

**AN EX VIVO FAMILIAL GENETICS STRATEGY  
FOR DETERMINING MECHANISM OF ACTION**

Venita Gresham Watson

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill  
in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the  
School of Pharmacy.

Chapel Hill  
2010

Approved by:

Howard McLeod, Pharm.D.

Stephen Frye, Ph.D.

Alison Motsinger-Reif, Ph.D.

Tim Wiltshire, Ph.D.

William Zamboni, Pharm.D., Ph.D.

© 2010  
Venita Gresham Watson  
ALL RIGHTS RESERVED

## ABSTRACT

VENITA GRESHAM WATSON:  
A Familial Genetic Strategy for Determining Mechanism of Action  
(Under the direction of Howard McLeod, Pharm.D.)

One of the greatest challenges in anticancer drug development is the discovery of molecular targets and biochemical interactions required for drug action. Lapses in drug efficacy and unanticipated toxicity, the two biggest causes of drug failure in clinical trials, are often attributed to our limited understanding of drug mechanism and cost the pharmaceutical industry millions. Genomics is rapidly emerging as tool for mechanism elucidation.

Our approach is one of the latest to link drugs to the genes which influence their activity. This *ex vivo* familial genetics strategy uses a collection of extensively genotyped, normal, healthy, human cell lines from multigenerational families. Cell lines are phenotyped for cytotoxic response to anticancer agents, heritability analysis gives a measure of the degree to which genetics influence response, and linkage analysis suggests regions of the genome which are associated with the observed variation in response.

To evaluate this strategy as method for mechanism elucidation, we first asked whether the system could produce pharmacological and genomic profiles related to a shared mechanism for a class of structurally related compounds. The *in vitro* sensitivity of CEPH cell lines to the camptothecins, a class of Topoisomerase 1 inhibitors (Top1),

was studied. Heritability analysis estimates that genetics accounts for as much 20% of the observed variation in cytotoxic response to these drugs. Linkage analysis revealed a pattern of seven quantitative trait loci (QTLs) that were shared by all of the camptothecins and independently replicated with a second set of camptothecin analogues. The pattern of QTLs observed with the camptothecins was compared to those of the indenisoquinolines, a structurally distinct class of Top1 inhibitors. The objective was to identify which if any QTLs are related to the general mechanism of Top1 inhibition or should be considered class-specific. Finally, the model was assessed for its ability to stratify compounds by mechanism based on their biological and genomic profiles. Cell lines were phenotyped for response to approximately 30 drugs belonging to 8 mechanistic classes. Intra-class biological and genomic profiles were more similar to each other than to compounds belonging to distinct mechanistic classes.

This work could have a significant impact on drug discovery and development as it provides a strategy for not only making predictions about mechanism of action for novel therapies, but for identifying genes involved in variable response to chemotherapeutic agents as well.

*To my husband, my family, and friends  
who, for years,  
have awaited the completion of this part of my journey,  
we made it!*

## ACKNOWLEDGEMENTS

It is an absolute pleasure to thank the many people who made this dissertation possible.

Foremost, I would like to thank my advisor, Dr. Howard McLeod, who shared with me a lot of his expertise and research insight. He quickly became for me the role model of a successful researcher and leader. I have valued his sound advice in both personal and professional matters and am deeply grateful for my time under his guidance.

I would like to thank my committee members: Drs. William Zamboni, Stephen Frye, and Tim Wiltshire for their support, advice, and insight. I cannot overstate my gratitude to Dr. Alison Motsinger-Reif, without whose efforts this project could not have been completed. (A heart felt thanks to her students, Nick Hardison and Zeke Harris, as well!)

I must also acknowledge the guidance and support from the past and present faculty and staff and the entire Eshelman School of Pharmacy. I would like to particularly thank Amber Allen, Dr. Kathryn Fiscelli, Angela Lyght, Kathy Maboll, Joann Davis, and Sherrie Settle for their help with administrative concerns throughout the years.

I have received an exceptional amount of guidance and support from past and present members of the McLeod lab. In particular, I am thankful to Eric Peters, Tammy Havener, Lorraine Everitt, Filipe Muhale, Michael Wagner, and Todd Auman. Thanks to

Bob He, Anne Dvorak, Leith Olson, and Pei-shi Ong for animated discussions about science and life (mostly about life).

I would like to thank the many people with whom I've shared meaningful life experiences who set me on this path. From Dr. Joanne Peart, who guided my first research efforts in high school and interested me in the study of Pharmaceutical Sciences, to those whose gift of companionship made my days at UNC more enjoyable. Roland Cheung, John An, Jin Lee, Enzo Palma, Leandra Miko, Diedre Washington, and Amica Simmons Yon deserve thanks for so much needed laughter and friendship.

I cannot finish without saying how truly thankful I am to my family. Most importantly, I wish to thank my parents, Larry Frazier and Veda Gresham Frazier. They have always encouraged me to do my very best in all areas of my life. Their belief in me kept me going even when self-doubt would have caused me to falter. To my husband, Roshawn, I could not have been blessed with a better partner to weather the highs and lows of pursuing a doctoral degree. For all of the time and hard work we have put in here, this thesis is as much yours as it is mine. I love you truly, madly, deeply.

Finally, thank you God, for this moment and this wonderful life.

## TABLE OF CONTENTS

LIST OF FIGURES.....	ix
LIST OF TABLES.....	xi
LIST OF ABBREVIATIONS AND SYMBOLS.....	xii
CHAPTER	
I. INTRODUCTION: APPLICATIONS IN MECHANISM ELUCIDATION.....	1
II. GENOME-WIDE APPROACH FOR DIFFERENTIATION OF THE CAMPTOTHECINS.....	29
III. BIOLOGICAL AND GENOMIC PROFILING IN CEPH DISTINGUISHES BETWEEN STRUCTURALLY DISTINCT CLASSES OF TOP1 INHIBITORS ....	62
IV. BIOLOGICAL AND GENOMIC PROFILING OF A MECHANISTIC SET IN CEPH CELL LINES .....	96
V. CONCLUSIONS AND FUTURE STUDIES.....	122
APPENDICES.....	135
APPENDIX 1: ASSAY OPTIMIZATION FOR CEPH PHENOTYPING .....	135
APPENDIX 2: SUPPORTING DATA FOR CAMPTOTHECIN ANALOGUES.....	152
APPENDIX 3: SUPPORTING DATA FOR ADDITIONAL TOP1 INHIBITORS .....	163
APPENDIX 4: SUPPORTING DATA FOR MECHANISTIC SET .....	167



## LIST OF FIGURES

Figure 1-1	Affinity chromatography based methods for target identification.....	19
Figure 1-2	Schematic overview of mechanism prediction model using NCI60 human cancer cell lines .....	20
Figure 1-3	Murine haplotype computational pharmacogenetic analysis.....	21
Figure 1-4	Discovery of genetic loci involved in the cytotoxic effect of chemotherapeutic agents.....	22
Figure 2-1	Structures of camptothecin analogues .....	48
Figure 2-2	Dose response curves and heritability estimates for the camptothecins .....	49
Figure 2-3	Hierarchal clustering of CEPH cell lines by camptothecin IC50s.....	50
Figure 2-4	Genome-wide pattern of QTLs for camptothecin analogues.....	51
Figure 2-5	QTL shared across all camptothecins on chromosome 20.....	52
Figure 2-6	Hierarchical clustering of CEPH cell lines for Top1 and Top2 drugs.....	53
Figure 2-7	Molecular pathways involved in cytotoxic response to camptothecins .....	54
Figure 3-1	Chemical structures of camptothecin and the indenoisoquinolines.....	79
Figure 3-2	Genomic profiling of Top1 inhibitors.....	80
Figure 3-3	Dose response curve for indenoisoquinolines.....	81
Figure 3-4	Boxplot of mean viability for CEPH cell lines at $\overline{GI50}$ .....	82
Figure 3-5	Sensitivity patterns for camptothecins and indenoisoquinolines in select CEPH pedigrees.....	83
Figure 3-6	Genome-wide pattern of QTLs for camptothecins and indenoisoquinolines ..	84
Figure 4-1	Distribution of mean viabilities for CEPH cell lines at $\overline{GI50}$ .....	109
Figure 4-2	Hierarchical clustering of 22 anticancer drugs based on cell viability at $\overline{GI50}$ .....	110

Figure 4-3	Results of a comparison analysis of biological activity profiles of CEPH cell lines at for 8 mechanistic classes .....	111
Figure 4-4	Genome wide pattern of QTLs for mechanistic set .....	112
Figure 4-5	Genome wide QTL map for oxaliplatin and carboplatin .....	113
Figure 4-6	Shared QTL on chromosome 3 for oxaliplatin and carboplatin .....	114
Figure A1-1	Optimizing CEPH phenotyping assay .....	144
Figure A1-2	Plate uniformity and signal variability.....	145
Figure A1-3	Identification of duration of drug exposure .....	146
Figure A1-4	Between run and intra- and interday variability .....	147
Figure A1-5	Effect of camptothecin stock solution pH on cell viability and assay stability .....	148
Figure A1-6	Effect of camptothecin stock solution pH on interday variability .....	149
Figure A1-7	Effect of camptothecin stock solution pH on week-to-week variability.....	150
Figure A1-8	Effect of camptothecin stock solution pH on month-to-month variability ....	151
Figure A3-1	Genomewide map of QTLs for camptothecins and indenoisoquinolines.....	164
Figure A4-1	Boxplots for mechanistic set.....	177

## LIST OF TABLES

Table 1-1	Comparison of mechanism elucidation strategies.....	23
Table 2-1	QTLs shared by camptothecins.....	55
Table 2-2	Similarity matrix for overall QTL patterns for each camptothecin .....	56
Table 2-3	Top overrepresented GO terms for QTL on chromosome 20.....	57
Table 3-1	Correlation analysis for mean viability at $\overline{GI50}$ for Top1 inhibitors.....	85
Table 3-2	Heritability estimates for indenoisoquinolines .....	86
Table 3-3	QTLs found in camptothecins and indenoisoquinolines.....	87
Table 3-4	QTLs shared by all indenoisoquinolines.....	88
Table 3-5	Genes under QTLs shared by camptothecins and indenoisoquinolines.....	89
Table 4-1	Chemotherapeutic agents used in this study .....	115
Table 4-2	Range of heritability estimates for sensitivity to anticancer agents.....	116
Table 4-3	Variation in peak location for significant QTLs on chromosome 20 Shared by all camptothecin analogues.....	117
Table A2-1	List of significant QTLs for camptothecin analogues.....	154
Table A2-2	Genes of interest and associated GO terms under chromosome 20.....	158
Table A2-3	List of cell lines sensitive or resistant to the camptothecins.....	159
Table A3-1	List of significant QTLs for indenoisoquinolines.....	165
Table A4-1	Drug concentrations and $\overline{GI50}$ for mechanistic set .....	168
Table A4-2	Cell lines sensitive and resistant to the mechanistic set, by family .....	169
Table A4-3	Heritability estimates for all drug-dose phenotypes of mechanistic set .....	172
Table A4-4	List of significant QTLs for mechanistic set.....	173

## LIST OF ABBREVIATIONS AND SYMBOLS

$\overline{GI50}$	Population mean IC50
9AC	9-aminocamptothecin
9NC	9-nitrocamptothecin
ADME	absorption, distribution, metabolism, elimination
AUC	area under the curve
CEPH	Centre d'Etude du Polymorphisme Humain
CES	carboxylesterase
chr	chromosome
cM	centimorgan
CICPT	7-chlorocamptothecin
CPT	camptothecin
CPT11	irinotecan
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DSB	double strand break
DTP	Developmental Therapeutics Program
EBV	Epstein-Barr virus
Em	emission
EtOH	ethanol
Ex	excitation
GEO	Gene Expression Omnibus

GI50	concentration which inhibits growth by 50%
GO	gene ontology
h	Hour
h <sup>2</sup>	heritability
hCPT	hydroxycamptothecin
HCS	high content screening
HTS	high throughput screening
Ind	indenoisoquinoline
IC50	concentration which inhibits biological process by 50%
LOD	logarithm (base 10) of odds
LCLs	Lymphoblastoid cell lines
M	Molar
mCPT	Methoxycamptothecin
mM	Millimolar
MOA	mechanism of action
NCI60	National Cancer Institute human cancer cell line panel
nM	Nanomolar
PBS	phosphate buffered saline
PD	pharmacodynamics
PK	pharmacokinetics
QTL	quantitative trait Loci
RNAi	RNA interference
SAR	structure activity relationship

shRNA	short hairpin RNA
siRNA	small interfering RNA
SN38	7-ethyl-10-hydroxy-camptothecin
SNP	single nucleotide polymorphism
SOM	self organizing map
Top	Topoisomerase
TPT	topotecan
uL	microliter
uM	micromolar
z'	z-factor

## **CHAPTER 1:**

### **Applications in Mechanism Elucidation**

This work has been published in part in *Advanced Drug Delivery Reviews*. 2009; 61(5): 369-374.

## **Abstract**

The inability to predict the pharmacology and toxicology of drug candidates in preclinical studies has led to the decline in the number of new drugs which make it to market and the rise in cost associated with drug development. Identifying molecular interactions associated with therapeutic and toxic drug effects early in development is a top priority. Traditional mechanism elucidation strategies are narrow, often focusing on the identification solely of the molecular target. Methods which can offer additional insight into wide-ranging molecular interactions required for drug effect and the biochemical consequences of these interactions are in demand. Genomic strategies have made impressive advances in defining a more global view of drug action and are expected to increasingly be used as a complimentary tool in drug discovery and development.

### **1. Predicting Drug Variability Requires Knowledge of Mechanism**

There is significant interpatient variability in response to anticancer agents; different patients may experience therapeutic benefit, no effect, or even life-threatening side effects from identical doses of the same drug. Very few methods are available to prospectively distinguish those who will benefit from those who may be harmed. Consequently, the number of adverse events associated with cancer therapy remains high. While clinical and environmental variables (e.g., age, gender, diet, organ function, concurrent medications) have been associated with variation in drug response, genetics has been estimated to account for as much as 20-95% of the variability in a broad range of drugs [1]. A drug's activity is the result of interactions with molecular targets and proteins involved in uptake, metabolism, and elimination. Genetic variations in any one of these proteins can have a significant affect on drug response.

The field of pharmacogenomics examines the inherited variations in genes that dictate drug response. It seeks to identify those variations associated with differential responses between



patients. In the past, pharmacogenomic studies used the candidate gene approach to identify factors responsible for variable response. These studies required some *a priori* knowledge about a drug's mechanism of action and the proteins it interacts with to elicit a pharmacological or toxic effect. For example, many cancers overexpress the epidermal growth factor receptor (EGFR) which, when ligand bound, triggers cell proliferation. Gefitinib was developed specifically to inhibit EGFR and suppress tumor growth. Early clinical trials revealed that most patients who received gefitinib saw no therapeutic effect [2]. However, 10% of the patients had a dramatic positive response to therapy [3]. It was subsequently discovered that the tumors of patients experiencing therapeutic benefit had specific activating mutations in the EGFR gene that made them susceptible to the chemotherapeutic agent. Understandably, it was concluded that administering this drug to patients whose tumors did not possess the EGFR mutations was neither medically or financially practical. Knowledge of the mechanism of action and protein interactions required for a drug's pharmacological effect can aid in the identification of patients likely to receive therapeutic benefit or suffer from adverse events.

Unfortunately, many drugs currently in use were developed without knowledge of their underlying molecular mechanisms. Predicting the mechanism of action has proven very difficult for both old and new drugs for several reasons. In many cases the target is unknown; as a result, the biochemical consequences of the drug-target interaction remain elusive. Even when the target is known, the cellular consequences of drug-target interactions remain vague. Moreover, drugs are often capable of binding to more than one target (considered off-target proteins), many of which have not been characterized. The end product is concurrent changes in many different known and unknown biochemical pathways. Our limited understanding is further confounded by unpredictable drug absorption, distribution, and metabolism. Drug action is a very complex process and clearly difficult to untangle. Our inability to elucidate a drug's mechanism is a significant cause for the high failure rates and high costs associated with drug development.

Methods that can provide information on direct targets, indirect targets, affected cellular pathways, and proteins involved in the uptake, metabolism, and elimination of a drug would be powerful tools in drug discovery and development.

A wide variety of technologies have been developed for unraveling drug mechanism (Table 1-1). Traditional approaches fall broadly into two categories: proteomic methods, which involve the identification of a target on the basis of direct binding, and biological strategies, which use bioactivity data to compare the profiles of compounds with known targets or mechanisms to those of the compound of interest. This review will summarize the strategies currently being employed and will use recent case studies to highlight the advantages and limitations of the different approaches. Based on this discussion, genomics will be presented as a new tool for circumventing some of the technical challenges associated with traditional mechanism elucidation strategies.

## **2. Limitations of Current Methods of Mechanism Elucidation**

While drug action is the result of complex biochemical cascades following interactions with a drug's interaction with metabolizing enzymes, transporters, and intracellular targets, traditional mechanism elucidation approaches tend to strictly focus on establishing a single molecular target responsible for a drug's therapeutic activity. Proteomic methods such as affinity chromatography, phage display, and protein microarray involve the direct identification of the target by binding to the compound of interest. *In silico* target prediction is an indirect method of protein target identification which suggests likely biological targets of small molecules via data mining in target-annotated chemical databases. Biological methods, such as cellular phenotyping, which compares the pharmacological profiles of compounds with known targets to those of the compound of interest, are also indirect methods of target identification.

## 2. 1. Proteomic Methods

### 2. 1. 1. Affinity purification of targets

Affinity chromatography is a powerful and classic method used to identify target proteins for small molecules [4-6]. While a number of successes have been reported, results are often variable. In this approach, a protein extract is passed over a packed column consisting of drug immobilized to a solid support (Figure 1-1). Following repeated washing to remove unbound proteins, the bound protein is eluted using denaturing conditions or elution with mobile ligand. In principle, this method is applicable only to small molecules that can be derivatized without disrupting their biological activity; biological activity and molecular targets have been shown to change with chemical modification for immobilization on solid supports [7]. For detection, this method requires high affinity ligands and a high abundance of the target protein in the cell extract. Compounds isolated following high throughput screens are typically not very potent, and low abundance target proteins are difficult to detect over background non-specific binders. Detection of weaker biologically relevant interactions is hampered by a number of factors. For example, many bioactive molecules are somewhat hydrophobic which predisposes them to non-specific binding when coupled at high density to a solid support. Consequently, highly stringent wash conditions are required to reduce the likelihood of detecting weak interactions.

The yeast three-hybrid system [8], phage display [9-11] and mRNA display [12-13] are three relatively new protein based methods for small molecule target discovery. They all utilize affinity chromatography but were developed to counter problems with low affinity ligands and the low abundance of the target protein in extracts. These methods involve the *in vitro* synthesis of a library or a pool of proteins which are then submitted to a selection process that entails repeated amplification and enrichment to isolate proteins of interest. The probability of finding a binding protein with high affinity increases as the library size increases in these systems. In addition, that

probability is also influenced by both the ability to diversify the library, and then isolate and characterize selected proteins from the library. A library lacking diversity or possessing underrepresented binding proteins may inadvertently be missing the target. The isolation and characterization of binding proteins in phage display or mRNA display can be difficult. Binding proteins must be expressed in *E. coli*, where they may fold improperly and form insoluble inactive aggregates or inclusion bodies, which can not be easily purified. The steps needed to solubilize and refold the protein can be highly variable and may not always result in high yields of active protein. The yeast three-hybrid system is limited to the study of proteins which can be expressed in yeast.

### 2. 1. 2. Protein microarray

Another method of identifying the molecular targets of small molecules is the protein microarray. Protein microarrays are prepared by spotting purified proteins on chemically derivatized glass slides. The binding profile for a small molecule across an entire proteome can be achieved by incubating the array with a fluorescent or radiolabelled form of the small molecule [14-15]. After rigorous washing, labeled proteins are identified. In a study to identify the cellular targets of SMIRs, small molecule inhibitors of rapamycin, which suppress rapamycin's inhibitory effects on cell growth, Huang et al. prepared a microarray of nearly the entire yeast proteome [16]. SMIRs were biotinylated and binding to protein targets was detected using fluorescently labeled streptavidin. Thirty binding proteins were identified, among them Ybr077cp, a protein of previously unknown function. Yeast strains with a Ybr077cp deletion (Ybr077cp $\Delta$ ) were hypersensitive to rapamycin. Ybr077cp-induced hypersensitivity was reversed when the Ybr077cp $\Delta$  cells were transfected with TOR1-1, a functional variant of the target of rapamycin protein which cannot bind rapamycin. The authors concluded Ybr077cp is likely a component of the TOR signaling network. Unfortunately, targets identified using the yeast proteome may not be relevant to human biology; for example there is no known human homolog for Ybr077cp.

The natural transition would be the use of arrays prepared from the human proteome. A number of technological advances are still needed before the whole human proteome can be applied to protein microarrays for drug discovery endeavors. To date only small scale human protein arrays have been used for study [15]. Moreover, while protein microarrays have advantages over affinity chromatography, there are limitations which can hinder target identification. For example, the discovery of potential targets by affinity chromatography is hampered by low levels of natively expressed proteins, while protein microarrays expose all proteins equally. However, placement on the array may result in steric hindrance which would prevent small molecule binding. In addition, small molecule target proteins might not be identified since the protein incorporated on arrays will lack post-translational modifications, or involvement in complex formation with other proteins which would contribute to their affinity for a ligand.

### *2. 1. 3. In silico target prediction*

Since it is currently unfeasible to screen all proteins expressed by the human genome, chemoinformatics has sought to develop computational methods which can predict the proteins to which a drug is likely to bind. One of the latest techniques towards this goal was reported by Keiser et al. [17]. They computationally screened the chemical structures of approximately 3,600 FDA approved and investigational drugs against the thousands of known ligands for 1,400 protein targets. Chemical similarities between drugs and ligands suggested thousands of known and novel drug-target associations. A subset of these drug-target associations were further examined (n = 184) and nearly 150 were predicted which had no literature precedent. The authors used in vitro binding assays to validate 30 of these new drug-target predictions. Twenty-three of these drugs (77%) bound with affinity less than 15  $\mu$ M and 5 of these had sub-100 nM affinities for the previously unknown targets. Some of these novel targets were suspected to contribute to drug action. For example, the sigma receptor ( $\sigma_1$ ) receptor was previously implicated as the target responsible for the hallucinogenic properties associated with N,N-dimethyltryptamine (DMT).

However, other non-hallucinogens bind the  $\sigma_1$  receptor with higher affinity than DMT and prior research indicates the hallucinogenic characteristics associated with DMT more closely resemble an interaction with serotonergic (5-HT) receptors. The study by Keiser et al. predicted multiple 5-HT interactions and binding studies confirmed DMT binds nine serotonin receptors with affinities ranging from 39 nM to 2.1  $\mu$ M. Unfortunately, there is no rapid and simple method of identifying and validating which, if any, of the nearly 4000 new predicted drug-target interactions are biologically relevant and associated with a drug's primary activity or side effects.

## **2. 2. Biological Methods**

### *2. 2. 1. Cellular Phenotyping*

A considerable number of drugs in use today were discovered by screening phenotypic changes induced by candidate drugs in cells, tissues, or model organisms. One of the most noteworthy examples of drug discovery and indirect target prediction was developed at the National Cancer Institute (NCI). Compounds are evaluated for their ability to inhibit cell growth in a panel of 60 human cancer cell lines. The COMPARE algorithm then matches a test compound's cell growth inhibition pattern across all 60 cell lines (referred to as the fingerprint) with one or more of the thousands of other compounds in the NCI database [18]. A high degree of correlation between two fingerprints suggests that the compounds share a molecular target. This model identified the novel drug kenpaullone as a cyclin dependent kinase (CDK) inhibitor when its fingerprint matched with other CDK inhibitors that had been through the screen previously [19]. The model is unsuccessful in assigning a mechanistic classification when the fingerprint for a drug candidate is too distinct from the patterns of other compounds with known mechanisms in the database, a scenario suggesting a novel molecular target.

### *2. 2. 2. High Content Screening*

Advances in automated microscopy and fluorescence have enabled the simultaneous screening of multiple cellular phenotypes in a method referred to as high content screening (HCS). Several phenotypic markers obtained from a single cell or cellular subpopulation can be combined to generate a multidimensional biological fingerprint. This fingerprint provides a more cohesive profile of the action of a drug and enables compounds of similar activity to be grouped together. Young and colleagues recently describe a methodology for the integration of HCS data with chemical structure information to make mechanisms of action inferences [20]. The authors screened a library of > 6,000 compounds using a cellular-proliferation assay which also measured 30 cytological phenotypes. Factor analysis was used to reduce these cytological features to six categories: nuclear size, DNA replication, mitosis, nuclear morphology, nuclear ellipticity, and EdU texture. A mean response score was calculated for each compound. Compounds with response factors in the upper 5% were considered biologically active. Hierarchical clustering of biologically active compounds by factor scores revealed seven broad clusters termed phenotypes. Since similar structures tend to possess similar mechanisms, phenotypic clusters were investigated for structurally related compounds. Approximately 96% of compounds with similar structures showed similar phenotypic readouts. Conversely, 4% of compounds with slight changes in structure showed large changes in phenotypic readouts and suggested a change in mechanism. Compounds from distinct structural classes which were known to share common molecular targets via the same or different binding sites, or perturbed different components of common pathways also produced similar phenotypic readouts. Further analysis using a subset of compounds with known shared targets revealed that phenotypes correlated better with predicted compound targets than with the compound structures themselves. Finally, the authors investigated whether the methodology could make predictions about molecular targets. A chemogenomic database with known-ligand target associations based on molecular substructure was used to generate a model which could predict the targets of the 211 active compounds. The authors focused on the predicted targets for four groups of phenotypically similar yet structurally distinct

compounds. The group consisted of colchicine derivatives, novel kinase inhibitors, a quinoline derivative and a pseudolarix acid B derivative. Colchicine is a well known microtubule inhibitor and the majority of the compounds were predicted to share the common target, tubulin. Follow-up in vitro assays confirmed all compounds caused microtubule depolymerization.

These phenotypic screening based approaches were designed to identify any compound which ameliorates a disease phenotype in an animal or cell-based model. Consequently, a vast number of compounds could be discovered to act on a substantial number of known and presently undiscovered targets and pathways associated with a disease. However, a fundamental limitation of these methods is that clustering or correlation on its own does not reveal the mechanism of action of compounds of interest. To infer mechanism, compounds with known mechanisms must be used to serve as markers for comparison. However, the number of compounds with clearly defined mechanisms of action is limited. Consequently, as a mechanism elucidation strategy, phenotypic screening is restricted to existing target knowledge. Moreover, many drugs have many targets and consequently a complicated resulting biochemical cascade. Finally, marketed drugs are only using a small portion of the potential protein targets of pharmacological interest [21]. A method built strictly on making predictions using compounds with known molecular targets is inherently limited.

### **3. Genomics Broadens Understanding of Drug Action**

To meet the challenging problem of identifying the MOA for drug candidates, novel methods are constantly being developed and old methods increasingly improved upon. Some impressive successes have been attributed to the use of genetics as a tool in the identification of mechanisms of action for drugs. The innovative genetic models that follow have several advantages over the target identification assays described above. Compounds with known molecular targets are not required as landmarks for mechanism elucidation. Moreover, they require no *a priori* knowledge



about the compound mechanism of action [22]. This allows the activities of novel drugs to be determined in a systematic and unbiased method. These processes allow the discovery of biological pathways involved in drug action (including proteins associated with metabolism, distribution, and off target effects) in addition to the precise mechanism of action to be determined. Traditional methods of elucidating mechanism are restricted by a static view of drug action: they oversimplify and focus the search on a single molecular target. By allowing the biology to reveal the genes influencing activity, genomic tools offer a more dynamic and global perspective of a drug's mechanism.

### **3. 1. Applications**

#### *3. 1. 1. Yeast genomics*

Enhanced knowledge of yeast genomics has enabled the use of the budding yeast *Saccharomyces cerevisiae* as a powerful tool for mechanistic discovery. There are different types of yeast mutant libraries which have been employed in mechanistic studies: heterozygous deletions, homozygous deletions, and overexpression libraries. An example of these libraries is the collection of genome-wide heterozygous deletion strains developed with molecular barcodes. When these libraries are grown in the presence of drug, the deletions that sensitize cells to a particular drug will cause a decrease in cell growth relative to control [23]. The barcode associated with each strain is used to quantitate growth and identify genes involved in the drug's mechanism. This method has been used to explore the cellular pathways and processes for a collection of compounds with known and unknown modes of action [23-24]. Hierarchical clustering of compounds with similar genomic profiles suggests common molecular targets and pathways [24]. For example, the genomic profiles of amiodarone, an antiarrhythmic drug, and the chemotherapeutic agent tamoxifen which targets the estrogen receptor, were quite similar. Amiodarone acts through perturbation of calcium homeostasis. In three independent validation assays, tamoxifen was also shown to disrupt

calcium homeostasis confirming published evidence that the drug increases calcium concentrations in a variety of mammalian cells [24]. Moreover, the system could be used to identify unknown targets for novel agents. The target of the antifungal, papuamide B (PapB), was identified by assessing both drug resistant and sensitive mutants. Sensitive mutants affected by exposure to PapB had gene deletions related to cell wall organization. A single gene, the enzyme required for synthesis of phosphatidylserine (PS) in yeast cell membranes, was associated with the growth of resistant wild type cells in the presence of PapB. Investigators proposed that papB binds PS and acts on membrane integrity and permeability. A comparison of the genomic profile of PapB with known membrane permeabilizers and agents which bind other phospholipids revealed a match. Yeast genomic profiling is exceptional because it can be used to identify primary and secondary targets via sensitivity as well as loss-of-function mutations that result in drug-resistance [25].

### *3. 1. 2. Human cancer cell lines*

Some of the earlier limitations of cell-based phenotypic screening have been circumvented with the incorporation of a genomics component. Scherf et al. were the first to use genomics to make predictions regarding targets critical to drug action in the cancer cell lines (Figure 1-2) [26]. They recognized that patterns in cellular sensitivity in the NCI60 could be linked to differences in gene expression between the cell lines, and might provide information about a compound's mechanism of action. The group measured gene expression levels in the untreated NCI60 cell lines. Patterns of gene expression across the NCI60 panel were then correlated to the biological activity of 1,400 compounds across the same cell line panel. Cells expressing higher levels of a gene were less sensitive to a compound and visa versa. For example, the expression of dihydropyrimidine dehydrogenase (DPYD) was negatively correlated ( $r = -0.53$ ) to the potency of 5-fluoracil (5-FU) across the NCI60 panel; cell lines which expressed low levels of DPYD, the enzyme which inactivates 5-FU, were more sensitive to the drug. Finally, since drug design and discovery is

often concerned with identifying which structural features might yield a particular mechanism, Blower et al. expanded on the work reported by Scherf [27]. The structural characteristics of compounds were linked to their biological fingerprints and the gene expression profiles of the NCI60. This model now serves as a valuable resource for identifying compounds which have anticancer activity and providing testable hypotheses about genes (and gene products) associated with that drug's mechanism of action.

### *3. 1. 3. Rodent models*

Another noteworthy example involves the use of mouse haplotype computational genetic analysis to identify genes that affect drug metabolism or response. This method was recently used to identify genes and the resulting proteins affecting the overall metabolism of warfarin in mice (Figure 1-3) [28]. Warfarin, a commonly prescribed anticoagulant, is metabolized by many different pathways and by a variety of enzymes into different metabolites. <sup>14</sup>C-labeled R-warfarin was administered to 13 inbred mouse strains and both parent compound and metabolites quantified in plasma for up to 24 h following dosing. Strain specific differences were observed in the production of warfarin metabolites. Of all the metabolites studied, inbred strains had the largest difference in the rate of formation of 7-hydroxylated warfarin. Computational genetic analysis was used to look for patterns of genetic variation that correlated with the observed differences in rate of formation of 7-hydroxywarfarin across mouse strains. The two strains of mice with the lowest rates of warfarin metabolism differed from the other strains in a region on chromosome 19 that encodes for the metabolizing enzyme cytochrome P450 2C (Cyp2c). To confirm the role of Cyp2c in murine warfarin metabolism, the formation of the major metabolite 7-hydroxywarfarin was inhibited in murine liver microsomes following the administration of a Cyp2c specific inhibitor [28]. Moreover, the expression of Cyp2c29 in liver extracts were 2-7.4

fold lower in the two strains with the low rate of metabolite generation. Although there are a number of genetic variables that contribute to the interstrain differences in R-warfarin metabolism, Cyp2c9 was identified because it was a rate-limiting enzyme in a major elimination pathway for warfarin.

The impact of this tool extends beyond evaluating drug metabolism. For example, this approach has also been used to link the beta-2-adrenergic receptor to increased pain sensitization, a side effect associated with the administration and subsequent withdrawal of opioids [29]. In this model, the pain response threshold was measured at baseline and following four days of exposure to morphine in 15 inbred mouse strains. There was a significant difference in the extent of pain sensitization observed between the mouse strains following morphine treatment and withdrawal. Computational genetics was used to identify genes which might be responsible for the observed differences. A haplotype block on chromosome 18, specifically within a region of the  $\beta_2$ -adrenergic receptor gene, was the most strongly correlated with the observed phenotypic difference. Administration of a  $\beta_2$  receptor antagonist caused a dose-dependent reversal of pain sensitization. Contrary to wild type mice, knockout mice for  $\beta_2$  receptor had no pain sensitization following morphine withdrawal. The authors concluded that the  $\beta_2$  adrenergic receptor was the likely receptor subtype responsible for the pain sensitization following morphine withdrawal. Studies like this which attempt to identify novel genetic factors affecting dependence on opioids are essential to the discovery of methods for the prevention or treatment of increased pain sensitivity and other symptom of opioid addiction.

#### **4. 1. Conclusions**

Drug activity is clearly a tangled and complex process. Gaining a clear understanding of drug action specifically a drug's interaction with direct and indirect targets and proteins involved in transport and metabolism remains a formidable task in drug discovery and development. The

analysis of the various proteomic and biological mechanism elucidation methods clearly demonstrates that each approach has its specific strengths and limitations. Genomics is an underutilized tool that can strengthen current efforts in identifying mechanisms of action. It has the potential to directly link drugs which produce a desired phenotype in a validated model to a more global view of their mechanism early on in the development process. However, the limitations associated with the above genomic strategies for mechanism elucidation are worth mentioning.

The most critical flaw is the fact that the extent to which these results can be translated to humans remains unclear. Yeast are primitive organisms whose intracellular conditions such as protein folding and post-translational modifications can differ significantly from mammalian cells. Likewise, some mammalian targets are absent in yeast and visa versa. Inconsistencies in the data between humans and yeast may even suggest compounds might affect an entirely different process in yeast [30]. Moreover, compounds can have a lower permeability in yeast cells as compared to mammalian cells. As a result, only a subset of compounds of interest may be evaluated in yeast. Similarly, genes and the resulting proteins identified as contributors to drug action in murine models may not always reflect events in humans. For example, drug metabolism in rodents can differ from humans due to major differences in P450 isoforms, expression, and catalytic activity.

Cancer cell lines are not always representative of primary tumors. Auman and McLeod compared genome wide expression data of human colorectal cancer cell lines to clinical colorectal tumors [31]. Hierarchical clustering on gene expression revealed that cancer cell lines formed a single cluster separate from clinical tumor samples. The group concluded that the cell lines did not accurately represent the genetic heterogeneity present in clinical tumor samples [31]. In another example, human prostate cancer cell lines do not exhibit features commonly seen in human prostate cancer. Prostate-specific antigen (PSA) is synthesized almost exclusively in the human prostate and is androgen regulated. High levels of PSA are indicative of prostate cancer. The two

prostate cancer cell lines used in the NCI60 panel, PC3 and DU145, do not express PSA and are not androgen sensitive [32]. Genomic studies based on gene expression patterns in cancer cell lines may suggest molecular mechanisms critical to drug action which are distinct from those in the clinical setting.

Moreover, *in vitro* cell line work using gene expression analysis is often inadequate. Microarray studies of gene expression will reveal some of the genes associated with drug effect but it may miss many other relationships. For example, not all of the changes will be under transcriptional control. Some will be controlled using post-translational modifications such as phosphorylation and glycosylation; gene expression techniques such as DNA microarray will be unable to detect these endpoints.

Considering these drawbacks, an ideal model for a genomic mechanism elucidation strategy would satisfy the following conditions:

1. performed in a human model
2. quantitative and reproducible
3. high throughput screening capability
4. amenable to database generation
5. can cover broad range of biology (targets, pathways, physiology, and diseases)
6. can provide links between chemistry and biology
7. mechanistically open such that changes MOA can be recognized with changes in structure
8. can give clinical indications about efficacy and toxicity
9. can give information about secondary or off-target activity and effects
10. can screen more than 1 concentration of a drug
11. can link differences in response to genetic heterogeneity

A new strategy for identifying genes critical to drug action has the potential to satisfy many of the aforementioned criteria. This tactic employs an *ex vivo* human familial genetic model to identify inherent mutations in genes involved in drug action which are associated with differences in response. The genes influencing the cytotoxicity of chemotherapeutic agents have been studied using immortalized lymphoblastoid cell lines (LCLs) derived from Centre d'Etude du

Polymorphisme Humain (CEPH) populations. The CEPH cell lines are a collection of multigenerational families that have been extensively genotyped. Cells from these families are phenotyped for response to a given anticancer agent then linkage analysis is used to correlate variation in response to variation in regions of the genome referred to as quantitative trait loci or QTLs (Figure 1-4). Watters et al. phenotyped sensitivity to increasing concentrations of 5-fluorouracil in 427 CEPH cell lines across 38 families [34]. Significant variation was noted across individual cell lines at each dose. Heritability, the degree to which a trait can be explained by genetic factors, ranged from 26% at the lowest concentration of 5-fluorouracil to 65% at the highest concentration. Dose-dependent QTLs associated with 5-fluorouracil cytotoxicity were observed on chromosomes 5 and 9. Further studies are expected to narrow these broad QTLs down to the genes which are involved in drug action. This model has also been applied to the study of the anticancer agents, cisplatin [35], etoposide [36], docetaxel [34], and daunorubicin [37]; cytotoxic response to each of these agents was a heritable trait in human families with genomic regions associated with observed differences in response. To date, this model has only been used as a pharmacogenomic tool to evaluate which genes and variations in the genome are responsible for the disparity in response to a *single* drug. Future studies are needed to evaluate the predictive genomic capacity of this model, specifically to use the pharmacological and genomic response profiles of numerous drugs, rather than those of one small molecule at a time, to provide incisive information about their mechanisms of action.

#### **4. 2. Introduction to Dissertation**

Tools for identifying genes and gene products critical to drug action early in development are in great demand. The primary objective of this dissertation project was to investigate the potential use of our *ex vivo* familial genetics model in CEPH cell lines as a tool for mechanism elucidation. The following chapters describe an initial investigation into using the CEPH cell lines to relate genes or regions of the genome identified as influencing the cytotoxicity of a compound across

compound classes to inherited response and mechanisms of action. The purpose of Chapter 2 was to investigate the model's ability to establish specific patterns of QTLs related to a shared mechanism for a class of structurally related compounds, the camptothecins, which are Topoisomerase 1 (Top1) inhibitors. In Chapter 3, the genomic profiles of structurally unrelated Top1 inhibitors were compared to those established for the camptothecins, to assess which regions might be associated with compound class versus the general mechanism of Top1 inhibition. The goal of Chapter 4 was to demonstrate that biological and genomic data generated from phenotyping the CEPH cell lines can be used to stratify compounds by mechanism of action. We predicted that intraclass pharmacological and genomic profiles would be more similar to each other than to compounds belonging to distinct mechanistic classes.



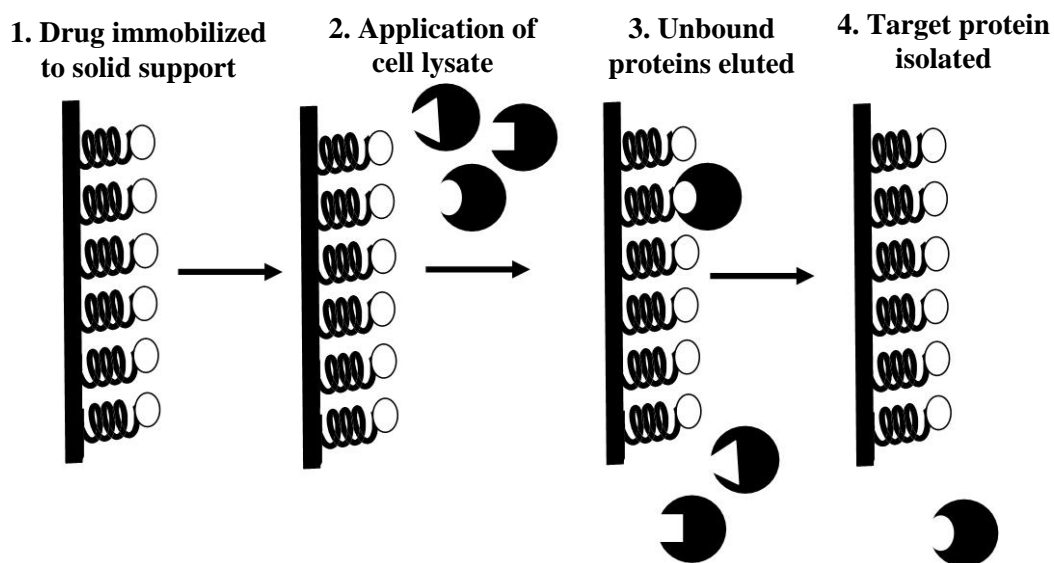


Figure 1-1. Affinity chromatography based methods for target identification. Affinity chromatography makes use of the highly specific binding sites usually present in biological macromolecules, separating molecules on their ability to bind a particular ligand. Covalent bonds attach the drug (ligand) to an insoluble, porous support in a manner that presents the ligand to the protein sample. The protein mixture is passed over the medium, and the target protein binds to the drug tethered to the solid support. A buffer is used to wash or remove impurities and unbound material. Finally, denaturing conditions are used elute the bound proteins from the column.

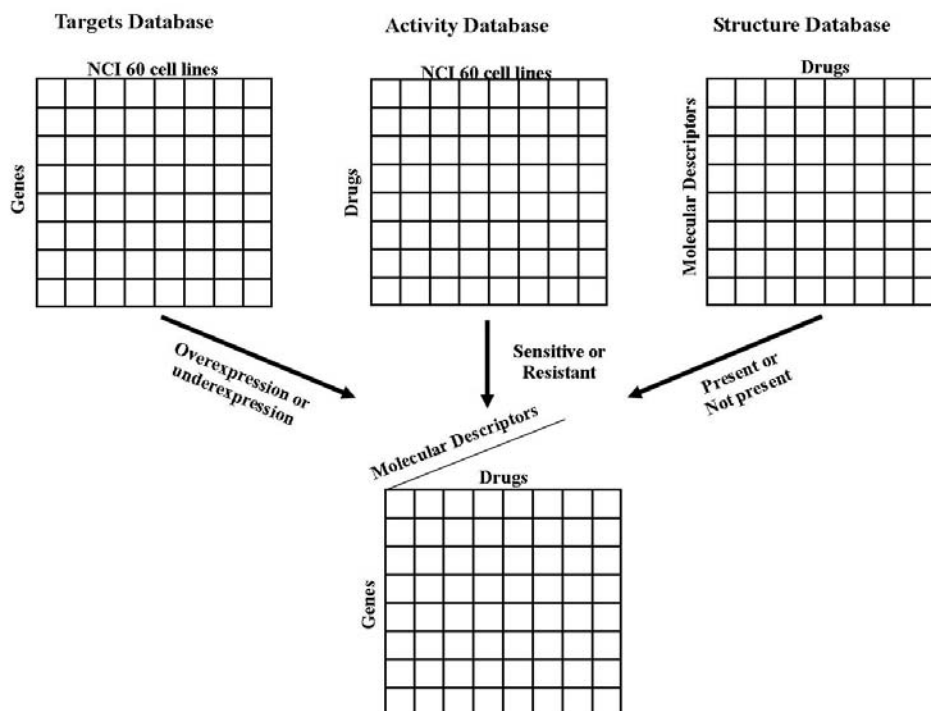


Figure 1-2. Schematic overview of mechanism prediction model using NCI60 human cancer cell lines. The target database is composed of relative gene expression measurements on the 60 human cancer cell lines. Each row of the activity database represents the pattern (sensitivity or resistant) of biological activity of a particular compound across the 60 cancer cell lines. The structure database contains the 2D or 3D chemical characteristics of the compounds investigated using the NCI60 cell line panel. Coupling of genomic, biological, and chemical information might allow genes that are selectively expressed in a tumor to be correlated not only with the compounds themselves but also with the subclasses and substructures of these compounds. The target, activity, and structure databases can then be used to make predictions about potential targets or activity patterns of a compound given its molecular substructure.



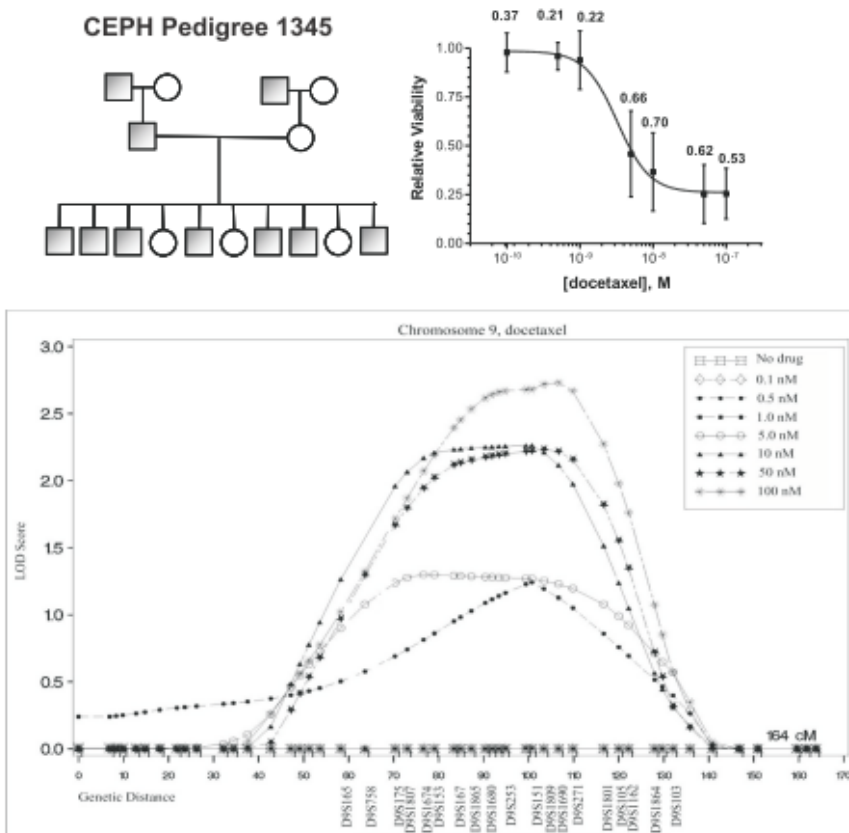


Figure 1-4. Discovery of genetic loci involved in the cytotoxic effect of chemotherapeutic agents. An *ex vivo* familial study was used to identify genes associated with docetaxel cytotoxicity. A) Variability at increasing concentrations of docetaxel was assessed in a collection of lymphoblastoid cell lines derived from CEPH pedigrees. Each pedigree consists of 5-10 offspring per family. B) A dose-response curve shows significant variation in cell viability at all concentrations across the entire CEPH population. Data points are mean cell viability and bars are standard deviations across the entire population. The degree to which observed variation in cell viability can be explained by genetic factors (heritability) is represented by numbers. C) Linkage analysis correlated a region on chromosome 9 with the observed variability in cytotoxic response. Furthermore, as drug dose increased the LOD score (probability the observed phenotype is related to the variation in a specific region) also increased.

Class	Technique	Overview	Advantages	Disadvantages	Ref
Proteomic	Protein microarrays	Proteins immobilized to chips are incubated with a labeled small molecule and repeatedly washed. Labeled drug remains bound to target proteins. Isolation of target protein reveals potential mechanisms.	Proteins expressed at low levels natively receive equal exposure to the small molecule on the microarray	If target proteins must complex with other proteins for drug interaction or require post-translational modification, the target may not be identified.	[16]
Biological	NCI60 IVCLSP using COMPARE	The cytotoxic response of 60 cancer cell lines to an agent is used in a pattern recognition algorithm to relate small molecules and their mechanisms.	Cytotoxicity data for compounds can be linked to chemical structure and differences in gene expression for each cancer cell line.	Chemosensitivity profiles do not always match existing small molecules with defined mechanisms of action.	[38]
Genomic	Yeast deletion library screening	Barcoded yeast strains possessing one copy of a specific gene have increased drug sensitivity to genes corresponding to that drug's mechanism of action.	Biological processes in addition to primary targets can be identified by this method	Compounds active in mammalian organisms may not be active in yeast.	[39]
	Gene expression in NCI60 cancer cell lines	Gene expression in treated cells is compared to that in untreated cells to identify genes whose expression is changed in response to a treatment.	Require no a priori knowledge about drug action	Not all changes in gene expression associated with drug effect are under transcriptional control	[27]
	Murine haplotype based genetic mapping	A phenotypic variable associated with drug treatment is measured across a panel of inbred mouse strains. Patterns of observable differences in the measured phenotype across strains are correlated with a pattern of genetic variation to identify genes associated with drug activity.	Can directly identify numerous individual genes responsible for drug action	Drug absorption, metabolism, distribution, and elimination in mice may not translate to the human setting	[40]

Table 1-1. Comparison of Mechanism Elucidation Strategies

Class	Technique	Overview	Advantages	Disadvantages	Ref
Proteomic	Affinity chromatography	Small molecule is immobilized to a solid support and a protein extract passed over the column. Following repeated washes to remove non-specific binding proteins, proteins which bind the small molecule are eluted using denaturing conditions. Isolation of target protein reveals potential mechanisms.	Target proteins maintains the 3D structure necessary for protein binding	Requires high affinity small molecules and an abundance of target protein	[6]
	Yeast three-hybrid system	A DNA binding site is placed upstream of a reporter in the yeast gene. A protein consisting of the DNA-binding domain linked to a ligand-binding domain interacts with the DNA-binding site on the yeast chromosome and a ligand covalently bound to a small molecule drug. The third component is a protein from a cDNA library (potential drug targets) which is fused to a transcriptional activation domain. When the trimeric complex is formed, the reporter is activated and targets may be identified.	A clone encoding the protein of interest is obtained directly in the screen	False positives may be identified if the reporter gene is activated without the small molecule-protein interaction.	[8]
	Phage display	Bacteriophage expressing a library of proteins are exposed to immobilized small molecules. The solid support is washed extensively to remove nonspecific binders, bound proteins eluted, and the process repeated for further enrichment of binding proteins and isolation of high-affinity sequences.	Repeated cycles of selection allow the detection of low abundance proteins.	Library must be sufficiently large and diverse or it may not contain the protein target responsible for investigated drug activity.	[11]
	mRNA display	A protein is chemically attached to its own mRNA at the 3' end through a puromycin linker. mRNA-protein fusions are subjected to repeated cycles of in vitro selection using an immobilized small molecule affinity column. Binding mRNA-protein fusions will be amplified to a cDNA library that codes for fewer proteins but with greater binding affinity for the small molecule.	Early rounds of selection can identify protein targets with low affinity binding.	Target proteins may not be in their native form or possess post-translational modifications.	[12]

Class	Technique	Overview	Advantages	Disadvantages	Ref
Genomic	Ex vivo familial genetics	Pedigrees of CEPH lymphoblasts are phenotyped for sensitivity to cytotoxic agents. Heritability and linkage analysis enables the identification of regions of the genome contributing to the variation in response and genes associated with a drug's mechanism.	CEPH cell lines have been extensively genotyped and the data is publicly available allowing the identification of many loci influencing drug action	Tissue specific phenotypic effects such as hepatotoxicity can not be studied in these lymphoblastoid cell lines	[33, 34]

## REFERENCES

1. Evans, W.E. and H.L. McLeod, *Pharmacogenomics--drug disposition, drug targets, and side effects*. N Engl J Med, 2003. **348**(6): p. 538-549.
2. Cohen, M.H., et al., *United States Food and Drug Administration Drug Approval summary: Gefitinib (ZD1839; Iressa) tablets*. Clin Cancer Res, 2004. **10**(4): p. 1212-1218.
3. Lynch, T.J., et al., *Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non-Small-Cell Lung Cancer to Gefitinib*. N Engl J Med, 2004. **350**(21): p. 2129-2139.
4. Yamaguchi, T., et al., *Identification of JTP-70902, a p15-INK4b-inductive compound, as a novel MEK1/2 inhibitor*. Cancer Science, 2007. **98**(11): p. 1809-1816.
5. Yoshida, M., et al., *A new mechanism of 6-((2-(dimethylamino)ethyl)amino)-3-hydroxy-7H-indeno(2,1-c)quinolin-7-one dihydrochloride (TAS-103) action discovered by target screening with drug-immobilized affinity beads*. Mol Pharmacol, 2008. **73**(3): p. 987-994.
6. Uga, H., et al., *A new mechanism of methotrexate action revealed by target screening with affinity beads*. Mol Pharmacol, 2006. **70**(5): p. 1832-9.
7. Prühs, C. and C. Kunick, *Darpones and water-soluble aminobutoxylated darpone derivatives are distinguished by matrix COMPARE analysis*. Bioorganic & Medicinal Chemistry Letters, 2007. **17**(7): p. 1850-1854.
8. Becker, F., et al., *A three-hybrid approach to scanning the proteome for targets of small molecule kinase inhibitors*. Chem Biol, 2004. **11**(2): p. 211-223.
9. Sidhu, S.S., et al., *Phage display for selection of novel binding peptides*. Methods Enzymol, 2000. **328**: p. 333-363.
10. Rodi, D.J., L. Makowski, and B.K. Kay, *One from column A and two from column B: the benefits of phage display in molecular-recognition studies*. Curr Opin Chem Biol, 2002. **6**(1): p. 92-96.
11. Jin, Y., J. Yu, and Y.G. Yu, *Identification of hNopp140 as a binding partner for doxorubicin with a phage display cloning method*. Chem Biol, 2002. **9**(2): p. 157-62.
12. McPherson, M., et al., *Drug receptor identification from multiple tissues using cellular-derived mRNA display libraries*. Chem Biol, 2002. **9**(6): p. 691-698.



13. Liu, R., et al., *Optimized synthesis of RNA-protein fusions for in vitro protein selection*. *Methods Enzymol*, 2000. **318**: p. 268-293.
14. Zhu, H., et al., *Global analysis of protein activities using proteome chips*. *Science*, 2001. **293**(5537): p. 2101-5.
15. MacBeath, G. and S.L. Schreiber, *Printing Proteins as Microarrays for High-Throughput Function Determination*. *Science*, 2000. **289**(5485): p. 1760-1763.
16. Huang, J., et al., *Finding new components of the target of rapamycin (TOR) signaling network through chemical genetics and proteome chips*. *Proc Natl Acad Sci U S A*, 2004. **101**(47): p. 16594-9.
17. Keiser, M.J., et al., *Predicting new molecular targets for known drugs*. *Nature*, 2009.
18. Paull, K.D., et al., *Display and analysis of patterns of differential activity of drugs against human tumor cell lines: development of mean graph and COMPARE algorithm*. *J Natl Cancer Inst*, 1989. **81**(14): p. 1088-1092.
19. Zaharevitz, D.W., et al., *Discovery and initial characterization of the paullones, a novel class of small-molecule inhibitors of cyclin-dependent kinases*. *Cancer Res*, 1999. **59**(11): p. 2566-2569.
20. Young, D.W., et al., *Integrating high-content screening and ligand-target prediction to identify mechanism of action*. *Nat Chem Biol*, 2008. **4**(1): p. 59-68.
21. Imming, P., C. Sinning, and A. Meyer, *Drugs, their targets and the nature and number of drug targets*. *Nature Reviews Drug Discovery*, 2006. **5**: p. 821-834.
22. Schriemer, D.C., D. Kemmer, and M. Roberge, *Design of phenotypic screens for bioactive chemicals and identification of their targets by genetic and proteomic approaches*. *Comb Chem High Throughput Screen*, 2008. **11**(8): p. 610-6.
23. Lum, P.Y., et al., *Discovering modes of action for therapeutic compounds using a genome-wide screen of yeast heterozygotes*. *Cell*, 2004. **116**(1): p. 121-137.
24. Parsons, A.B., et al., *Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast*. *Cell*, 2006. **126**(3): p. 611-625.
25. Winzeler, E.A., et al., *Functional Characterization of the *S. cerevisiae* Genome by Gene Deletion and Parallel Analysis*. *Science*, 1999. **285**(5429): p. 901-906.
26. Scherf, U., et al., *A gene expression database for the molecular pharmacology of cancer*. *Nat Genet*, 2000. **24**(3): p. 236-44.

27. Blower, P.E., et al., *Pharmacogenomic analysis: correlating molecular substructure classes with microarray gene expression data*. Pharmacogenomics J, 2002. **2**(4): p. 259-71.
28. Guo, Y., et al., *In silico pharmacogenetics of warfarin metabolism*. Nat Biotechnol, 2006. **24**(5): p. 531-536.
29. Liang, D.Y., et al., *A genetic analysis of opioid-induced hyperalgesia in mice*. Anesthesiology, 2006. **104**(5): p. 1054-1062.
30. Luesch, H., et al., *A functional genomics approach to the mode of action of apratoxin A*. Nat Chem Biol, 2006. **2**(3): p. 158-67.
31. Auman, J.T. and H.L. McLeod, *Colorectal cancer cell lines lack the molecular heterogeneity of clinical colorectal tumors*. Clin Colorectal Cancer, 2010. **9**(1): p. 40-7.
32. Navone, N.M., et al., *Model Systems of Prostate Cancer: Uses and Limitations*. Cancer and Metastasis Reviews, 1998. **17**(4): p. 361-371.
33. Dolan, M.E., et al., *Heritability and linkage analysis of sensitivity to cisplatin-induced cytotoxicity*. Cancer Res, 2004. **64**(12): p. 4353-4356.
34. Watters, J.W., et al., *Genome-wide discovery of loci influencing chemotherapy cytotoxicity*. Proc Natl Acad Sci U S A, 2004. **101**(32): p. 11809-11814.
35. Huang, R.S., et al., *Identification of genetic variants contributing to cisplatin-induced cytotoxicity by use of a genomewide approach*. Am J Hum Genet, 2007. **81**(3): p. 427-437.
36. Huang, R.S., et al., *A genome-wide approach to identify genetic variants that contribute to etoposide-induced cytotoxicity*. Proc Natl Acad Sci U S A, 2007. **104**(23): p. 9758-9763.
37. Huang, R.S., et al., *Genetic variants contributing to daunorubicin-induced cytotoxicity*. Cancer Res, 2008. **68**(9): p. 3161-3168.
38. Kohlhagen, G., et al., *Protein-linked DNA strand breaks induced by NSC 314622, a novel noncamptothecin topoisomerase I poison*. Mol Pharmacol, 1998. **54**(1): p. 50-8.
39. Baetz, K., et al., *Yeast genome-wide drug-induced haploinsufficiency screen to determine drug mode of action*. Proc Natl Acad Sci U S A, 2004. **101**(13): p. 4525-30.
40. Liao, G., et al., *In silico genetics: identification of a functional element regulating H2-Ealpha gene expression*. Science, 2004. **306**(5696): p. 690-695.

## **CHAPTER 2:**

### **Genome-wide Approach for Differentiation of the Camptothecins**

## **ABSTRACT**

Differences in biological activity and ADME profiles are often observed for compounds within a mechanistic class bearing slight modifications on a structural theme. We propose that these changes can be detected by examining the changes in genes which influence the cytotoxicity of these compounds using HTS in collections of genotyped human familial (CEPH) cell lines. Moreover, this genomic strategy can be used to establish a specific pattern of genes related to the shared mechanism for a class of structurally related compounds. The camptothecins were chosen as model drugs since extensive studies reveal differences in antitumor activity, metabolism, and transport with changes in structure. A simultaneous screen of six camptothecin analogues resulted in cytotoxicity profiles and orders of potency which were in agreement with the literature. We estimated the heritability for cytotoxic response to the camptothecins to be approximately 0.23. Nonparametric linkage analysis was used to identify a relationship between genetic markers and the response to camptothecins. An initial screen of the six camptothecin analogues revealed ten shared quantitative trait loci (QTL) on chromosomes 1, 3, 5, 6, 11, 12, 16 and 20. In a separate validation experiment with 3 additional camptothecins, nine of the ten QTLs were replicated. Subtle distinctions in significant and suggestive QTLs were also observed between drugs. These results provide a step towards streamlining the anticancer drug development process by simultaneously enabling phenotypic screening and identifying genes critical to drug action which impact patient sensitivity or toxicity.

## INTRODUCTION

Prior to the 1990s, the phenotypic based drug discovery approach dominated the pharmaceutical industry. In this approach, small molecules were screened against cells, tissues, or even whole organisms for their ability to enhance or suppress a specific phenotype desired in humans. The apparent advantages of this method over the existing target-based drug discovery paradigm have resulted in a renewed interest in phenotypic screening. One of the greatest advantages of this approach is that it enables the discovery of novel therapeutic targets for a disease. Compounds are screened for a biological effect rather than perturbation of a single molecular target, linking chemistry with biology and driving the serendipitous discovery of numerous structures with novel mechanisms of action (MOA).

Despite the recent revival in phenotypic screening, there are noteworthy limitations which can create a considerable bottleneck in the drug discovery process. Mechanism elucidation following the identification of biologically active compounds remains the most important weakness. A number of methods are being developed and optimized for mechanism elucidation; however, they are fraught with limitations which have been reviewed extensively elsewhere [1]. Since the typical phenotypic screening methods are unable to suggest key information about the mechanism of biologically active compounds, there is no way to distinguish between them other than by potency. Without a clear understanding of MOA, problems arise in lead optimization, drug safety, and efficacy. Structure activity relationship (SAR) studies for lead optimization become quite complicated with phenotypic screens. Binding to an unknown target can be influenced by cell absorption and transport, additional protein binding, secondary target interactions, drug metabolism, etc. These sites of compound loss can vary significantly within a series of structurally related compounds. Most current methods of mechanism elucidation are also unable to account for or convey changes in mechanism (ie primary and secondary targets) with changes

in structure. As a result SAR patterns become difficult to interpret and use during lead optimization. Finally, when mechanism is unclear, our ability to assess the risks of mechanism-based toxicity, side effects associated with secondary targets, or lapses in efficacy are also quite limited.

Genetic and genomic methods which screen all possible targets of compounds of interest are being developed to surmount issues associated with target identification following phenotypic screens. These methods which simultaneously screen compounds for a desired biological effect and provide information about molecular targets and SAR patterns are rising as powerful tools in drug discovery and development. Some of the most prominent examples of this approach use the budding yeast *Saccharomyces cerevisiae* [2, 3] or human cancer cell lines [4, 5] as *in vitro* model systems. In both cases, inconsistencies in data between humans and the model are a significant drawback. An ideal genomic strategy would investigate drug activity in a normal healthy human model. Recently, an *ex vivo* familial genetic strategy involving lymphoblastoid cells lines (LCLs) derived from Centre d'Etude du Polymorphisme Humain (CEPH) reference pedigrees was employed to quantify the impact of genetics on drug response and to identify quantitative trait loci (QTLs) harboring genes critical to drug action [6, 7]. Here we asked whether this *ex vivo* familial genetics model could be used to establish specific patterns of QTLs related to a shared mechanism for a class of structurally related compounds.

The camptothecins were chosen as a model class of compounds to investigate for a number of reasons (Figure 2-1). Intensive efforts in medicinal chemistry have led to the generation of a large number of camptothecin derivatives. Two of these, topotecan and irinotecan, are being used in the clinic as antitumor agents, and many are in preclinical and clinical development. In spite of the identification of a number of analogs with improved therapeutic activity, (intrinsic and acquired) resistance and toxicity remain major limitations to camptothecin therapy. While extensively studied, the mechanisms of resistance and toxicity remain unclear [8]. In addition,

though it is firmly established that the key molecular target of all of the camptothecins is Topoisomerase 1 (Top1), the post target interaction events responsible for antitumor activity are vague [9]. It is reasonable to suggest that a clearer understanding of the biochemical cascade associated with camptothecin cytotoxicity might lend answers to the questions surrounding mechanisms of activity, toxicity, and resistance. To this end, the CEPH model system was used to a) assess variation in response to the camptothecins across normal healthy human LCLs, b) evaluate the genetic contribution to variation in response and c) establish a pattern of multiple QTLs common to a class of compounds suggesting a shared mechanism of action.

## **MATERIALS AND METHODS**

**Cell lines.** One hundred twenty-five Epstein-Barr virus-immortalized lymphoblastoid cells derived from 14 CEPH reference pedigrees (35, 45, 1334, 1340, 1341, 1350, 1362, 1408, 1420, 1447, 1451, 1454, 1459) were purchased from Coriell Cell Repositories (Camden, New Jersey). Cells were maintained in RPMI medium 1640 (Invitrogen, Rockville, MD) supplemented with 15% fetal bovine serum, incubated in a 5% CO<sub>2</sub> atmosphere at 37°C, and passaged 2-3 times per week. Exponentially growing lymphoblastoid cell lines with greater than 85% viability, at passages 3-7 were used for experimentation.

**Drugs.** The following panel of camptothecins was purchased from LKT Labs (St Paul, MN): camptothecin (CPT), irinotecan (CPT11), 7-ethyl-10-hydroxycamptothecin (SN38), topotecan (TPT), 9-aminocamptothecin (9AC) and 9-nitrocamptothecin (9NC). Dr. Daniel Comins (North Carolina State University, Raleigh, NC) kindly provided 10-methoxycamptothecin (mCPT), 10-hydroxycamptothecin (hCPT), and 7-chlorocamptothecin (ClCPT). All camptothecins were prepared in 10 mM working solutions of DMSO (Sigma-Aldrich, St Louis, MO). Since camptothecins have a labile lactone form that exists in a pH dependent equilibrium with the

inactive carboxy form (present at basic pH), drugs were serially diluted in citrate-phosphate buffer at pH 3. Final concentrations of DMSO were 0.1% in all experiments.

**Cytotoxicity Profiling.** The conditions for drug preparation and phenotyping were optimized prior to this investigation (Appendix 1). The cytotoxic effect of the panel of camptothecins was determined by using the nontoxic colorimetric-based assay, alamar blue [6]. Plates (384 well, Corning, Corning, NY) were preloaded with vehicle (citrate-PBS, 0.1% DMSO), 10% DMSO, and increasing concentrations of each drug (n = 9 doses per drug). Each plate contained 6 replicates for each drug-dose combination. Cells were then plated at a density of 4000 cells in 45  $\mu$ l per well. Following 72 h incubation, 5  $\mu$ l alamar blue was added. Fluorescence was read at Ex 535nm and Em 595nm using a DTX880 plate reader (Beckman Coulter) at 96 h drug exposure. Raw fluorescence values for each set of replicates of a drug-dose combination were considered outliers if there was more than a ten-fold increase or decrease in the fluorescence signal of a single replicate. Cell viability (survival) relative to untreated controls was determined according to the manufacturer's protocol. The final percent survival at each concentration was averaged from six replicates of two independently plated experiments (n = 12). Additionally, growth rate in vehicle was calculated as previously described [10]. The IC<sub>50</sub> (the dose needed to inhibit cell viability by 50%), was calculated based on a sigmoidal dose-response curve using the nls package in R ([www.r-project.org](http://www.r-project.org)) [11].

**Hierarchical Clustering.** LogIC<sub>50</sub>s for each cell line-drug combination were z-score transformed prior to clustering. The data was loaded into Cluster 3.0 (<http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/>) and clustered using uncentered correlation and complete linkage. To stabilize clusters, a self organizing map (SOM) was calculated using 100,000 iterations for cell lines and 20,000 iterations for drugs. Clusters were visualized using Java TreeView.



**Heritability Analysis.** Heritability estimates of the proportion of variation in cytotoxic response due to inherited factors were calculated using variance components analysis using MERLIN 1.1.2 [12] (<http://www.sph.umich.edu/csg/abecasis/Merlin/index.html>). The degree of heritability associated with growth rate in vehicle was also calculated, and the heritability calculation for each drug-dose combination was adjusted using growth rate as a covariate in the variance components analysis [12].

**Genotype Data and Error Checking.** Genotype data for each cell line were downloaded from V10 of the CEPH database ([ftp://ftp.cephb.fr/ceph\\_genotype\\_db/ceph\\_db/Ver\\_10/mkr/](ftp://ftp.cephb.fr/ceph_genotype_db/ceph_db/Ver_10/mkr/)) [13] using error checked markers. Genetic map information was downloaded from the Marshfield database (<http://research.marshfieldclinic.org/genetics>) [14]. Error checking for Mendelian incompatibility, misspecified relationships and unlikely recombinations was performed, as previously described [14]. A combined total of 8269 single nucleotide polymorphisms (SNPs) and microsatellite markers were used for linkage analysis.

**Linkage Analysis.** Drug-dose combinations were considered the phenotypes of interest for linkage analysis ( $n = 54$ ). For each phenotype, non-parametric linkage analysis was performed using MERLIN which constructs a likelihood ratio test for linkage based on inheritance vectors. For quantitative traits, scores used to calculate the likelihood ratio test are defined as follows:

$$S(v) = \sum \text{founder alleles } S_{\text{allele}}(v)^2,$$

$$S_{\text{allele}}(v) = \sum \text{all carriers of allele } (y_i - \mu),$$

where  $S(v)$  is the score for each inheritance vector,  $S_{\text{allele}}(v)$  is the score for each founder allele,  $y_i$  is the phenotype for each individual,  $\mu$  is the mean phenotype for the population, and  $v$  is the list of individuals who carry a specific founder allele such that the score for each inheritance vector is the summation of the squared score for each founder allele, and the score for each found

allele is the sum of square deviation from all individuals that carry that allele. For each phenotype of interest, QTL maps were generated by displaying the logarithm of odds scores from the likelihood ratio tests across each chromosome.

**Peak Identification, Prioritization, and Replication.** Guidelines for interpreting LOD scores have suggested viewing LOD scores of 2.2 as suggestive and 3.6 as significant [15]. However, since such a categorization is inexact, the data in this study was used to dictate at which threshold results would no longer be considered due to chance and most likely occur as a result of linkage. To estimate the probability of obtaining false-positive evidence of linkage for each drug-dose combination under the null hypothesis of no linkage to observed phenotypes, gene-dropping permutations were conducted using Merlin [12]. Marker data were simulated under the null hypothesis of no linkage or association to the observed phenotypes while retaining the same pedigree structures, maps, marker allele frequencies, and missing data patterns. Ten thousand replicates were simulated for each of the 54 phenotypes, resulting in a total of 54,000 simulated datasets. Nonparametric linkage analysis was conducted as described above for each replicate set. Based on these simulations, permutation distributions were generated across the chromosomes for each phenotype and then used to determine genome-wide LOD score cut-offs for statistical significance corresponding to p-values less than or equal to .05 for each phenotype. Additionally, cutoffs for suggestive linkage were determined for each drug-dose combination for an alpha of 0.05 for each chromosome. The start and end of a peak was defined as the region with LOD scores that were above the suggestive threshold or LOD minus one, whichever was greater. QTLs observed for a drug-dose phenotype were considered significant if the highest LOD score in that region surpassed the LOD score threshold for significance for that drug-dose phenotype. QTLs observed for a drug-dose phenotype were considered suggestive if the highest LOD score in that region surpassed the suggestive LOD score threshold for that drug-dose phenotype on that chromosome. For all statistically significant QTLs, the total number of replications was

calculated. Across drug-dose phenotypes, the number of replications was calculated as the total number of drug-dose phenotypes that demonstrate evidence of linkage at any point of the peak regions at either the significant or suggestive level. Table A2-1 contains a list of QTLs identified as significant for camptothecin drug-dose phenotypes.

## RESULTS

**Variation in Camptothecin-Induced Cytotoxicity.** Sensitivity to the camptothecins was assessed in 125 lymphoblastoid cell lines derived from 14 CEPH pedigrees. Cells were exposed to increasing concentrations (9 doses) of each camptothecin for 96 h and cell viability relative to vehicle control was determined. Variation in cytotoxic response to each camptothecin within and between the CEPH pedigrees was observed (Figure 2-2). For example, 9AC, which had the widest range of IC50s, had a population mean IC50 of 93 nM and the IC50 ranged from 7 nM to 4  $\mu$ M. Both the order of potency and IC50s in the CEPH cell lines are consistent with literature values in cancer cell lines such as the NCI60 cell line panel (NCI Developmental Therapeutics Human Tumor Cell Line Screening data, <http://dtp.nci.nih.gov/dtpstandard/InvivoSummary/index.jsp> [16]).

This data was also used to identify individuals and/or families which were hypersensitive or resistant to the camptothecins. Further genetic and genomic studies with these individuals might lend insight into mechanisms of activity and resistance. A hierarchical clustering analysis of z-score transformed logIC50 values (where IC50 is the concentration required to inhibit viability by 50%) was performed keeping family structure intact (Figure 2-3). The clusters matched the overall potency (SN38>CPT>9NC>TPT>9AC>CPT11) in the cell lines studied. CPT11 is most divergent from the other camptothecins studied. Since CPT11 is the prodrug of SN38 and requires submicromolar concentrations for effective cell kill, IC50s across the panel of CEPH cell lines are considerably higher for CPT11 than other camptothecins investigated. Of note, there are

individuals who are sensitive to some but not all camptothecins and whole families which are resistant or sensitive to all camptothecins. For example, pedigree 1408 appears resistant to all camptothecins with the exception of 9AC. All but two members of pedigree 1362 are sensitive to all camptothecins; two offspring (11982 and 11983) are resistant to all camptothecins.

**Heritability Analysis.** Heritability was estimated to quantify the impact of genetic factors on the cytotoxic response to each of the camptothecins at each dose. There is a known correlation between cellular sensitivity to many chemotherapeutic agents and growth rate [10, 17]. As a control, heritability was calculated for growth rate in the presence of vehicle. The heritability estimate for growth rate was low (1.60%) which suggests that environmental factors play a much larger role than genetics in growth rate [17]. For each drug, the growth-rate adjusted heritability estimates at each dose are featured in Figure 2-2. Heritability estimates at the asymptotes of the sigmoidal dose-response curve are low as there is little to no variability in cytotoxic response at these points. For all camptothecins studied heritability estimates averaged  $23.1 \pm 2.6$  % for doses within the linear phase of the sigmoid curve. Since heritability estimates were approximately 20% for all camptothecins this reinforces the idea that inherited genetic variation is an important determinant of the cytotoxic response to camptothecins. The heritability associated with the cytotoxicity of these compounds is analogous to heritabilities reported for other common human phenotypes such as systolic and diastolic blood pressures [18], and for the cytotoxic response to daunorubicin in CEPH cell lines [19].

**Genome-wide Linkage Analysis.** Nonparametric linkage analysis was performed using mean cell viability at each dose for each drug, which is referred to as the drug-dose phenotype. For each drug-dose phenotype statistically significant LOD score thresholds corresponding to a genome-wide p-values less than or equal to 0.05 were determined using gene-dropping permutations under the null hypothesis that no linkage exists. Regions of the genome referred to as quantitative trait loci (QTLs) were considered significant if the highest LOD score in the region was greater than

or equal to the predetermined LOD score cut-offs for each drug-dose combination on a given chromosome. The mean LOD score cut-off across all phenotypes and chromosomes indicating significant linkage was 1.37 (range: 0.83-1.72). Additionally, cutoffs for suggestive linkage were determined for each drug-dose combination for an alpha of 0.05 for each chromosome. A region identified as significant to one drug-dose phenotype was considered replicated in another drug-dose phenotype if the maximum LOD score in that region surpassed the suggestive LOD score threshold. The mean LOD score cut-off across all phenotypes and chromosomes indicating suggestive linkage was 0.59 (range: 0.41-0.72).

To establish a pattern of QTLs significant to a class of compounds, regions of the genome which were overrepresented across the camptothecins were examined. Figure 2-4 illustrates significant and suggestive QTLs identified in one camptothecin which were replicated in other camptothecins. The results of a sign test ( $p < 0.5$ ) indicated there was a significant overrepresentation of overlapping QTLs compared to the null hypothesis that QTLs were randomly distributed across the genome amongst all drugs. Ten linkage peaks were identified as significant in a given drug-dose combination and replicated in all of the camptothecins at a number of concentrations (Table 2-1). This implies that the same linkage regions influence the cytotoxic response to all camptothecins over a broad range of concentrations. The highest LOD score with genomic significance (2.13) was observed with the 8.0 nM SN38 phenotype and was located on chromosome 20 between 42 and 101 cM (20p12.1-20q13.32), and presumably associated with Top1 (20q12-q13.1) the primary target of the camptothecins. All camptothecin analogues studied (at multiple dosages for each drug) had a peak at chromosome 20 centered around 49 cM (Table 2-1, Figure 2-5). Unlike the other significant linkage peaks, the QTL on chromosome 6 from 0 to 29 cM is only associated with high concentrations of the camptothecins, resulting in greater than 80% growth inhibition.

Moreover, notable distinctions between significant QTLs associated with camptothecin analogues have been observed and are summarized in Figure 2-4. For example, TPT is the only camptothecin with a linkage peak extending from 0 to 19.6 cM on chromosome 13 (LOD = 1.365). Interestingly, 9NC is considered to be the prodrug of 9AC and there is one linkage peak which was identified exclusively in these drugs on chromosome 5 [20]. Chromosome 1 has two QTLs centered at 70 and 129 cM respectively which are shared exclusively by camptothecins possessing a nitrogen bearing substituent on carbon 9: 9AC, 9NC, and TPT. No peaks were identified which were unique solely to SN38 and its prodrug CPT11. However, a QTL on chromosome 4 is only present in CPT11 and 9AC. Regions suggested to influence the cytotoxic response to CPT11 were not always replicated in SN38 or vice versa. This was also observed for 9AC and 9NC. This is unsurprising since for example, the prodrug CPT11 must undergo activation by carboxylesterases (CESs) to the active SN38 and SN38 is not subsequently metabolized by CES. Only suggestive QTLs for CPT11 were located on chromosome 16 from 1-69 cM; CES1 and CES2 are centered around 73 cM on chromosome 16. Finally, to compare the overall QTL patterns a similarity matrix was constructed using a binary assessment of peaks present at either the significant or suggestive level for each camptothecin. R squared correlations are bound by 0 and 1 and the greater the value the more related the patterns are to each other (Table 2). The majority of the correlations are above 0.5, indicating a strong association between overall QTL patterns and similar mechanisms of action. The highest correlations (highest degree of similarity) are between the 9AC and 9NC, CPT11 and CPT, and CPT11 and SN38. While the biological profile of CPT11 appears different from the remaining camptothecins, the genomic profile of TPT appears most distinct.

**Independent Validation of Shared QTLs.** We next asked whether QTLs identified as shared among all camptothecins could be replicated independently. In a separate phenotyping experiment, the same 14 CEPH pedigrees were exposed to a dosing spectrum of a second set of

camptothecins: mCPT, hCPT, and CICPT. Variation in sensitivity to this set of camptothecins was then used to calculate heritability estimates at each drug-dose phenotype. Heritability estimates were highest in the linear phase of the sigmoid curve. Growth rate adjusted heritability estimates for doses within the sigmoid curve for mCPT, hCPT, and CICPT were comparable to estimates for the first set of camptothecin analogues. The highest heritability estimates for mCPT, hCPT, and CICPT were 20.2%, 18.7%, and 20.7% respectively. Linkage analysis, peak prioritization, and peak replication assessment were repeated with this second set of camptothecins. Nine of the ten QTLs identified as characteristic of camptothecin activity were subsequently validated in multiple doses of mCPT, hCPT, and CICPT (Figure 2-4). While no concentrations of mCPT or hCPT possessed the shared QTL on chromosome 6 from 0-29 cM, seven of the nine doses of CICPT possessed this shared QTL. Variation in response across the broad dosing spectrum for mCPT, hCPT, and CICPT was not linked to the QTL on chromosome 12 from 0-6 cM.

**Comparison to Topoisomerase 2 Inhibitors.** To illustrate class specific patterns could be established, the same cell lines were phenotyped for sensitivity to the Topoisomerase 2 (Top2) inhibitors, etoposide and teniposide. Genetics plays a greater role in cytotoxic response to the Top2 inhibitors compared to Top1 inhibitors. The maximum heritability estimates for a Top1 inhibitor (TPT) was 25.9% compared to 42.4 and 32.9% for etoposide and teniposide respectively. When comparing cytotoxic response to the camptothecins across the entire population of CEPH cell lines IC50s were used to visualize patterns of sensitivity and resistance. We chose another mode of comparison between the Top1 and Top2 inhibitors since IC50s could not be obtained for more than 80% of the cell lines treated with teniposide. Hierarchical clustering using the dose which yields a population mean viability closest to 50%, ( $\overline{GI50}$ ), for each drug reveals that overall patterns of sensitivity and resistance between the Top1 and Top2 inhibitors are indeed distinct and form two clusters corresponding to differences in mechanism (Figure 2-6).

The same is true for the dose which yields a population mean viability of 40 and 60% (data not shown). Nonparametric linkage analysis was performed using cell viability at each drug-dose combination of the Top2 inhibitors. Four QTLs present on chromosomes 6 (32-113 cM), 12 (9-30 cM), 13 (0-25 cM), and 18 (58-76 cM) were identified as significant and replicated (considered replicated if  $LOD > \text{suggestive threshold}$ ) in both Top2 inhibitors at multiple dosages. This pattern of QTLs for the Top2 inhibitors was quite distinct from those established for the camptothecins (Table 2-1).

## **DISCUSSION**

Early models for chemogenomic studies have used cancer cell lines [4, 5], mutant yeast strains [2, 3], and rodents [21, 22]. The biggest limitation with these systems is that data does not always correlate to humans. For example, some mammalian targets are absent in yeast and vice versa. Targets which produce a desired phenotype in rodents may not exhibit the same phenotype in man [23]. In addition, cancer cell lines can differ morphologically and genetically from primary tissues [24].

This is one of the first genomic studies to use a healthy human cell line model to identify class specific pharmacological and genomic profiles. While cancer cell line panels such as the NCI60 are prepared from 4-5 cell lines of a given tissue origin, this study uses a large collection of cell lines of the same type. Moreover, just as genetic heterogeneity across the cancer cell lines has been used to stratify compounds by mechanistic class, natural genetic variation in the CEPH cell lines can be used to identify a class-specific profile for the camptothecins [4, 25]. In fact, heritability analysis demonstrates that  $23.1 \pm 2.6$  % of human variation in sensitivity to the camptothecins is due to a genetic component. Not only were heritability estimates replicated across multiple doses, but they were replicated across multiple camptothecin analogues and experiments.



Using this system to investigate compounds within a structural class and sharing the same mechanism one would expect a pattern of QTLs related to the cytotoxic activity of all compounds within that class. Indeed, ten QTLs across eight chromosomes were replicated in the first six camptothecin analogues studied suggesting a pattern of QTLs associated with a general and shared mechanism of action. We consider the fact that these QTLs were replicated across multiple analogues and doses within the first screen a form of internal validation. In a separate phenotyping experiment using three additional camptothecins, nine of those ten QTLs were again independently replicated. The remaining QTL on chromosome 12 shared by all camptothecins in the initial screen was only present in CICPT, but present in seven of the nine doses. Replication of QTLs both across multiple drug-dose combinations within the first screen and multiple drug-dose combinations within a second screen of different camptothecins is very exciting. Finally, both the biological and genomic profiles generated in CEPH for the camptothecins and the Topoisomerase 2 inhibitors, etoposide and teniposide were very distinct. Hierarchical clustering on biological data generated two clusters in agreement with the two distinct mechanisms of action. Moreover, the overall pattern of shared QTLs differed significantly between the two groups; no QTLs were present in the same regions for the two classes.

Figure 2-4 highlights regions which might contain genes that contribute to the cytotoxic activity of all of the camptothecins. There are more than 4000 candidate genes for follow-up under the nine QTLs shared by all nine camptothecin analogues alone. Identifying which of these genes are critical to camptothecin-induced cytotoxicity can be a challenging and time-consuming process. To maximize success, a tiered approach is recommended when choosing QTLs for further investigation. QTLs shared by all nine camptothecins are considered the most promising (Table 2-1). QTLs shared by the first set of six camptothecins should be investigated next, followed by the QTLs identified as significant and shared by all three camptothecins in the validation set. Those significant QTLs which have been identified as unique to 1 or more drugs

but are not replicated even at the suggestive level in all camptothecins should be considered next. Examples of this class include the QTL on chromosome 1 at 145-168 cM which is present at the significance level in multiple concentrations of 9AC, and at the suggestive level in 1-2 doses of 9NC and TPT and the linkage peak on chromosome 13 (0-19 cM) that is observed solely with the 10 nM TPT phenotype. Finally, since the average LOD score threshold for a suggestive QTL is 0.59, suggestive QTLs present in all 9 camptothecins at multiple doses should be pursued last.

Using these prioritization criteria, the QTL on chromosome 20 which is common to all camptothecins is considered the most important for follow-up investigations. There are 453 protein coding genes within this region of the genome. The functional annotation clustering tool from the web-accessible program Database for Annotation, Visualization, and Integrated Discovery (DAVID) was applied to identify over-represented gene ontology terms (GO) and KEGG pathways within this gene list (Table 2-3) [26, 27]. Out of the over-represented biological processes in this gene list, five of the top eight biological processes may be related to camptothecin activity: response to stimulus, bioregulation, protein binding, cell differentiation, and regulation of cell growth. Table A2-2 lists the genes related to these GO terms that are present under the QTL on chromosome 20. Association studies could be used to fine map QTLs and pinpoint genes associated with drug response; however, limited statistical power prevents us from doing so here. The presence of Top1, the sole molecular target of all camptothecins, in this region is encouraging. Top1 expression levels have previously been correlated with cellular sensitivity to the camptothecins; low levels of Top1 confer resistance to cancer cell lines such as lymphomas [9]. Smirnov et al. performed microarray experiments to measure human gene expression levels in CEPH [28] (data accessible at NCBI GEO database [29], accession GSE12626). Baseline measures of Top1 gene expression varied as much as 2 fold in this dataset. (Limited overlap between cell lines used in the studies prevented direction association analysis in the current study.) Admittedly, since linkage analysis produced a broad QTL spanning hundreds

of genes, it cannot be assumed that a single gene under this QTL is influencing the activity of these compounds. Bcl-xl, is another promising gene within this region. Down-regulation of Bcl-xl, which inhibits apoptosis, has been shown to enhance cytotoxic response to the camptothecins [30, 31].

Unlike the other QTLs which are shared between the camptothecins, the peak on chr12 (0-6 cM) is relatively small and consists of only 24 protein coding genes. The most intriguing genes in this region are FBXL14 and RAD52. FBXL14 is a member of the F-box protein family. Proteasomal degradation of Top1 has been implicated as a major pathway in the repair of Top1 mediated DNA damage. Resistance to camptothecin can occur following overexpression of a component of the SCF complex (SCF components: Skp, Cullin, F-box) leading to increased ubiquitination and proteasomal degradation of Top I [32, 33]. RAD52 is involved in DNA double-strand break repair and homologous recombination and hypersensitizes cells to camptothecins [34].

Observing significant or suggestive LOD scores for a given drug across a number of doses has been previously reported as replication and suggestive of a shared genetic component contributing to the cytotoxic effect at all concentrations [6, 19]. The same regions of interest were not identified as significant or suggestive for all drug-dose combinations of the camptothecins. In fact, some QTLs were apparent only in the higher concentrations of the camptothecins. One plausible explanation for changes in patterns of observed QTLs with differences in dose might be different mechanisms of action predominating at different concentrations. The cytotoxic effect of camptothecins in cancer cells and yeast are typically related to replication-mediated DNA lesions. However, protection by aphidicolin, a DNA polymerases inhibitor associated with replication, is only apparent at the lowest concentrations of camptothecin (submicromolar) in cancer cell lines [35]. When DNA replication is blocked, cell death at higher concentrations of camptothecin is suggested to occur by transcription-mediated DNA lesions. It has been reported that the anticancer activity of the camptothecins can switch from replication-dependent to transcription-

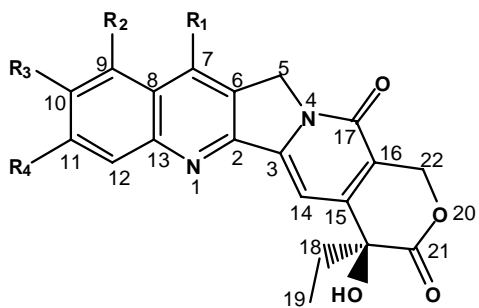
dependent solely at higher concentrations in normal lymphoblasts and other highly proliferative cell lines [9]. Also different DNA repair, cell cycle checkpoint, and cell-death signaling pathways have been implicated following DNA damage at different doses (Figure 2-7). These dose dependent effects have been associated with differences in gene expression profiles [36] and cell cycle responses [37]. For example, low doses of camptothecin result in reversible G2 delay while high doses cause S-phase delay and G2 arrest [37]. Without a doubt, there a number of complex mechanisms associated with the cytotoxic activity of the camptothecins that can occur simultaneously or selectively given certain intracellular conditions [38]. Work is ongoing to identify the conditions that dictate which pathways are preferred and why.

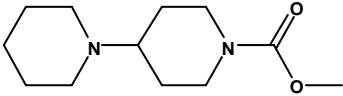
Future studies with this dataset should begin to ascertain which genes under shared QTLs are influencing the cytotoxic activity of the camptothecins. Recently, RNA interference (RNAi) screens in model organisms and human cells have successfully identified genes that modulate cell growth, apoptosis, chemoresistance, and chemosensitivity [39-43]. Large scale RNAi in the form of high throughput screens using small interfering RNAs (siRNA) can be used to systematically screen all genes under a QTL of interest. Known and novel genes whose loss of function confers decreased or increased sensitivity to the camptothecins can be identified.

We have demonstrated that specific patterns of biological response and QTLs could be established for a class of structurally related compounds sharing a mechanism of action. Most importantly, we have identified a set of QTLs associated with the sensitivity to the camptothecins, validated those QTLs internally, and with a second replication set. Future studies should compare the genomic profiles of the camptothecins with other structurally unrelated Top1 inhibitors and ascertain which if any shared QTLs are the result of the general mechanism of Top1 inhibition. Moreover, as the ultimate goal of this research is correlate biological response to genes involved in drug action, work is needed to pinpoint the genes under these QTLs which are influencing response. Thousands of genes are present in the six QTLs shared by all of the camptothecins.

RNAi mediated inhibition of gene expression can be used to prove the contribution of a gene to these QTLs. Taken together, these results lay the groundwork for using the *ex vivo* familial genetic strategy in CEPH cell lines for drug discovery efforts.

Figure 2-1. Camptothecin analogues.



Name	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>
Camptothecin (CPT)	H	H	H	H
Topotecan (TPT)	H	CH <sub>2</sub> NH(CH <sub>3</sub> ) <sub>2</sub>	OH	H
Irinotecan (CPT11)	C <sub>2</sub> H <sub>5</sub>	H		H
SN38	C <sub>2</sub> H <sub>5</sub>	H	OH	H
IDEC-132 (9AC)	H	NH <sub>2</sub>	H	H
Rubitecan (9NC)	H	NO <sub>2</sub>	H	H

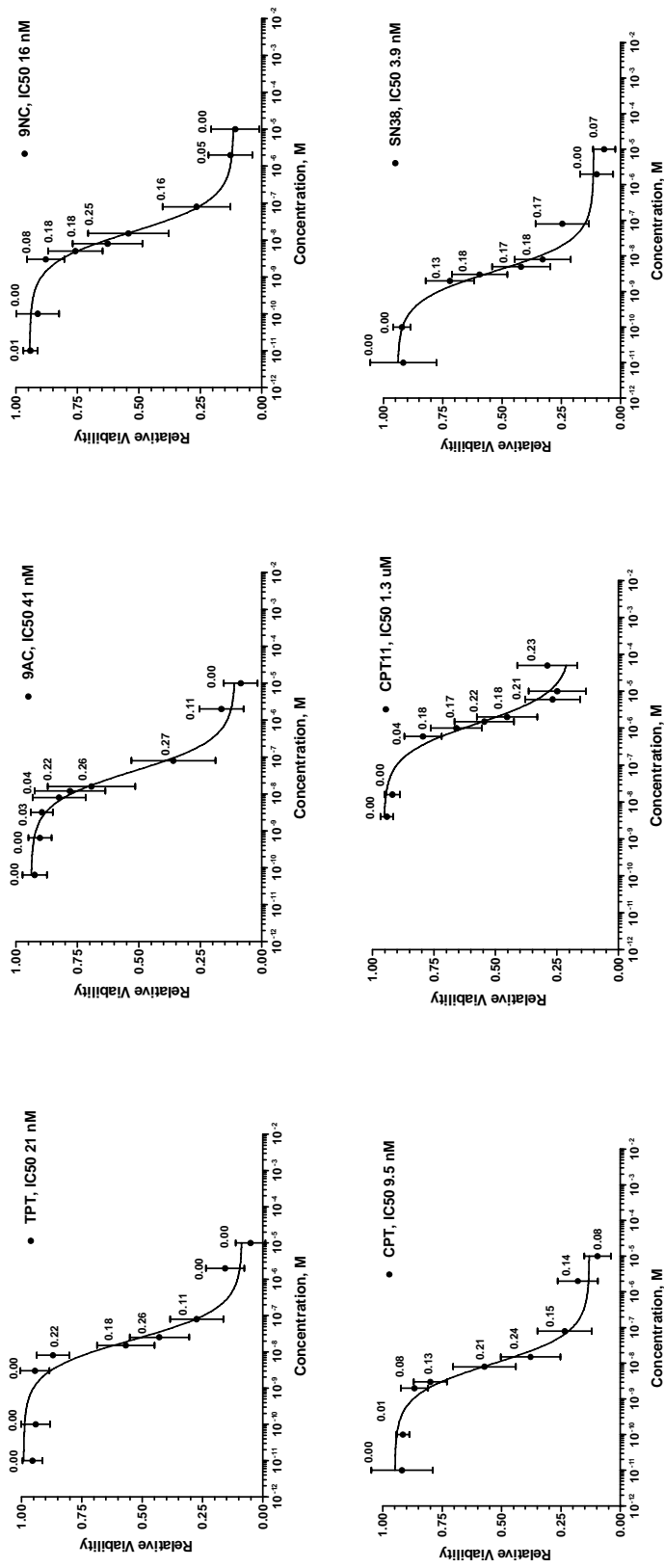


Figure 2-2. Dose-response curves and heritability estimates for the camptothecins. Data points represent the overall population mean (n=126) for viability relative to untreated controls at each dose. Vertical bars represent the standard deviation for cell viability across the population. Numbers are the growth-rate adjusted heritability estimates for each concentration. IC50 represents overall population IC50.

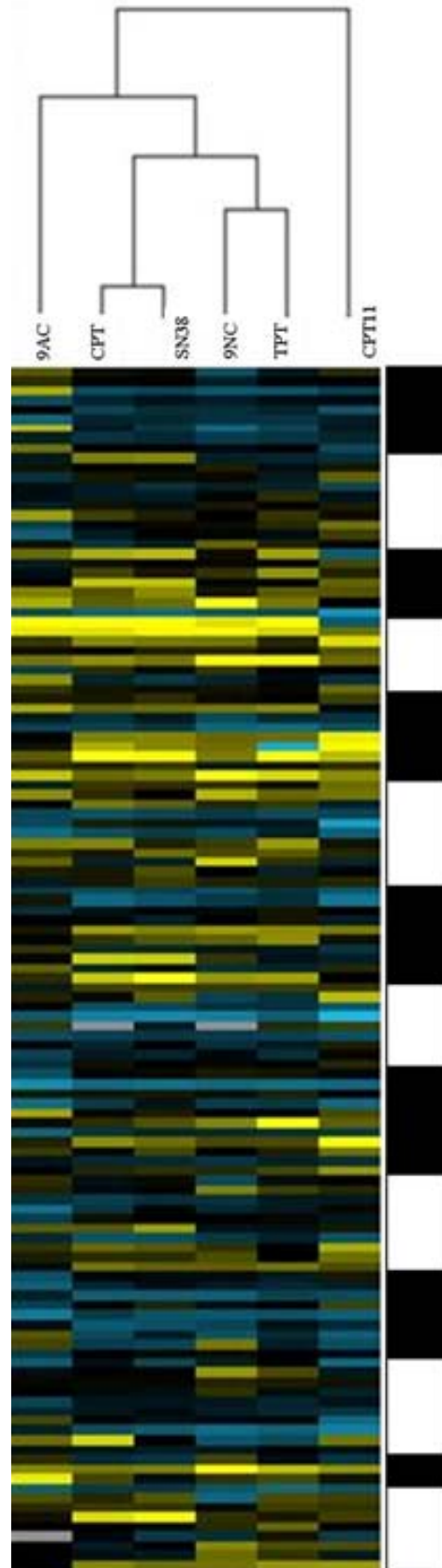


Figure 2-3. Hierarchical clustering of CEPH cell lines by camptothecin logIC50s. LogIC50s were z-score transformed. Clustering based on drugs holding family structure intact (black and white bar indicates each of 14 families). Yellow indicates positive Z-scores (resistance), blue indicates negative Z-scores (sensitive), black indicates Z-score = 0 (median resistance value). The brighter the color the greater the value from 0, with max



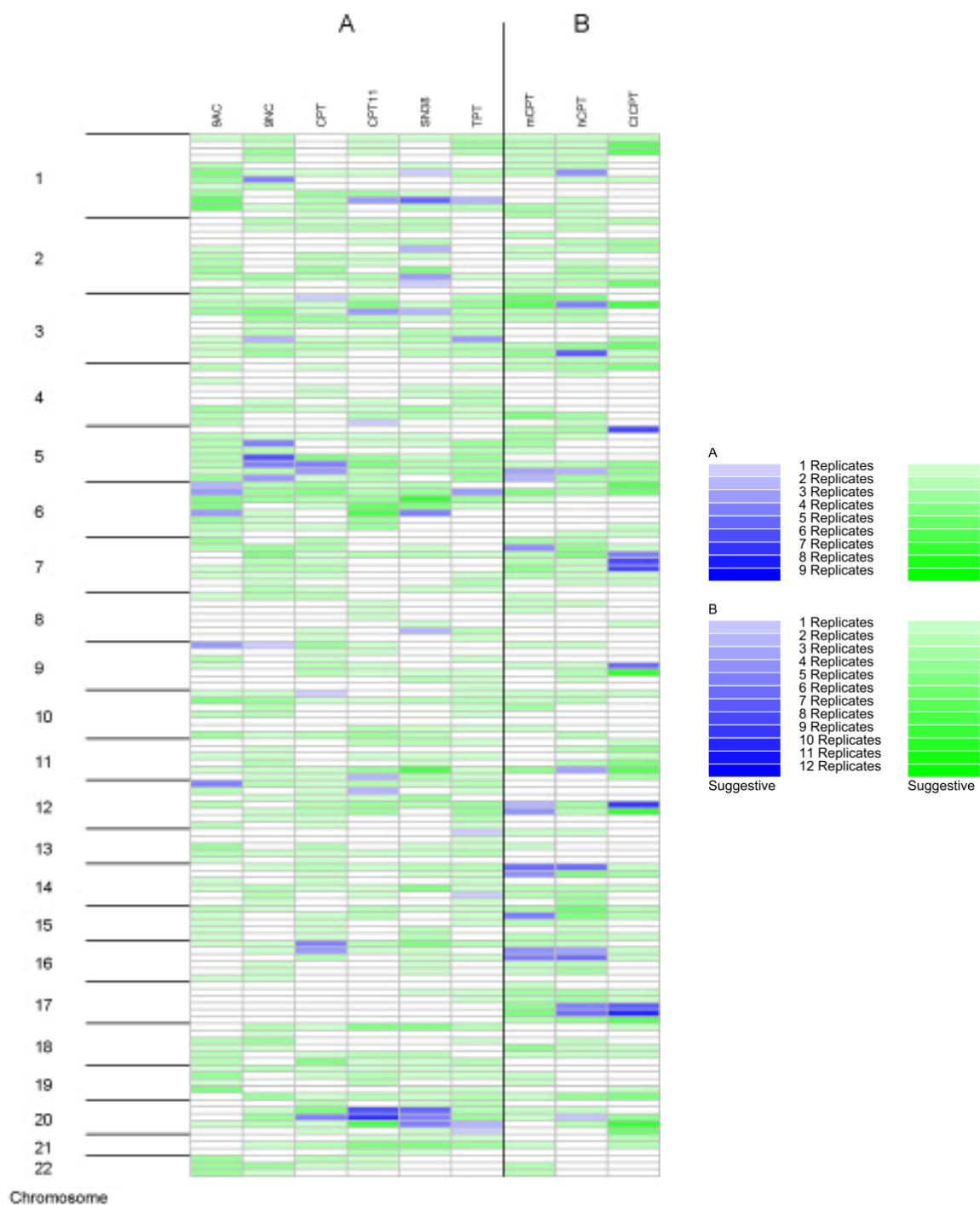


Figure 2-4. Genome wide pattern of QTLs for camptothecin analogues. Each chromosome was partitioned into 25cM regions. Each drug-dose combination that resulted in a significant QTL (LOD > significance threshold) is indicated in blue. Intensity of the shading indicates the number of doses replicating that QTL at either the suggestive or significant level. Regions which also had a suggestive QTL (LOD > suggestive threshold) are indicated in green with color intensity referring to the number of doses replicating this peak.

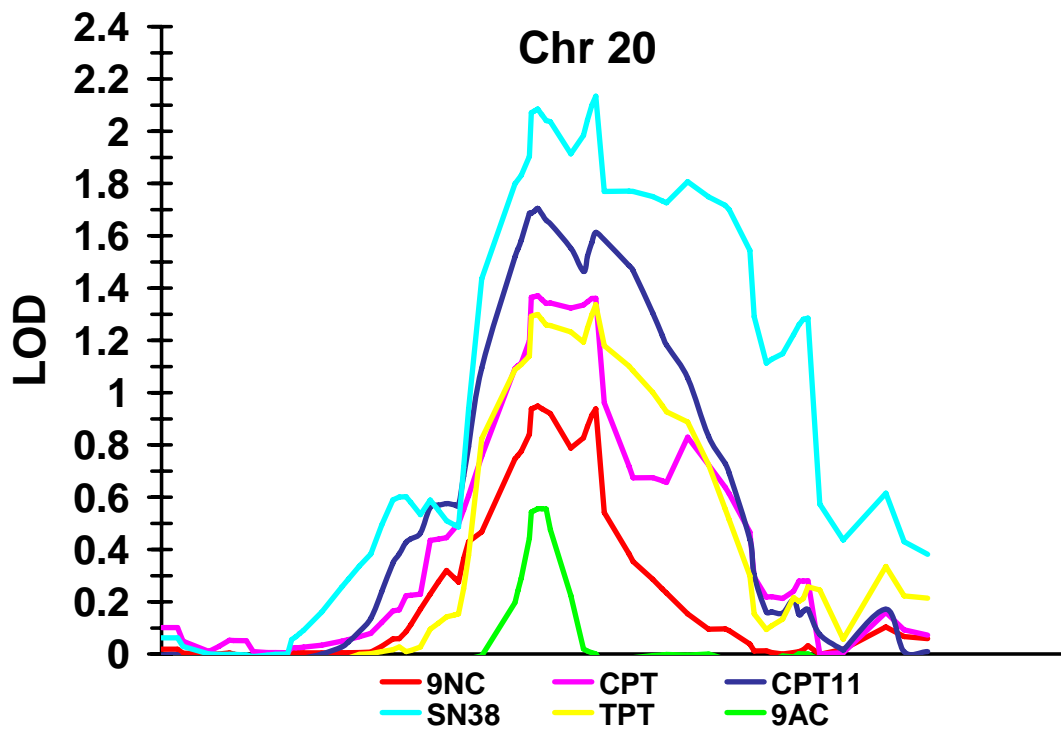


Figure 2-5. QTL shared across all camptothecins on chromosome 20. The QTL on chromosome 20 contains the gene for Top1, the sole molecular target of all camptothecins. Each drug is represented by a different color. Multiple doses for each drug were identified as significant and suggestive at this location. The drug-dose combinations with the highest LOD score are represented here.

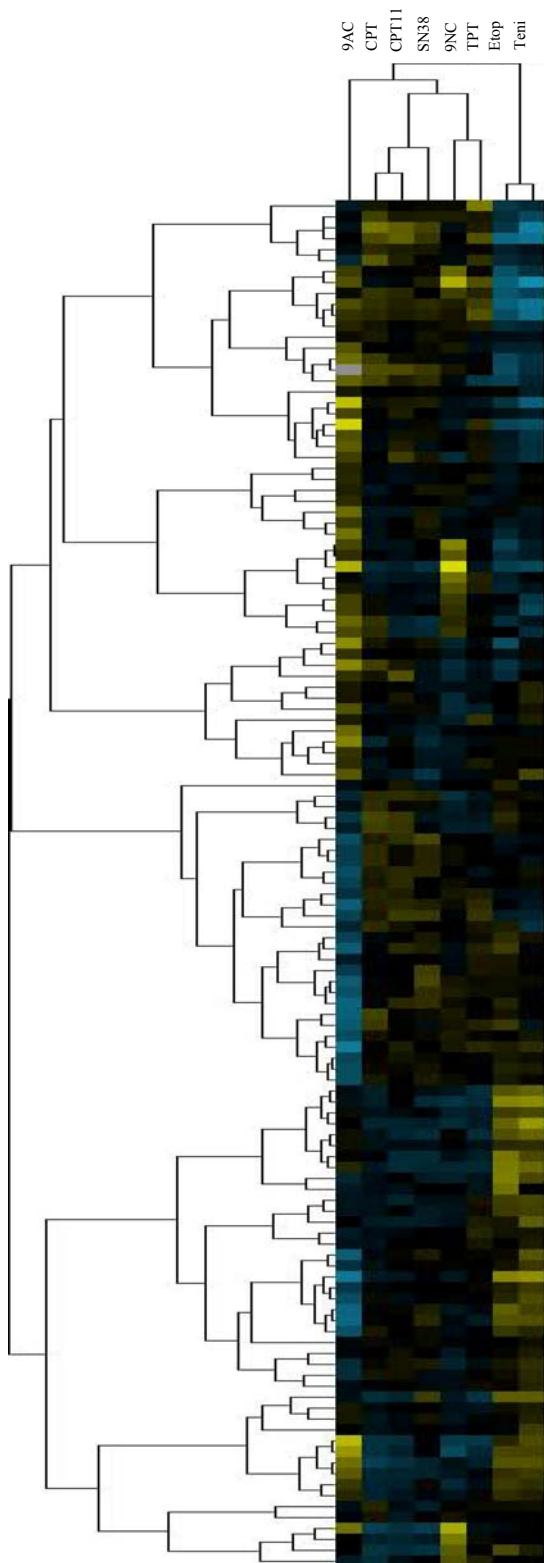


Figure 2-6. Hierarchical clustering of Top1 and Top2 inhibitors based on cell viability at the  $GI_{50}$ , the dose which yields a population mean closest to 50%. Clustering was done on both drugs and cell lines.

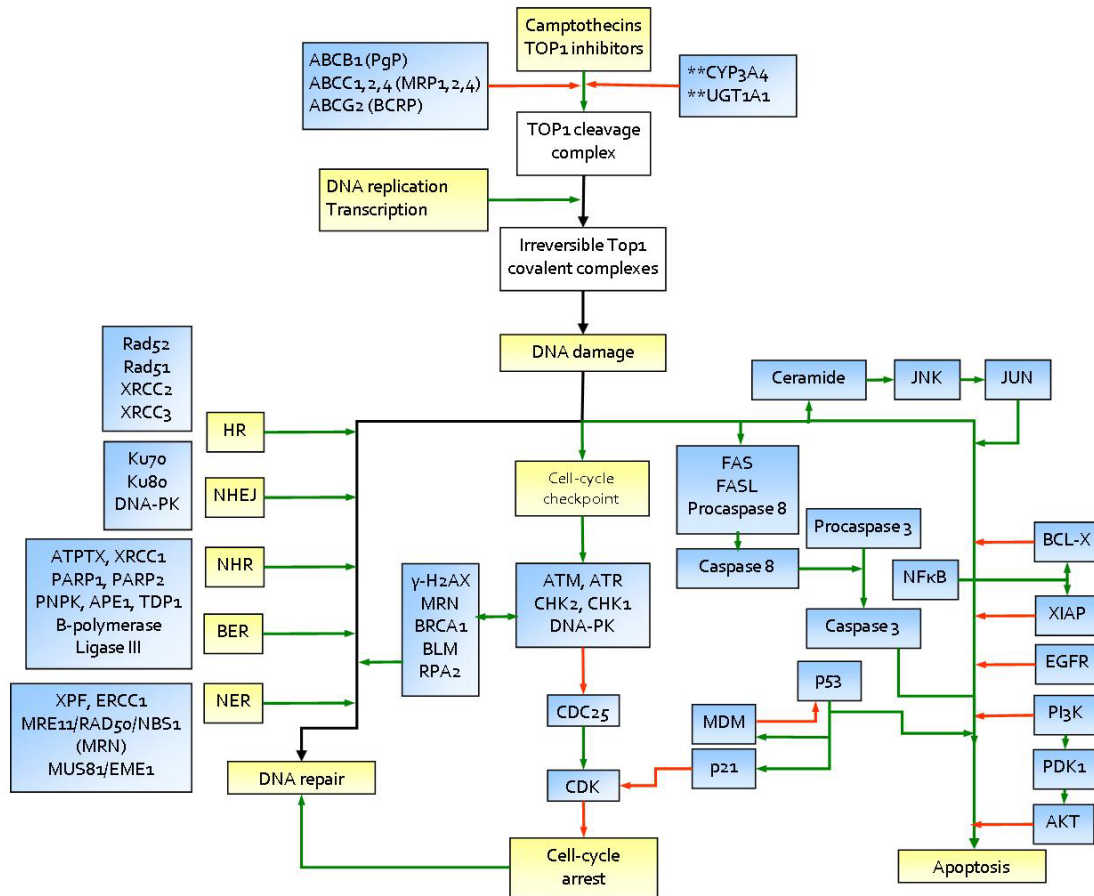


Figure 2-7. Molecular pathways involved in cytotoxic response to camptothecins. Green lines refer to activation, red lines to inhibition and black lines to conversion. Molecular processes are indicated in yellow boxes. Proteins implicated in cytotoxic response to the camptothecins are in blue boxes. Camptothecins are actively pumped out of cells by the ATP-binding cassette transporters Pgp and MRP. Several camptothecin analogues are actively metabolized by CYP3A. Only irinotecan is metabolized to the active SN38 by UGT1A1. Reversible Top1 cleavage complexes are induced by camptothecins and DNA damage. These cleavage complexes are converted to irreversible TOP1 covalent complexes by replication and transcription. DNA damage resulting from irreversible TOP1 covalent complexes can be repaired, induce cell-cycle arrest, or apoptosis. DNA repair involves five pathways: base excision repair (BER), homologous recombination (HR), non-homologous end joining (NHEJ), non-homologous recombination (NHR), and nucleotide excision repair (NER). Cell-cycle arrest involves checkpoint kinases such as ataxia telangiectasia mutated (ATM), ataxia telangiectasia and RAD3 (ATR), and DNA-dependent protein kinase (DNA-PK). Regulators of the checkpoint kinase pathway include histone H2AX, RPA2, MRE11–RAD50–NBS1 (MRN complex) and BRCA1. The Bloom syndrome helicases (BLM)–TOP3 $\alpha$  pathway suppresses homologous recombination when activated by the checkpoint kinase pathway. Cell-cycle arrest facilitates DNA repair. p53 activates apoptosis both directly and by transactivating pro-apoptotic genes. TOP1 inhibitors tend to suppress apoptosis by activating nuclear factor  $\kappa$ B (NF $\kappa$ B) which increases the expression of anti-apoptotic genes such as XIAP (chromosome X-linked inhibitor of apoptosis) and BCL-X. Blocking of the PI3K/Akt pathway enhances apoptosis induced by camptothecins. Camptothecins also increase de novo synthesis of FAS/FAS ligand and ceramide which leads to apoptosis. *Modified from [9]*

Table 2-1. QTLs shared by camptothecins.

Chromosome	Peak Start (cM)	Peak End (cM)	LOD*
1 <sup>†</sup>	229	252	1.855
3 <sup>†</sup>	36	82	1.682
3 <sup>†</sup>	137	180	1.638
5 <sup>†</sup>	104	194	1.709
6 <sup>†</sup>	0	29	1.528
6 <sup>†</sup>	42	65	1.652
11 <sup>†</sup>	115	131	1.352
12	0	6	1.705
16 <sup>†</sup>	0	75	1.345
20 <sup>†</sup>	42	101	2.134

\*Maximum LOD score observed in region.

<sup>†</sup>QTLs which were also present in the second set of camptothecins

Table 2-2. Similarity matrix of overall QTL patterns for each camptothecin.

	9AC	9NC	CPT	CPT11	SN38	TPT
9AC	1.000	0.724	0.517	0.483	0.552	0.483
9NC		1.000	0.517	0.483	0.621	0.483
CPT			1.000	0.759	0.690	0.448
CPT11				1.000	0.655	0.517
SN38					1.000	0.586
TPT						1.000

Constructed from binary evaluation of presence of significant and suggestive QTLs.

Table 2-3. Top overrepresented GO terms for the QTL on chromosome 20 (42-101 cM).

Over-represented biological processes	ES
Enzyme inhibitor activity	8.24
<u>Response to stimulus</u>	3.61
Response to chemical stimulus	
Response to drug	
Response to DNA damage	
Embryonic development	2.33
<u>Biological regulation</u>	2.11
Nucleic acid binding	
DNA binding	
Transcription	
Gene expression	
Protein binding	1.93
Reproduction	1.53
<u>Cell differentiation</u>	1.24
Apoptosis	
Programmed cell death	
Cell cycle	
Regulation of cell growth	0.74

ES: Enrichment Score. Underlined terms are parent terms of related GO terms beneath.

## REFERENCES

1. Gresham, V. and H.L. McLeod, *Genomics: Applications in mechanism elucidation*. Advanced Drug Delivery Reviews, 2009. **61**(5): p. 369-374.
2. Baetz, K., et al., *Yeast genome-wide drug-induced haploinsufficiency screen to determine drug mode of action*. Proc Natl Acad Sci U S A, 2004. **101**(13): p. 4525-30.
3. Parsons, A.B., et al., *Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast*. Cell, 2006. **126**(3): p. 611-25.
4. Scherf, U., et al., *A gene expression database for the molecular pharmacology of cancer*. Nat Genet, 2000. **24**(3): p. 236-44.
5. Blower, P.E., et al., *Pharmacogenomic analysis: correlating molecular substructure classes with microarray gene expression data*. Pharmacogenomics J, 2002. **2**(4): p. 259-71.
6. Watters, J.W., et al., *Genome-wide discovery of loci influencing chemotherapy cytotoxicity*. Proc Natl Acad Sci U S A, 2004. **101**(32): p. 11809-14.
7. Peters, E.J., et al., *Association of thymidylate synthase variants with 5-fluorouracil cytotoxicity*. Pharmacogenet Genomics, 2009. **19**(5): p. 399-401.
8. Beretta, G.L., P. Perego, and F. Zunino, *Mechanisms of Cellular Resistance to Camptothecins*, in *Curr Med Chem Anticancer Agents*. 2006, Bentham Science Publishers Ltd. p. 3291-3305.
9. Pommier, Y., *Topoisomerase I inhibitors: camptothecins and beyond*. Nat Rev Cancer, 2006. **6**(10): p. 789-802.
10. Choy, E., et al., *Genetic analysis of human traits in vitro: drug response and gene expression in lymphoblastoid cell lines*. PLoS Genet, 2008. **4**(11): p. e1000287.
11. Team, R.C.D., *R: A language and environment for statistical computing*. 2005, Vienna, Austria: R Foundation for Statistical Computing.
12. Abecasis, G.R., et al., *Merlin--rapid analysis of dense genetic maps using sparse gene flow trees*. Nat Genet, 2002. **30**(1): p. 97-101.
13. Cohen, D., I. Chumakov, and J. Weissenbach, *A first-generation physical map of the human genome*. Nature, 1993. **366**(6456): p. 698-701.
14. Broman, K.W., et al., *Comprehensive human genetic maps: individual and sex-specific variation in recombination*. Am J Hum Genet, 1998. **63**(3): p. 861-9.



15. Lander, E. and L. Kruglyak, *Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results*. Nat Genet, 1995. **11**(3): p. 241-7.
16. Tanizawa, A., et al., *Comparison of topoisomerase I inhibition, DNA damage, and cytotoxicity of camptothecin derivatives presently in clinical trials*. J Natl Cancer Inst, 1994. **86**(11): p. 836-42.
17. Welsh, M., et al., *Pharmacogenomic discovery using cell-based models*. Pharmacol Rev, 2009. **61**(4): p. 413-29.
18. Pardanani, A., et al., *Host genetic variation contributes to phenotypic diversity in myeloproliferative disorders*. Blood, 2008. **111**(5): p. 2785.
19. Duan, S., et al., *Mapping genes that contribute to daunorubicin-induced cytotoxicity*. Cancer Res, 2007. **67**(11): p. 5425-33.
20. Ulukan, H. and P.W. Swaan, *Camptothecins: a review of their chemotherapeutic potential*. Drugs, 2002. **62**(14): p. 2039-57.
21. Zembutsu, H., et al., *Genome-wide cDNA microarray screening to correlate gene expression profiles with sensitivity of 85 human cancer xenografts to anticancer drugs*. Cancer Res, 2002. **62**(2): p. 518-27.
22. Guo, Y., et al., *In silico pharmacogenetics of warfarin metabolism*. Nat Biotechnol, 2006. **24**(5): p. 531-6.
23. Hill, R., *NK1 (substance P) receptor antagonists - why are they not analgesic in humans?* Trends Pharmacol Sci, 2000. **21**(7): p. 244-246.
24. Borrell, B., *How accurate are cancer cell lines?* Nature, 2010. **463**(7283): p. 858.
25. Nakatsu, N., et al., *Chemosensitivity profile of cancer cell lines and identification of genes determining chemosensitivity by an integrated bioinformatical approach using cDNA arrays*. Molecular Cancer Therapeutics, 2005. **4**(3): p. 399-412.
26. Dennis, G., Jr., et al., *DAVID: Database for Annotation, Visualization, and Integrated Discovery*. Genome Biol, 2003. **4**(5): p. P3.
27. Huang, D.W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources*. Nat Protoc, 2009. **4**(1): p. 44-57.
28. Smirnov, D.A., et al., *Genetic analysis of radiation-induced changes in human gene expression*. Nature, 2009.
29. Edgar, R., M. Domrachev, and A.E. Lash, *Gene Expression Omnibus: NCBI gene expression and hybridization array data repository*. Nucleic Acids Res, 2002. **30**(1): p. 207-10.

30. Hayward, R.L., et al., *Antisense Bcl-xl Down-Regulation Switches the Response to Topoisomerase I Inhibition from Senescence to Apoptosis in Colorectal Cancer Cells, Enhancing Global Cytotoxicity*. *Clinical Cancer Research*, 2003. **9**(7): p. 2856-2865.
31. Takeba, Y., et al., *Irinotecan activates p53 with its active metabolite, resulting in human hepatocellular carcinoma apoptosis*. *J Pharmacol Sci*, 2007. **104**(3): p. 232-42.
32. Zhang, H.F., et al., *Cullin 3 promotes proteasomal degradation of the topoisomerase I-DNA covalent complex*. *Cancer Res*, 2004. **64**(3): p. 1114-21.
33. Lin, C.P., et al., *A ubiquitin-proteasome pathway for the repair of topoisomerase I-DNA covalent complexes*. *J Biol Chem*, 2008. **283**(30): p. 21074-83.
34. Malik, M. and J.L. Nitiss, *DNA repair functions that control sensitivity to topoisomerase-targeting drugs*. *Eukaryot Cell*, 2004. **3**(1): p. 82-90.
35. Borovitskaya, A.E. and P. D'Arpa, *Replication-dependent and -independent camptothecin cytotoxicity of seven human colon tumor cell lines*. *Oncol Res*, 1998. **10**(5): p. 271-6.
36. Daoud, S.S., et al., *Impact of p53 knockout and topotecan treatment on gene expression profiles in human colon carcinoma cells: a pharmacogenomic study*. *Cancer Res*, 2003. **63**(11): p. 2782-93.
37. Zhou, Y., et al., *Transcriptional regulation of mitotic genes by camptothecin-induced DNA damage: microarray analysis of dose- and time-dependent effects*. *Cancer Res*, 2002. **62**(6): p. 1688-95.
38. Beretta, G.L., P. Perego, and F. Zunino, *Targeting topoisomerase I: molecular mechanisms and cellular determinants of response to topoisomerase I inhibitors*. *Expert Opin Ther Targets*, 2008. **12**(10): p. 1243-56.
39. Bartz, S.R., et al., *Small interfering RNA screens reveal enhanced cisplatin cytotoxicity in tumor cells having both BRCA network and TP53 disruptions*. *Mol Cell Biol*, 2006. **26**(24): p. 9377-86.
40. Huang, R.Y., et al., *Genome-wide screen identifies genes whose inactivation confer resistance to cisplatin in *Saccharomyces cerevisiae**. *Cancer Res*, 2005. **65**(13): p. 5890-7.
41. MacKeigan, J.P., L.O. Murphy, and J. Blenis, *Sensitized RNAi screen of human kinases and phosphatases identifies new regulators of apoptosis and chemoresistance*. *Nat Cell Biol*, 2005. **7**(6): p. 591-600.
42. Silva, J.M., et al., *Profiling essential genes in human mammary cells by multiplex RNAi screening*. *Science*, 2008. **319**(5863): p. 617-20.

43. Schlabach, M.R., et al., *Cancer proliferation gene discovery through functional genomics*. Science, 2008. **319**(5863): p. 620-4.

## **CHAPTER 3:**

**Biological and Genomic Profiling in CEPH Distinguishes  
Between Structurally Distinct Topoisomerase 1 Inhibitors**

## ABSTRACT

We have attempted to use an *ex vivo* familial genetics strategy to study mechanisms of Topoisomerase 1 (Top1) inhibition. Investigations have steadily been chipping away at the pathways involved in cellular response following Top1 inhibition for more than 20 years. Our system-wide approach, which phenotypes a collection of genotyped human cell lines for sensitivity to drugs, interrogates all possible targets and cellular pathways simultaneously. Previously, we characterized the *in vitro* sensitivity of several families of CEPH cell lines to nine camptothecin analogues, including the FDA approved drugs topotecan and irinotecan. Linkage analysis revealed a pattern of nine quantitative trait loci (QTLs), regions of DNA which were associated with the observed variation in cytotoxic response, and these loci were shared by all of the camptothecins. In this study, we repeated phenotyping using seven of the original camptothecin analogues, to determine whether these QTLs could be independently validated. Seven of the nine QTLs on chromosomes 1, 5, 6, 11, 16, and 20 were replicated. Finally, to identify which, if any, QTLs are related to the general mechanism of Top1 inhibition or should be considered camptothecin-specific, we characterized the *in vitro* sensitivity of the CEPH cell lines to the indenisoquinolones, a structurally distinct class of Top 1 inhibitors. Four QTLs on chromosomes 1, 5, 11, and 16 were shared by both the camptothecins and the indenisoquinolines and are considered associated with the general mechanism of Top1 inhibition. The remaining three QTLs (chromosomes 6 and 20) are considered specific to camptothecin-induced cytotoxicity. Finally, eight QTLs were identified which were unique to the indenisoquinolines.

## INTRODUCTION

Much of the cost of drug development can be attributed to a poor understanding of a drug's mechanism of action. The practice of target-based drug discovery has produced many promising drug candidates. Unfortunately, most ultimately fail in the clinic due to lapses in efficacy and unanticipated side-effects. As a result, screening based on phenotypic changes induced by candidate drugs in cells or model organisms, is experiencing a resurgence. The greatest limitation of phenotypic screening is determining the mechanism for biologically active compounds. Genomic tools that interrogate all targets and cellular pathways simultaneously are being included with the phenotypic screen to provide insight into a drug's mechanism.

Our own *ex vivo* familial genetics strategy is one of the latest examples of these phenotypic screens. In this model, genes influencing drug action are studied using the Centre d'Etude Polymorphisme Humain (CEPH) cell lines, a collection of extensively genotyped, immortalized human cell lines from multigenerational families. These lymphoblastoid cell lines (LCLs) are first phenotyped for sensitivity to a drug. Linkage analysis is then used to identify regions of the genome referred to as quantitative trait loci (QTLs) where genetic variation correlates with the observed variation in response.

Most recently, we have used this model to establish a specific pattern of QTLs related to a shared mechanism for a class of structurally related compounds, the camptothecins (Chapter 2). Nine QTLs were shared by six camptothecin analogues and then independently validated in a second set of three additional camptothecins (Chapter 2). The camptothecins are excellent model for study. More than 25 years after the discovery that the primary molecular target for the camptothecins is Topoisomerase 1 (Top1), investigations are still underway to define the pathways involved in cellular response [1-4]. Two of these analogues are FDA approved for use

in cancer chemotherapy. A clearer understanding of mechanism of action can be used to find genetic markers of drug susceptibility and resistance in patient populations.

The objective of this study was to determine which QTLs are specific to camptothecin-induced cytotoxicity and which are related to the general mechanism of Top1 inhibition. Drug-to-drug comparative studies between structurally distinct compounds sharing a molecular target have been used to identify genes and pathways which are involved in a shared general mechanism and those which are class-specific and suggest subtle differences in mechanism [5]. We chose to phenotype the CEPH LCLs for sensitivity to the indenoisoquinolines, a new class of non-camptothecin Top1 inhibitors, and compare the resulting genomic profile with the camptothecins (Figure 3-1). While sharing the same molecular target, there are notable pharmacological distinctions between the two classes [6]. The indenoisoquinolines are chemically stable. They trap Top1cc at different genomic sites which suggests potentially different gene targeting [7-9]. Cleavage complexes stabilized by indenoisoquinolines are more persistent than those induced by the camptothecins [7, 10]. Additionally, they are not substrates for ABC transporters which may be useful in the treatment of camptothecin resistant malignancies [11]. By comparing the biological and genomic profiles of the indenoisoquinolines with the camptothecins we can demonstrate that this model can establish a pattern of QTLs (a) related to the general mechanism of Top1 inhibition, (b) specific to camptothecin-induced cytotoxicity, and (c) associated with indenoisoquinoline activity.

## **MATERIALS AND METHODS**

**Chemicals.** The indenoisoquinolines, NS 706744 (Ind1), NSC 725776 (Ind2) and NSC 724998 (Ind3), were generously supplied by Drs. Stephen Frye and Jian Jin of the Center for Integrative Chemical Biology and Drug Discovery (University of North Carolina, Chapel Hill, NC). All drugs were dissolved in DMSO. Final concentrations for working solutions were 1% DMSO. For

validation purposes, cellular phenotyping using the following camptothecins was repeated in this study (referred to as Camptothecin set 1+2):

*Camptothecin set 1:* 9-aminocamptothecin (9AC), 9-nitrocamptothecin (9NC), topotecan (TPT), camptothecin (CPT), and 7-ethyl-10-hydroxycamptothecin (SN38)

*Camptothecin set 2:* 10-methoxycamptothecin (mCPT) and 10-hydroxycamptothecin (hCPT)

The camptothecins were diluted to the same concentrations outlined in Chapter 2.

**Cell lines.** CEPH cell lines (Corriell Repositories) from the following pedigrees were used for this study: 35, 45, 1334, 1340, 1341, 1345, 1350, 1362, 1408, 1420, 1447, 1451, 1454, 1459, 1463. The cells were cultured in RPMI 1640 supplemented with 10% fetal bovine serum at 37°C in humidified air containing 5% CO<sub>2</sub>.

**Measurements of Cell Growth Inhibition and Data Analysis.** Experimental plates were prepared containing a panel of both the indenoisoquinolines and the camptothecins. Drugs were serially diluted and plated in 384 well plates and frozen at -20 C prior to experimentation. Each plate contained four replicates of each drug-dose combination. On the day of experimentation, cells were plated at a density of 4000 cells/well in RPMI 1640 with 10% FBS. Alamar blue was added at 72 h and fluorescence read after 24 h alamar blue exposure (96 h exposure to drug, Ex 535 nm, Em 595 nm). Raw fluorescence values for each set of replicates were considered outliers if there was a ten-fold change in signal (in either directions) for a single replicate. Cell viability relative to untreated controls was determined using the manufacturer's protocol. Final percent survival at each drug-dose combination was averaged from 4 replicates of 2 independently plated experiments (n =8 replicates). Pearson correlation coefficients were used to determine the degree of similarity between biological activity profiles of the Top1 inhibitors.



**Genomic Profiling.** Phenotypes were defined as cytotoxic response at each drug-dose combination. Heritability estimates, which quantify the genetic contribution to observed variation in cytotoxic response at each drug-dose combination, were calculated using MERLIN as described in Chapter 2 [12]. The degree of heritability associated with growth rate was calculated and all heritability estimates for all drug-dose phenotypes were adjusted for growth rate. The genotype data for each cell line were downloaded from V10 of the CEPH database using error checked markers. The genetic map information was downloaded from the Marshfield database. Error checking and nonparametric linkage analysis was performed as previously described (Chapter 2). Quantitative trait loci (QTLs), regions of the genome linked to each phenotype, were identified using the same criteria as previously reported (Chapter 2). Gene-dropping permutation testing was used to identify LOD score thresholds for significant and suggestive QTLs for each drug-dose combination on each chromosome. QTLs observed for a drug-dose phenotype were considered significant if the highest LOD score in that region surpassed the significance LOD score threshold for that drug-dose phenotype. The start and end of a significant QTL region was defined as regions with LOD scores that were above either the suggestive LOD score threshold or peak LOD score minus one, whichever was greater. QTLs observed for a drug-dose phenotype were considered suggestive if the highest LOD score in that region surpassed the suggestive LOD score threshold for that drug-dose phenotype on that chromosome. QTLs which were identified as significant in one drug-dose phenotype were considered replicated in the remaining drug-dose phenotypes if they surpassed the LOD score suggestive threshold. A complete list of QTLs which were significant in the indenoisoquinolines can be found in Tables A3-1.

The goal of this study was to establish a pattern of QTLs (a) related to the general mechanism of Top1 inhibition, (b) specific to camptothecin-induced cytotoxicity, and (c) associated with indenoisoquinoline activity. To achieve this, the following analysis plan was undertaken. In Chapter 2, ten QTLs were identified as shared among all of the camptothecins in the first panel

(Camptothecins set 1: 9AC, 9NC, TPT, CPT, CPT11, SN38). These QTLs were then queried for replication in a second set of 3 additional camptothecins (Camptothecins set 2: mCPT, hCPT, Cl-CPT). QTLs which weren't replicated in the Camptothecins 2 data set were excluded from further analysis. In this study, drugs from both Camptothecin set 1 and Camptothecin set 2 were reevaluated for sensitivity in the CEPH cell lines; the aim was to determine if any of the remaining QTLs could be replicated independently. QTLs which weren't replicated across both data sets (Camptothecin set 1 + Camptothecin set 2) were excluded from further analysis in this study. To identify QTLs associated with the general mechanism of Top1 inhibition, the remaining QTLs were filtered for replication in the indenoisoquinolines, a structurally unique class of Top1 inhibitors. Finally, QTLs which had been removed from this last stage of replication were considered specific to camptothecin-induced cytotoxicity. Figure 3-2 provides a detailed summary of this analysis plan. Rules for replication were as follows:

Within a data set: A QTL identified as significant for one drug-dose phenotype must be replicated at the significant or suggestive level in all drugs within that data set

Between data sets: A QTL identified as significant and replicated in all drugs of one data set must be replicated at the significant or suggestive level in all drugs within a second data set. When there are drugs which are shared between data sets, a QTL which was identified as significant for a given drug-dose phenotype does not need to be detected at the significance level for that same drug-dose phenotype within the second data set. To be considered replicated in the second data set, the QTL must be present at the significant or suggestive level in all drugs belonging to that data set.

## RESULTS

**Variation in Biological Response Between Top1 Inhibitors.** Some have argued that molecular & pharmacological distinctions between the camptothecins and the indenoisoquinolines could result in clinical differences [6]. In this study, we used EBV-transformed LCLs derived from 15 CEPH pedigrees to evaluate differences in cytotoxic activity between the indenoisoquinolines and the camptothecins. Sensitivity to the indenoisoquinolines, Ind1 (NSC70644), Ind2 (NCS725776), and Ind3 (NSC724998), were evaluated in CEPH LCLs (n = 142) using the alamar blue assay. Ind1, Ind2, and Ind3 all showed dose-dependent cytotoxicity (Figure 3-3). We also observed considerable interindividual variation in sensitivity to the indenoisoquinolines (Figure 3-3).

Concentrations were insufficient to reach below the IC<sub>50</sub>, the concentration which inhibits viability by 50%, for 30% of the cell lines exposed to each of the idenoisoquinolines. As a result, a new parameter was chosen to compare variability in cytotoxic response between the camptothecins and the indenoisoquinolines. The cell viability at the dose which yielded a population mean viability closest to 50% was compared for all cell lines across each drug (referred to as the  $\overline{GI50}$ ). Boxplots of cell viability at the  $\overline{GI50}$  for each drug are depicted in Figure 3-4. Cell lines which had a mean viability greater than the 90<sup>th</sup> percentile of viabilities at the  $\overline{GI50}$  were considered resistant. Cell lines which had a mean viability less than the 10<sup>th</sup> percentile of viabilities at the  $\overline{GI50}$  were considered sensitive. Viability data for the six camptothecin analogues were handled in the same manner. Using there parameters, we were able to observe changes in sensitivity patterns between the two classes of Top1 inhibitors for some CEPH pedigrees (Figure 3-5). In pedigree 1451, six out of ten individuals were resistant to all three indenoisoquinolines. Five of these same individuals were sensitive to four or more camptothecins. Likewise, pedigree 1463 possessed six out of sixteen individuals who were sensitive to all three indenoisoquinolines; those cell lines were also resistant to 5 or more

camptothecin analogues. A single individual within this pedigree (GM12885) was resistant to all three indenoisoquinolines and sensitive to all six camptothecin analogues. This same pattern was observed for GM10846. Members of pedigree 1340 appeared sensitive to all six camptothecin analogues and all three indenoisoquinolines. While both the indenoisoquinolines and camptothecins act by a similar mechanism, i.e. Top1 inhibition, a correlation matrix on cell viability at the  $(GI50)$  for each drug suggests distinctions between the two classes. The three indenoisoquinolines are highly correlated with each other ( $PCC > 0.8$ ) and distinct from the camptothecins ( $PCC < 0$ ) (Table 3-1).

**Heritability analysis.** There exists a significant relationship between growth rate and sensitivity to chemotherapeutic agents; cell lines which grow more rapidly are more sensitive to these cytotoxic agents [13]. As a control, we estimated the genetic contribution to growth rate in the presence of vehicle alone. Growth rate was found to be heritable ( $h^2 = 8.65\%$ ). The growth rate adjusted heritability estimates for each drug-dose phenotype are indicated in Table 3-2. Overall, 21 of the drug-dose phenotypes had a maximum heritability at or below that observed for growth rate. Sensitivity to Ind1 (100 nM) and Ind3 (5  $\mu$ M) were found to be heritable traits with estimates of 15.7 and 14.1% respectively. A number of doses were found not to be heritable in this sample ( $h^2 = 0$ ). Taken together, these heritability estimates were lower than those reported for the camptothecin analogues which averaged  $23.1\% \pm 2.6\%$  at doses within the linear phase of the sigmoid curve (Chapter 2). While the heritability estimates for the indenoisoquinolines varied considerably by dose (Table 3-2), heritability estimates for multiple doses and a number of camptothecin analogues were consistently approximately 20% for doses within the linear phase of the sigmoid curve.

**Independent Replication of QTLs from Camptothecin data sets 1 and 2.** Ten QTLs were identified as replicated within all six camptothecin analogues of Camptothecin set 1 (Chapter 2).

In a separate validation experiment using 3 additional camptothecin analogues, Camptothecin set 2, nine of those QTLs were replicated (Chapter 2). In this study, we sought to further refine this list by performing a third independent validation study with nine drugs from Camptothecin sets 1 and 2 (Camptothecin set 1+2). Any QTLs which could not be replicated at the significant or suggestive level for all drugs within this independent validation study were excluded from further analysis. Seven of the nine QTLs were replicated in Camptothecin set 1+2. The peak on chromosome 16 was identified as significant in multiple doses of CPT just as previously reported (Chapter 2). In fact, this QTL was replicated at the significance level in all camptothecins (for  $n \geq 1$  doses) from Camptothecin set 1+2. The QTL on chromosome 11 was only present in two doses of CPT11 when studying Camptothecin set 1. In this study, it was present in all camptothecins at the significance level. Multiple doses of both SN38 and CPT11 (Camptothecin set 1) had QTLs on chromosome 20 which surpassed the significance LOD score thresholds in our earlier work. The QTL on chromosome 20 was replicated at the significance level in multiple doses of SN38 and at the suggestive level of all other camptothecin analogues in the Camptothecin set 1+2. In our earlier work, a QTL on chromosome 5 was significant in multiple doses of 9NC, CPT, mCPT, and hCPT. In this study, that same region was replicated at the significance level for multiple doses of hCPT, SN38, CPT, 9AC, and 9NC. Figure A3-1 is a genome-wide map detailing these findings. Table 3-3 lists the QTLs shared by all camptothecin analogues which were replicated in this study.

**QTLs associated with Top1 inhibition.** The next objective was to identify which peaks might be class-specific and which might be associated with the general mechanism of Top1 inhibition. Seven QTLs were replicated across all drugs in Camptothecin set 1, Camptothecin set 2, and Camptothecin set 1+2 (Table 3-3). We compared this pattern of QTLs to the QTLs associated with indenoisoquinoline sensitivity. A peak was considered replicated if it was found in all three indenoisoquinolines at the significant or suggestive level. Four of the seven predefined QTLs

were shared by all of the camptothecins and all of the indenoisoquinolines. These QTLs are summarized in Table 3-3. Peaks on chromosomes 11 and 16 were present at the significance level in multiple doses of all three indenoisoquinolines. The QTL on chromosome 5 was significant at multiple doses of Ind2 and met the suggestive threshold for multiple doses of all three indenoisoquinolines. Finally, the QTL on chromosome 1 was shared by multiples doses of all three indenoisoquinolines at the suggestive level only. Most notably, the peak on chromosome 20, which contains Top1 and is shared by all of the camptothecins, was not present at the significant level in any indenoisoquinoline. However, it was present at the suggestive level in at least one dose of Ind1 and Ind3. These results are illustrated in Figure 3-6, a genome wide map of QTLs at the significant and suggestive level for both the camptothecins and indenoisoquinolines. We consider the remaining QTLs on chromosomes 6 and 20, which were not replicated in the indenoisoquinoline, camptothecin-specific (Table 3-3).

**QTLs associated with indenoisoquinoline-induced cytotoxicity.** There are also QTLs which appear specific to the indenoisoquinolines alone. These QTLs are not replicated in the camptothecins and are summarized in Table 3-4. Of note, the QTLs on chromosomes 6 and 16 were identified in multiple doses of all three indenoisoquinolines with LOD scores that exceeded the significance thresholds (maximum LOD score 2.286 for Ind3 at 7uM and 2.139 for 3 uM Ind2). Similarly, the QTL on chromosome 13 was also present in multiple doses of all three indenoisoquinolines (maximum LOD score 1.44 for 10 uM Ind3). At least two indenoisoquinolines had significant peaks on chromosomes 4 and 10 at multiple doses which were replicated at the suggestive level in  $n \geq 1$  doses of the third indenoisoquinoline. There are also subtle distinctions between the indenoisoquinolines. Multiple doses of Ind1 and Ind3 had a QTL with significant LOD scores on chromosome 14 (98-134 cM); this QTL was not present in drug-dose combinations of Ind2. A QTL on chromosome 19 (52-77 cM) surpasses the significance threshold for multiple doses of Ind1 but is not shared by the other members of that

class. Regions with significant peak LOD scores for growth in vehicle did not overlap peaks with any of the significant QTLs associated with sensitivity to the indenoisoquinolines (Figure 3-6) with the exception of a peak on chromosome 7 that is present in all three indenoisoquinolines. This suggests that the remaining QTLs are associated with cytotoxicity rather than the genetic effects from growth rate.

## **DISCUSSION**

Drugs with common molecular targets can have very different therapeutic activities. For example, while the vinca alkaloids and colchicines are structurally distinct classes of tubulin destabilizing drugs, they are used in the treatment of cancer and gout respectively. The vinca alkaloids and taxanes target microtubules, but have distinct antitumor profiles, post-target interaction events, and mechanisms of resistance [14]. Differences in the expression of the molecular direct and indirect targets, metabolizing enzymes, and transporters between tumor types have been also been attributed to these observed differences in activity [14]. The similarities and differences in mechanisms between classes of compounds sharing the same molecular target have been characterized using genomic profiling [5]. In the present study, we used pharmacologic and genomic profiling in the CEPH cell lines to investigate two classes of Top1 inhibitors, the camptothecins and the indenoisoquinolines. Reports indicate that while sharing the same molecular target, the indenoisoquinolines exhibit unique properties which may set them apart clinically from the camptothecins. We previously used our *ex vivo* familial genetics model to study mechanisms of camptothecin-induced cytotoxicity (Chapter 2). The goal of this study was identify regions of the genome which were correlated to class-specific cytotoxicity and shared mechanisms of action.

For profiling, we evaluated the biological activity of each class of drugs using a growth inhibitory assay. We compared the sensitivity of the CEPH cell lines to the indenoisoquinolines and the

camptothecins. Cell viability at the dose which produced 50% population mean viability ( $\overline{GI50}$ ) for each drug was used as an endpoint. Interindividual variation in response to the indenoisoquinolines was found to be positively correlated (Table 3-1). An inverse correlation was noted when viability at ( $\overline{GI50}$ ) for each cell line was compared between the two structural classes. A number of cell lines which were resistant to all camptothecins were sensitive to all indenoisoquinolines and vice versa (Figure 3-5). These differences in cytotoxic activity suggest that while they share the same molecular target, there may be subtle distinctions in the biochemical cascade required for drug action (e.g. uptake mechanisms, metabolism, and secondary molecular interactions). Pommier suggests that (a) the differential genomic targeting of Top1 cleavage complexes by the camptothecins and indenoisoquinolines, (b) the differences in chemical structure and chemical stability of the indenoisoquinolines compared with camptothecins, and (c) the low cross-resistance to camptothecins based on drug efflux and Top1 point mutations, make it likely that indenoisoquinolines will exhibit unique clinical and molecular properties which distinguish them from the camptothecins [6]. Even with a shared mechanism of action, mutations may render a cell sensitive or resistant to these distinct structural classes. For example, Antony et al. reported that human leukemic cells which were resistant to the camptothecins as a result of a Top1 point mutation were sensitive to Ind1 [9]. Not only can mutations in genes shared between these two classes change sensitivity, but this data suggest that there might be subtle distinctions in genes critical for action between the two classes (Table 3-4).

While heritability analysis was not essential to the objective of this study, the results merit discussion. Sensitivity to individual concentrations of the indenoisoquinolines seems to be under some genetic control with a maximum heritability estimates of 0.15 (Table 3-2). Heritability estimates for drug-induced cytotoxicity by dose in CEPH cell lines have ranged from 0.06-0.70 [15-17]. A number of the indenoisoquinoline drug-dose combinations have heritability estimates that are low ( $0 \leq h^2 \leq 10$ ) which had significant linkage peaks. The overall heritability of a trait is



typically considered evidence of a genetic contribution to phenotypic variation and implies that regions of the genome might be mapped which are responsible for this variation. In turn, this suggests that traits with low or no heritability values will not possess significant linkage. This was not the case in this study. Heritability has been described as an inconsistent predictor of significant linkage in the CEPH LCL model before [18-20]. In a study to identify regions of the genome influencing the expression profiles of CEPH cell lines, Huang et al uncovered 422 expression traits with significant linkage. Of these 89 (21%) traits had low estimates ( $h^2 \leq 10$ ), 23 of which had  $h^2 = 0$ . QTLs for gene expression traits with  $h^2 = 0$  were not false positive at the 0.5 significance level.

Nonparametric linkage analysis was used to identify regions of the genome which are specific to the general mechanism of Top1 inhibition, and distinct regions associated with the drug-induced cytotoxicity of camptothecins and indenoisoquinolines. We used permutation testing to determine empirical LOD score thresholds for significant and suggestive linkage for each drug-dose combination on each chromosome. Four QTLs were shared by all camptothecin analogues and replicated at the significant and/or suggestive level in all indenoisoquinolines (chr 1, 5, 11, 16; Table 3-3). We consider these QTLs specific to the general mechanism of Top1 inhibition, while QTLs in Table 3-4 are considered specific to the indenoisoquinolines. Different DNA cleavage patterns, biological activity within the CEPH, and unique QTLs patterns suggest some genes may be more selectively targeted by one compound class than the other [7, 10]. While Ind1, Ind2, and Ind3 depend on Top1 for cytotoxic effect, siRNA knockdown of Top1 (at 80-90% efficiency) does not completely reverse growth inhibition which suggests additional targets [9, 10]. Moreover, Ind2 and Ind3 have weak activity against Top2 [10]. It's too early to begin making hypotheses about what genes under those class-specific QTLs might be critical to the activity of the indenoisoquinolines. Since the discovery that this class of compounds acts on Top1

in 1998, our search of Pubmed reveals only 34 articles about the class. These articles predominately involve studies to optimize Top1 binding and potency.

This study does open the doorway for the investigation of molecular pathways associated with cytotoxicity which are shared by the two structural classes. We used the functional categorization tool from the web-accessible program Database for Annotation, Visualization, and Integrated Discovery (DAVID) to stratify the protein-coding genes under each shared QTL from Table 3-3 according to gene ontology (GO) terms. Table 3-5 lists genes under each QTL associated with GO terms related to camptothecin activity. The bolded gene names have been previously associated with camptothecin activity in yeast and/or mammalian cell lines. For example, the QTL on chromosome 1 contains PARP1 (poly(ADP-ribose) polymerase 1), an abundant nuclear protein that is activated by DNA strand breaks to modify proteins with poly(ADP-ribose) (PAR). PARP1 recruits DNA repair and checkpoint proteins to sites of DNA damage. Knockdown of PARP1 using siRNA sensitizes cells to camptothecin-induced cytotoxicity [21]. The SCF (Skp, Cullin, F-box) ubiquitin ligase complex degrades Top1 following Top1 mediated DNA damage. Overexpression of components of the SCF complex, such as Cul3 (on chromosome 1) and Skp2 (on chromosome 5) have been associated with resistance [22, 23]. Inhibition of Chk1 (chromosome 11), a serine/threonine kinase that is activated following DNA strand breakage from Top1 inhibitors, sensitizes cells to camptothecin analogues [24-26]. Rad50 (chromosome 5) belongs to the DNA repair complex MRE11-RAD50-NBS1 (MRN) complex which is activated in the presence of camptothecin-induced DNA double strand breaks [27, 28]. The overexpression of NDRG1 (chromosome 16) suppresses camptothecin-induced apoptosis, while inhibition of expression potentiated apoptosis in camptothecin treated cells [29]. An shRNA gene knockdown experiment using this gene list would make an excellent follow-up study for clarifying the molecular consequences of Top1 inhibition.

The identification of QTLs on chromosomes 6 and 20, which are specific to the camptothecins, is also intriguing. The fact that the QTLs were shared by all camptothecins but not present at the significant or suggestive level in any of the indenoisoquinolines also suggests distinctions in mechanism. The QTLs on chromosomes 20 and 6 have been validated in three independent studies. In Chapter 2, we used the functional annotation clustering tool from DAVID to identify Gene ontology (GO) terms which were enriched for genes within the QTL on chromosome 20 (Table 2-3, Table A2-2) [30, 31]. The top over-represented GO terms were: response to stimulus, DNA binding, transcription, protein binding, cell differentiation, and regulation of cell growth. Some of the genes associated with these GO terms may be relevant to the camptothecin mechanism (Table 2-3, Table A2-2). A literature search reveals that many have yet to be evaluated for their effect on camptothecin-induced cytotoxicity. In Chapter 2, we indicated that Bcl-xl, an anti-apoptotic protein which promotes cell survival, and Top1 were two genes with known involvement in camptothecin cytotoxicity that were present in the QTL on chromosome 20. It is possible that neither gene is responsible for this QTL. This QTL was not shared by the indenoisoquinolines, which also target Top1. Downregulation of Bcl-xl using siRNA prior to treatment increases sensitivity to SN38 and other camptothecin analogues, while overexpression of Bcl-xl results in resistance to camptothecins analogues [32] [33]. There are also contradictory reports regarding changes in Bcl-xl expression following treatment with the camptothecins. While Bcl-xl expression was induced in MCF-7 cells treated with increasing concentrations of camptothecin, expression was down-regulated in HepG2 cells with increasing concentrations of TPT [34, 35]. The most overrepresented GO terms for genes under the QTL on chromosome 6 at 0-29 cM were cell differentiation (subcategory GO terms: apoptosis and cell death), protease inhibition, cell proliferation, and regulation of biological processes (subcategory GO terms: transcription, gene expression, DNA binding). Of the genes associated with these GO terms, there have been reports for the involvement of both WRN and FOXQ1 in camptothecin-induced cytotoxicity [36, 37]. The QTL on chromosome 6 from 42-65 cM is enriched with genes

associated with chromatin assembly, response to chemical stimulus, and NF-kappa binding. NF-kappa B interferes with the effect of most anti-cancer drugs through induction of anti-apoptotic genes. Blocking NF-kappa B activation sensitizes cells to camptothecin analogues [38]. NF-kappa B tumor expression seems to be negatively correlated with response to irinotecan [39]. A number of genes involved in the NF-kappa B signaling pathway are present in this QTL. This QTL also contains 22 members of the histone H1 family. The histones are proteins which package and order DNA into nucleosomes. Phosphorylation of histone H2AX is an extremely sensitive marker for double strand breaks induced by DNA damaging agents, such as the camptothecins and indenoisoquinolines [10, 40]. It has been suggested that other histones are involved in DNA repair and cell proliferation [41]. Studies indicate that histones are ADP-ribosylated *in vivo* in response to DNA damage [42]. All core histones, including H1, can be mono-ADP ribosylated [41].

We have characterized a method for clarifying the mechanism of action for Top1 inhibition associated with the camptothecins and indenoisoquinolines. More than 25 years of research have been devoted to identifying the molecular pathways associated with camptothecin-induced cytotoxicity. Regions of the genome which contain known and potentially novel genes critical to their action have identified and validated. The QTLs shared by both the indenoisoquinolines and camptothecins (Table 3-3) are considered the most important for followup, followed by the camptothecin-specific QTLs, and finally the QTLs unique to the indenoisoquinolines. QTLs related to the general mechanism of Top1 inhibition may not only lead to predictions about camptothecin sensitivity and resistance in patient populations but could provide valuable insight into proposed functions of mammalian Top1 [43]. Whole genome association studies using LCLs should be used next to narrow down these broad QTLs. Most importantly, we've described a system-wide genomic method for delving into drug action which could surmount some of the limitations of existing mechanism elucidation strategies.

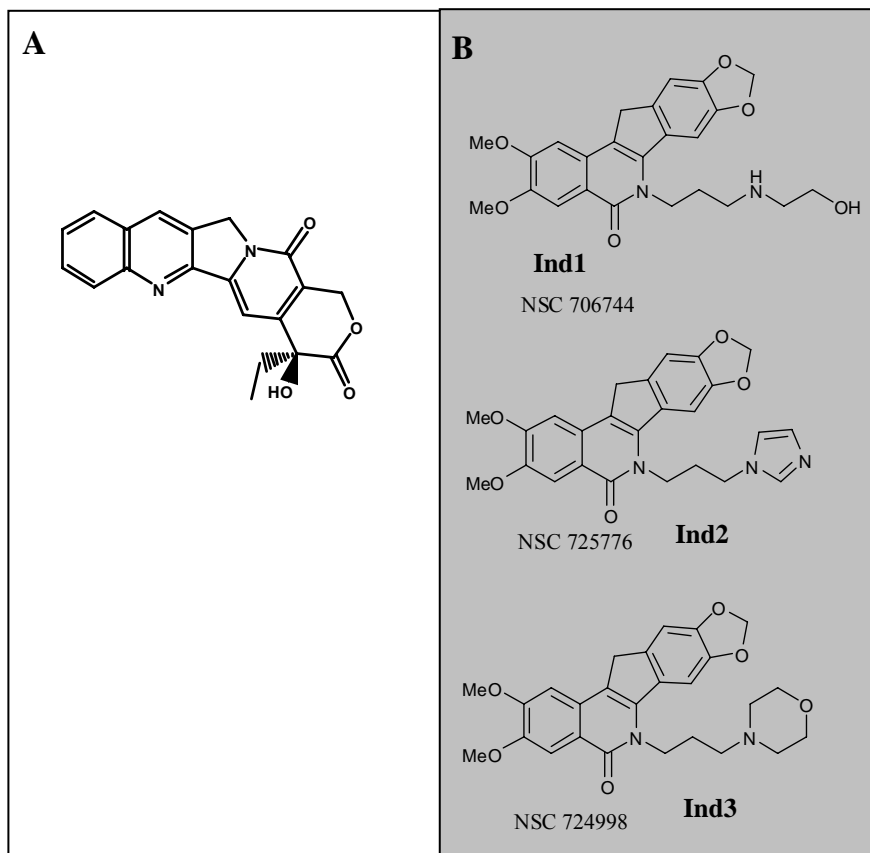


Figure 3-1. Chemical structures of camptothecin (A) and the indenoisoquinolines (B). The lead compound NSC 706744 (Ind1) led to two derivatives, NSC 725776 (Ind2) and 725998 (Ind3), which are being pursued for clinical development.

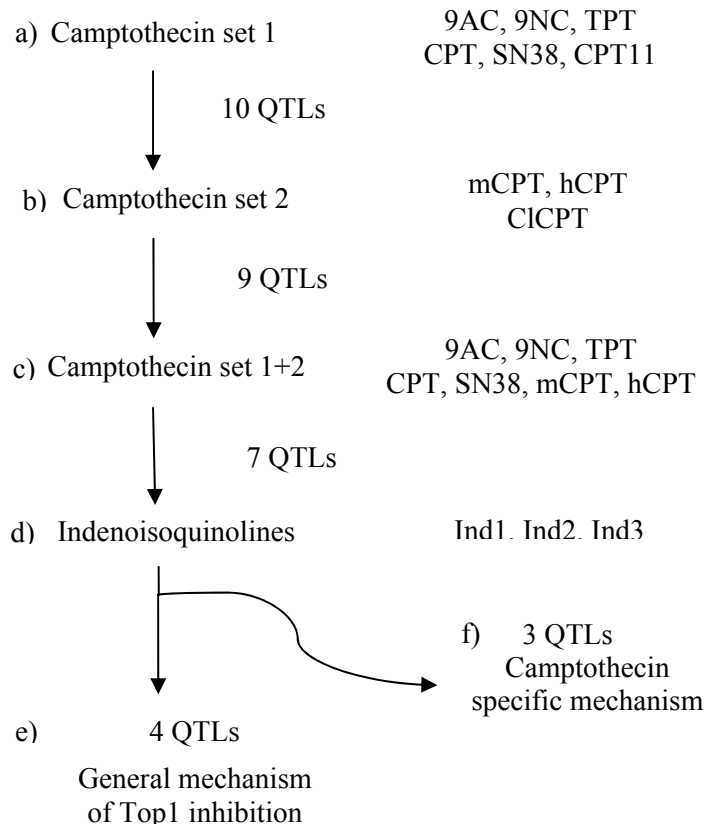


Figure 3-2. Genomic profiling of Topoisomerase 1 inhibitors. A) The CEPH cell lines were phenotyped for sensitivity to a panel of six camptothecin analogues (Camptothecin set 1). Nonparametric linkage analysis was used to identify 10 QTLs which were shared by all six camptothecin analogues. B) A phenotyping study with three distinct camptothecin analogues (Camptothecin set 2) was performed. QTLs which were shared in Camptothecin set 1 were queried for replication in Camptothecin set 2. QTLs which were not shared by all drugs in Camptothecin set 2 were excluded. C) For additional validation, a third independent phenotyping experiment was performed using seven of the nine camptothecin (Camptothecin set 1+2). Seven QTLs were independently validated in this study. QTLs which were not shared by all drugs in Camptothecin set 1+2 were excluded. D) To stratify QTLs as specific to camptothecin-induced cytotoxicity or the general mechanism of Top1 inhibition, CEPH cell lines were phenotyped for sensitivity to the indenoisoquinolines, noncamptothecin Top1 inhibitors. E) Four of the seven QTLs were shared by both the camptothecins and indenoisoquinolines and considered to be associated with the general mechanism of Top1 inhibition. F) Three QTLs which were shared by all camptothecins were not replicated in the indenoisoquinolines.

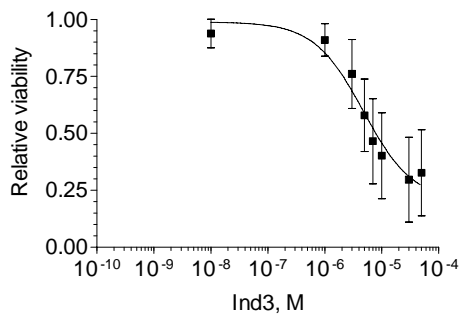
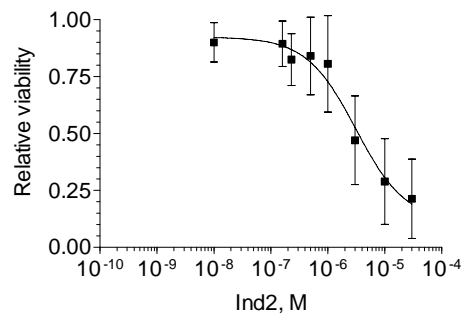
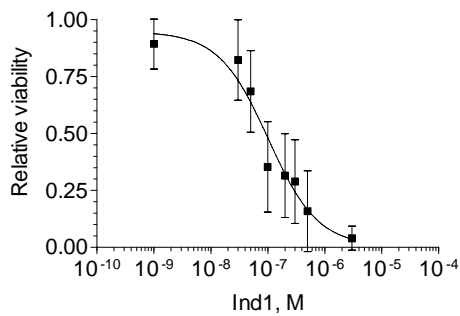


Figure 3-3. Dose–response curve for (A) Ind1, (B) Ind2, and (C) Ind3. Data points represent the overall population mean (n=142) for viability relative to untreated controls at each dose. Vertical bars represent the standard deviation for cell viability across the population.

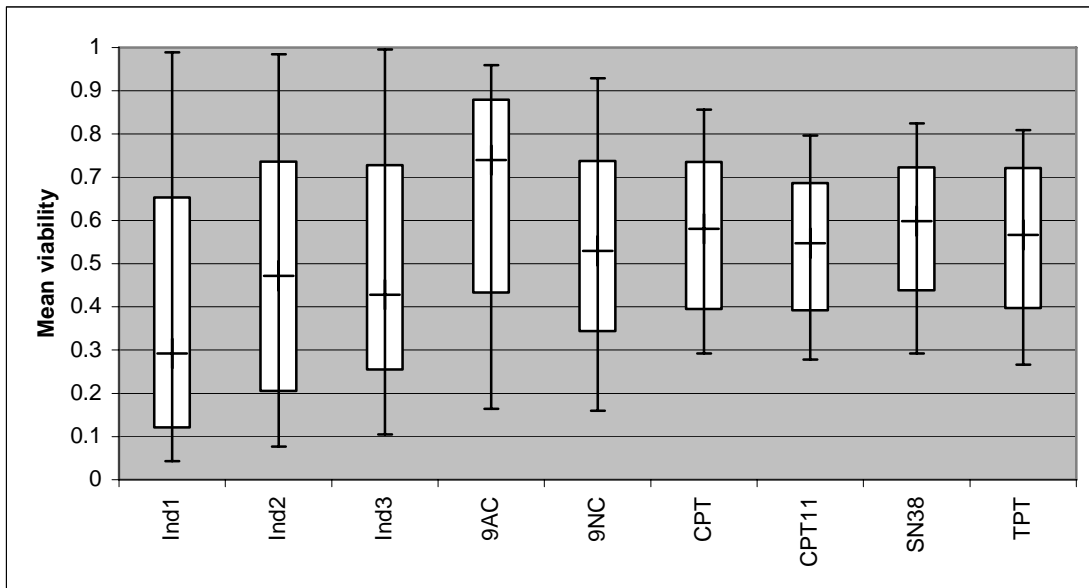


Figure 3-4. Boxplot of mean viabilities for CEPH cell lines at  $(\overline{GI50})$ , the dose closest to a population mean viability of 50%. Line represents mean viability across cell population, box represents the 90<sup>th</sup> and 10<sup>th</sup> percentiles, and whiskers are the maximum and minimum values.



Pedigree	Cell ID	Ind1	Ind2	Ind3	9AC	9NC	TPT	CPT	SN38	CPT11
1451	12772									
	12767									
	12771									
	12773									
	12769									
	12766									
	12770									
	12768									
	12774									
	12848									
1463	12877									
	12879									
	12880									
	12882									
	12883									
	12885									
	12886									
	12887									
	12888									
	12889									
	12890									
	12892									
	12884									
12893										
1340	11821									
	7008									
	7053									
	7062									
	7342									
	7027									
	7029									
	7019									

Figure 3-5. Sensitivity patterns for camptothecins and indenoisoquinolines in select CEPH pedigrees. Red indicates resistance ie. mean viability > 90<sup>th</sup> percentile of viabilities at the ( $\overline{GI50}$ ), green indicates sensitivity ie. mean viability < 10<sup>th</sup> percentile of viabilities at the ( $\overline{GI50}$ ), and black indicates mean viability between 10-90<sup>th</sup> percentile at the ( $\overline{GI50}$ ).

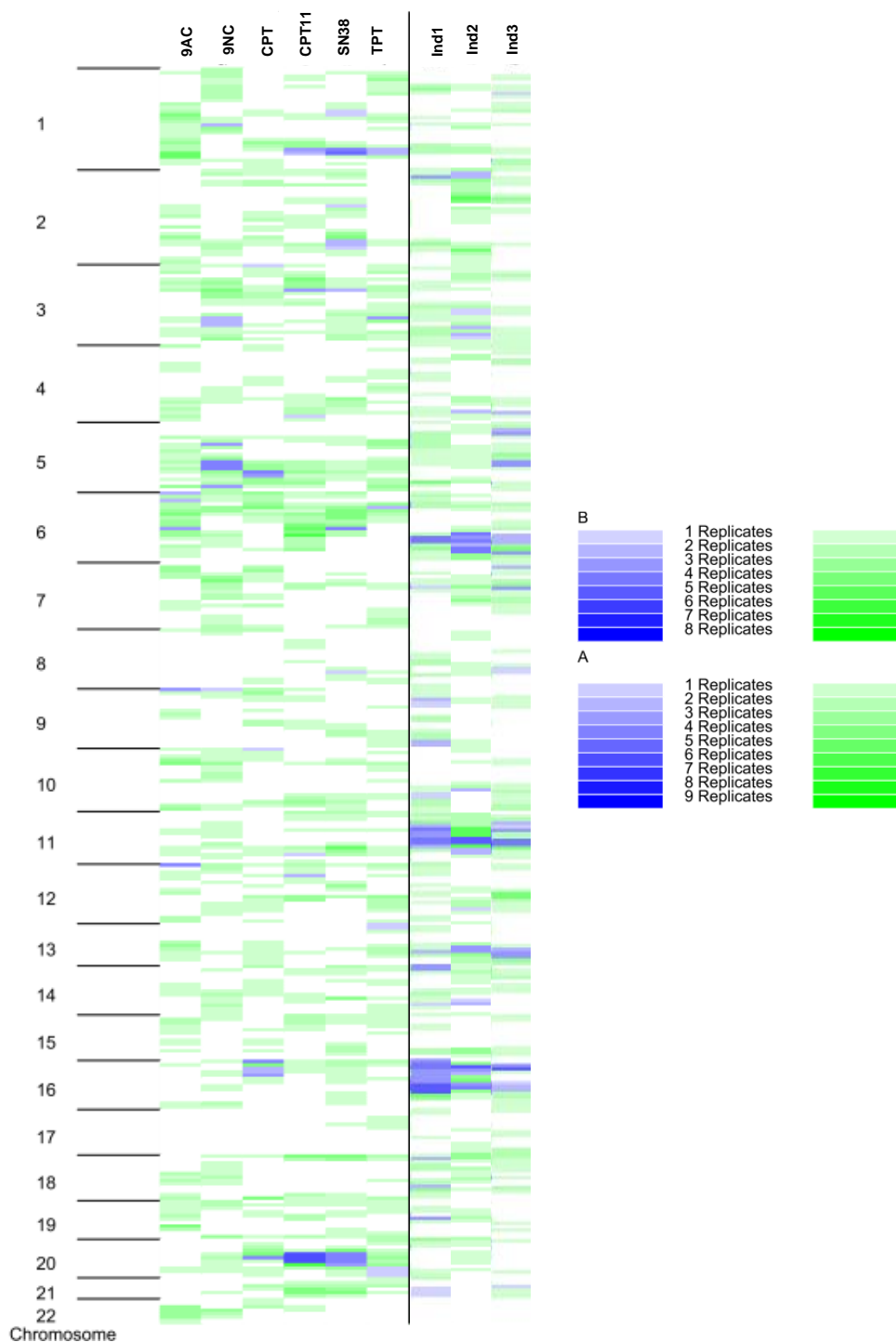


Figure 3-6. Genome wide pattern of QTLs for the camptothecins and indenoisoquinolines. Each chromosome was partitioned into 10 cM regions. For each drug, any significant QTL regions ( $\text{LOD} > \text{permutation threshold for significance}$ ) is indicated in blue. The intensity of the shading indicates the total number of additional doses that were replicated at either the suggestive or significant level. Regions which also had at least a suggestive QTL are shown in green. Color intensity of the color represents the number of dose replications. Drugs are grouped according to their structural class, where the groups are labeled as follows: Group A: Camptothecin set 1 and Group B: indenoisoquinolines.

Table 3-1. Results of correlation analysis for cell viability at GI50 for camptothecins and indenoisoquinolines

PCC	Ind1	Ind2	Ind3	9AC	9NC	CPT	CPT11	SN38	TPT
Ind1	1	0.88	0.84	-0.14	-0.14	-0.12	-0.15	-0.17	-0.14
Ind2		1	0.82	-0.11	-0.09	-0.10	-0.13	-0.14	-0.10
Ind3			1	-0.09	-0.05	-0.06	-0.11	-0.15	-0.10
9AC				1	0.65	0.65	0.65	0.65	0.65
9NC					1	0.77	0.76	0.73	0.78
CPT							0.94	0.88	0.86
CPT11							1	0.91	0.86
SN38								1	0.85
TPT									1

Pearson Correlation Coefficient, (PCC)

Table 3-2. Heritability estimates for the indenoisoquinolines

Drug	Concentration (M)	Growth rate adjusted heritability estimate
Ind1	3.00E-06	2.21
	5.00E-07	0.00
	3.00E-07	11.75
	2.00E-07	3.18
	1.00E-07	15.68
	5.00E-08	0.45
	3.00E-08	0.00
	1.00E-09	0.00
Ind2	3.00E-05	0.00
	1.00E-05	4.42
	3.00E-06	6.72
	1.00E-06	2.24
	5.00E-07	0.00
	2.30E-07	0.00
	1.60E-07	0.00
	1.00E-08	0.00
Ind3	5.00E-05	6.84
	3.00E-05	6.85
	1.00E-05	6.86
	7.00E-06	0.00
	5.00E-06	14.10
	3.00E-06	4.13
	1.00E-06	0.00
	1.00E-08	0.00

Table 3-3. QTLs found in Camptothecin set 1+2 and indenoisoquinolines

Chromosome	Peak Start (cM)	Peak End (cM)	LOD <sup>b</sup>	LOD <sup>c</sup>
1 <sup>a</sup>	229	252	1.855	0.870
5 <sup>a</sup>	125	194	1.709	1.551
6	0	29	1.528	-
6	42	65	1.652	-
11 <sup>a</sup>	115	131	1.352	2.421
16 <sup>a</sup>	0	75	1.345	2.139
20	42	101	2.134	-

<sup>a</sup> QTLs which were also shared by the indenoisoquinolines

<sup>b</sup> Maximum LOD score observed in this region associated with the camptothecins

<sup>c</sup> Maximum LOD score observed in this region associated with the indenoisoquinolines

Table 3-4. QTLs shared by all Indenoisoquinolines

Chromosome	Peak Start (cM)	Peak End (cM)	LOD*
2	1	34	2.024
4	192	211	1.483
6	100	192	2.286
10	118	155	1.775
13	67	114	1.44
16	27	108	2.139
18	0	31	1.574
18	90	96	1.554

\* Maximum LOD score observed in this region

Table 3-5: Genes under QTLs shared by camptothecins and indenoisoquinolines\*

chr	GO terms	GENES
1	GO:0016055	Wnt receptor signaling pathway <b>WNT3A, WNT10A, WNT6, WNT9A,</b>
	GO:0006974	response to DNA damage stimulus TNP1, <b>NHEJ1</b> , C1orf124, <b>PARP1, BARD1, DTL, XRCC5,</b>
	GO:0003677	DNA binding MIXL1, BATF3, PROX1, TAF1A, SP110, ZNF142, ESRRG, <b>HIST3H3, IKZF2, SMARCAL1, HIST3H2BB, ZNF678, TSNAX, TAF5L, LBR, HIST3H2A, FEV, RCOR3, TIGD1, PAX3, HLX, SP100, ATF3, SP140, NCL,</b>
	GO:0006350	transcription KCNH1, RCOR3, PTMA, HLX, EGLN1, IHH, STK36
	GO:0006915	apoptosis COL4A3, <b>CUL3</b> , SCG2, IL8RB, <b>AGT</b> , INHA, INPP5D, SPATA3, <b>TP53BP2</b> , PECR, PSEN2, <b>TGFB2</b> , TRAF5,
	GO:0030154	cell differentiation USH2A, ERBB4, MREG, NGEF, EFHD1, ITPKB, ACTA1, SERPINE2, OBSCN, SCG2, COL4A4, DNER, SPEG, LEFTY2,
5	GO:0030154	cell differentiation NRG2, NME5, ISL1, NAIP, HSPA9, DND1, <b>MAP3K1</b> , CYFIP2, ZNF346, AGGF1, UNC5A, TSSK1B, CTNNA1, NPM1, DRD1, SLIT3, PDLIM7, FST, POU4F3, PURA, F2R, AP3B1, <b>PIK3R1</b> , PRKAA1, NKX2-5, GPR98, CD74, PRDM6, NEUROG1, PPP2CB, NR2F1, JMY, MEF2C, CSF2, GZMA, MAP1B, ADRB2, FGF10, VDAC1, <b>UBE2B</b> , IL12B, CARD6, CARTPT, ATG12, IL3, MRPS30, SEMA6A, SMAD5, EFNA5, HDAC3, SMN2, PROP1, AFF4, ERAP1, C9, GDNF, SPINK5, EGR1, PPP2R2B, DDX41, PRLR, FGF1, HAND1, SFXN1, HNRPAB, HSD17B4, RASA1, IL4, ACSL6, FGF18, CD14, DBN1, GFRA3,
	GO:0006915	apoptosis PROP1, PPP2CB, ADRB2,
	GO:0008283	cell proliferation FABP6, <b>SKP2</b> , APC, CNOT8, TGFBI, B4GALT7, DAB2, IL9, HTR1A, L6ST, ADRA1B, SPOCK1, PPAP2A, LIFR, FGF1, CSF1R, HBEGF,
	GO:0006512	ubiquitin cycle <b>FBXL17, FBXO4</b> , RNF14, TRIM23, <b>SKP1</b> , ENC1, CDC23, RNF145, TSPAN17, ERCC8, <b>FBXO38</b> , MARCH3, KLHL3, <b>FBXW11</b> , UBE2D2, <b>FBXL21</b> ,
	GO:0000079	regulation of cyclin-dependent protein kinase activity <b>CDC25C, CDK7, CCNG1, CCNH,</b>
	GO:0006974	response to DNA damage stimulus CCNO, POLK, <b>RAD17</b> , PTTG1, <b>XRCC4, RAD50</b> , GTF2H2, MSH3,

chr	GO terms		GENES
5	GO:0043122	regulation of NF-kappaB cascade	F2R, PLK2, CXXC5, NDFIP1, TICAM2,
	GO:0007049	cell cycle	PAM, MCC, IRF1, CETN3, NIPBL, DUSP1, <b>CCNB1</b> , PFDN1, ERBB2IP, SEPT8, CDC20B, <b>RASA1</b>
	GO:0003682	chromatin binding	NSD1, <b>H2AFY</b> , CHD1,
11	GO:0050896	response to stimulus	TIRAP,CHEK1,OR10G4,MFRP,BSX,C1QTNF5,OR10G7,OR10G9,PVRL1,POU2F3,R8A1,ROBO3,HSPA8,CRTAM,OR10S1,OR8G5,OR6T1,OR4D5,OR6M1,EI24,OR8B8,OR8B12,OR8B2,OR8B3,OR8D1,OR8B4,OR8D2,OR8D4,PATE4,OR6X1
	GO:0030154	cell differentiation	PVRL1,POU2F3, <b>DDX25</b> ,CDON,ROBO4,TIRAP,ROBO3,THY1,FEZ1
	GO:0022402	cell cycle process	HEPACAM,HEPN1,TBRG1, <b>CHK1</b>
	GO:0048522	positive regulation of cellular process	BSX,EI24,CRTAM,POU2F3,CBL,ARHGEF12,THY1
	GO:0003677	DNA binding	PKNOX2,BSX,POU2F3,TRIM29,CBL,TBRG1,ZNF202
16	GO:0005524	ATP binding	ERN2, PDPK1, <b>PLK1</b> , ATP2A1, NOD2, ATP6V0C, CREBBP, DNAH3, <b>TAOK2</b> , NME3, <b>PRKCB1</b> , NUBP2, KIFC3, WDR51B, NUBP1, PHKG2, RAB26, CIITA, EEF2K, MYH11, EARS2, PKMYT1, SEPHS2, TRAP1, BCKDK, CHTF18, NLRC5, ACSM1,
	GO:0007049	cell cycle	AXIN1, CORO1A, CYLD, CP110, SEPT1, CSNK2A2, DNAJA2, <b>RBL2</b> , TSC2, PAPD5, PRM2, KATNB1, PRM1, SIAH1, <b>CCNF</b> , GSPT1, <b>MAPK3</b> ,
	GO:0051092	activation of NF-kappaB transcription factor	PYCARD, NLRC3
	GO:0006974	response to DNA damage stimulus	C16ORF35, NTHL1, SMG1, ERCC4, KCTD13, KIF22, MPG, GIYD2,
	GO:0042493	response to drug	ABCA3, <b>ABCC6</b> , <b>ABCC1</b> , <b>MVP</b> ,
	GO:0030154	cell differentiation	SOX8, TRAF7, NTN2L, CACNA1H, TBX6, SALL1, IRX5, GNAO1, NUPR1, BBS2, METRN, TNP2, MT3, MKL2, MYST1, IL27, <b>NDRG4</b> ,
	GO:0008219	cell death	CLN3, TNFRSF12A, CIAPIN1, DNAJA3, LITAF, PDIA2, EMP2, BFAR, DNASE1, SPN,
	GO:0006512	ubiquitin cycle	RAB40C, TCEB2, SOCS1, AMFR, <b>FBXL16</b> , STUB1, RNF40, USP7, <b>FBXL19</b> , AKTIP,
chr	GO terms		GENES



16	GO:0003677	DNA binding	ZNF646, GTF3C1, ZNF213, HN1L, ORC6L, ZNF174, CARHSP1, MAZ, UBN1, IRX3, ALG1, IRX6, ZNF597, ZNF75A, ZNF434, FUS, SRCAP, TFAP4, CHD9, ZNF263, POLR2C, DNASE1L2, ZNF500, ZNF267, POLR2K, MRPL28, ZNF319, ZNF205, E4F1,
	GO:0006350	transcription	RPUSD1, NUDT21, TUFM, RPL3L, EIF3C, TRAF7, RSL1D1, RNPS1, RPS2, TBL3, POLR3E, ZNF720, C16ORF33, RPS15A, SRRM2,
	GO:0010467	gene expression	ZNF200, RRN3,

\*Gene names in bold indicate genes which have previously been associated with camptothecin-induced cytotoxicity.

## REFERENCES

1. Bjornsti, M.A., et al., *Expression of human DNA Topoisomerase-I in yeast cells lacking yeast DNA Topoisomerase-I: Restoration of sensitivity of the cells to the antitumor drug camptothecin*. *Cancer Research*, 1989. **49**(22): p. 6318-6323.
2. Eng, W.K., et al., *Evidence that DNA topoisomerase I is necessary for the cytotoxic effects of camptothecin*. *Molecular Pharmacology*, 1988. **34**(6): p. 755-760.
3. Hsiang, Y.H., et al., *Camptothecin induces protein-linked DNA breaks via mammalian DNA Topoisomerase-I*. *Journal of Biological Chemistry*, 1985. **260**(27): p. 4873-4878.
4. Zhang, P., et al., *MEPE/OF45 as a new target for sensitizing human tumour cells to DNA damage inducers*. *Br J Cancer*. **102**(5): p. 862-6.
5. Shimoyama, T., et al., *Reference profiling of the genomic response induced by an antimicrotubule agent, TZT-1027 (Soblidotin), in vitro*. *Pharmacogenomics J*, 2006. **6**(6): p. 388-96.
6. Pommier, Y. and M. Cushman, *The indenoisoquinoline noncamptothecin topoisomerase I inhibitors: update and perspectives*. *Mol Cancer Ther*, 2009.
7. Antony, S., et al., *Differential Induction of Topoisomerase I-DNA Cleavage Complexes by the Indenoisoquinoline MJ-III-65 (NSC 706744) and Camptothecin: Base Sequence Analysis and Activity against Camptothecin-Resistant Topoisomerases I*. *Cancer Res*, 2003. **63**(21): p. 7428-7435.
8. Strumberg, D., et al., *Synthesis of cytotoxic indenoisoquinoline topoisomerase I poisons*. *J Med Chem*, 1999. **42**(3): p. 446-57.
9. Antony, S., et al., *Cellular topoisomerase I inhibition and antiproliferative activity by MJ-III-65 (NSC 706744), an indenoisoquinoline topoisomerase I poison*. *Mol Pharmacol*, 2005. **67**(2): p. 523-30.
10. Antony, S., et al., *Novel indenoisoquinolines NSC 725776 and NSC 724998 produce persistent topoisomerase I cleavage complexes and overcome multidrug resistance*. *Cancer Res*, 2007. **67**(21): p. 10397-405.
11. Pommier, Y., *DNA topoisomerase I inhibitors: chemistry, biology, and interfacial inhibition*. *Chem Rev*, 2009. **109**(7): p. 2894-902.
12. Abecasis, G.R., et al., *Merlin--rapid analysis of dense genetic maps using sparse gene flow trees*. *Nat Genet*, 2002. **30**(1): p. 97-101.

13. Stark, A.L., et al., *Heritable and non-genetic factors as variables of pharmacologic phenotypes in lymphoblastoid cell lines*. *Pharmacogenomics J*, 2010.
14. Jordan, M.A. and L. Wilson, *Microtubules as a target for anticancer drugs*. *Nat Rev Cancer*, 2004. **4**(4): p. 253-265.
15. Duan, S., et al., *Mapping genes that contribute to daunorubicin-induced cytotoxicity*. *Cancer Res*, 2007. **67**(11): p. 5425-33.
16. Dolan, M.E., et al., *Heritability and linkage analysis of sensitivity to cisplatin-induced cytotoxicity*. *Cancer Res*, 2004. **64**(12): p. 4353-6.
17. Watters, J.W., et al., *Genome-wide discovery of loci influencing chemotherapy cytotoxicity*. *Proc Natl Acad Sci U S A*, 2004. **101**(32): p. 11809-14.
18. Huang, S., D. Ballard, and H. Zhao, *The role of heritability in mapping expression quantitative trait loci*. *BMC Proc*, 2007. **1 Suppl 1**: p. S86.
19. Wijsman, E.M., et al., *Summary of Genetic Analysis Workshop 15: Group 9 linkage analysis of the CEPH expression data*. *Genet Epidemiol*, 2007. **31 Suppl 1**: p. S75-85.
20. Kan, D., R. Cooper, and X. Zhu, *A genome-wide linkage study of GAW15 gene expression data*. *BMC Proc*, 2007. **1 Suppl 1**: p. S87.
21. Loseva, O., et al., *PARP-3 Is a Mono-ADP-ribosylase That Activates PARP-1 in the Absence of DNA*. *Journal of Biological Chemistry*. **285**(11): p. 8054-8060.
22. Zhang, H.F., et al., *Cullin 3 promotes proteasomal degradation of the topoisomerase I-DNA covalent complex*. *Cancer Res*, 2004. **64**(3): p. 1114-21.
23. Ishii, T., et al., *The Effects of S-Phase Kinase-Associated Protein 2 (SKP2) on Cell Cycle Status, Viability, and Chemoresistance in A549 Lung Adenocarcinoma Cells*, in *Experimental Lung Research*. 2004, Taylor & Francis Ltd. p. 687-703.
24. Zhang, Y.W., et al., *Implication of checkpoint kinase-dependent up-regulation of ribonucleotide reductase R2 in DNA damage response*. *J Biol Chem*, 2009. **284**(27): p. 18085-95.
25. Huang, M., et al., *Chk1 and Chk2 are differentially involved in homologous recombination repair and cell cycle arrest in response to DNA double-strand breaks induced by camptothecins*. *Mol Cancer Ther*, 2008. **7**(6): p. 1440-9.
26. Tse, A.N., et al., *CHIR-124, a novel potent inhibitor of Chk1, potentiates the cytotoxicity of topoisomerase I poisons in vitro and in vivo*. *Clin Cancer Res*, 2007. **13**(2 Pt 1): p. 591-602.

27. Furuta, T., et al., *Phosphorylation of histone H2AX and activation of Mre11, Rad50, and Nbs1 in response to replication-dependent DNA double-strand breaks induced by mammalian DNA topoisomerase I cleavage complexes*. J Biol Chem, 2003. **278**(22): p. 20303-12.
28. Wu, J., et al., *Induction of biphasic DNA double strand breaks and activation of multiple repair protein complexes by DNA topoisomerase I drug 7-ethyl-10-hydroxy-camptothecin*. Mol Pharmacol, 2002. **61**(4): p. 742-8.
29. Zheng, Y., et al., *NDRG1 is down-regulated in the early apoptotic event induced by camptothecin analogs: the potential role in proteolytic activation of PKC delta and apoptosis*. Proteomics, 2009. **9**(8): p. 2064-75.
30. Dennis, G., Jr., et al., *DAVID: Database for Annotation, Visualization, and Integrated Discovery*. Genome Biol, 2003. **4**(5): p. P3.
31. Huang, D.W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources*. Nat Protoc, 2009. **4**(1): p. 44-57.
32. Guichard, S.M., et al., *Short hairpin RNAs targeting Bcl-xL modulate senescence and apoptosis following SN-38 and irinotecan exposure in a colon cancer model*. Cancer Chemother Pharmacol, 2007. **60**(5): p. 651-60.
33. Schulze-Bergkamen, H., et al., *Bcl-x(L) and Myeloid cell leukaemia-1 contribute to apoptosis resistance of colorectal cancer cells*. World J Gastroenterol, 2008. **14**(24): p. 3829-40.
34. Hayward, R.L., et al., *Antisense Bcl-xl down-regulation switches the response to topoisomerase I inhibition from senescence to apoptosis in colorectal cancer cells, enhancing global cytotoxicity*. Clin Cancer Res, 2003. **9**(7): p. 2856-65.
35. Zhang, J., et al., *The expression of Bcl-XL, Bcl-XS and p27Kip1 in topotecan-induced apoptosis in hepatoblastoma HepG2 cell line*. Cancer Invest, 2008. **26**(5): p. 456-63.
36. Poot, M., K.A. Gollahon, and P.S. Rabinovitch, *Werner syndrome lymphoblastoid cells are sensitive to camptothecin-induced apoptosis in S-phase*. Hum Genet, 1999. **104**(1): p. 10-4.
37. Kaneda, H., et al., *FOXQ1 is overexpressed in colorectal cancer and enhances tumorigenicity and tumor growth*. Cancer Res. **70**(5): p. 2053-63.
38. Lagadec, P., et al., *Pharmacological targeting of NF-kappaB potentiates the effect of the topoisomerase inhibitor CPT-11 on colon cancer cells*. Br J Cancer, 2008. **98**(2): p. 335-44.

39. Scartozzi, M., et al., *Nuclear factor-kB tumor expression predicts response and survival in irinotecan-refractory metastatic colorectal cancer treated with cetuximab-irinotecan therapy*. J Clin Oncol, 2007. **25**(25): p. 3930-5.
40. Huang, X., F. Traganos, and Z. Darzynkiewicz, *DNA damage induced by DNA topoisomerase I- and topoisomerase II-inhibitors detected by histone H2AX phosphorylation in relation to the cell cycle phase and apoptosis*. Cell Cycle, 2003. **2**(6): p. 614-9.
41. Misri, S., et al., *Telomeres, histone code, and DNA damage response*. Cytogenet Genome Res, 2008. **122**(3-4): p. 297-307.
42. Hassa, P.O., et al., *Nuclear ADP-ribosylation reactions in mammalian cells: where are we today and where are we going?* Microbiol Mol Biol Rev, 2006. **70**(3): p. 789-829.
43. Miao, Z.H., et al., *Nonclassic functions of human topoisomerase I: genome-wide and pharmacologic analyses*. Cancer Res, 2007. **67**(18): p. 8752-61.

## ABSTRACT

The objective of this work was to demonstrate pharmacological and genomic profiling in CEPH cell lines can be used to stratify chemicals according to mechanism of action. A panel of 22 anticancer agents belonging to 8 major mechanistic classes was evaluated for cell growth inhibition in the CEPH cell lines. We propose that intraclass biological and genomic profiles will be more similar to each other than to compounds belonging to distinct mechanistic classes. Considerable phenotypic variation was observed across and within families. For each compound, cell viability at the dose closest to a population mean viability of 50%, ( $\overline{GI50}$ ) was analyzed by hierarchal clustering and resulted in clusters in agreement with the distinct modes of action. Pearson correlation coefficients (PCC) were calculated for all compounds based on cell viability at the  $\overline{GI50}$ ; while compounds within a class were highly correlated (PCC>0.65), PCC ranged from 0.00-0.55 between compound classes. The cytotoxic response to each agent was shown to be a heritable trait with genetics estimated to account for 5-60% of the observed variation in response. Genome-wide linkage analysis was then used to identify QTLs influencing the effect of each of the anticancer agents. QTLs with moderate peak LOD scores (maximum LOD scores = 1.5- 2.3) were identified which were unique to each of the mechanistic classes. Results suggest that compounds belonging to distinct mechanistic classes can be distinguished on the basis of pharmacological and genomic profiling in CEPH cell lines.

## INTRODUCTION

Predicting the sensitivity and toxicity of individual patients is important in improving the safety and efficacy of cancer chemotherapy. An approach to this end is to understand the genes that determine sensitivity. Many genes which have been indicated to influence the response to chemotherapeutic agents include drug transporters [1, 2], metabolizing enzymes [3-5], and molecular targets [6]. An analysis of genes which are already known to be involved drug action is not sufficient to explain all of the observed variation in patient response. Our understanding of the global action of an anticancer agent, all of its direct targets, indirect targets, affected cellular pathways, and proteins involved in ADME, is severely lacking. Increasing our knowledge of the global mechanism can lead to the identification of additional genes which can predict sensitivity and toxicity to chemotherapeutic agents.

Recent attempts to predict chemosensitivity and identify genes critical to drug action have involved high throughput phenotypic screening and genome-wide expression profiling in cancer cell lines. The most prominent approach enables mechanism oriented evaluation of anticancer agents in the National Cancer Institute's panel of human cancer cell lines (NCI60). Cell lines are screened for sensitivity or resistance to a compound whose mechanism is unknown. The GI50 (molar concentration inhibiting 50% cell growth) is reported for each of the cell lines and used to generate an activity profile, referred to as the compound's fingerprint. The COMPARE algorithm then enables mode of action predictions by comparing the fingerprint of a compound of interest to others of known mechanisms [7, 8]. A high degree of correlation suggests that the two agents share a similar antiproliferative mechanism [9]. Since its beginning, thousands of compounds have been evaluated for cytotoxicity in these cell lines. Compounds have been successfully classified by their mechanisms of action by hierarchal clustering using their activity profiles [1]. Recognizing that variation in genes critical to drug action might be responsible for the variation

in susceptibility to anticancer agents, the genome-wide expression profiles of all 60 cell lines were generated [2]. The gene expression database was combined with the activity database and Pearson correlation analysis was used to identify genes with expression patterns that showed significant correlation to patterns of sensitivity. Indeed, variations in the transcript levels of particular genes correlated well to cell line sensitivity or resistance. Subsequently, patterns of gene expression and compound activity have been used to provide incisive information about the mechanism of action of compounds of interest, and guide our understanding of which genes influence response.

While extremely powerful, the limitations of the NCI60 genomic model necessitate the exploration of additional supportive strategies for mechanism elucidation. It has been fairly controversial whether tumor cell activities can predict human patient chemotherapeutic responses [10]. Furthermore, all important tumor types were not included in the NCI-60. For example, there are no lymphomas, sarcomas, head and neck tumors, or small cell lung cancers. Even if these types could be added to the panel now, all compounds screened over the past 20 years would have to be tested again in the updated panel to gain the full predictive power of the database. Moreover, RNA levels are not always correlated with the expression data for the protein it encodes [11]. It is widely appreciated that protein levels can vary significantly among genes that have similar mRNA-expression profiles, and, that there can be a significant variation in the mRNA levels of proteins that are expressed with comparable abundance [11, 12]. Finally, post-translational modifications, such as phosphorylation and acetylation, which may be required for drug-target interaction would not be detected. Taken together, this suggests that the linking of expression profiles and compound activity in cancer cell lines may miss meaningful information for mechanism elucidation and for predicting the susceptibility of a cell or an organism to a compound.



Additional methods for indentifying genes influencing drug action are needed to support existing mechanism elucidation strategies such as the NCI60 cell line model. An *ex vivo* familial genetics model using lymphoblastoid cell lines (LCLs) derived from multigenerational human families has been used as a model for the discovery of genes involved in the cytotoxic action of anticancer agents [13, 14]. We have illustrated that a pattern of quantitative trait loci (QTLs) associated with the activity of a mechanistic class of drugs, the camptothecins, a group of Topoisomerase 1 inhibitors, could be established (Chapter 2). Moreover, both the biological and genomic profiles of the camptothecins could be used to distinguish these compounds from Topoisomerase 2 inhibitors, a mechanistically distinct class of compounds (Chapter 2). The objective of this study was to evaluate whether this model is generally effective for stratifying anticancer agents according to mechanism of action using biological and genomic profiling.

## **MATERIALS & METHODS**

**Cell lines and Culture Conditions.** One hundred twenty-five Epstein-Barr virus-immortalized LCLs derived from 14 CEPH pedigrees (35, 45, 1334, 1340, 1341, 1350, 1362, 1408, 1420, 1447, 1451, 1454, 1459) were purchased from Coriell Cell Repositories (Camden, New Jersey). Cells were incubated in a 5% CO<sub>2</sub> atmosphere at 37°C in RPMI medium 1640 (Invitrogen, Rockville, MD) supplemented with 15% fetal bovine serum. Cells were passaged 2-3 times per week and used for experimentation at passages 3-7.

**Drugs.** For the *in vitro* drug sensitivity test 22 drugs in 8 mechanistic classes were used (summarized in Table 4-1). All the drugs were dissolved in dimethyl sulfoxide (DMSO) or water and printed on 96 well plates using a Biomek 3000 fluid dispenser robot (Beckman). The same robot was then used to generate 384 well experimental plates (Corning, Corning, NY) containing vehicles (water and 0.1% DMSO), 10% DMSO, and increasing concentrations of each drug in quadruplicate. For each drug, four different concentrations were chosen to yield the slope of the

dose-response curve (approximately 20, 40, 60, and 80% viability) and are reported in Table A4-1. Final concentrations of DMSO for all drug solutions in experimental plates were no more than 0.1%.

**Cytotoxicity Profiling.** Growth inhibition experiments were performed as previously described. Briefly, cells were plated at a density of 4000 cells in 45 ul media per well in 384 well experimental plates preloaded with drug. Following 72 h incubation, the non-toxic colorimetric dye, alamar blue, was added and fluorescence read at 96 h. The cell count at each replicate was screened for outliers, where an outlier was defined as more than a ten-fold increase or decrease in cell count of a single replicate. Cell viability (survival) relative to untreated controls was determined according to the manufacturer's protocol. The final percent survival at each concentration was averaged from four replicates of two independently plated experiments (n = 8). Additionally, growth rate in vehicle was calculated as previously described (Chapter 2).

**Hierarchical Clustering on Biological Activity.** Viability for each cell line was assessed at the dose closest to the population mean of 50%, referred to as the  $\overline{GI50}$ . Raw viability scores were z-score transformed and loaded into Cluster 3.0 (<http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/>) [15, 16]. A self organizing map (SOM) was calculated using 100,000 iterations for cell lines and 20,000 iterations for drugs to stabilize clusters and then the data clustered using uncentered correlation and complete linkage clusters. Clusters were visualized using Java TreeView.

**Genomic profiling.** Genomic profiling, which consisted of heritability estimates, linkage analysis, peak identification and peak replication analysis, was performed as described in (Chapters 2 and 3). Briefly, heritability estimates ( $h^2$ ), the degree of variation in cytotoxic response which can be explained by genetics, were calculated for each drug-dose phenotype using MERLIN [17]. Heritability estimates were calculated using cellular growth rate as a covariate.

The genotype data were downloaded from CEPH [18] and Marshfield [19] databases and used for nonparametric linkage analysis in MERLIN [17]. Permutation testing was used to identify LOD score thresholds for significant and suggestive QTLs for each drug-dose combination at each chromosome (Chapters 2 and 3). A QTL was defined as significant if the maximum LOD score within that region surpassed the LOD score significance threshold for that drug-dose combination on that chromosome. Regions identified as significant peaks for a given drug-dose combination were evaluated for replication at the suggestive or significant level for all dosages of that drug. A region was considered replicated in another dose if any LOD score in that region surpassed the suggestive LOD score threshold. Finally, to establish a pattern of QTLs specific to a mechanistic class, QTLs which were identified as significant in one drug were queried for replication at the suggestive level in all drug-dose combinations belonging to that mechanistic class.

## **RESULTS**

**Biological activity profiling.** CEPH cell lines ( $n = 125$ ) belonging to 14 pedigrees were tested for sensitivity to 22 anticancer agents belonging to 8 mechanistic classes. With the exception of the camptothecins, four doses were chosen for each drug to capture the anticipated most variable region within the linear portion of the sigmoid curve. Boxplots illustrating variation across individual cell lines at each drug-dose phenotype can be found in Figure A4-1. The antimicrotubule drugs vincristine, docetaxel, and vinorelbine were only active at the highest concentrations investigated. The anthracyclines, TYMS inhibitors, Top1 and Top2 inhibitors were highly effective across all concentrations. The observed sensitivity patterns are consistent with those reported previously in lymphoblastoid cell lines with a few exceptions [13, 20, 21]. Bleomycin, vinblastine, and topotecan were reported as inactive in healthy normal lymphoblast; however, the concentrations used were 10-1000 fold lower than those used in this study [20].

Typical studies stratifying drugs based on their biological activity in a cell panel use the GI50 [22-28]. However, a GI50 could not be determined for 61% of the drugs in at least 93 (75%) of the CEPH cell lines. Hierarchical clustering analysis was performed using the dose which yielded a population mean viability closest to 50% for each drug, which will be referred to as  $\overline{GI50}$  (Table A4-1). The mean viability for each cell line was compared at the  $\overline{GI50}$  of each drug. A boxplot of cell viability at the  $\overline{GI50}$  for the anticancer agents shows that the data appears to be evenly distributed (Figure 4-1). The 90<sup>th</sup> and 10<sup>th</sup> percentile viability thresholds were calculated for each drug. Cell lines which had a mean viability greater than the 90<sup>th</sup> percentile for a drug at  $\overline{GI50}$  were identified as resistant. Similarly, cell lines with a mean viability less than the 10<sup>th</sup> percentile were considered sensitive. Some families and individuals were consistently sensitive or resistant to the test compounds (Table A4-2). Pedigree 1408 has six out of 10 members who were sensitive to more than 50% of the drugs and they were consistently sensitive to the same drugs. Five out of 10 members of pedigree 35 are also sensitive to approximately 50% of the anticancer drugs studied and resistant to another 45%. The individual CEPH cell lines 12766 and 12767 (pedigree 1451) was resistant to 16 (72%) of the drugs.

An analysis of z-score transformed viabilities at  $\overline{GI50}$  by hierarchical clustering on drugs and cell lines was performed (Figure 4-2). Most anticancer agents tested clustered by their mechanistic class. The camptothecin analogues all clustered together and share the same molecular target, Topoisomerase 1 (Top1). A cluster was generated which contained the anthracyclines (daunorubicin, doxorubicin, idarubicin, and epirubicin) and the podophyllotoxins (etoposide and teniposide). The podophyllotoxins are Top2 inhibitors and the anthracyclines are DNA intercalators which are also reported to inhibit Top2 [29]. Agents which act on microtubules, specifically docetaxel, paclitaxel, vinorelbine and vincristine, all belong to the same cluster. Interestingly, vinblastine, a microtubule destabilizing agent, sorts distinctly from the other

antimicrotubule drugs. This is in accordance with the boxplot shown in Figure 4-1. Bleomycin which is mechanistically distinct from the other agents is in a cluster of its own. Floxuridine exceptionally belonged to the cluster bearing the antimicrotubule agents, not the cluster with 5FU, although both are considered thymidylate synthase (TYMS) inhibitors. Pearson correlation coefficients (PCC) were calculated for all compounds based on cell viability at  $\overline{GI50}$ ; while compounds within a class were highly correlated (PCC>0.65), PCC ranged from 0.00-0.55 between compound classes (Figure 4-3).

**Genomic profiling.** The family data demonstrates a genetic component in the explanation of inter-individual differences for sensitivity to chemotherapeutic agents. Growth-rate adjusted heritability estimates were calculated for each drug-dose phenotype and ranged from 0-64% (Table 4-2, Table A4-3). The heritability estimates for cellular growth rate did not exceed 14%. Estimates for 5-fluoruracil, docetaxel, and daunorubicin are comparable to previously reported values (Table 4-2). Heritability estimates for compounds within a mechanistic class were also similar. For example, heritability estimates for doses within the linear portion of the sigmoid curve averaged  $23.1 \pm 2.6$  % for the Top1 inhibitors.

Nonparametric linkage analysis was used to identify QTLs for each drug-dose phenotype. Permutation testing was used to identify statistically significant LOD score thresholds for each drug-dose phenotype on each chromosome. Thresholds indicating suggestive linkage across each chromosome for each drug-dose combination were also calculated using permutation testing. The average significant and suggestive LOD score thresholds for drug-dose phenotypes across all chromosomes were 1.43 (range: 0.94-1.74) and 0.59 (range: 0.46-0.64) respectively. Peaks identified as significant at one drug-dose phenotype were queried for replication at the suggestive level in other doses of that drug. A peak was considered replicated if the maximum LOD score in that region surpassed the suggestive threshold for that drug-dose phenotype on that chromosome.

To establish a pattern of QTLs related to a mechanistic class, peaks identified as significant for one member of a class were queried at the suggestive level for all other drug-dose phenotypes within that mechanistic class.

Testing for overrepresentation of QTLs within mechanistic classes suggests that compounds which share similar chemical structures shared similar genomic profiles (Figure 4-4). A list of all significant QTLs can be found in Table A4-4. Genome-wide patterns of QTLs were distinct for each mechanistic class. For example, QTLs in a similar location on chromosome 3, 5, 14, and 19 were shared by the platinum analogs, carboplatin and oxaliplatin (Figure 4-5 and Figure 4-6). This pattern was distinct from the shared QTLs on chromosomes 1 and 11 for the tubulin stabilizing agents, paclitaxel and docetaxel. (Figure 4-4) depicts the overall patterns of significant and suggestive QTLs for each drug by mechanistic class. Significant QTLs identified for cellular growth in each vehicle did not overlap with any significant drug QTLs. In addition, with the exception of the camptothecins all drug classes shared a broad QTL on chromosome 7. Chemical-specific QTLs were also identified as significant to one drug within a mechanistic class. For example, 2 out of 4 doses of oxaliplatin possessed a significant QTL on chromosome 12 which is not shared by carboplatin. Similarly, carboplatin has a significant QTL on chromosome 6 which was replicated in all four doses but not present in oxaliplatin.

## **DISCUSSION**

A system for determining key molecular targets and genes related to the activity of chemicals using a panel of cancer cell lines was first developed in the National Cancer Institute [7, 9, 22, 30]. It was later suggested that a good correlation between a drug's mechanism of action and its fingerprint, biological and/or genomic profile(s), could be observed in any panel of cell lines with diverse chemosensitivities [28]. We explored the potential of natural genetic variation within the CEPH cell lines to stratify anticancer agents according to mechanism of action using biological

and genomic profiling. We examined the antiproliferative activity of 22 anticancer agents against 125 CEPH LCLs derived from 14 pedigrees and observed differential activity across the whole cell panel. Linkage analysis revealed genomic regions related to the observed inter-individual differences in sensitivity to each drug. Hierarchical clustering on the biological activity profiles in CEPH classified drugs with a similar mode of action (such as a tubulin binders or Top1 inhibitors) into the same cluster (Figure 4-2), which were the same as the clusters established for NCI60 [9, 22]. Moreover, patterns of QTLs shared among compounds belonging to the same mechanistic class were distinct from compounds belonging to other mechanistic classes.

While we have identified class-specific QTLs, we note that for most mechanistic classes significant linkage peaks were not observed in the genomic region bearing the primary targets. This is not alarming. Peters et al. commented on the failure of an earlier linkage analysis study of 5-fluoruracil toxicity to identify a significant linkage peak on chromosome 18 around thymidylate synthetase (TYMS), the presumed primary target of 5FU [13, 31]. Genetic variants in TYMS were subsequently correlated to cytotoxic response in a subset of the HAPMAP LCLs [13, 31]. The region containing TYMS may not have been identified because of the density of genotype data used in linkage analysis and low to moderate effect sizes of genetic variants TYMS. The genotype density improved when going from the microsatellite markers used in the preliminary linkage analysis study of 5FU to the SNP data available for HAPMAP cell lines; the HAPMAP data enabled the detection of an association between 5-FU cytotoxicity and TYMS.

QTLs common to multiple classes as well as chemical-specific QTLs were also observed. QTLs common to multiple classes such as the one present on chromosome 7 are likely to possess genes which are not drug specific but are important in cell proliferation, survival and death. It is also plausible that distinct but closely linked genes are contributing to the cytotoxic response to these drugs. We have also observed QTLs which are unique to compounds within a mechanistic class, ie. chemical-specific QTLs. It is generally accepted that subtle changes in structure within a

mechanistic class do result in changes in the metabolism, antitumor profiles, mechanisms of resistance, and even molecular targets [32, 33]. While highly correlated, oxaliplatin clustered away from carboplatin according to biological activity in the CEPH (PCC = 0.69). Moreover, there were subtle distinctions in QTLs which were identified as significant to cytotoxic response (Figure 4-4). Oxaliplatin and carboplatin clustered separately according to biological activity profiles in two different cancer line panels [28, 34]. Subtle distinctions between the biological activity and genomic profiles of oxaliplatin, carboplatin, and cisplatin have been attributed to differences in mechanisms of activity and resistance [28, 34-36].

Peters et al. chose to only examine QTLs which were associated with the highest heritable doses [37]. More highly heritable traits have a greater proportion of phenotypic variation explained by genetic effect and tend to have significant linkage scores. However, drug-dose phenotypes with lower heritability estimates did have significant linkage peaks associated with response. For example, the 10 uM 9NC drug-dose phenotype ( $h^2 = 0$ ) had a significant QTL on chromosome 6 from 0-29 cM. In these cases, often the pedigrees contribute to overall significance and can explain why LOD scores can be high for lower heritability traits. Three pedigrees contribute considerably to the overall score with LODs of 0.77, 0.44, and 0.59 while the LOD scores from the other pedigrees range from -0.07 to 0.12. In many cases, these QTLs have been replicated in additional drugs within a mechanistic class. For example, the QTL on chromosome 6 for 9AC is replicated at the suggestive level in multiple doses of 5 out of six of the camptothecins. Earlier reports involving the identification of genomic regions influencing the response to anticancer agents in CEPH, have observed differences in QTL patterns for a single drug at different doses [13, 21]. It is certainly plausible, given different mechanisms may predominate at different doses. For example, genes related to cell survival mechanisms may be critical at lower doses while at higher doses genes related to cell death and apoptosis might be more important. Filtering on the



highest heritable dose may result in a loss of important information regarding drug mechanism. All significant QTLs identified for all drug-dose phenotypes were used in this study.

Peak position (start and end cM) varied as much as 5-20 cM for QTLs which were considered significant and replicated across numerous drug-dose combinations within a class. For example, significant QTLs on chromosome 20 were detected in at least four of the six camptothecins and also replicated in all six camptothecins at the suggestive level. Peak start and end positions for these peaks varied, but shared considerably overlap (Table 4-3). Consequently, the whole region from 40-101 cM is considered shared and replicated by the camptothecins as a class. With such variation in the position the question becomes: (a) Can these findings result from the same gene or different genetic factors and (b) can we truly consider linkage at these QTLs “replicated” multiple times? This issue is not limited to our study. Many independent studies of complex traits report evidence for linkage in nearby regions on the same chromosome: chromosome 6p for schizophrenia [38-40] and chromosome 12p for Alzheimer disease [41-43].

Substantial variation in the location of a linkage signal has been shown in simulations for complex phenotypes even when linkage was the result of a single gene [44]. The authors studied frequency distributions and the variation in peak location when increasing family size in a study. As the number of families increased, the distribution becomes taller and more narrow (ie. variation in peak location decreases and the frequency of observations increases in the central part of the distribution). In fact, variation in position covered as much as 10s of cM with family sizes of 200 and 400. Our own linkage study was performed with considerably less than 200 families ( $n = 14$ ); the number of families picked for this study was powered to detect heritability not linkage. The degree of variation observed is considered consistent with linkage studies where the effect size is weak. Moreover, every genetic mapping study is specific to the environment in which the experiment was conducted. Despite our best efforts to control environmental variability, there are environmental factors which can result in differences in cytotoxic response

[45, 46]. This emphasizes the need for complimentary studies to narrow down these chromosomal regions of interest identified in our linkage studies. On the other hand, it seems at least plausible that QTLs identified across all compounds within a mechanistic class may in fact be the result of the same gene despite variation in location.

Taken together these results are encouraging. Preliminary evidence suggests that the biological and genomic profiles established in the CEPH cell lines may stratify compounds by mechanism. The ultimate goal of this project is to use the biological and genomic profiles of compounds as a predictive model for drug discovery. An expansion of the database, validation of QTLs and biological profiles identified as class specific, and progress in data-mining methodology are necessary to reach this end. Additional compounds with known mechanisms can be seeded into the chemical library. For example, this process could be repeated with the NCI Developmental Therapeutics Approved Oncology Plated drug set which includes 88 compounds (some of which were covered in this study) ([http://dtp.nci.nih.gov/branches/dscb/oncology\\_drugset\\_explanation.html](http://dtp.nci.nih.gov/branches/dscb/oncology_drugset_explanation.html)). Additionally, while compounds could be stratified by mean viability at the  $\overline{GI50}$  future studies should include a full-dose response curve to ascertain GI50s for individual cell lines. An informatics model which could rapidly assess and compare patterns of QTLs associated with a drug to others previously evaluated in this system is highly desirable. Moreover, integration of the *in vitro* sensitivity data such as the GI50 or AUC associated with each cell line and each drug could further serve to stratify compounds by mechanism. These advancements would further support investigations to identify genes involved in drug mechanism which influence cytotoxic response and make predictions about the activity and action of novel compounds as they become available.

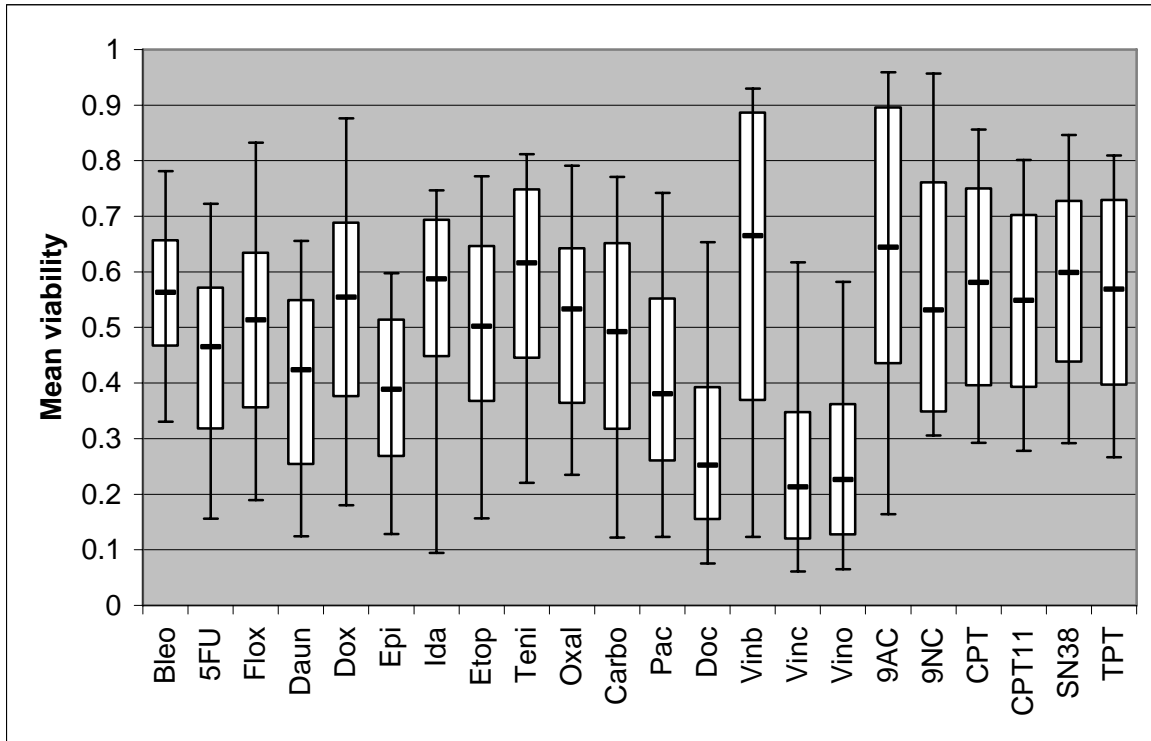
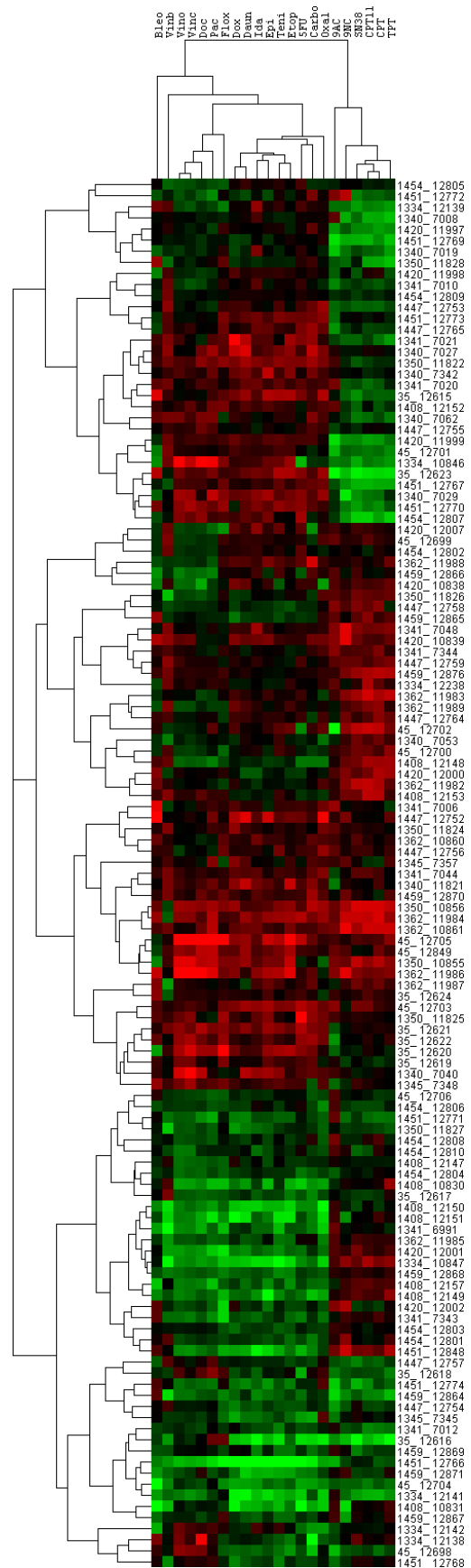


Figure 4-1. Distribution of mean viabilities for CEPH cell lines at  $\overline{GI50}$ . Line represents median viability across population, box represents 90 and 10th percentiles, and whiskers represent maximum and minimum viabilities.

Figure 4-2. Hierarchical clustering of 22 anticancer drugs based on cell viability at *GI50*. Cell viability for each cell line was z-score transformed prior to cluster analysis using the complete linkage method using the Pearson correlation as distance. On the color scale, red represents resistance (positive z-score), green represents sensitivity (negative z-score), and black color indicates Z-score = 0 (median resistance value). The brighter the color the greater the value from 0, with max brightness set at 2.5.



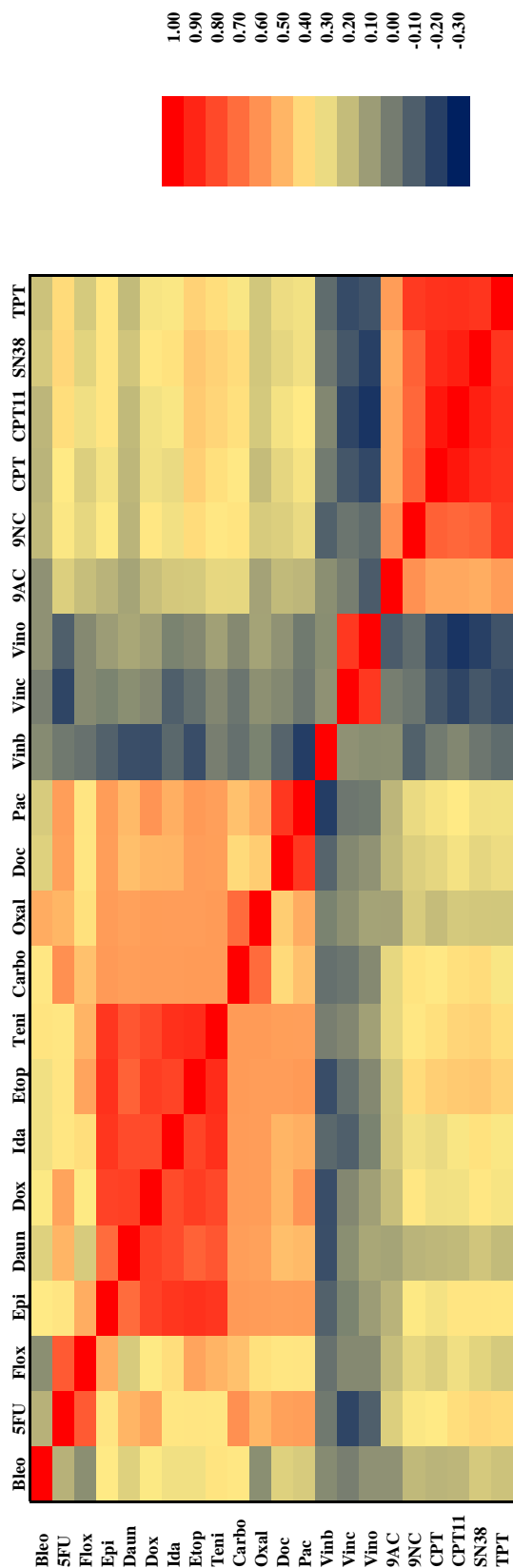


Figure 4-3. Results of a comparison analysis of biological activity profiles of CEPH cell lines for 8 mechanistic drug classes. The color gradient represents the Pearson correlation coefficient between the distinct compounds as calculated.

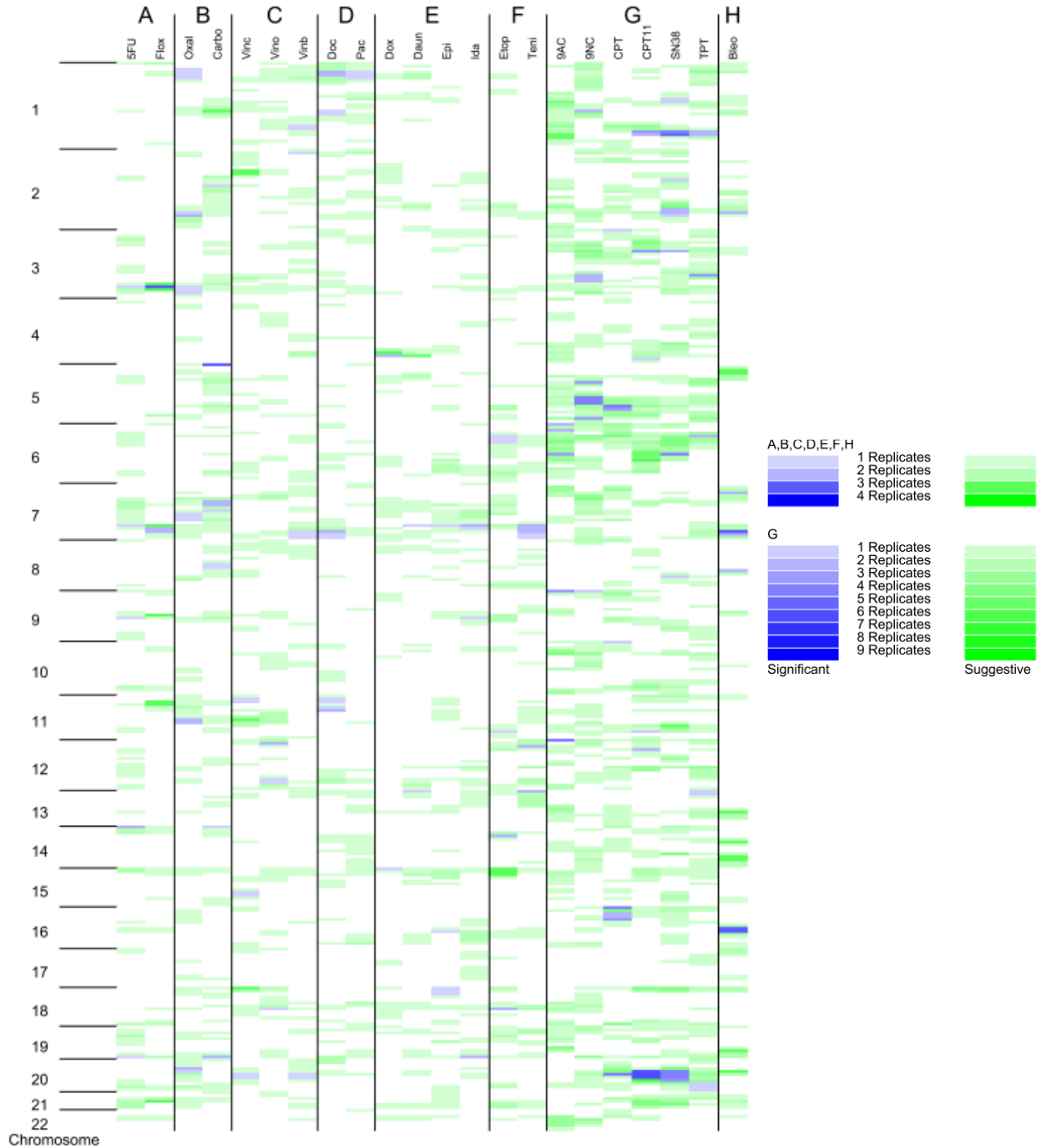


Figure 4-4. Genome wide pattern of QTLs for mechanistic set. Drugs are divided into mechanistic classes A-H as described in Table 4-1. Each chromosome was partitioned into 10 cM regions. Each drug-dose combination that resulted in a significant QTL (LOD > significance threshold) is indicated in blue. Intensity of the shading indicates the number of doses replicating that QTL at either the suggestive or significant level. Regions which also had a suggestive QTL (LOD > suggestive threshold) are indicated in green with color intensity referring to the number of doses replicating this peak.

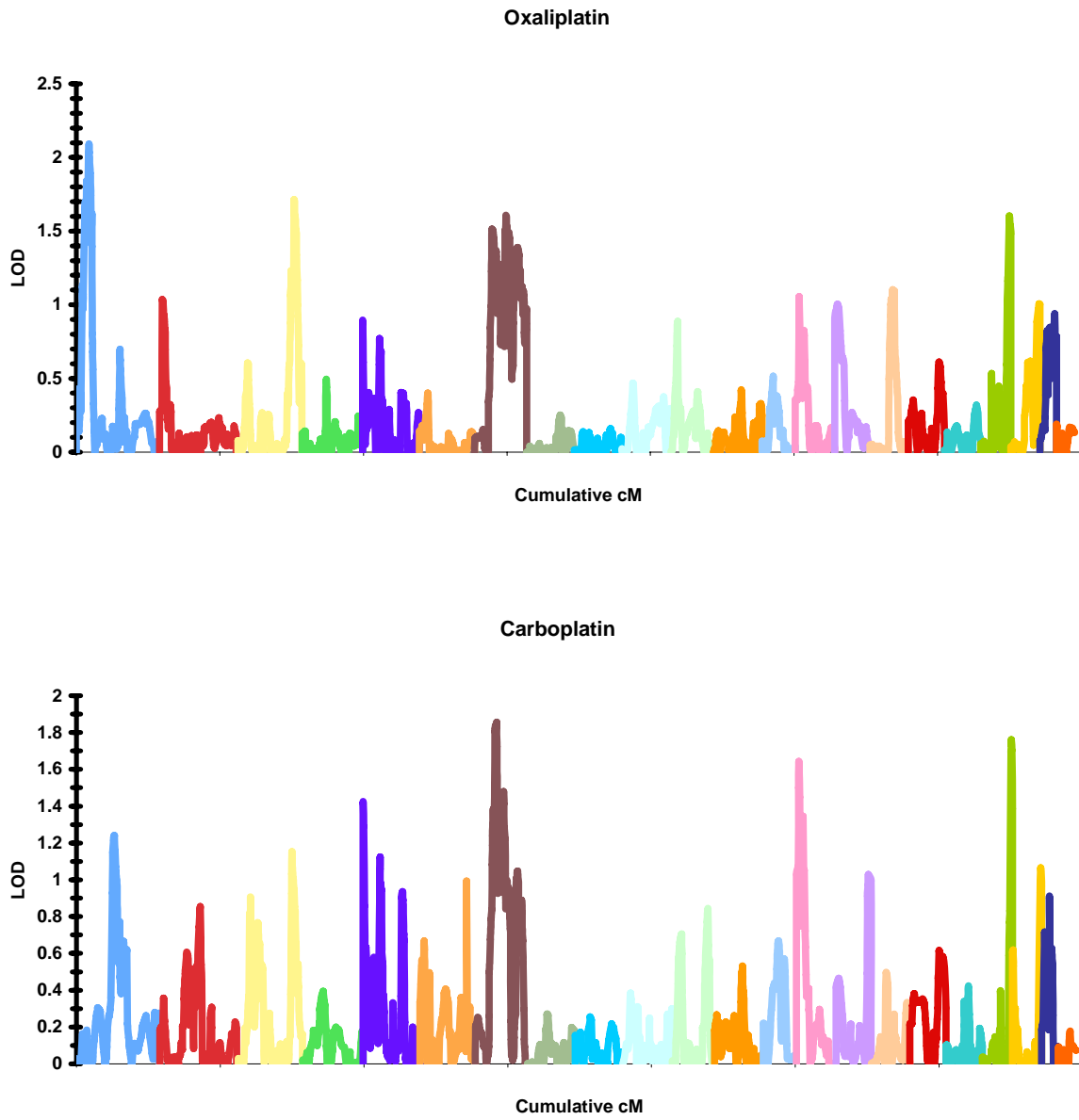


Figure 4-5. Genome-wide QTL map for oxaliplatin and carboplatin. These platinum analogs are two structurally and functionally related drugs. Each chromosome is represented by a different color.

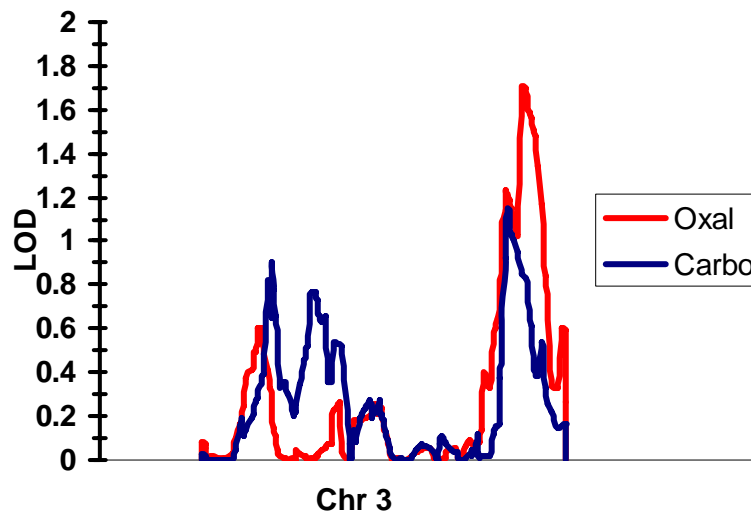


Figure 4-6. Shared QTL on chromosome 3 for oxaliplatin and carboplatin.



Table 4-1. Chemotherapeutic agents used in this study.

<b>Drug Name</b>	<b>Abbreviation</b>	<b>Mechanism of action</b>
5-Fluorouracil	5FU	Thymidylate synthase inhibitor
Floxuridine	Flox	Thymidylate synthase inhibitor
Epirubicin	Epi	DNA intercalators and topoisomerase 2 inhibitors
Doxorubicin	Dox	DNA intercalators and topoisomerase 2 inhibitors
Daunorubicin	Daun	DNA intercalators and topoisomerase 2 inhibitors
Idarubicin	Ida	DNA intercalators and topoisomerase 2 inhibitors
Vincristine	Vinc	Microtubule destabilizers
Vinorelbine	Vino	Microtubule destabilizers
Vinblastine	Vinb	Microtubule destabilizers
Docetaxel	Doc	Microtubule stabilizers
Paclitaxel	Pac	Microtubule stabilizers
Oxaliplatin	Oxal	DNA crosslinkers
Carboplatin	Carbo	DNA crosslinkers
Etoposide	Etop	Topoisomerase 2 inhibitors
Teniposide	Teni	Topoisomerase 2 inhibitors
Topotecan	TPT	Topoisomerase 1 inhibitors
Camptothecin	CPT	Topoisomerase 1 inhibitors
Irinotecan	CPT-11	Topoisomerase 1 inhibitors
7-ethyl-10-hydroxy-camptothecin	SN38	Topoisomerase 1 inhibitors
9-aminocamptothecins	9AC	Topoisomerase 1 inhibitors
9-nitrocamptothecin	9NC	Topoisomerase 1 inhibitors
Bleomycin	Bleo	Other

Table 4-2. Heritability estimates and shared QTLs for mechanistic sets.

Drug	Identified in this study		Previously reported		Ref.
	$h^2$ (%)	shared QTLs (Chr)	$h^2$ (%)	shared QTLs (Chr)	
Bleo	0.00-15.09	2,7,8,16			
5FU	0.00-30.85	3,7,9,13,19	26-65	9,16	[13]
Flox	11.41-25.54				
Epi	2.30-62.40	1-6, 8, 11, 13,16	18-63	4,16	[21]
Daun	13.40-37.12				
Dox	0.01-34.12				
Ida	3.93-48.95				
Carbo	13.51-40.57	2,3,5,7,11,20	38-47**	1,12**	[14]
Oxal	13.03-49.74				
Pac	7.77-53.24	1,7,11	21-70	5,6,9	[13]
Doc	13.45-31.17				
Etop	23.76-42.42	6,7,12,13,18			
Teni	19.94-37.72				
Vinb	0.01-35.86	1,7			
Vino	0.00-36.75				
Vinc	9.00-24.40				
9AC	0.00-22.72	1,5,6,11,16,20			
9NC	0.00-24.72				
CPT	0.00-23.49				
CPT11	0.00-23.79				
SN38	0.00-18.24				
TPT	0.00-25.92				

Growth-rate adjusted heritability estimates were calculated for each drug-dose phenotype and a listed as a range. Drugs are arranged by mechanistic class.

\*In some cases, heritability estimates were previously reported following evaluation in CEPH cell lines.

\*\*Heritability estimates for increasing concentrations of cisplatin.

Table 4-3. Variation in peak location for significant QTLs on chromosome 20 shared by camptothecins.

Drug	Dose (M)	Peak start cM	Peak end cM	LOD
SN38	8.00E-06	42.28	86.98	2.134
SN38	5.00E-06	42.28	77.75	1.731
SN38	8.00E-05	40.55	72.27	1.673
CPT11	0.002	46.71	78.29	1.704
CPT11	0.006	42.28	72.27	1.577
CPT11	0.01	42.28	72.27	1.705
TPT	1.00E-08	42.28	46.71	1.033
TPT	1.00E-08	72.91	101.22	1.749
CPT	8.00E-05	47.52	54.09	1.106
CPT	8.00E-05	55.74	61.77	1.273

## REFERENCES

1. Plasschaert, S.L., et al., *Influence of functional polymorphisms of the MDR1 gene on vincristine pharmacokinetics in childhood acute lymphoblastic leukemia*. Clin Pharmacol Ther, 2004. **76**(3): p. 220-9.
2. Yoh, K., et al., *Breast cancer resistance protein impacts clinical outcome in platinum-based chemotherapy for advanced non-small cell lung cancer*. Clin Cancer Res, 2004. **10**(5): p. 1691-7.
3. Goetz, M.P., A. Kamal, and M.M. Ames, *Tamoxifen pharmacogenomics: the role of CYP2D6 as a predictor of drug response*. Clin Pharmacol Ther, 2008. **83**(1): p. 160-6.
4. Evans, W.E., *Pharmacogenetics of thiopurine S-methyltransferase and thiopurine therapy*. Ther Drug Monit, 2004. **26**(2): p. 186-91.
5. van Kuilenburg, A.B., et al., *Dihydropyrimidinase deficiency and severe 5-fluorouracil toxicity*. Clin Cancer Res, 2003. **9**(12): p. 4363-7.
6. Lynch, T.J., et al., *Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non-Small-Cell Lung Cancer to Gefitinib*. N Engl J Med, 2004. **350**(21): p. 2129-2139.
7. Paull, K.D., et al., *Display and Analysis of Patterns of Differential Activity of Drugs Against Human Tumor Cell Lines: Development of Mean Graph and COMPARE Algorithm*. Journal of National Cancer Institute, 1989. **81**(14): p. 1088-1092.
8. Zaharevitz, D.W., et al., *Discovery and initial characterization of the paullones, a novel class of small-molecule inhibitors of cyclin-dependent kinases*. Cancer Res, 1999. **59**(11): p. 2566-9.
9. Weinstein, J.N., et al., *An information-intensive approach to the molecular pharmacology of cancer*. Science, 1997. **275**(5298): p. 343-9.
10. Hayon, T., et al., *Appraisal of the MTT-based assay as a useful tool for predicting drug chemosensitivity in leukemia*. Leuk Lymphoma, 2003. **44**(11): p. 1957-62.
11. Nie, L., et al., *Integrative analysis of transcriptomic and proteomic data: challenges, solutions and applications*. Crit Rev Biotechnol, 2007. **27**(2): p. 63-75.
12. Gygi, S.P., et al., *Correlation between protein and mRNA abundance in yeast*. Mol Cell Biol, 1999. **19**(3): p. 1720-30.
13. Watters, J.W., et al., *Genome-wide discovery of loci influencing chemotherapy cytotoxicity*. Proc Natl Acad Sci U S A, 2004. **101**(32): p. 11809-14.

14. Dolan, M.E., et al., *Heritability and linkage analysis of sensitivity to cisplatin-induced cytotoxicity*. *Cancer Res*, 2004. **64**(12): p. 4353-6.
15. Eisen, M.B., et al., *Cluster analysis and display of genome-wide expression patterns*. *Proc Natl Acad Sci U S A*, 1998. **95**(25): p. 14863-8.
16. Hoon, M.J.L.D., et al., *Open source clustering software*. *Bioinformatics*, 2004. **20**(9): p. 1453-1454.
17. Abecasis, G.R., et al., *Merlin--rapid analysis of dense genetic maps using sparse gene flow trees*. *Nat Genet*, 2002. **30**(1): p. 97-101.
18. Cohen, D., I. Chumakov, and J. Weissenbach, *A first-generation physical map of the human genome*. *Nature*, 1993. **366**(6456): p. 698-701.
19. Broman, K.W., et al., *Comprehensive human genetic maps: individual and sex-specific variation in recombination*. *Am J Hum Genet*, 1998. **63**(3): p. 861-9.
20. Markasz, L., et al., *Cytotoxic drug sensitivity of Epstein-Barr virus transformed lymphoblastoid B-cells*. *BMC Cancer*, 2006. **6**(1): p. 265.
21. Duan, S., et al., *Mapping genes that contribute to daunorubicin-induced cytotoxicity*. *Cancer Res*, 2007. **67**(11): p. 5425-33.
22. Scherf, U., et al., *A gene expression database for the molecular pharmacology of cancer*. *Nat Genet*, 2000. **24**(3): p. 236-44.
23. Perlstein, E.O., et al., *Genetic basis of individual differences in the response to small-molecule drugs in yeast*. *Nat Genet*, 2007. **39**(4): p. 496-502.
24. Dan, S., et al., *Identification of candidate predictive markers of anticancer drug sensitivity using a panel of human cancer cell lines*. *Cancer Science*, 2003. **94**(12): p. 1074-1082.
25. Dan, S., et al., *An integrated database of chemosensitivity to 55 anticancer drugs and gene expression profiles of 39 human cancer cell lines*. *Cancer Research*, 2002. **62**(4): p. 1139-1147.
26. Nakatsu, N., et al., *Evaluation of action mechanisms of toxic chemicals using JFCR39, a panel of human cancer cell lines*. *Mol Pharmacol*, 2007. **72**(5): p. 1171-80.
27. Nakatsu, N., et al., *Chemosensitivity profile of cancer cell lines and identification of genes determining chemosensitivity by an integrated bioinformatical approach using cDNA arrays*. *Molecular Cancer Therapeutics*, 2005. **4**(3): p. 399-412.

28. Yamori, T., *Panel of human cancer cell lines provides valuable database for drug discovery and bioinformatics*. *Cancer Chemother Pharmacol*, 2003. **52 Suppl 1**: p. S74-9.
29. Minotti, G., et al., *Anthracyclines: Molecular Advances and Pharmacologic Developments in Antitumor Activity and Cardiotoxicity*. *Pharmacol Rev*, 2004. **56(2)**: p. 185-229.
30. Weinstein, J.N., et al., *Neural computing in cancer drug development: predicting mechanism of action*. *Science*, 1992. **258(5081)**: p. 447-51.
31. Peters, E.J., et al., *Association of thymidylate synthase variants with 5-fluorouracil cytotoxicity*. *Pharmacogenet Genomics*, 2009. **19(5)**: p. 399-401.
32. Kubinyi, H., *Similarity and dissimilarity: A medicinal chemist's view*. *Perspectives in Drug Discovery and Design*, 1998. **9-11**: p. 225-252.
33. Kubinyi, H., *Chemical similarity and biological activities*. *Journal of the Brazilian Chemical Society*, 2002. **13(6)**: p. 717-726.
34. Fojo, T., et al., *Identification of non-cross-resistant platinum compounds with novel cytotoxicity profiles using the NCI anticancer drug screen and clustered image map visualizations*. *Crit Rev Oncol Hematol*, 2005. **53(1)**: p. 25-34.
35. Meynard, D., et al., *Functional analysis of the gene expression profiles of colorectal cancer cell lines in relation to oxaliplatin and cisplatin cytotoxicity*. *Oncology Reports*, 2007. **17(5)**: p. 1213-1221.
36. Rabik, C.A. and M.E. Dolan, *Molecular mechanisms of resistance and toxicity associated with platinating agents*. *Cancer Treatment Reviews*, 2007. **33(1)**: p. 9-23.
37. Peters, E., et al., *Pharmacogenomic characterization of FDA-approved cytotoxic drugs submitted to Cancer Cell; Manuscript No: CANCER-CELL-D-10-00201*. 2010, UNC Institute for Pharmacogenomics and Individualized Therapy.
38. Brzustowicz, L.M., et al., *Use of a quantitative trait to map a locus associated with severity of positive symptoms in familial schizophrenia to chromosome 6p*. *Am J Hum Genet*, 1997. **61(6)**: p. 1388-96.
39. Daniels, J.K., et al., *Linkage study of chromosome 6p in sib-pairs with schizophrenia*. *Am J Med Genet*, 1997. **74(3)**: p. 319-23.
40. Cao, Q., et al., *Suggestive evidence for a schizophrenia susceptibility locus on chromosome 6q and a confirmation in an independent series of pedigrees*. *Genomics*, 1997. **43(1)**: p. 1-8.

41. Pericak-Vance, M.A., et al., *Complete genomic screen in late-onset familial Alzheimer disease. Evidence for a new locus on chromosome 12*. JAMA, 1997. **278**(15): p. 1237-41.
42. Rogaeva, E., et al., *Evidence for an Alzheimer disease susceptibility locus on chromosome 12 and for further locus heterogeneity*. JAMA, 1998. **280**(7): p. 614-8.
43. Mayeux, R., et al., *Chromosome-12 mapping of late-onset Alzheimer disease among Caribbean Hispanics*. Am J Hum Genet, 2002. **70**(1): p. 237-43.
44. Roberts, S.B., et al., *Replication of linkage studies of complex traits: an examination of variation in location estimates*. Am J Hum Genet, 1999. **65**(3): p. 876-84.
45. Choy, E., et al., *Genetic analysis of human traits in vitro: drug response and gene expression in lymphoblastoid cell lines*. PLoS Genet, 2008. **4**(11): p. e1000287.
46. Zhang, W. and M.E. Dolan, *Use of cell lines in the investigation of pharmacogenetic loci*. Curr Pharm Des, 2009. **15**(32): p. 3782-95.

## **CHAPTER 5:**

### **Summary & Future Directions**



## 1. SUMMARY

The pharmaceutical industry is suffering from a severe productivity crisis. The number of new drugs entering clinical development, being submitted to regulatory agencies, and being made available to patients appears to be declining at the same time that the cost to bring therapeutics to market is escalating. Bringing a new molecular entity (NME) to market is associated with considerable expense; an estimated \$800 million is spent on drug discovery, evaluation and clinical trials for every NME which enters the market. This represents an estimated a 54% increase in the cost from 1995-2002 [1]. Unfortunately, the number of drugs reaching market is not proportionate to the exponential rise in pharmaceutical research and development costs. In fact, the last decade has seen a decline in the number of new drugs. In 2006, only 18 drug approvals for new molecular entities were granted by the FDA, an 84% drop from a peak of 110 new drugs in 1996 [2]. Finally, the drug discovery process has generally been rather slow. It has typically taken 10-15 years, and often very much longer, to progress from the start of the discovery phase to the launch of a successful new drug product. Drug development is becoming increasingly challenging, inefficient, and costly.

Many argue that the decline in productivity coincides with the industry's switch from the phenotypic to the target based drug discovery approach. The limitations of this approach are significant [2-5]. The greatest weakness is our inability to predict safety and efficacy in man, the most frequent cause of drug failure [6]. If productivity is to improve, the FDA suggests new and innovative drug discovery tools are needed:

“Not enough applied scientific work has been done to create new tools to get fundamentally better answers about how the safety and effectiveness of new products can be demonstrated in faster time frames, with more certainty, and lower costs. In many cases, developers are using the tools and concepts of last

century to assess this century's candidates. As a result, the vast majority of investigational products fail that enter clinical trials [7]."

The FDA tasks the industry with developing highly effective drugs with minimum or predictable side effects and toxicity, cheaper and faster. Consequently, tools which give us greater insight into a drug's mechanism of action, toxicity, and resistance are in high demand.

The use of genomics as a tool to identify pharmacokinetic and pharmacodynamic genes which are associated with lapses in drug safety and/or efficacy is on the rise [8-13]. However, many genes and their gene products are investigated based on *a priori* knowledge of their involvement in drug action. Our limited knowledge of drug action is insufficient and will be unable to explain all of the observed differences in toxicity and efficacy between patients. Research is ongoing to identify genomic tools which might be able to offer incisive information about a drug's global mechanism of action, proteins involved in its ADME, primary and secondary targets associated with efficacy, and all affected molecular pathways. The most prominent example of this effort is the chemogenomic database developed with the NCI. More than 100,000 compounds have been screened for biological activity in the NCI60, a panel of human cancer cell lines. Compounds have been broken down into their molecular scaffolds and their biological activity linked to the gene expression profiles of the NCI60. This enables researchers to make predictions about which key structural features will be biologically active, the genes which might be critical to their activity, and the cancers which are sensitive or resistant to active compounds.

Our lab is in the preliminary stages of developing a drug discovery tool that uses natural genetic variation in normal healthy human cell lines to make inferences about drug mechanism. The objective of Chapter 2 was to demonstrate that pharmacological and genomic profiles in the CEPH cell lines can be established for compounds sharing a mechanism of action. To this end, cell growth inhibition in 125 CEPH cell lines (14 families) was evaluated following treatment

with a panel of camptothecin analogues. Considerable phenotypic variation was observed across and within families. The cytotoxic response to each agent was shown to be a heritable trait. We estimated genetics accounts for  $23.1 \pm 2.6\%$  of the variation in response to the camptothecins. Linkage analysis was used to identify a relationship between genetic markers and the response to camptothecins. The camptothecins shared ten quantitative trait loci (QTL) on chromosomes 1, 3, 5, 6, 11, 12, 16 and 20. Nine of these QTLs were independently validated in a second set of camptothecin analogues. Moreover, the genomic profiles of the camptothecins were quite distinct from those of the podophyllotoxins, Top2 inhibitors.

In Chapter 3, we sought to independently validate these QTLs and identify which might be associated with the general mechanism of Top1 inhibition versus camptothecin-specific induced cytotoxicity. We repeated the phenotyping study for seven of the original camptothecins from Chapter 2. Seven of the nine QTLs were reproduced. CEPH were then phenotyped for sensitivity to the indenoisoquinolines, Top1 inhibitors, and nonparametric linkage analysis was used to identify QTLs associated with response. Four QTLs were shared between all of the camptothecins and the indenoisoquinolines; the remaining three were specific to the camptothecins. Eight were linked to the indenoisoquinolines alone.

In Chapter 4, we tested whether specific patterns of biological and genomic profiles could be defined for mechanistic classes of compounds. The goal was to evaluate whether pharmacological and genomic profiles in CEPH cell lines could be used to stratify drugs by their mechanism. Twenty-two anticancer agents belonging to eight mechanistic classes were evaluated for cellular sensitivity in a panel of CEPH cell lines. Heritability estimates ranged from 5-64%. Then, linkage analysis was used to correlate variation in response to variation in regions of the genome for each drug investigated. We were able to show using cluster analysis, correlation coefficients, pattern recognition, and principle components analysis that these biological and genomic profiles in CEPH are unique to each mechanistic class.

Taken together, we have shown illustrative studies in which this model can be used to (i) establish a pharmacological and genomic profile for compounds sharing a mechanism, (ii) refine regions of the genome influencing drug activity to those which are associated with the general mechanism and those which are class-specific and (iii) stratify compounds by mechanism using biological and genomic profiling. Our results may have a positive impact on the drug discovery process which has been tasked with developing safer and more effective drugs by cheaper and faster means. The conventional phenotypic based screening approach to drug discovery is hindered tremendously by target identification which can take as much as 10 years or more [14]. Likewise, the ability to predict off target effects for small molecules discovered through the target-based approach is limited [2]. Without a clear understanding of mechanism, drugs derived from these routes have an increased risk of failure due to unanticipated toxicity. The nature of our assay permits mechanism elucidation by simultaneously screening of all possible genes and pathways in a single experiment, consequently streamlining the drug discovery process. Moreover, the genomics component of this strategy makes it possible for us to make predictions about which patients would benefit or be harmed by a given therapy prior to its use in costly clinical trials.

## **2. FUTURE DIRECTIONS**

Regardless of the method used to identify potential genes and proteins involved in drug activity, the most important step is to confirm their involvement in mechanism. In Chapters 2 and 3, QTLs likely to contain the genes important for the observed differences in response to all Top1 inhibitors, were identified using genome-wide linkage analysis. This method has narrowed the search for the causative genes from the entire genome to a significantly smaller and more manageable subset. Specifically, the effects of camptothecins and indenoisoquinolines were mapped to chromosomes 1, 5, 11, and 16. By narrowing these regions of interest further to genes

which are expressed in the CEPH cell lines, an unbiased candidate gene list can be formed (Table 3-5). The whole genome linkage studies, such as those described in this dissertation, require family based samples, are typically powerful for genes with large effect sizes and identify large regions of the genome linked to the phenotype of interest. Alternatively, association studies use a large collection of unrelated individuals are powerful for detecting genes with small effect sizes and will pinpoint small intervals of the genome associated with a phenotype. Cytotoxic response to chemotherapeutic agents is dependent on multiple genes, each with have a small independent effect. Follow-up genome wide association studies (GWAS) in an independent sample could not only be used to independently replicate regions of the genome identified through linkage analysis, but to minimize false discovery, and narrow these broad QTLs down to smaller intervals. Both the Children's Hospital of Oakland Research Institute (CHORI) ([http://www.chori.org/Principal\\_Investigators/Krauss\\_Ronald/krauss\\_research.html](http://www.chori.org/Principal_Investigators/Krauss_Ronald/krauss_research.html)) and the National Heart, Lung, and Blood Institute's Framingham Heart Study (<http://www.framinghamheartstudy.org/participants/index.html>) have collected and genotyped large collections of lymphoblastoid cell lines from unrelated Caucasian individuals which could be used for drug response profiling and genome wide association studies.

Any candidate genes identified through GWAS or from linkage analysis which are attributed to the general mechanism of Top1 inhibition and camptothecin-specific induced cytotoxicity (Table 3-5) can also be pursued using RNA interference (RNAi). Functional validation of candidate genes from CEPH studies have previously been reported using gene knockdown by RNAi. For example, in a study to identify genes influencing sensitivity to carboplatin knockdown of CD44 expression through small interfering RNA (siRNA) resulted in increased cellular sensitivity in CEPH cell lines [15]. Also in the CEPH model, down regulation of NT5C3 and FKBP5 by siRNA was associated with altered sensitivity to gemcitabine and cytosine arabinoside (AraC) respectively [16]. Our lab is working to develop a high throughput method to study all potential

genes under QTLs of interest using gene expression knockdown with short hairpin RNAs (shRNA). Additionally, individual CEPH LCLs used in this study which were highly sensitive and resistant to the camptothecins (upper and lower 10% of IC50s) have been identified (Table A2-3). Candidate genes identified from gene knockdown experiments can be sequenced in this subpanel to ascertain whether genetic variants correlate to the observed variation in response. We should also examine whether gene expression correlates to genotype using real-time quantitative RT-PCR. This same process can be applied to the QTLs unique to the other mechanistic classes in this study.

The ultimate goal of this project is to use the biological and genomic profiles of compounds as a model for getting at the mechanism of action for drugs. In Chapter 4, the preliminary ground work was laid for this objective. An expansion of the database (ie. more drugs), validation of QTLs and biological profiles identified as class specific, and progress in data-mining methodology are necessary to reach this end. Currently, investigations of genomic relationships with cellular sensitivity are often completed by analyzing a summary measure of the dose-response curve such as the GI50. A considerable number of significant QTLs were identified across multiple doses of each drug investigated, more than were identified by linkage analysis for variation in GI50 for each drug. An informatics model which could rapidly assess and compare patterns of significant QTLs associated with all doses of a drug to the others previously evaluated would be helpful. Moreover, a number of systems have been described which summarize the biological performance of a compound in a cell line panel by a single parameter such as the GI50. Investigating other parameters such as slope or area under the curve (AUC) might lend more insight into drug action. Not only might they prove ideal biological parameters for classifying drugs according to mechanism, but linkage analysis on slope or AUC may identify genomic regions directly related to the differences in therapeutic indexes (ratio of amount of drug required to produce a therapeutic effect to amount which causes a toxic effect) among individuals. Both

slope and AUC would also be rapidly amenable to database generation for biological profiling of compound libraries.

In addition to validating the reported genomic regions and genes influencing the activity of the drugs in this investigation, some technical limitations of this study need to be addressed. The technique used to immortalize LCLs has the potential to confuse results. LCLs used in this study are derived from normal B-lymphocytes which have been immortalized using the Epstein Barr virus (EBV), a member of the herpesviruses family. EBV is not unusual as it infects over 90% of the human population and persists in the latent phase for the lifetime of the host (<http://www.cdc.gov/ncidod/diseases/ebv.htm>). EBV genes critical for the immortalization of B cells have been identified and extensively studied. Some of these gene products can induce differential expression in host cells [17]. To date, it is unclear whether EBV transformation has affected gene expression and drug sensitivity in CEPH LCLs. Welsh points out natural variation in gene expression in CEPH LCLs clusters by family which suggests that gene expression is directed by genetic factors rather than EBV transformation [18, 19]. The EBV genome contains the capacity to express at least 80 gene products, only 11 of which are expressed in LCLs [20, 21]. Treatment with sodium butyrate or phorbol esters can induce the lytic cycle and expression of the EBV genome in EBV-immortalized B cells [22]. Drugs used in this study could also activate the viral lytic phase, resulting in genetic changes which may confound cellular sensitivity results [23]. The expression of EBV viral proteins has also been linked to the inactivation of cell cycle checkpoints and DNA repair machinery induced by drug exposure in B-cell lymphomas [24-26]. The apoptotic response to some but not all anticancer drugs has been altered in EBV-positive lymphomas [19, 27, 28]. More research is required to ascertain whether this also occurs in normal immortalized B-lymphocytes. Moreover, differences in sensitivity to several anticancer agents were noted pre- and post-immortalization of LCLs [29]. Clearly, this could pose a problem in the identification of genes responsible for the therapeutic activity of many anticancer agents.

A thorough investigation of the effect of EBV transformation on gene expression and drug sensitivity in the CEPH cell lines is essential. The presence of the viral proteins BZLF1 and BRLF1 indicate activation of the viral lytic phase. To rule out the contribution of EBV to cellular sensitivity, the expression levels of these genes can be quantified by real-time RT-PCR before and after treatment with increasing concentrations of each drug used in this study [30]. Treatment with sodium butyrate or tetradecanoyl phorbol acetate (TPA) activate the viral lytic phase in Daudi, Akata or Raji cell lines (EBV-positive B cell lymphomas) and could be used as positive controls. EBV-negative B lymphocytes such as, DG75 [31], BJAB, or Ramos could be used as negative controls. Epstein-Barr virus (EBV) genome-chips have also been employed to examine EBV gene expression patterns in tumor biopsies [32]. Finally, the UNC Lineberger Comprehensive Cancer Center Tissue Culture Facility uses a protocol for EBV-immortalization of B-lymphocytes which is similar to the methods outlined in publications for the collection of CEPH LCLs ([http://www.unclineberger.org/tcf/protocols\\_GPI.asp](http://www.unclineberger.org/tcf/protocols_GPI.asp)) [33]. It may be possible to perform a genome wide analysis of cellular genes differentially expressed pre and post-EBV immortalization of primary human B-lymphocytes [34].

Another potential confounder is the variation in LCL growth rate observed over time and between individuals. One study reported that EBV copy number, ATP levels, and growth rate were confounders in cellular sensitivity to a number of drugs including the anticancer agents: 5-fluouracil, methotrexate, and 6-mecapopturine [35]. While the Dolan laboratory found no correlation between EBV copy number or ATP levels and sensitivity to chemotherapeutic agents, cellular growth rate did appear to have a direct relationship with cellular sensitivity (Dolan, ME, communication IPIT visiting scholar seminar series). This was not unexpected. Many antitumor agents are preferentially toxic to rapidly proliferating cells, therefore the rate of cellular proliferation is related to cellular sensitivity [36]. Dolan et al., also reported a strong correlation between cell proliferation rate and cytotoxic response to the anticancer agents, carboplatin and



cytarabine arabinoside (ara-C) [37, 38]. Heritability estimates of cellular growth rate in CEPH LCLs across vehicles in these studies ranged from 1-14%. To account for the effect of growth rate on response to chemotherapeutic drugs, dose-response curves can be adjusted for growth rate [35].

## **CONCLUDING REMARKS**

High throughput screening in the CEPH cell lines has the potential to be an awesome predictive tool for drug discovery. Biologic and genomic profiles could be established for compounds belonging to the same mechanistic class. Results suggest that intraclass pharmacological and genomic profiles in the CEPH are more similar to each other than to compounds belonging to distinct mechanistic classes. Structurally unrelated compound belonging to the same mechanistic class produced similar but distinct profiles. These results suggest that with further work, profiling might not only suggest mechanisms of action for novel compounds (based on biological and genomic profiles), but point to the genes which are critical to drug action, and enable investigators to predict which patients will be sensitive or resistant to that drug.

## REFERENCES

1. Gilbert, J., H. Preston, and A. Singh, *Rebuilding big pharma's business model*. Business and Medical Report. *In vivo.*, 2003. **21**(10): p. 73-82.
2. Brown, D., *Unfinished business: target-based drug discovery*. Drug Discovery Today, 2007. **12**(23-24): p. 1007-1012.
3. Chanda, S.K. and J.S. Caldwell, *Fulfilling the promise: drug discovery in the post-genomic era*. Drug Discovery Today, 2003. **8**(4): p. 168-174.
4. Sams-Dodd, F., *Target-based drug discovery: is something wrong?* Drug Discovery Today, 2005. **10**(2): p. 139-147.
5. Sams-Dodd, F., *Drug discovery: selecting the optimal approach*. Drug Discov Today, 2006. **11**(9-10): p. 465-72.
6. Schuster, D., C. Laggner, and T. Langer, *Why drugs fail--a study on side effects in new chemical entities*. Curr Pharm Des, 2005. **11**(27): p. 3545-59.
7. Administration, F.a.D., *Innovation or Stagnation: Challenge and Opportunity on the Critical Path to New Medical Products*, U.D.o.H.a.H. Services, Editor. 2004.
8. Goetz, M.P., A. Kamal, and M.M. Ames, *Tamoxifen pharmacogenomics: the role of CYP2D6 as a predictor of drug response*. Clin Pharmacol Ther, 2008. **83**(1): p. 160-6.
9. Evans, W.E., *Pharmacogenetics of thiopurine S-methyltransferase and thiopurine therapy*. Ther Drug Monit, 2004. **26**(2): p. 186-91.
10. Zhou, S., *Clinical pharmacogenomics of thiopurine S-methyltransferase*. Curr Clin Pharmacol, 2006. **1**(1): p. 119-28.
11. Lynch, T.J., et al., *Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non-Small-Cell Lung Cancer to Gefitinib*. N Engl J Med, 2004. **350**(21): p. 2129-2139.
12. Takeuchi, F., et al., *A genome-wide association study confirms VKORC1, CYP2C9, and CYP4F2 as principal genetic determinants of warfarin dose*. PLoS Genet, 2009. **5**(3): p. e1000433.
13. Guo, Y., et al., *In silico pharmacogenetics of warfarin metabolism*. Nat Biotechnol, 2006. **24**(5): p. 531-6.
14. Bredel, M. and E. Jacoby, *Chemogenomics: an emerging strategy for rapid target and drug discovery*. Nat Rev Genet, 2004. **5**(4): p. 262-75.

15. Shukla, S.J., et al., *Whole-genome approach implicates CD44 in cellular resistance to carboplatin*. Hum Genomics, 2009. **3**(2): p. 128-42.
16. Li, L., et al., *Gemcitabine and cytosine arabinoside cytotoxicity: association with lymphoblastoid cell expression*. Cancer Res, 2008. **68**(17): p. 7050-8.
17. Zhang, W. and M.E. Dolan, *Use of cell lines in the investigation of pharmacogenetic loci*. Curr Pharm Des, 2009. **15**(32): p. 3782-95.
18. Cheung, V.G., et al., *Natural variation in human gene expression assessed in lymphoblastoid cells*. Nat Genet, 2003. **33**(3): p. 422-5.
19. Welsh, M., et al., *Pharmacogenomic discovery using cell-based models*. Pharmacol Rev, 2009. **61**(4): p. 413-29.
20. Allday, M.J., et al., *Epstein-Barr virus efficiently immortalizes human B cells without neutralizing the function of p53*. Embo J, 1995. **14**(7): p. 1382-91.
21. Farrell, P.J., *Epstein-Barr virus immortalizing genes*. Trends in Microbiology, 1995. **3**(3): p. 105-109.
22. Matusali, G., et al., *Inhibition of p38 MAP kinase pathway induces apoptosis and prevents Epstein Barr virus reactivation in Raji cells exposed to lytic cycle inducing compounds*. Mol Cancer, 2009. **8**: p. 18.
23. Wang, P., et al., *Topoisomerase I and RecQL1 function in Epstein-Barr virus lytic reactivation*. J Virol, 2009. **83**(16): p. 8090-8.
24. Gruhne, B., R. Sompallae, and M.G. Masucci, *Three Epstein-Barr virus latency proteins independently promote genomic instability by inducing DNA damage, inhibiting DNA repair and inactivating cell cycle checkpoints*. Oncogene, 2009. **28**(45): p. 3997-4008.
25. Barozzi, P., et al., *B cells and herpesviruses: a model of lymphoproliferation*. Autoimmun Rev, 2007. **7**(2): p. 132-6.
26. Leao, M., et al., *Epstein-barr virus-induced resistance to drugs that activate the mitotic spindle assembly checkpoint in Burkitt's lymphoma cells*. J Virol, 2007. **81**(1): p. 248-60.
27. Feng, W.H., et al., *Lytic induction therapy for Epstein-Barr virus-positive B-cell lymphomas*. J Virol, 2004. **78**(4): p. 1893-902.
28. Liu, M.T., et al., *Epstein-Barr virus latent membrane protein 1 induces micronucleus formation, represses DNA repair and enhances sensitivity to DNA-damaging agents in human epithelial cells*. Oncogene, 2004. **23**(14): p. 2531-9.

29. Sawada, K., et al., *Differential cytotoxicity of anticancer agents in pre- and post-immortal lymphoblastoid cell lines*. Biol Pharm Bull, 2005. **28**(7): p. 1202-7.
30. Iwata, S., et al., *Quantitative analysis of Epstein-Barr virus (EBV)-related gene expression in patients with chronic active EBV infection*. J Gen Virol. **91**(Pt 1): p. 42-50.
31. Ben-Bassat, H., et al., *Establishment in continuous culture of a new type of lymphocyte from a "Burkitt like" malignant lymphoma (line D.G.-75)*. Int J Cancer, 1977. **19**(1): p. 27-33.
32. Li, C., et al., *Detection of Epstein-Barr virus infection and gene expression in human tumors by microarray analysis*. J Virol Methods, 2006. **133**(2): p. 158-66.
33. Pelloquin, F., J.P. Lamelin, and G.M. Lenoir, *Human B lymphocytes immortalization by Epstein-Barr virus in the presence of cyclosporin A*. In Vitro Cell Dev Biol, 1986. **22**(12): p. 689-94.
34. Hertle, M.L., et al., *Differential Gene Expression Patterns of EBV Infected EBNA-3A Positive and Negative Human B Lymphocytes*. PLoS Pathog, 2009. **5**(7): p. e1000506.
35. Choy, E., et al., *Genetic analysis of human traits in vitro: drug response and gene expression in lymphoblastoid cell lines*. PLoS Genet, 2008. **4**(11): p. e1000287.
36. Valeriote, F. and L. van Putten, *Proliferation-dependent Cytotoxicity of Anticancer Agents: A Review*. Cancer Res, 1975. **35**(10): p. 2619-2630.
37. Hartford, C.M., et al., *Population-specific genetic variants important in susceptibility to cytarabine arabinoside cytotoxicity*. Blood, 2009. **113**(10): p. 2145-2153.
38. Huang, R.S., et al., *Genetic variants associated with carboplatin-induced cytotoxicity in cell lines derived from Africans*. Mol Cancer Ther, 2008. **7**(9): p. 3038-3046.

## **APPENDIX 1:**

### **Development and Optimization of High-Throughput CEPH Phenotyping Assay**

## INTRODUCTION

Current therapies for the majority of cancers consist of radiation chemotherapy, and small molecule therapies including DNA alkylating agents, antimetabolic agents, DNA antimetabolites, and Topoisomerase inhibitors. Unfortunately, the response to these agents varies widely. Commonly used chemotherapeutic drugs provide a cure for some, confer no therapeutic benefit or even trigger severe side effects in others. The pharmacological effect of a drug, its mechanism of action, is determined by its interplay with numerous proteins involved in pathways of drug absorption, distribution, metabolism, and effect; variants in any one of these processes may affect drug response. In addition, most proteins involved in these processes function in complex networks with several mechanisms of regulation. Being able to predict the therapeutic response and potentially life threatening toxic effects associated with a particular chemotherapeutic agent requires extensive knowledge of the direct target effects and cellular consequences associated with a drug's activity.

A multigenerational collection of lymphoblastoid cell lines derived from the Centre d'Etude Polymorphisme Humain (CEPH) have been used in a global approach to identifying genes and gene products responsible for a drug's cytotoxic effect. Waters et al. (2004) and Dolan et al. (2004) have used CEPH families to identify genes influencing the cytotoxicity activity of 5-fluorouracil and cisplatin respectively. In each case, CEPH pedigrees were phenotyped for sensitivity to an anticancer agent. Phenotyping allowed researchers to establish heritability, the proportion of variation in response resulting from a genetic component, and then regions of the genome responsible for cytotoxicity.

In this work we will use familial genetics to identify genes linked to the cytotoxicity of structurally similar pharmacologically active agents and use those genes to further elucidate mechanisms of action. We will establish a correlation between chemical substructure, mechanism

of action, and inherited response. The following is a detailed method to design, and optimize an in vitro CEPH cytotoxicity phenotyping assay. This assay will be used to conduct high throughput screens to establish mechanisms of action for large compound libraries.

## **MATERIALS AND METHODS**

**Cell lines.** Epstein-Barr virus-immortalized lymphoblastoid cells derived from 14 CEPH reference pedigrees (35, 45, 1334, 1340, 1341, 1350, 1362, 1408, 1420, 1447, 1451, 1454, 1459) were purchased from Coriell Cell Repositories (Camden, New Jersey). Cells were maintained in RPMI medium 1640 (Invitrogen, Rockville, MD) supplemented with 15% fetal bovine serum, incubated in a 5% CO<sub>2</sub> atmosphere at 37°C, and passaged 2 times per week. Exponentially growing lymphoblastoid cell lines with greater than 85% viability, at passages 3-7 were used for experimentation.

**Length of Alamar Blue Exposure and Optimal Cell Count.** Black-clear bottomed 384 well plates obtained from Corning (Corning, NY) were used for these and all other experiments. Cells were counted in the log phase of growth using a Z1 Coulter Particle Counter (Beckman, Fullerton, CA) and plated at densities ranging from 500-40,000 cells per well (45 ul/ well). Control wells of medium alone were used to provide the background signal for fluorescence readings. Alamar Blue (5 ul, Invitrogen) was then added to each well and plates incubated in 5% CO<sub>2</sub> at 37°C for 4, 6, 12, or 24 h before reading the plate using a Beckman DTX 880 fluorescence microplate reader (Ex = 535 nm and Em = 590 nm).

To evaluate the quality of the assay for high-throughput screens (HTS), the coefficient of variance (CV), signal-to-background ratio (S:B) and the Z'-factor were determined using the following formulas [1] (Zhang et al., 1999):

$$(1) \quad CV = \sigma_n / \mu_n$$

$$(2) \quad S:B = \mu_s / \mu_b$$

$$(3) \quad Z' = 1 - [(3\sigma_s + 3\sigma_b) / (\mu_s - \mu_b)]$$

where the subscript  $n$  refers to the replicate viability measures,  $s$  refers to the maximum assay signal and subscript  $b$  is the minimum signal,  $\sigma$  is the standard deviation, and  $\mu$  is the mean signal in each condition. An assay is considered robust enough for HTS when conditions result in CVs for replicates  $< 10\%$ ,  $Z' > 0.5$  and  $S:B > 5$ .

**Plate Uniformity and Signal Variability Assessment.** Uniformity tests were conducted on three types of signals varied systematically so that over all plates on a given day each signal is observed in every well. The maximum signal, cells treated with vehicle (0.1% DMSO), was considered 100% cell viability. Minimum signal or background was achieved by treating cells with 10% DMSO. Camptothecin (16 nM) was selected to give values between the maximum and minimum signals. Cells (4000 cells/well) were exposed to treatment for 72 h, alamar blue added, and the plate read at 96 h. The plate uniformity assay was performed twice over 2 consecutive days. Cell viability was assessed using the following equation:

$$(4) \quad Viability(\%) = 100 \left( \frac{RFU_{drug} - RFU_{back}}{RFU_{veh} - RFU_{back}} \right)$$

where  $RFU_{drug}$  is the average relative fluorescence units (RFU) of cells in the presence of test compounds (i.e., camptothecin),  $RFU_{back}$  is the average RFU in the presence of 10% DMSO,  $RFU_{veh}$  is the average RFU of cells treated with PBS (0.1% DMSO).

**Automated Pipetting Accuracy.** Several programs designed for the delivery of 5  $\mu$ l (the volume used for both drug and alamar blue dispensing) to each well on 384 well plate using the Biomek



3000 were compared for accuracy. For each program, fluorescein (1mM, 5 ul/well) was dispensed into all wells of two replicates plates and fluorescence read using a DTX 880 fluorescence microplate reader (Ex = 535 nm and Em = 590 nm) to determine dispensing CVs for each plate.

**Between-run, Intra- and Inter-day Variability.** To examine between-run variation, two separate 10 mM stock solutions of camptothecin in DMSO were prepared, serially diluted (0.01 nM–100 uM camptothecin, final concentration 0.1% DMSO), and added to two separate 96 well master plates. Drug solutions (5ul) from master plates were transferred in quadruplicate to 384 well plates using the Biomek 3000. Cells (4000 cells/well) were exposed to drug for 72 h, alamar blue added, and the plates read at 96 h. Percent viabilities and IC50s were compared between the different drug preparations. For intraday (or plate-to-plate) and interday (or day-to-day) variability, a single set of drug serial dilutions was prepared in a 96 well master plate and transferred to four 384 well plates. Half of the 384 well plates were used for experimentation that day. The remaining half of the 384 well plates were stored at -20 C and used the following day for interday variability assessment of the cytotoxicity assay. These experiments were performed using all camptothecins and evaluated in 3 different CEPH cell lines.

**pH and Long Term Stability.** Stock solutions were made by dissolving the pure drugs in DMSO to 10 mM. Working solutions were made by diluting the stock solutions in PBS at pH 7 or citrate-phosphate buffer at pH 3.0. Five 384 well plates were prepared which contained working solutions of camptothecin, 9-aminocamptothecin, and SN38 in both pH 7 and pH 3 buffer. Three of these plates were frozen and used for interday, week-to-week, and month-to-month stability studies. Cells (4000 cells/well) were added to two -384 well plates containing increasing concentrations of drug and incubated for 72 h. Alamar blue was added, fluorescence read after 24 h exposure, and intraday variability assessed. This experiment was evaluated using three different cell lines.

## RESULTS AND DISCUSSION

Cell-based assay optimization begins with optimizing the number of cells per well and maximizing the difference between maximum and minimum raw fluorescence signals. Optimization of specific assay conditions (drug concentrations, pH, DMSO compatibility, etc) follows. Once the assay is developed, it is assessed for HTS compatibility by testing fluorescent signal stability, drug stability, and assay variability. These are critical factors for determining throughput and capacity of the assay and the ability to reliably identify distinctions between compound sensitivity. Finally, the fully developed and optimized high-throughput assay can be used for the screening of compound libraries in CEPH cell lines.

**Cell Number and Length of Alamar Blue Exposure.** There are many commercially available vital dyes used for cell viability and cytotoxicity assays. The generic alamar blue was chosen because it is non-toxic, readily detectable by fluorescence and absorbance, and has increased sensitivity in cytotoxicity assays when compared to other vital dyes such as MTT [2]. Moreover, the lack of cell washing, media removal, and multiple pipetting steps makes the assay ideal for HTS cell viability and cytotoxicity measurements. An assay is considered robust enough for high throughput screening if  $z'$  is greater than 0.5 and S:B is greater than 5. Plating 4000 cells/well with 24 h alamar blue exposure reproducibly met these conditions.

Cells were seeded into 384 well plates at densities ranging from 500-40,000 cells per well (45  $\mu$ l/well) and fluorescence was measured following 4, 6, 12, or 24 h exposure to alamar blue. At each time point, relative fluorescence increased with increase in cell number. Higher cell densities and longer Alamar blue exposure resulted in a departure from linearity (Figure A1-1a). The data in Figures 1 and 2 were used to calculate S:B and  $Z'$ -factor values as an indicator of stability of the assay for HTS.  $Z'$ -factor and S:B improve for longer incubation at lower cell numbers leading to an increase in assay sensitivity (Figure A1-1). Growth rate for cell lines does

vary (Figure A1-1b), results in a difference in relative fluorescent intensities and has an impact on the assay Z'-factor (Figure A1-1c). The average Z'-factor for 24 h incubation with alamar blue at 5000 cells/well was 0.67 and 0.83 for the slow growing CEPH cell line 12771 and the fast growing cell line 12141 respectively. For 12771, S:B and Z'-factor improve significantly with increase in cell number at 24 h Alamar Blue exposure (Figure A1-1c).

**Plate Uniformity and Signal Variability.** Percent viability should be tight at each signal (maximum, midpoint, or minimum) across all well locations on a plate. A scatter plot of viability for each signal type against well number (Figure A1-2) can indicate sources of systematic variability such as drift or edge effects within an assay. Values are sufficiently tight across plates for mid and min signals with no apparent drift. However, signals in the outer perimeter of the plate are consistently greater than those from interior wells (Figure A1-2). These edge effects are often the result of evaporation or plate stacking. When designing drug plate formats, this information was taken into account. Negative control (10% DMSO) and the highest drug concentrations were applied to the plate perimeter and concentrations likely to yield the most variation for our study were placed in the interior.

**Automated Pipetting Accuracy and Drug Plate Preparation.** This assay depends on accurate and reproducible delivery of drug and alamar blue with a liquid handler. Because the concentrations of drugs are volume dependent, inaccurate volume transfers will directly impact assay results. We used the fluorescent dye, fluorescein, to assess dispensing reproducibility *J Biomol Screen* 2004; 9; 726. Two different programs were designed to prepare 384 well plates with 4 replicate dispenses of 5 ul. For each program, fluorescein to all wells of two replicates plates. Program 1, which aspirated 25 ul of dye and dispensed 5 ul four times to four wells, had an average CV of 18.6% for each plate. Program 2, which was a single 5 ul aspirate followed by a single 5 ul dispense over four replicate wells, resulted in average CVs of 5.3%. Programs using a

single aspirate and dispense for delivering drug or alamar blue to assay plates were subsequently used for all experiments.

**Cytotoxicity Assay and Concentration Profiling.** To identify the appropriate drug exposure time for CEPH cell lines, cytotoxicity was evaluated using the camptothecins over a wide concentration range (0.1 nM-10 uM). Cell viability was monitored after 48, 72, and 96 h exposure to drug (Figure A1-3). The 96 h drug incubation was considered appropriate for our purposes since 100% viability was observed at the lowest concentrations and the curve went below 20% at the highest concentrations for all drugs studied. For all drug studies, cells were plated and treated with drug for 72 h, alamar blue added, and the plates read at the 96 h time point.

**Assessment of Between-run, Intra- and Inter-day Variability.** If the concentration of buffer components or drug serial dilutions prepared in stock plates differs between runs then assay results could be affected. This difference will only show up when comparing one run to another run and is defined as between-run variation. To examine between-run variation, two separate 10 mM stock solutions of camptothecin in DMSO were serially diluted to the same concentrations and percent viability assessed. There was no observed difference between the two runs (Figure A1-4a). Similarly, plates prepared and drug cytotoxicity assessed on Day 1 were compared to plates thawed and used two days later (Figure A1-4b). No significant difference was observed between days.

**Effect of pH on Viability and Assay Stability.** Camptothecins have a labile lactone form that exists in a pH dependent equilibrium with the inactive carboxy form. At basic pH, the carboxy form predominates. Changing the pH of working solutions buffers from pH 7 to pH 3 was expected to increase the concentration of active camptothecin. Greater cell kill is observed at lower concentrations when pH 3 buffer is used and the dose-response curve shifts to the left when compared to pH 7 (Figure A1-5a). There appears to be little intraday variability for both working

solutions at pH 3 and pH 7 within the slope and right tail of the curves (Figure A1-5b and A2-5c). To examine long term stability of these pH 3 and pH 7 working solutions cytotoxicity was examined 2 days (Figure A1-6), 1 week (Figure A1-7), and 1 month (Figure A1-8) after -20 C storage. Both preparations appear quite stable over all several freeze thaws and time frames studied. For HTS, working solutions prepared using citrate-phosphate buffer at pH 3 will be used.

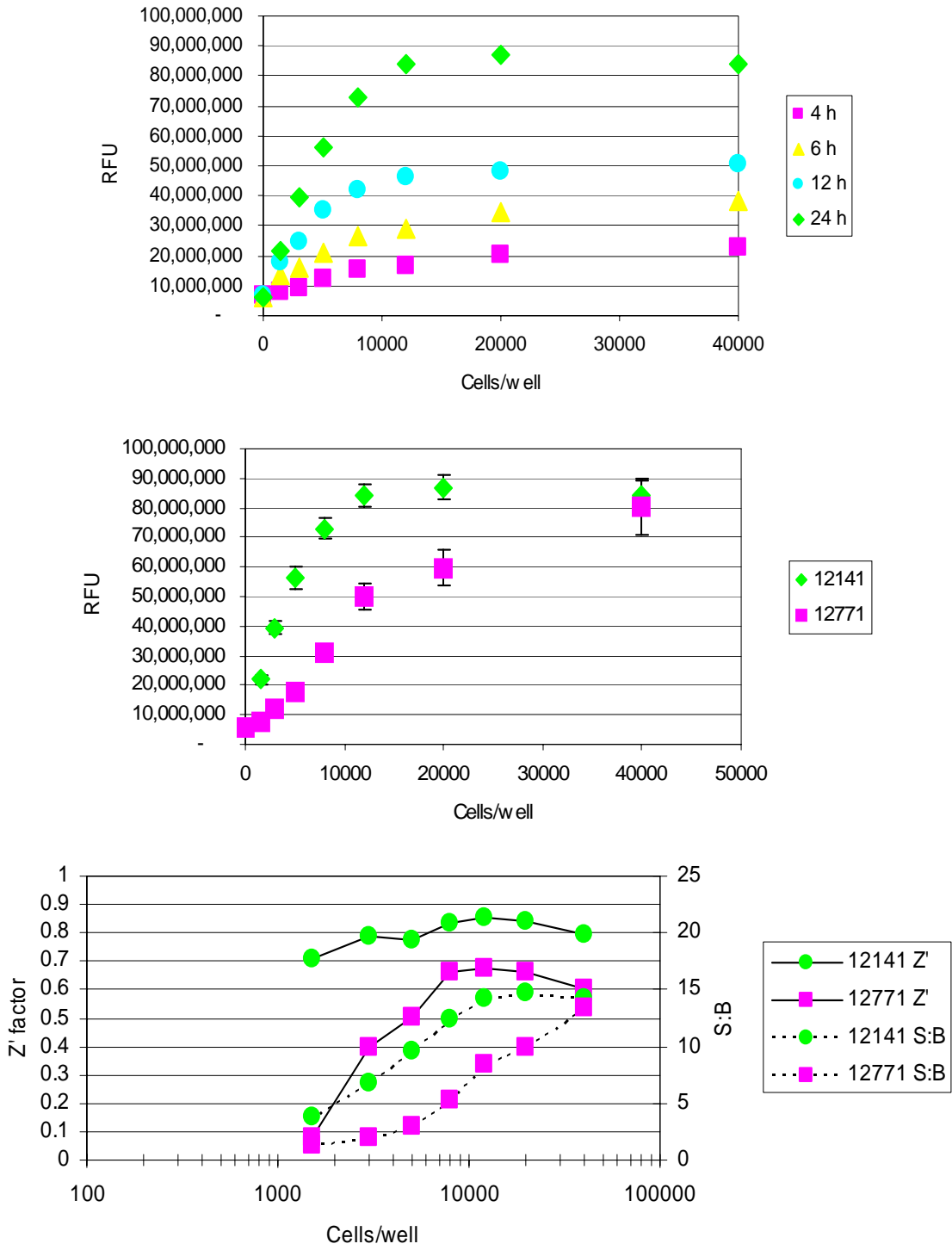


Figure A1-1. Optimizing CEPH Phenotyping assay. (A) RFU as a function of cells plated per well following 4, 6, 12, or 24 h exposure to alamar blue. (B) Fluorescence signal following 24 h exposure to alamar blue for fast (12141) and slow growing (12771) cell lines. (C) Z'-factor and S:B values as a function of cell number.

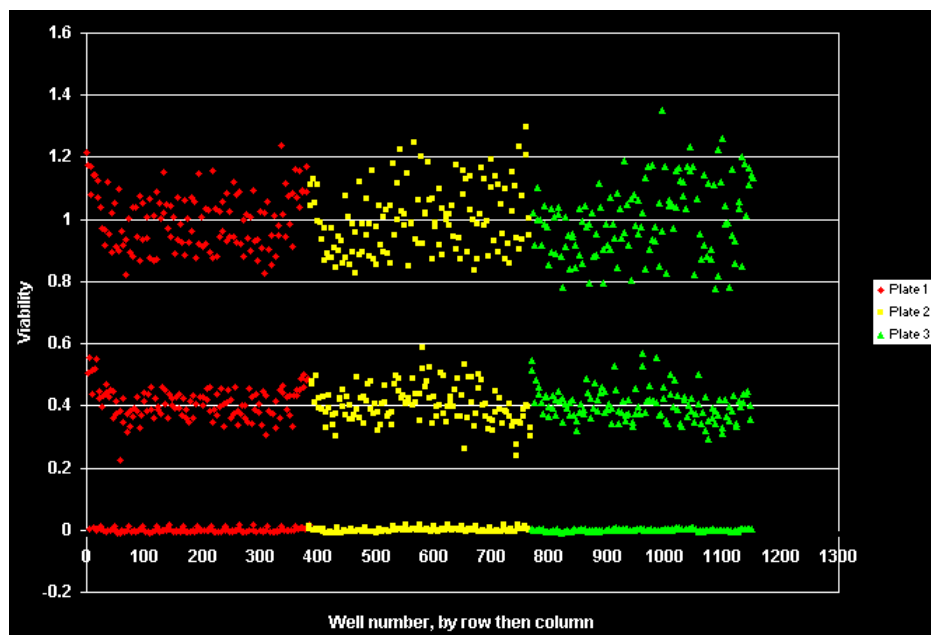


Figure A1-2. Optimizing CEPH Phenotyping assay. Plate Uniformity and Signal Variability. Cell viability (%) following 96 h exposure to vehicle, drug, or 10% DMSO. All treatments were applied such that signal for every treatment was observed in every well on a 384 well plate (n=3 plates).

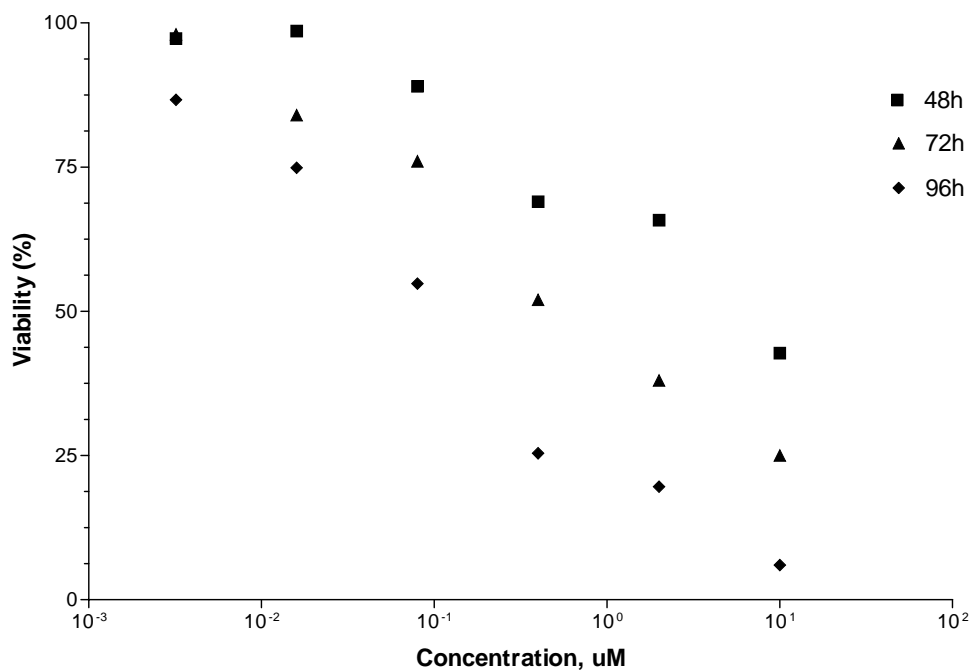


Figure A1-3. Optimizing CEPH Phenotyping assay. Identifying duration of Drug exposure. Cell viability (%) following 48, 72, or 96 h exposure to increasing concentrations of camptothecin. The same cell line was used in each case.



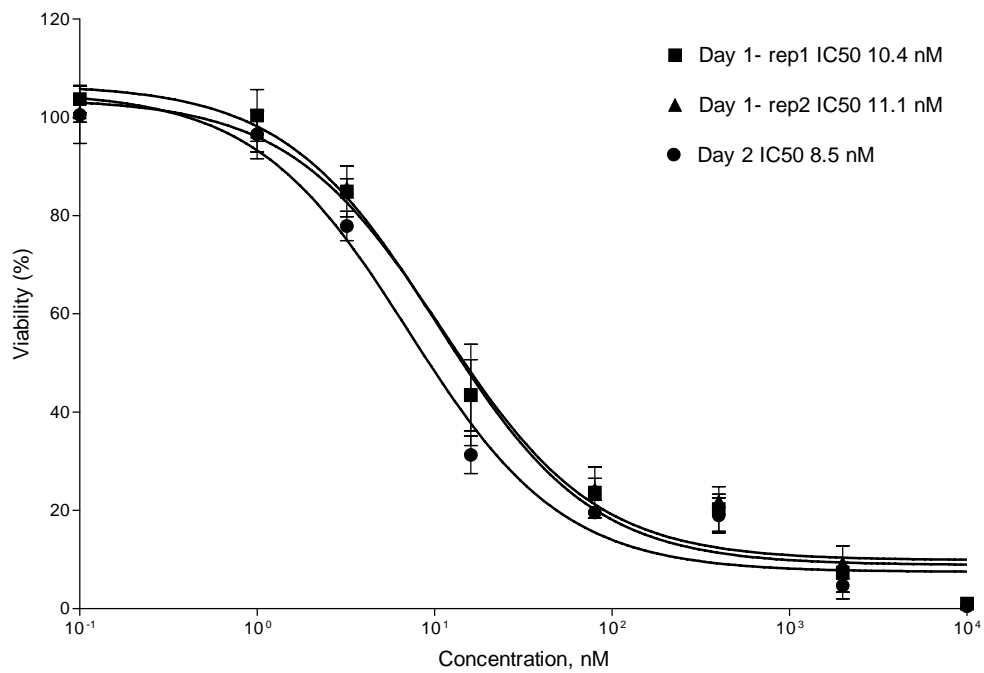
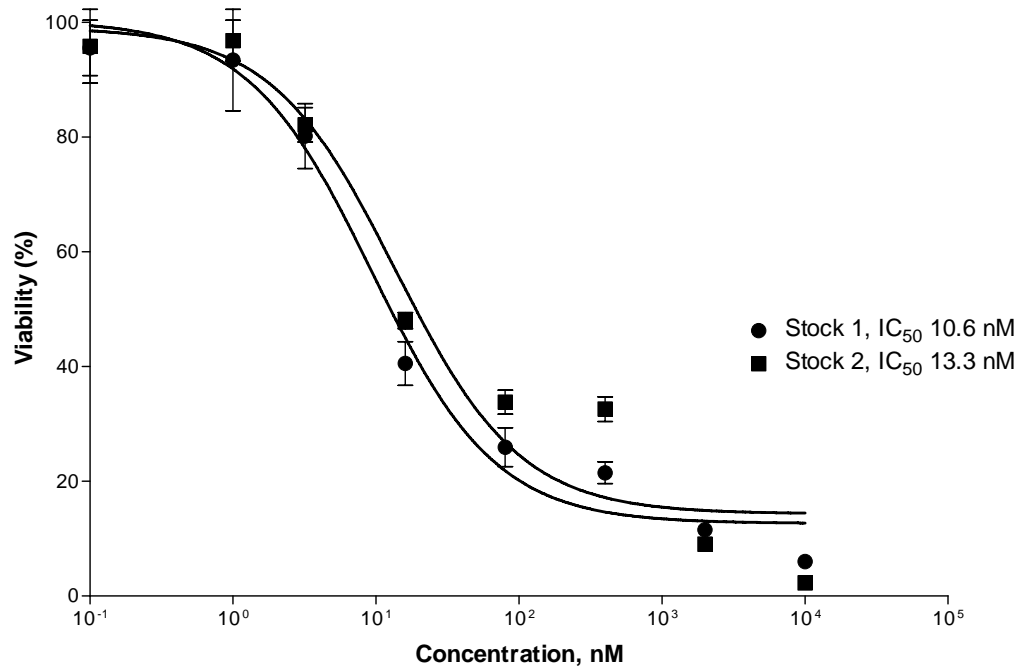


Figure A1-4. (A) Between-run variability in response to camptothecin ( $r = 0.95$ ). (B) Comparison of intra- and interday variability in cytotoxic response to increasing concentrations of camptothecin. (intraday  $r = 0.98$ ; interday  $r = 0.96$ ).

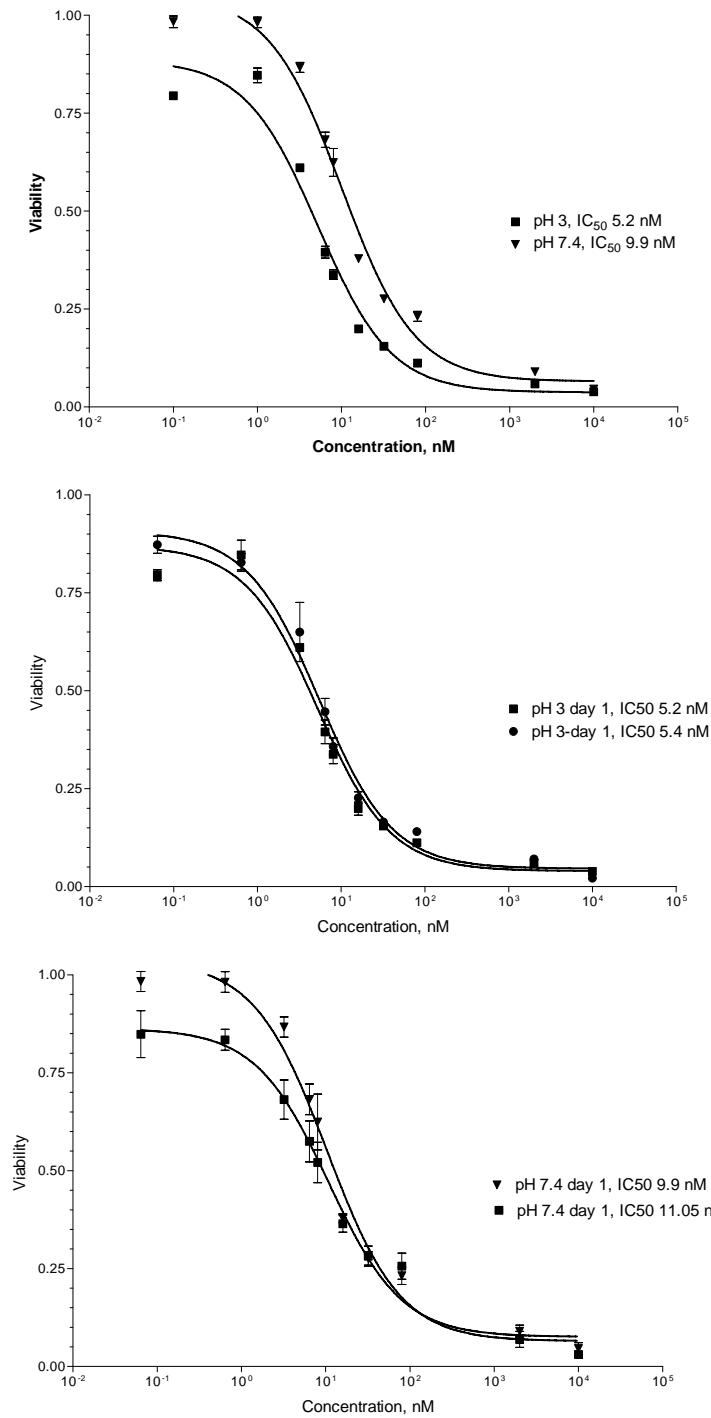


Figure A1-5. Representative examples of the effect of camptothecin stock solution pH on cell viability and assay stability. The same cell line was exposed to increasing concentrations of camptothecin in each case. (A) Camptothecin was diluted in PBS at pH 7 or citrate-PBS at pH 3. (B) Intraday variability in cytotoxic response at pH 3. (C) Intraday variability in cytotoxic response at pH 7.

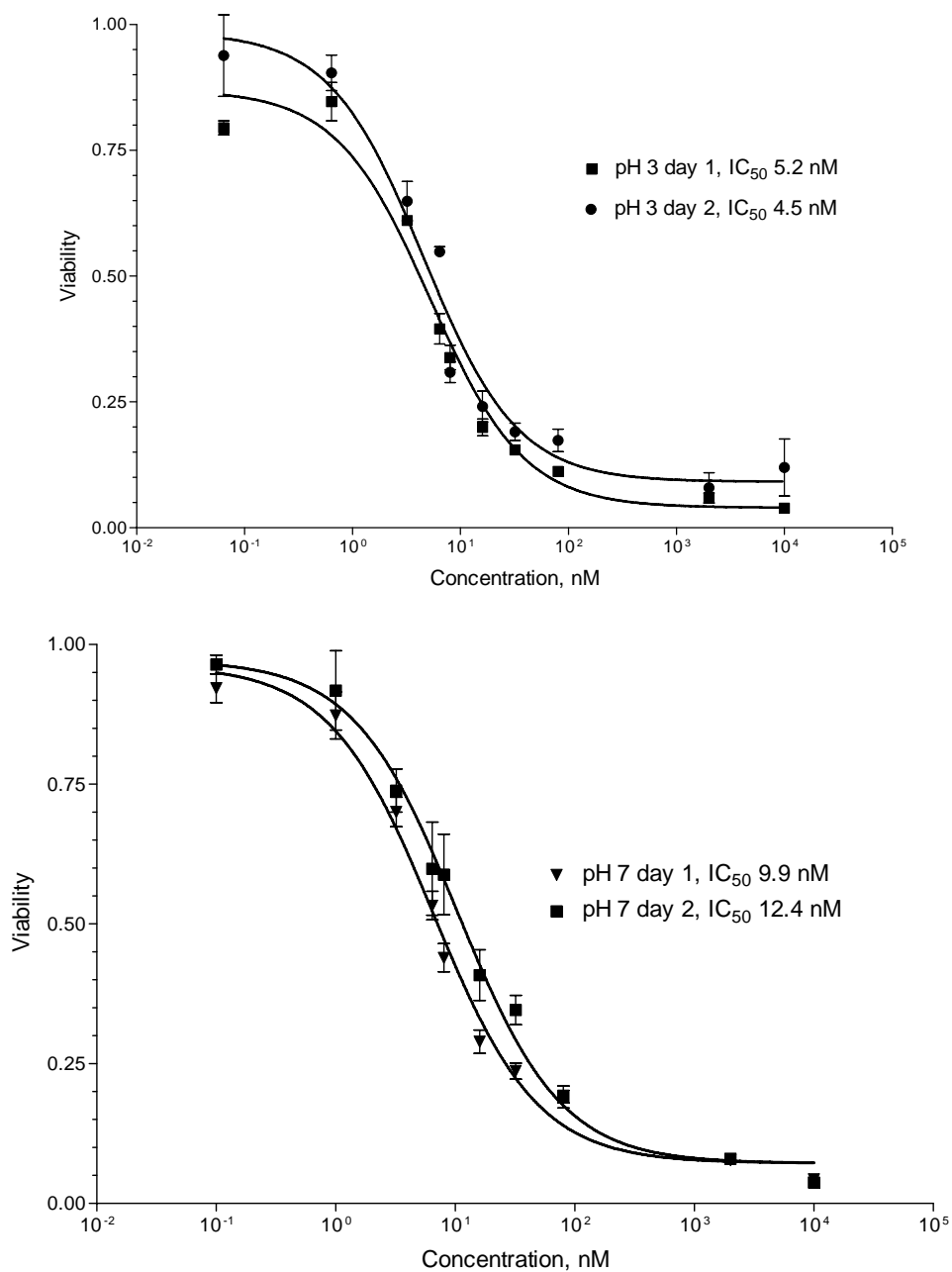


Figure A1-6. Representative examples of the effect of camptothecin stock solution pH on assay stability. The same cell line was exposed to increasing concentrations of camptothecin in each case. (A) Camptothecin was diluted in citrate-PBS at pH 3 and interday variability between plates compared. (B) Camptothecin was diluted in PBS at pH 7 and interday variability between plates compared.

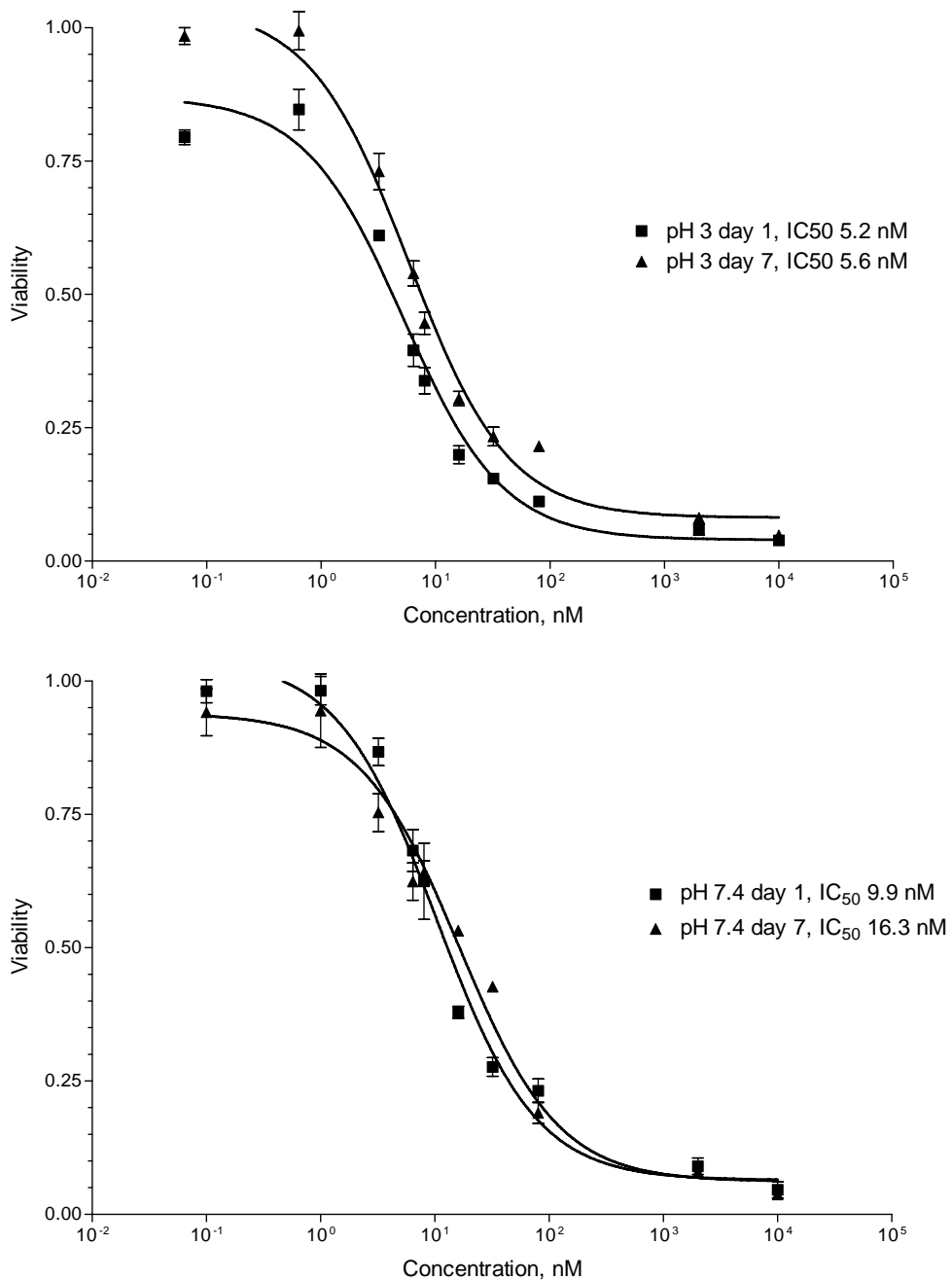


Figure A1-7. Representative examples of the effect of camptothecin stock solution pH on assay stability. The same cell line was exposed to increasing concentrations of camptothecin in each case. (A) Camptothecin was diluted in citrate-PBS at pH 3 and week to week differences in cell viability assessed. (B) Camptothecin was diluted in PBS at pH 7 and week to week differences in cell viability assessed.

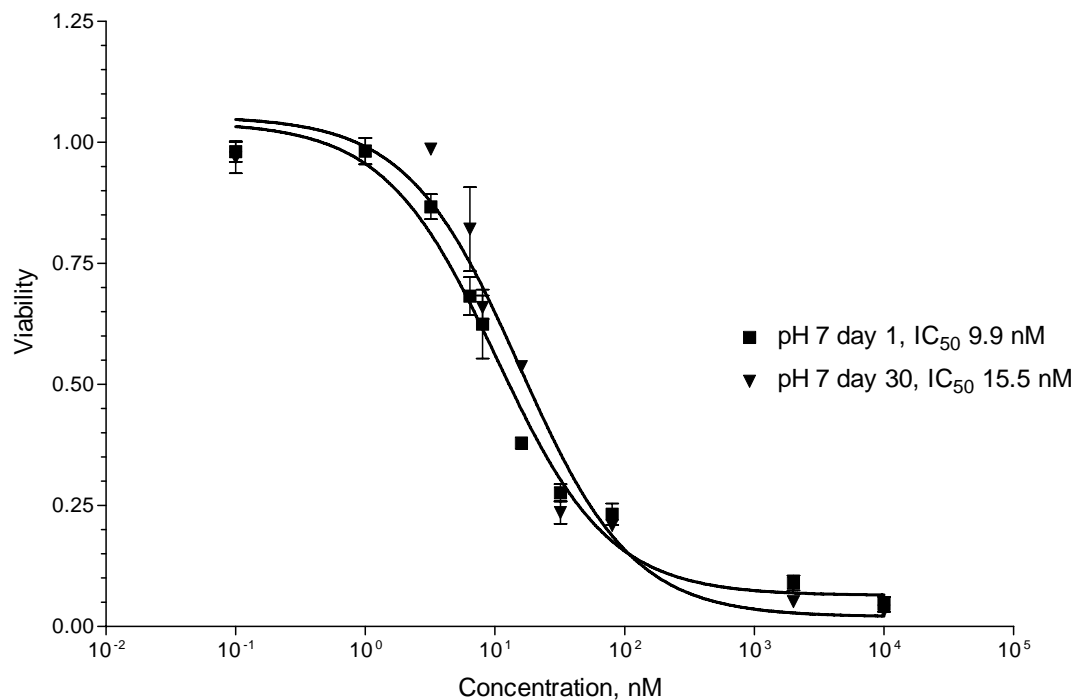
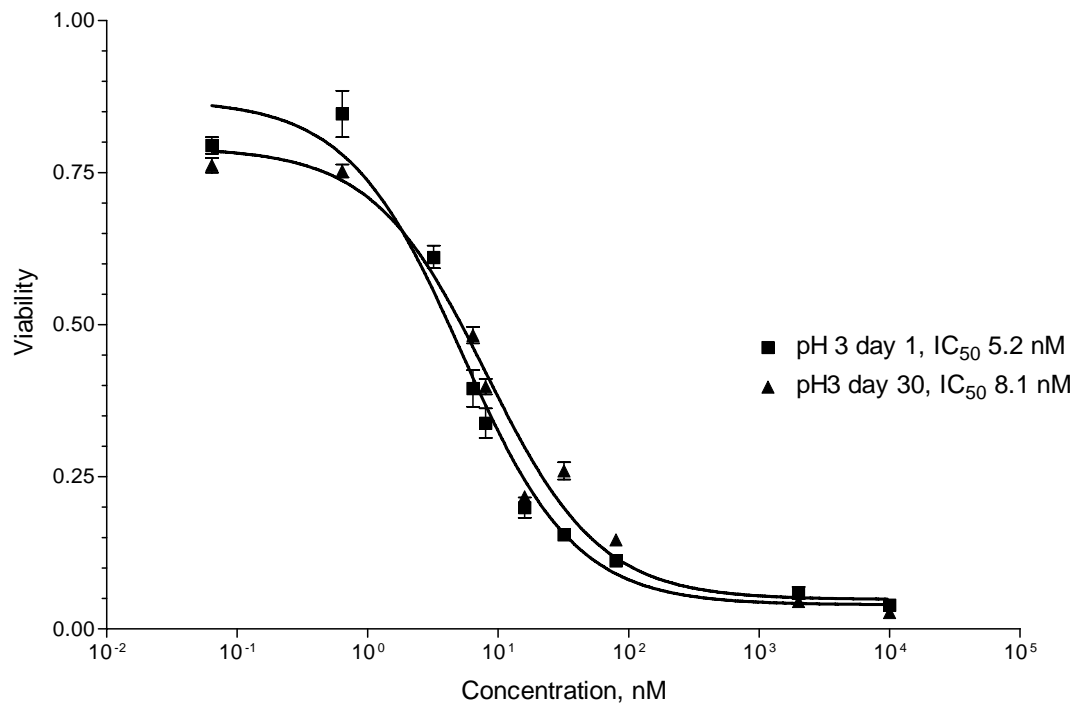


Figure A1-8. Representative examples of the effect of camptothecin stock solution pH on assay stability. The same cell line was exposed to increasing concentrations of camptothecin in each case. (A) Camptothecin was diluted in citrate-PBS at pH 3 and month to month differences in cell viability assessed. (B) Camptothecin was diluted in PBS at pH 7 and month to month differences in cell viability assessed.

1. Zhang, H.F., et al., *Cullin 3 promotes proteasomal degradation of the topoisomerase I-DNA covalent complex*. *Cancer Res*, 2004. 64(3): p. 1114-21.
2. Hamid, R., et al., Comparison of alamar blue and MTT assays for high through-put screening. *Toxicol In Vitro*, 2004. 18(5): p. 703-10.

**APPENDIX 2:**

**SUPPORTING DATA FOR  
CAMPTOTHECIN ANALOGUES**

Table A2-1. List of significant QTLs for the camptothecins

<b>Chr</b>	<b>Drug</b>	<b>Dose (mM)</b>	<b>Peak Start (cM)</b>	<b>Peak End (cM)</b>	<b>Max Peak LOD</b>
1	9NC	1.00E-02	67.22	70.41	1.115
1	9NC	1.00E-02	71.13	73.21	1.185
1	SN38	2.00E-03	116.72	142.24	1.363
1	hCPT	1.00E-03	125.51	136.34	1.495
1	9NC	1.00E-02	145.45	147.6	1.073
1	9NC	1.00E-02	149.2	150.27	1.184
1	9NC	1.00E-02	162.57	168.52	1.269
1	SN38	2.00E-03	229.13	252.12	1.389
1	TPT	8.00E-05	232.81	252.12	1.855
1	SN38	8.00E-05	233.38	252.12	1.501
1	CPT11	6.00E-03	237.73	247.23	1.411
1	CPT11	1.00E-02	237.73	247.23	1.295
1	CPT11	5.00E-02	240.19	252.12	1.457
2	IC50 hCPT	1.00E+00	215.78	216.31	1.031
3	CPT	8.00E-05	5.54	8.31	1.216
3	hCPT	2.00E-03	22.33	50.25	2.042
3	SN38	3.01E-06	44.81	47.44	1.072
3	SN38	3.01E-06	48.09	49.18	1.032
3	SN38	3.01E-06	50.94	61.52	1.115
3	SN38	3.01E-06	62.05	78.64	1.682
3	CPT11	1.50E-03	65.26	75.41	1.675
3	IC50 mCPT	1.00E+00	84.11	97.75	1.315
3	IC50 mCPT	1.00E+00	99.38	101.55	1.103
3	9NC	1.00E-02	151.49	173.34	1.638
3	TPT	1.00E-02	151.49	158.38	1.375
3	9NC	1.00E-02	173.76	174.94	1.088
3	hCPT	5.00E-03	198.68	207.73	1.686
3	9NC	1.00E-02	209.41	213.64	1.193
4	CPT11	1.60E-05	161.04	166.85	1.112
4	CPT11	1.60E-05	199.93	211.65	1.251
5	CICPT	4.00E-01	0	6.67	1.961
5	CICPT	2.00E+00	0	6.67	2.272
5	CICPT	3.00E+00	0	6.67	1.864
5	CICPT	4.00E+00	0	5.43	1.626
5	mCPT	2.50E-02	49.54	51.99	1.174
5	9NC	1.50E-05	54.79	74.07	1.853
5	9NC	8.00E-06	63.6	68.63	1.242
5	9NC	1.50E-05	92.38	93.59	1.045
5	9NC	1.50E-05	102.62	123.45	1.54
5	9NC	8.00E-06	111.97	140.72	1.709
5	9NC	1.50E-05	124.47	136.33	1.394
5	CPT	2.00E-06	137.39	161.4	1.685
5	9NC	8.00E-06	141.82	155.92	1.19
5	hCPT	8.00E-02	146.73	152.62	1.101



Chr	Drug	Dose (mM)	Peak Start (cM)	Peak End (cM)	Max Peak LOD
5	mCPT	8.00E-02	148.63	149.48	1.022
5	hCPT	8.00E-02	162.16	174.8	1.384
5	mCPT	8.00E-02	167.69	174.8	1.408
5	mCPT	2.50E-02	171.06	177.92	1.746
5	hCPT	8.00E-02	177.06	184.66	1.246
5	mCPT	8.00E-02	177.06	187.81	1.305
5	9NC	8.00E-06	182.35	189.23	1.356
6	9AC	2.00E-03	0	29.93	1.528
6	9AC	1.00E-02	9.18	11.89	1.304
6	IC50 hCPT	1.00E+00	32.62	33.43	1.202
6	IC50 hCPT	1.00E+00	42.27	42.98	1.008
6	TPT	1.50E-05	42.27	53.81	1.652
6	TPT	1.50E-05	57.96	65.14	1.106
6	SN38	8.00E-06	68.65	80.99	1.376
6	9AC	2.00E-03	76.62	78.85	1.112
6	9AC	8.00E-06	87.29	89.23	1.125
6	SN38	8.00E-06	87.29	89.23	1.195
6	SN38	8.00E-06	91.34	117.29	1.613
6	9AC	8.00E-06	92.25	107.25	1.659
6	9AC	8.00E-06	112.2	124.64	1.252
7	mCPT	1.00E-03	30.9	38.48	1.281
7	CICPT	2.00E+00	58.86	62.87	1.085
7	IC50 CPT11	1.00E+00	59.39	62.07	1.073
7	CICPT	7.00E-01	61	125.15	2.604
7	CICPT	5.00E-01	62.87	120.61	1.662
7	IC50 CPT11	1.00E+00	62.87	72.78	1.611
7	CICPT	1.00E+00	67.43	78.65	1.211
7	CICPT	4.00E-01	70.64	73.84	1.198
7	IC50 CPT11	1.00E+00	73.84	74.38	1.108
7	CICPT	2.00E+00	77.91	78.65	1.087
7	CICPT	1.00E+00	80.42	122.48	1.657
7	CICPT	2.00E+00	80.42	125.15	2.351
7	CICPT	3.00E+00	93.1	123.01	2.023
7	CICPT	4.00E-01	97.38	120.61	1.62
7	CICPT	4.00E+00	97.89	123.01	1.702
9	9AC	1.60E-05	0	12.78	1.588
9	9NC	1.00E-08	0	18.06	1.211
9	CICPT	4.00E-01	83.41	104.48	1.772
10	CPT	1.00E-07	2.13	13.49	1.475
10	CPT	1.00E-07	32.8	37.9	1.111
10	IC50 CPT11	1.00E+00	46.23	49.43	1.213
10	IC50 CPT11	1.00E+00	56.89	69.7	1.626
11	CPT11	1.00E-02	2.11	4.84	1.049
11	IC50 9AC	1.00E+00	54.75	58.4	1.299
11	IC50 9AC	1.00E+00	61.78	93.12	1.842

Chr	Drug	Dose (mM)	Peak Start (cM)	Peak End (cM)	Max Peak LOD
11	hCPT	8.00E-03	104.03	105.74	1.082
11	IC50_hCPT	1.00E+00	104.03	105.74	1.051
11	hCPT	8.00E-03	110.73	123	1.896
11	CPT11	1.00E-02	126.21	129.02	1.352
12	9AC	1.20E-05	0	6.42	1.705
12	CPT11	4.00E-06	19.68	38.5	1.277
12	CICPT	7.00E-01	95.56	104.65	1.627
12	CICPT	1.00E+00	95.56	101.98	1.442
12	CICPT	2.00E+00	95.56	104.12	1.646
12	CICPT	3.00E+00	95.56	106.52	1.714
12	mCPT	2.00E+00	95.56	119.55	2.052
13	TPT	1.00E-08	0	19.36	1.365
14	hCPT	5.00E-03	0	21.51	1.747
14	TPT	8.00E-06	0	13.89	1.286
14	hCPT	3.00E-03	8.28	34.43	1.673
14	IC50_9AC	1.00E+00	8.28	9.36	1.029
14	IC50_9AC	1.00E+00	12.46	13.89	1.014
14	mCPT	2.00E-03	21.51	36.76	1.973
14	hCPT	5.00E-03	26.59	27.01	1.026
14	hCPT	3.00E-03	92.69	95.89	1.094
14	hCPT	3.00E-03	98.96	111.27	1.15
14	TPT	8.00E-06	98.96	111.27	1.677
14	IC50_9AC	1.00E+00	105	134.3	2.354
14	hCPT	3.00E-03	113.17	114.81	1.043
14	hCPT	3.00E-03	115.2	125.88	1.242
14	hCPT	3.00E-03	134.3	138.18	1.177
15	mCPT	2.00E+00	20.24	39.72	1.81
16	CPT	3.01E-06	1.08	10.36	1.345
16	CPT	3.01E-06	29.97	34.22	1.019
16	mCPT	3.00E-03	38.51	59.68	1.537
16	mCPT	1.00E-03	40.65	66.1	1.318
16	hCPT	1.00E-03	44.45	69.05	1.769
16	mCPT	5.00E-03	57.79	65.77	2.287
17	CICPT	3.00E+00	93.98	126.46	2.198
17	CICPT	2.00E+00	95.99	126.46	1.985
17	CICPT	7.00E-01	97.56	126.46	1.583
17	hCPT	1.00E-04	98.14	114.41	1.557
17	CICPT	5.00E-01	99.21	126.46	1.599
17	CICPT	4.00E+00	99.21	126.46	1.684
18	IC50_mCPT	1.00E+00	95.46	99.04	1.2
20	SN38	8.00E-05	40.55	72.27	1.673
20	CPT11	6.00E-03	42.28	72.27	1.577
20	CPT11	1.00E-02	42.28	72.27	1.705
20	SN38	5.00E-06	42.28	77.75	1.731
20	SN38	8.00E-06	42.28	86.98	2.134

<b>Chr</b>	<b>Drug</b>	<b>Dose (mM)</b>	<b>Peak Start (cM)</b>	<b>Peak End (cM)</b>	<b>Max Peak LOD</b>
20	TPT	1.00E-08	42.28	46.71	1.033
20	CPT11	2.00E-03	46.71	78.29	1.704
20	hCPT	1.00E+01	46.71	48.85	1.144
20	CPT	8.00E-05	47.52	54.09	1.106
20	hCPT	1.00E+01	49.71	79.91	1.508
20	CPT	8.00E-05	55.74	61.77	1.273
20	TPT	1.00E-08	79.91	101.22	1.749

Table A2-2. Genes of interest and associated GO terms under chromosome 20

GO terms	Genes
GO:0003677: DNA binding	MYBL2, TAF4, HBF4A, CTCFL, SALL4, SYCP2, VSX1, PHF0, CHD5, TOP1, CST1, TFAP23, ZNF341, SPO11, ZHX3, ZNF337, TCEA2, SOX18, INSM1, TGIF2, TCFL5, CST2, FOXA2, ZNF334, ZBP1, GZF1, MYT1, ZNFX1, FKHL18, SNA1, NKX2-2, MAFB, L3MBTL, SCNAD1, TSHZ2, GMEB2, E2F1, PRIC285, ADNP, GATA5, PAX1, RTEL1, ZFP64, RBPJL, ZNF217, PLAGL2, RP11-227D2.4, CBFA2T2, WISP2, ZNF335, DNMT3B, HMG1L1, CEPB
GO:0030154: cell differentiation	NKX2-2, SPATA2, MAFB, BCL2L1, SCAND1, MYBL2, NCO86, DIDO1, PARD6B, CD40, E2F1, ELMO2, RPS21, NEURL2, PTGIS, TOP1, TFPA2C, EE1A2, PAX1, LAMA5, EYA2, TGM2, CDH4, TNFRSF6B, PLAGL2, SERINC3, SNTA1, SOX18, CDK5RAP1, NDRG3, CTNBL1, INSM1, STK4, TCFL5, FOXA2, GDF5, CEPB, ZNF313, BMP7, MYT1, TMEM189-UBE2V1, BIRC7, MMP9
GO:0008219: cell death	TGMS, BCL2L1, TNFRSF6B, SCAND1, MYBL2, PLAGL2, SERINC3, CTNBL1, DIDO1, CD40, STK4, E2F1, ELMO2, GDF5, PTGIS, CEPB, TOP1, EE1A2, BIRC7, MMP9
GO:0007049: cell cycle	DSN1, CEP250, RBL1, CTCFL, TPX2, MAPRE1, UBE2C, CABLES2, AURKA
GO:0006350: transcription	RBM39, C20ORF20, TAF4, POFUT1, NCOA3, ZNF341, TH1L, TCEA2, ADRM1, RP5-890O15.2, ZNFX1, FKHL18, L3MBTL, XRN2, TSHZ2, GMEB2, ASLX1, PRIC285, ZMYND8, ADNP, GATA5, ZFP64, ZBTB46, ZNNF217, ID1, NFATC2, HMG1L1, SLA2
GO:0040007: regulation of growth	OGFR, C20ORF10, GINS1, GHRH, ITCH
GO:0050896: response to stimulus	ASIP, ADA, PLUNC, SGK2, PROCR, VAPB, DEFB123, WFDC12, DEFB124, GNAS, TOP1, CHRNA4, TGM2, DEFB118, CD93, NTSR1, DEFB115, BPI, GSS, DEFB116, DEFB121, C20ORF185, CST7, LBP, THBD, DYNLRB1, SAMHD1, DEFB119, CST11, LIME1

Table A2-3. List of cell lines sensitive or resistant to the camptothecins

Family	cell line	# Drugs cell lines are		sum
		resistant*	sensitive**	
35	12615	1	0	1
35	12616	0	2	2
35	12617	1	3	4
35	12618	1	4	5
35	12619	1	3	4
35	12620	0	3	3
35	12621	0	4	4
35	12622	1	4	5
35	12623	0	3	3
35	12624	1	3	4
45	12698	0	1	1
45	12699	1	2	3
45	12700	0	1	1
45	12701	0	3	3
45	12702	1	3	4
45	12703	1	2	3
45	12704	2	0	2
45	12705	1	2	3
45	12706	1	3	4
45	12849	1	0	1
1334	10846	3	1	4
1334	10847	0	5	5
1334	12138	2	0	2
1334	12139	1	2	3
1334	12141	2	3	5
1334	12142	1	1	2
1334	12238	3	0	3
1340	7008	0	3	3
1340	7019	1	0	1
1340	7027	0	1	1
1340	7029	1	0	1
1340	7040	0	3	3
1340	7053	1	3	4
1340	7062	2	0	2
1340	7342	1	0	1
1340	11821	1	3	4

Family	cell line	resistant*	sensitive**	sum
1341	6991	3	1	4
1341	7006	1	1	2
1341	7010	0	1	1
1341	7012	0	4	4
1341	7020	1	0	1
1341	7021	2	3	5
1341	7044	0	3	3
1341	7048	1	3	4
1341	7343	0	3	3
1341	7344	1	1	2
1345	7345	0	2	2
1345	7348	3	1	4
1345	7357	1	1	2
1350	10855	0	5	5
1350	10856	1	1	2
1350	11822	1	4	5
1350	11824	2	2	4
1350	11825	1	3	4
1350	11826	1	0	1
1350	11827	1	4	5
1350	11828	2	3	5
1362	10860	1	0	1
1362	10861	2	0	2
1362	11982	1	2	3
1362	11983	1	3	4
1362	11984	0	3	3
1362	11985	0	3	3
1362	11986	1	3	4
1362	11987	0	4	4
1362	11988	0	2	2
1362	11989	1	3	4
1408	10830	1	2	3
1408	10831	1	0	1
1408	12147	1	0	1
1408	12148	2	1	3
1408	12149	0	3	3
1408	12150	0	4	4
1408	12151	1	0	1

Family	cell line	resistant*	sensitive**	sum
1408	12152	0	1	1
1408	12153	2	0	2
1408	12157	1	0	1
1420	10838	1	0	1
1420	10839	1	0	1
1420	11997	1	3	4
1420	11998	3	0	3
1420	11999	2	1	3
1420	12000	4	1	5
1420	12001	4	0	4
1420	12002	1	0	1
1420	12007	1	1	2
1447	12752	1	3	4
1447	12753	1	0	1
1447	12754	0	1	1
1447	12755	0	3	3
1447	12756	0	5	5
1447	12757	1	3	4
1447	12758	0	3	3
1447	12759	0	2	2
1447	12764	0	3	3
1447	12765	1	2	3
1451	12766	1	3	4
1451	12767	0	5	5
1451	12768	1	1	2
1451	12769	0	4	4
1451	12770	2	1	3
1451	12771	2	0	2
1451	12772	0	4	4
1451	12773	1	0	1
1451	12774	1	1	2
1451	12848	1	0	1
1454	12801	0	3	3
1454	12802	1	0	1
1454	12803	1	3	4
1454	12804	1	2	3
1454	12805	0	3	3
1454	12806	0	4	4

Family	cell line	resistant*	sensitive**	sum
1454	12807	1	3	4
1454	12808	1	0	1
1454	12809	0	3	3
1454	12810	1	0	1
1459	12864	0	2	2
1459	12865	0	4	4
1459	12866	0	4	4
1459	12867	1	1	2
1459	12868	0	5	5
1459	12869	0	3	3
1459	12870	0	4	4
1459	12871	0	3	3
1459	12876	1	3	4
# cell lines resistant to >=3 drugs		7		
# cell lines sensitive to > 4 drugs			19	
# cell lines sens or resist to at least 4 camptos				47

\*Resistance defined as IC50 > 90th percentile of all IC50s

\*\*Sensitivity defined as IC50 < 10th percentile of all IC50s



**APPENDIX 3:**

**SUPPORTING DATA FOR  
TOPOISOMERASE 1 INHIBITORS**

Figure A3-1. Genomewide map of QTLs for camptothecins set 1 (A), camptothecin set 1+2 (B) and indenoisoquinolines (B). Each chromosome was partitioned into 10 cM regions. Each drug-dose combination that resulted in a significant QTL (LOD > significance threshold) is indicated in blue. Intensity of the shading indicates the number of doses replicating that QTL at either the suggestive or significant level. Regions which also had a suggestive QTL (LOD > suggestive threshold) are indicated in green with color intensity referring to the number of doses replicating this peak

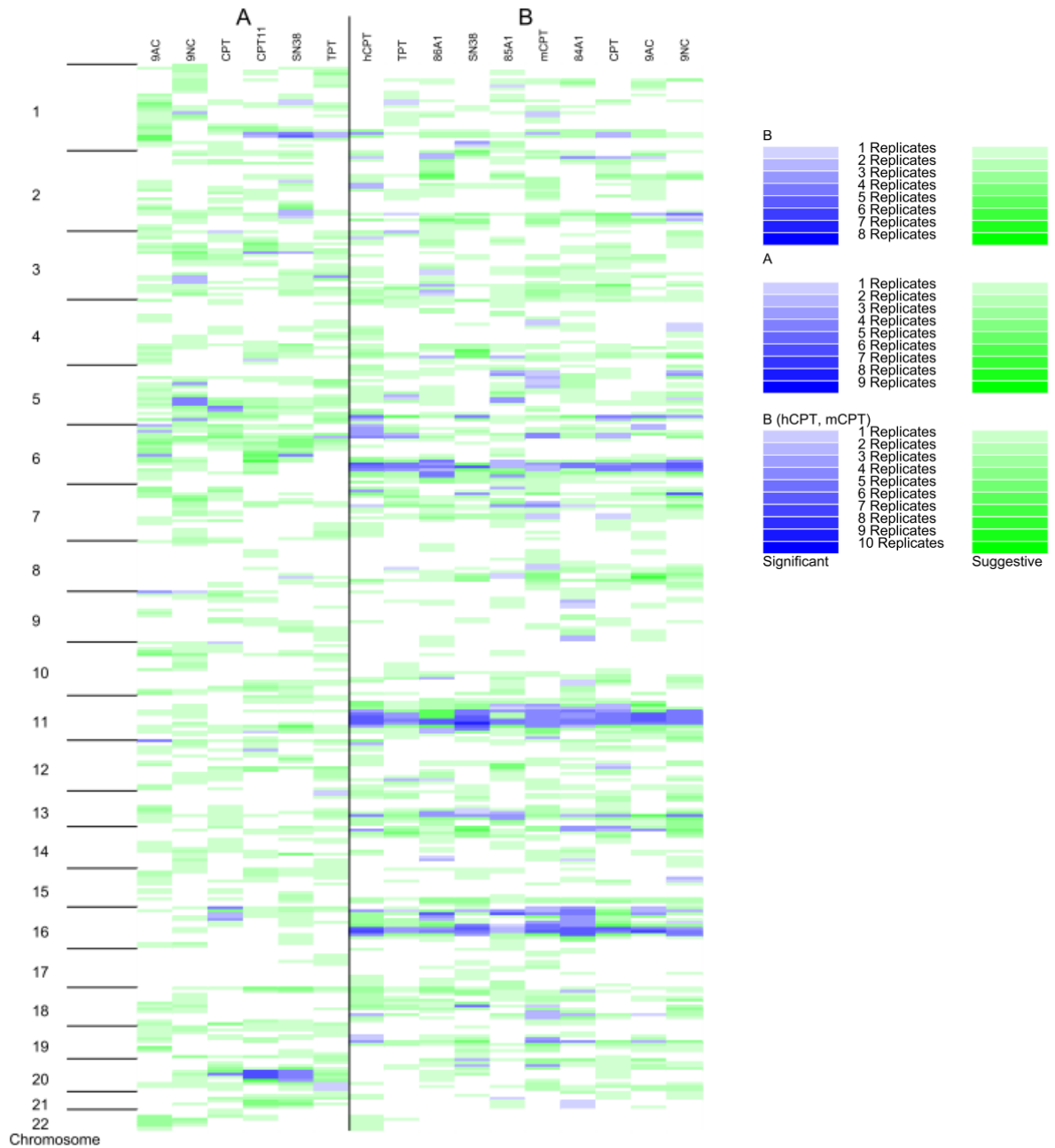


Table A3-1. List of significant QTLs for indenoisoquinolines

Chr	Drug	Dose	Peak Start cm	Peak End cm	LOD	# replicate* doses
2	Ind1	0.2	26.52	31.05	1.129	2
6	Ind1	0.3	124.11	153.04	1.964	3
6	Ind1	0.1	130	153.04	1.734	3
6	Ind1	0.2	136.97	149.8	1.28	3
7	Ind1	0.001	74.38	86.12	1.348	0
9	Ind1	0.03	27.32	58.26	1.558	1
9	Ind1	0.001	159.61	163.84	1.331	1
10	Ind1	0.05	131.8	153.78	1.775	0
11	Ind1	0.1	35.21	110.73	1.797	4
11	Ind1	0.3	53.87	93.12	1.714	4
11	Ind1	3	72.82	89.69	1.591	3
13	Ind1	0.1	84.87	90.27	1.21	0
14	Ind1	3	105.53	121.95	1.447	0
16	Ind1	0.1	70.69	84.75	1.474	4
18	Ind1	0.03	91.62	96.48	1.554	0
19	Ind1	0.1	58.69	67.37	1.574	1
1	Ind2	0.01	72.59	82.41	1.345	0
4	Ind2	1	195.06	203.77	1.365	0
5	Ind2	0.23	24.48	37.32	1.114	1
5	Ind2	0.16	119.5	123.45	1.551	2
6	Ind2	10	124.64	153.04	1.694	3
6	Ind2	30	173.31	177.88	1.273	1
7	Ind2	0.16	17.17	27.66	1.143	0
7	Ind2	1	72.78	75.44	1.517	1
8	Ind2	1	106.9	121.9	1.683	0
11	Ind2	1	71.6	88.49	1.593	3
16	Ind2	3	48.53	93.78	2.139	2
2	Ind3	50	14.1	33.31	2.024	2
3	Ind3	3	131.83	146.6	1.148	0
3	Ind3	1	180.8	181.87	1.212	1
3	Ind3	3	180.8	187.49	1.362	2
3	Ind3	5	205.56	222.83	1.877	1
3	Ind3	3	207.73	214.45	1.121	1
4	Ind3	3	192.12	206.98	1.483	1
6	Ind3	7	124.64	159.98	2.286	4
6	Ind3	10	124.64	159.98	2.195	4
6	Ind3	5	133.18	159.44	1.805	4
6	Ind3	30	142.86	146.06	1.012	4
6	Ind3	10	161.59	164.78	1.509	4
6	Ind3	10	173.31	177.88	1.352	4
10	Ind3	1	121.98	128.73	1.43	1
11	Ind3	10	87.89	101.75	1.592	5
11	Ind3	50	104.03	138.56	2.421	5

Chr	Drug	Dose	Peak Start cm	Peak End cm	LOD	# replicate doses
12	Ind3	1	130.94	144.83	1.472	0
13	Ind3	10	84.87	90.27	1.44	2
14	Ind3	5	98.96	125.88	2.022	2
16	Ind3	10	75.34	84.75	1.228	2
16	Ind3	10	87.06	99.44	1.227	3

\* number of doses for this drug which replicated this QTL

**APPENDIX 4:**

**SUPPORTING DATA FOR  
MECHANISTIC SET**

Table A4-1 Drug concentrations and  $\overline{GI50}$  for mechanistic set

<u>Drug Name</u>	<u>Abbreviation</u>	<u>Dose 1 (mM)</u>	<u>Dose 2 (mM)</u>	<u>Dose 3 (mM)</u>	<u>Dose 4 (mM)</u>
5-Fluorouracil	5FU	1.10E-03	4.94E-03	2.22E-02	1.00E-01
Floxuridine	Flox	2.33E-03	8.16E-03	2.86E-02	1.00E-01
Epirubicin	Epi	6.30E-06	2.50E-05	1.00E-04	4.00E-04
Doxorubicin	Dox	6.00E-06	1.30E-05	2.50E-05	5.00E-05
Daunorubicin	Daun	1.25E-05	2.50E-05	5.00E-05	1.00E-04
Idarubicin	Ida	5.00E-06	1.23E-05	2.78E-05	6.25E-05
Vincristine	Vinc	1.25E-05	2.50E-05	5.00E-05	1.00E-04
Vinorelbine	Vino	5.93E-05	8.89E-05	1.33E-04	2.00E-04
Vinblastine	Vinb	5.00E-06	1.00E-05	2.00E-05	4.00E-05
Docetaxel	Doc	1.28E-05	3.20E-05	8.00E-05	2.00E-04
Paclitaxel	Pac	1.85E-05	5.56E-05	1.67E-04	5.00E-04
Oxaliplatin	Oxal	3.70E-03	1.11E-02	3.33E-02	1.00E-01
Carboplatin	Carbo	1.28E-03	3.20E-03	8.00E-03	2.00E-02
Etoposide	Etop	1.17E-04	4.08E-04	1.43E-03	5.00E-03
Teniposide	Teni	2.56E-04	6.40E-04	1.60E-03	4.00E-03
Bleomycin	Bleo	8.00E-05	4.00E-04	2.00E-03	1.00E-02

$\overline{GI50}$

Table A4-2. Cell lines sensitive and resistant to the mechanistic set, by family

Family	cell line	Drugs	
		# Resistant*	# Sensitive**
35	12615	7	0
35	12616	0	11
35	12617	1	11
35	12618	1	0
35	12619	1	1
35	12620	9	9
35	12621	9	15
35	12622	6	13
35	12623	6	0
35	12624	1	0
45	12698	1	7
45	12699	1	2
45	12700	0	0
45	12701	1	0
45	12702	0	0
45	12703	8	0
45	12704	0	0
45	12705	12	9
45	12706	0	1
45	12849	7	2
1334	10846	6	3
1334	10847	0	0
1334	12138	1	0
1334	12139	1	1
1334	12141	0	2
1334	12142	0	0
1334	12238	0	0
1340	7008	0	2
1340	7019	1	1
1340	7027	8	0
1340	7029	10	0
1340	7040	7	3
1340	7053	0	0
1340	7062	1	1
1340	7342	1	0
1340	11821	7	0
1341	6991	0	0
1341	7006	1	10
1341	7010	0	1
1341	7012	0	0
1341	7020	0	0
1341	7021	6	0
1341	7044	1	0
1341	7048	3	0
1341	7343	0	7
1341	7344	0	0

Family	cell line	Drugs	
		# Resistant	# Sensitive
1345	7345	0	0
1345	7348	2	0
1345	7357	3	2
1350	10855	8	0
1350	10856	11	0
1350	11822	10	0
1350	11824	3	0
1350	11825	4	0
1350	11826	0	0
1350	11827	0	9
1350	11828	1	0
1362	10860	0	0
1362	10861	3	0
1362	11982	0	0
1362	11983	1	0
1362	11984	9	0
1362	11985	0	0
1362	11986	11	0
1362	11987	1	0
1362	11988	2	3
1362	11989	2	0
1408	10830	0	15
1408	10831	0	3
1408	12147	1	0
1408	12148	0	0
1408	12149	0	0
1408	12150	0	0
1408	12151	0	0
1408	12152	0	1
1408	12153	1	0
1408	12157	0	1
1420	10838	1	0
1420	10839	4	12
1420	11997	0	5
1420	11998	0	0
1420	11999	1	0
1420	12000	1	1
1420	12001	0	0
1420	12002	0	0
1420	12007	2	0
1447	12752	6	2
1447	12753	3	0
1447	12754	0	0
1447	12755	0	1
1447	12756	1	0
1447	12757	0	0
1447	12758	0	0
1447	12759	1	0
1447	12764	0	0



Family	cell line	Drugs	
		# Resistant	# Sensitive
1447	12765	1	0
1451	12766	0	0
1451	12767	0	16
1451	12768	0	0
1451	12769	0	1
1451	12770	7	0
1451	12771	0	0
1451	12772	0	2
1451	12773	3	1
1451	12774	0	0
1451	12848	0	0
1454	12801	0	4
1454	12802	0	10
1454	12803	0	0
1454	12804	0	0
1454	12805	0	0
1454	12806	1	0
1454	12807	2	0
1454	12808	0	1
1454	12809	0	0
1454	12810	0	0
1459	12864	0	4
1459	12865	1	0
1459	12866	0	4
1459	12867	0	0
1459	12868	0	3
1459	12869	0	1
1459	12870	1	0
1459	12871	0	7
1459	12876	0	0

\*Resistance defined as mean viability > 90th percentile of mean viability at  $\overline{GI50}$

\*\*Sensitivity defined as mean viability < 10th percentile of mean viability at  $\overline{GI50}$

Table A4-3. Heritability estimates for all drug-dose phenotypes of mechanistic set

Growth-rate adjusted h <sup>2</sup>	Dose 1	Dose 2	Dose 3	Dose 4
Bleo	15.09	9.49	5.74	0.00
5FU	23.84	30.85	22.91	0.00
Flox	25.03	25.54	16.03	11.41
Carbo	37.17	40.57	16.74	13.51
Oxal	49.74	30.92	13.03	20.61
Epi	62.40	53.68	43.70	2.30
Daun	37.12	22.98	13.40	17.51
Dox	34.12	28.71	0.01	0.01
Ida	48.95	27.85	6.00	3.93
Doc	23.04	24.26	31.17	13.45
Pac	53.24	34.23	7.77	27.57
Etop	42.42	38.10	33.26	23.76
Teni	32.95	29.13	37.72	19.94
Topo	45.78	26.14	0.00	0.71
Vinb	0.07	0.01	13.76	35.86
Vino	36.75	2.83	8.19	0.00
Vinc	24.40	9.00	23.39	10.93

\* refer to Table A4-1 for doses

Table A4-4. List of significant QTLs for mechanistic set

Chr	Drug	Dose (mM)	Peak Start cm	Peak End cm	LOD
1	Doc	8.00E-05	8.85	16.22	1.231
1	Doc	0.0002	33.75	52.7	1.341
1	Doc	8.00E-05	168.52	190.98	1.623
1	Oxal	0.1	20.61	23.35	1.12
1	Oxal	0.1	24.68	56.19	2.091
1	Pac	0.0001667	23.35	60.01	1.921
1	Pac	0.0005	33.75	54.3	1.685
1	Vinb	1.00E-05	191.52	201.58	1.208
1	Vinb	1.00E-05	214.08	237.73	1.596
2	Bleo	0.01	218.45	229.14	1.503
2	Carbo	0.0032	94.05	103.16	1.354
2	Carbo	0.0032	106.84	119.22	1.164
2	Carbo	0.0032	120.29	137.93	1.701
2	Carbo	0.0032	185.13	195.65	1.246
2	Carbo	0.0032	196.85	204.53	1.375
2	Oxal	0.0037037	215.78	227	1.511
2	Vinb	2.00E-05	1.95	14.1	1.59
2	Vinb	2.00E-05	137.93	147.4	1.257
2	Vinb	2.00E-05	149.89	161.26	1.196
3	5FU	0.004938	190.43	207.73	1.648
3	5FU	0.004938	212.61	214.45	1.124
3	Flox	0.0023324	187.49	198.68	1.941
3	Flox	0.0285714	187.49	206.43	2.094
3	Flox	0.0081633	189	201.14	1.801
3	Flox	0.1	190.43	203.28	1.571
3	Oxal	0.0333333	187.49	228.14	2.179
3	Oxal	0.1	190.43	214.45	1.712
4	Dox	2.50E-05	180.01	199.93	1.47
5	Carbo	0.00128	0	1.72	1.126
5	Carbo	0.008	0	5.43	1.48
5	Carbo	0.008	140.72	144.06	1.073
5	Carbo	0.00128	152.62	155.92	1.058
5	GR_1_DM SO	1	142.92	155.92	1.5
5	GR_1_DM SO	1	159.77	162.47	1.061
5	GR_10_D MSO	1	134.72	156.47	2.13
5	GR_10_D MSO	1	157.02	167.69	1.25
5	GR_H20	1	144.06	153.17	1.502
5	GR_H20	1	160.87	162.16	1.027
6	Etop	0.000408	37.79	74.28	1.493
6	Etop	0.000408	76.62	87.29	1.217
6	Etop	0.000408	100.91	103.45	1.045
7	5FU	0.1	97.89	98.44	1.103

Chr	Drug	Dose (mM)	Peak Start cm	Peak End cm	LOD
7	5FU	0.1	103.63	113.92	1.173
7	5FU	0.1	137.83	155.1	1.971
7	5FU	0.022222	140.63	155.1	1.41
7	5FU	0.022222	159.53	168.98	1.195
7	Bleo	0.0004	23.29	47.08	1.929
7	Bleo	0.0004	48.69	52.7	1.064
7	Bleo	0.0004	155.1	181.97	1.635
7	Carbo	0.02	57.79	78.65	1.854
7	Carbo	0.008	58.86	73.84	1.385
7	Carbo	0.0032	61	76.71	1.647
7	Carbo	0.008	74.38	78.65	1.471
7	Carbo	0.02	79.24	90.42	1.363
7	Carbo	0.0032	83.99	86.12	1.038
7	Carbo	0.008	84.52	86.12	1.019
7	Carbo	0.02	91.67	104.86	1.48
7	Carbo	0.008	94.87	100.05	1.061
7	Carbo	0.008	140.63	149.9	1.313
7	Carbo	0.02	141.29	148.11	1.047
7	Daun	0.0001	143.33	149.9	1.257
7	Daun	0.0001	160.09	161.21	1.007
7	Doc	8.00E-05	128.41	181.97	1.652
7	Doc	3.20E-05	168.98	181.97	1.595
7	Epi	0.0001	112.32	113.92	1.026
7	Epi	0.0001	140.63	155.1	1.443
7	Flox	0.0285714	146.28	155.1	1.227
7	Flox	0.0081633	156.33	170.94	2.013
7	Flox	0.0285714	159.53	170.94	1.521
7	Flox	0.0285714	178.41	181.97	1.032
7	Ida	2.78E-05	104.86	105.39	1.045
7	Ida	2.78E-05	106.46	123.01	1.286
7	Ida	6.25E-05	112.32	113.92	1.033
7	Ida	2.78E-05	134.55	165.18	2.285
7	Ida	6.25E-05	138.42	162.33	1.839
7	Oxal	0.0333333	54.11	67.43	1.374
7	Oxal	0.0333333	75.98	77.91	1.038
7	Oxal	0.0333333	103.63	126.75	1.857
7	Teni	0.004	140.63	161.21	1.97
7	Teni	0.000256	162.33	181.97	1.887
7	Vinb	5.00E-06	163.03	181.97	2.061
8	Bleo	0.0004	103.69	111.68	1.55
8	Carbo	0.00128	87.52	102.62	1.325
9	5FU	0.004938	81.92	96.46	1.5
9	5FU	0.004938	150.92	159.61	1.075
9	Ida	1.23E-05	88.16	88.92	1.023
9	Ida	1.23E-05	90.4	105.82	1.643
9	Ida	1.23E-05	110.91	116.67	1.33
11	Doc	0.0002	11.05	33.02	1.485
11	Doc	0.0002	35.21	50.88	1.185

Chr	Drug	Dose (mM)	Peak Start cm	Peak End cm	LOD
11	Doc	3.20E-05	45.94	54.75	2
11	Doc	3.20E-05	91.47	98.45	1.439
11	Etop	0.000117	110.73	112.87	1.067
11	Etop	0.000117	116.07	127.33	1.186
11	GR_1_DM SO	1	90.89	100.05	2.088
11	GR_1_DM SO	1	100.62	101.75	1.043
11	GR_10_D MSO	1	91.47	100.05	1.852
11	GR_Blank	1	89.69	101.75	1.751
11	GR_H20	1	90.29	100.05	2.067
11	GR_H20	1	100.62	101.75	1.053
11	Oxal	0.0037037	51.42	53.02	1.067
11	Oxal	0.0037037	83.83	89.69	1.229
11	Oxal	0.0111111	88.49	96.85	1.334
11	Vinc	0.0001	14.52	50.88	1.343
11	Vino	0.0002	12.92	51.95	1.541
12	Teni	0.0016	17.72	29.73	1.499
12	Teni	0.0016	90.77	94.49	1.151
12	Teni	0.0016	139.61	144.83	1.121
12	Vino	8.89E-05	9.52	19.68	1.581
12	Vino	8.89E-05	125.31	153.33	1.685
13	Daun	5.00E-05	0	17.21	1.362
13	Teni	0.00064	0	16.2	1.622
14	5FU	0.004938	0	13.89	1.728
14	Carbo	0.02	0	13.89	1.643
14	Carbo	0.02	21.51	31.13	1.345
14	Etop	0.001429	21.51	25.87	1.124
14	Etop	0.001429	26.59	36.76	1.454
14	GR_1_DM SO	1	93.76	118.68	1.42
14	GR_10_D MSO	1	53.19	63.25	1.209
14	GR_10_D MSO	1	78.2	84.16	1.066
14	GR_10_D MSO	1	92.69	125.88	1.801
14	GR_H20	1	57.43	57.98	1.012
14	GR_H20	1	93.76	118.68	1.414
14	GR_Media FCS	1	93.76	117.3	1.469
15	Dox	5.00E-05	6.11	19.12	1.108
15	Vinc	5.00E-05	82.84	102.21	1.524
16	Bleo	0.0004	73.2	85.94	1.471
16	Epi	6.30E-06	73.2	87.06	1.536
18	Epi	6.30E-06	0	35.46	1.907
18	Epi	6.30E-06	66.66	68.3	1.253
18	Etop	0.001429	64.48	66.66	1.05

Chr	Drug	Dose (mM)	Peak Start cm	Peak End cm	LOD
18	Etop	0.001429	67.21	74.93	1.356
18	Vino	8.89E-05	9.26	16.54	1.194
18	Vino	8.89E-05	35.46	49.55	1.065
18	Vino	8.89E-05	66.66	76.15	1.588
19	5FU	0.1	100.01	105.02	1.713
19	Carbo	0.008	100.01	105.02	1.568
19	Carbo	0.02	100.01	105.02	1.761
19	Dox	5.00E-05	100.01	105.02	1.338
19	Ida	2.78E-05	89.73	105.02	1.965
19	Ida	6.25E-05	100.01	105.02	1.404
20	Oxal	0.0111111	36.58	48.85	1.559
20	Oxal	0.0111111	58.48	61.77	1.327
20	Vinb	2.00E-05	6.25	8.97	1.325
20	Vinb	2.00E-05	9.53	17.19	1.504
20	Vinb	2.00E-05	37.65	47.52	1.357
20	Vinb	2.00E-05	55.74	74.47	2.08
20	Vinc	5.00E-05	42.28	77.75	1.858
21	Pac	0.0005	9.72	40.43	1.685

Figure A4-1. Boxplots for variability in response for the mechanistic drug set. The CEPH cell lines (n =125) were treated with increasing concentrations of each drug (4 doses) and mean viability measured relative to control. The line within each box represents the median (50%) viability for the population (n = 125) at the specified dose, the upper edge of the box indicates the 75th percentile of the data set, and the lower edge indicates the 25th percentile. The range of the middle two quartiles is the inter-quartile range (IQR). The whiskers are 1.5 times the IQR. Circles outside the whiskers are considered outliers.

