., Shi, L., Beck, A. H., Aerts, H. J., Quackenbush, J., and Haibe-Kains, B. (2016b). Revisiting inconsistency in large pharmacogenomic studies. *F1000Research*, 5:2333.
- [Schork, 2015] Schork, N. J. (2015). Personalized medicine: Time for one-person trials. *Nature*, 520(7549):609–611.
- [Shoemaker, 2006] Shoemaker, R. H. (2006). The NCI60 human tumour cell line anticancer drug screen. *Nature Reviews Cancer*, 6(10):813–823.
- [Shrager and Tenenbaum, 2014] Shrager, J. and Tenenbaum, J. M. (2014). Rapid learning for precision oncology. *Nature Reviews Clinical Oncology*, 11(2):109–118.
- [Snider et al., 2015] Snider, J., Kotlyar, M., Saraon, P., Yao, Z., Jurisica, I., and Stagljar, I. (2015). Fundamentals of protein interaction network mapping. *Molecular Systems Biology*, 11(12).

- [St Onge et al., 2007] St Onge, R. P., Mani, R., Oh, J., Proctor, M., Fung, E., Davis, R. W., Nislow, C., Roth, F. P., and Giaever, G. (2007). Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nature Genetics*, 39(2):199–206.
- [Strong et al., 2008] Strong, R., Miller, R. A., Astle, C. M., Floyd, R. A., Flurkey, K., Hensley, K. L., Javors, M. A., Leeuwenburgh, C., Nelson, J. F., Ongini, E., Nadon, N. L., Warner, H. R., and Harrison, D. E. (2008). Nordihydroguaiaretic acid and aspirin increase lifespan of genetically heterogeneous male mice. *Aging cell*, 7(5):641–650.
- [Subramanian et al., 2005] Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, 102(43):15545–15550.
- [Taylor, 2010] Taylor, D. L. (2010). A Personal Perspective on High-Content Screening (HCS) From the Beginning. *Journal of Biomolecular Screening*, 15(7):720–725.
- [The Cancer Genome Atlas Research Network et al., 2013] The Cancer Genome Atlas Research Network, Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J. M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nature Genetics*, 45(10):1113–1120.
- [The CCLE and The GDSC Consortia, 2015] The CCLE and The GDSC Consortia (2015). Pharmacogenomic agreement between two cancer cell line data sets. *Nature*, 528(7580):84–87.
- [Torres et al., 2013] Torres, N. P., Lee, A. Y., Giaever, G., Nislow, C., and Brown, G. W. (2013). A high-throughput yeast assay identifies synergistic drug combinations. *Assay and Drug Development Technologies*, 11(5):299–307.
- [UniProt Consortium, 2015] UniProt Consortium (2015). UniProt: a hub for protein information. *Nucleic Acids Research*, 43(Database issue):D204–212.
- [Wagner et al., 2016] Wagner, A. H., Coffman, A. C., Ainscough, B. J., Spies, N. C., Skidmore, Z. L., Campbell, K. M., Krysiak, K., Pan, D., McMichael, J. F., Eldred, J. M., Walker, J. R., Wilson, R. K., Mardis, E. R., Griffith, M., and Griffith, O. L. (2016). DGIdb 2.0: mining clinically relevant drug-gene interactions. *Nucleic Acids Research*, 44(D1):D1036–1044.

- [Wang et al., 2016] Wang, Y., Fang, J., and Chen, S. (2016). Inferences of drug responses in cancer cells from cancer genomic features and compound chemical and therapeutic properties. *Scientific Reports*, 6:32679.
- [Warner, 2006] Warner, H. R. (2006). Longevity Regulation and Aging in Animal Models. *The Gerontologist*, 46(6):844–847.
- [Warren et al., 2016] Warren, W. C., Hillier, L. W., Tomlinson, C., Minx, P., Kremitzki, M., Graves, T., Markovic, C., Bouk, N., Pruitt, K. D., Thibaud-Nissen, F., Schneider, V., Mansour, T. A., Brown, C. T., Zimin, A., Hawken, R., Abrahamsen, M., Pyrkosz, A. B., Morisson, M., Fillon, V., Vignal, A., Chow, W., Howe, K., Fulton, J. E., Miller, M. M., Lovell, P., Mello, C. V., Wirthlin, M., Mason, A. S., Kuo, R., Burt, D. W., Dodgson, J. B., and Cheng, H. H. (2016). A New Chicken Genome Assembly Provides Insight into Avian Genome Structure. *G3 (Bethesda, Md.)*.
- [Winter et al., 2012] Winter, G. E., Rix, U., Carlson, S. M., Gleixner, K. V., Grebien, F., Gridling, M., Mller, A. C., Breitwieser, F. P., Bilban, M., Colinge, J., Valent, P., Bennett, K. L., White, F. M., and Superti-Furga, G. (2012). Systems-pharmacology dissection of a drug synergy in imatinib-resistant CML. *Nature Chemical Biology*, 8(11):905–912.
- [Yang et al., 2013] Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J. A., Thompson, I. R., Ramaswamy, S., Futreal, P. A., Haber, D. A., Stratton, M. R., Benes, C., McDermott, U., and Garnett, M. J. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Research*, 41(D1):D955–D961.
- [Yang et al., 2014] Yang, Y., Han, L., Yuan, Y., Li, J., Hei, N., and Liang, H. (2014). Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. *Nature Communications*, 5:3231.
- [Ye et al., 2014] Ye, X., Linton, J. M., Schork, N. J., Buck, L. B., and Petrascheck, M. (2014). A pharmacological network for lifespan extension in *Caenorhabditis elegans*. *Aging Cell*, 13(2):206–215.