IDENTIFICATION AND VALIDATION OF GERMLINE GENETIC VARIANTS THAT ASSOCIATE WITH SORAFENIB CLINICAL OUTCOMES AND CYTOTOXICITY

Daniel James Crona

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 Eshelman School of Pharmacy (Pharmacotherapy and Experimental Therapeutics).

Chapel Hill
2015

Approved by:

Federico Innocenti

William Kim

Karen Mohlke

John Valgus

Timothy Wiltshire

ABSTRACT

Daniel James Crona: Identification and validation of germline variations that associate with overall survival in metastatic renal cell carcinoma patients treated with sorafenib
(Under the direction of Federico Innocenti)

Sorafenib is a potent inhibitor of multiple oncogenic, stromal and angiogenic receptor tyrosine kinases. Germline variants in VEGF-pathway genes and in sorafenib pharmacology genes might associate with prognosis and/or sorafenib efficacy in metastatic renal cell carcinoma (mRCC patients). A total of 295 mRCC patients from the phase III TARGET trial were genotyped using candidate germline variants from 56 candidate genes implicated in angiogenesis, sorafenib pharmacology and/or RCC prognosis/pathogenesis. Seven variants that significantly associated with overall survival (OS) in mRCC patients treated with sorafenib, and an additional two variants associated with OS in a combined analysis of both treatment arms.

Statistical associations between genetic variants and outcomes in cancer studies should be supported with molecular mechanistic evidence of variant function to aid in biomarker validation. Variants identified in Aim 1 that significantly associated with OS were analyzed using *in silico* bioinformatic tools to prioritize *in vitro* validation assays. Cell viability assays validated one non-synonymous variant in *FLT-4*, and dual reporter gene luciferase assays validated two intronic *VEGFA* variants in three different cell lines.

Novel pathways and targets of sorafenib activity remain to be identified. Primary mouse embryonic fibroblasts (MEFs) from 32 inbred strains were profiled for sorafenib

cytotoxicity utilizing high content imaging and simultaneous evaluation of cell health parameters (cell viability, membrane permeability, mitochondrial membrane potential, and cytochrome C release). One quantitative locus (QTL) on chromosome 9, which reached genome-wide significance and significantly associated with cytochrome C release, was identified. A total of nine genes, expressed in MEF cells at mRNA level, were present in this QTL. A second QTL associated with cell viability was also identified. A total of 13 candidate genes, expressed in MEF cells at mRNA level, were present in this QTL. In the future, functional validation of candidate genes under these two identified QTLs, using knockdown and overexpression approaches, will be conducted in MEF and human cell lines.

I would like to dedicate this work to my one and only true love: Dr. Lana Crona. Thank you for sticking by me through all of this, and for being my rock. I love you, my dear Nannybelle. Without you, I am nothing.

ACKNOWLEDGEMENTS

Federico Innocenti, you were precisely what I needed in a mentor. You taught me self-reliance, critical thinking skills, and self-confidence. You helped me see the big picture of why I am here and why I want to continue as a clinical pharmacologist and pharmacogeneticist in oncology: to make impactful changes and advancements that better the lives of our patients. I truly value your hands-off mentoring approach and that you did not micro-manage my every move. I appreciate that, from the beginning, you treated me like you would a post-doctoral researcher and gave me the latitude to pursue new ideas and test new hypotheses. As a result, you engendered within me a sense of independence. I hope I met your high expectations for a first graduate student. Amy Etheridge, you have certainly been a great teacher and mentor to me as well. Thank you for helping me design and develop my assays, for making sure supplies were always available, for babysitting my cells, for providing me with sage career advice, for talking about biking and running, and for listening to all of my thoughts and concerns on so many different subjects. I literally could not have done this without you. My friend and colleague from "down under," Dylan Glubb, I only have one thing to say to you: shot bro. I could not have asked for a better colleague, and role model when you moved here from Chicago. I appreciate all of your advice on pharmacogenetics, genetics, and molecular biology. Moreover, thank you for being a really great friend. And Eric Seiser, thank you so much for being a great resource since you joined the lab. I am so glad I can count on you for your expertise in bioinformatics, and biostatistics. I am so excited for you to get to start your life in Washington DC. I would also like to thank

one of our collaborators from University of Chicago, Andrew Skol. You are an amazing statistician, and were integral to the analysis of the data and my overall learning in the early stages of my dissertation. I cannot thank you enough. I also would like to thank Carol Peña from Bayer for providing me access to the TARGET patient samples, as well as to thank Habibul Ahsan at the University of Chicago and Jason Luo from UNC for their assistance with the genotyping portion of my research.

Tim Wiltshire, I sincerely could not have asked for a better Chair of my Dissertation Committee. You are so kind, intelligent, easy to talk with, and truly personify the words "mentor" and "teacher." I would not be at UNC if not for you because, as you know, you literally convinced me to attend UNC while I was waiting in the lift line while snowboarding in Colorado. You were my first lab PI here, and I am so thankful that I will soon get the opportunity to be your collaborator and colleague. A million times thank you. I cannot say enough good things about two more members of the Wiltshire lab: Oscar Suzuki and Amber Frick. Oscar, you are simply amazing. I wish I knew half as much about genetics, bioinformatics, and molecular biology (you are also one heck of a swimmer). And Amber, you are no less amazing. You are so smart, so kind, are an incredible teacher, and wonderful at explaining difficult concepts (and you are also on heck of a baker). I am also very grateful to Rusty Thomas, Bethany Parks, Joe Trask, and all of the members of the Hamner Institute for collaborating with me on the MEF study. And, thank you very much to Tammy Havener for helping guide me on cell viability assays, and molecular biology. I learned so much from each and every one of you. Thank you for being integral to my learning.

John Valgus, of all my mentors and Dissertation Committee members, I relate best to you because we are both pharmacists. I value every single one of our discussions on GU

oncology topics, your advice on my different HOPS presentations, you advice on my dissertation, your explanations regarding supportive care topics, as well as how you've advised me extensively on career and on life. You have truly made this training experience unique and special for me. You are an oncology pharmacist of the highest quality, an insightful pharmacogenetics researcher, and have incredible perspective. You are everything I want to be as a clinician, as a clinical researcher and as a person. Thank you so very much for taking a chance on me.

Billy Kim, I wanted to thank you for being such a great part of my Dissertation Committee and for being a mentor to me in clinic. I will always remember the first time you introduced me to one of your patients as "Dr. Crona." That was such a validating moment for me. Thank you for challenging me to think deeply in my Dissertation Committee meetings, and for all of our discussions about science while in clinic. Thank you so much for your generosity. Giving me the Caki-1 cells REALLY helped me, and all of your lab members were so kind and giving of their time. You are clearly one of the most intelligent people I have ever met, and I really look forward to working together in the future, both in clinic and hopefully through lab collaborations.

Karen Mohlke, I sincerely appreciate all of your contributions to my research as part of my Dissertation Committee. You are simply amazing when it comes to your knowledge of human genetics. I hope someday that I understand even a fraction of what you know. You and your lab members, most notably Tamara Roman, have been essential to the directions that I took in designing, implementing and analyzing my validation experiments. I also appreciate how you challenged me on difficult genetics and statistics issues during my Dissertation Committee meetings. You have definitely helped me grow as a researcher. I

hope that you will continue to mentor me in the future. I value you and your expertise so very much.

As part of my education, I have been lucky enough to engage in a longitudinal clinic experience. I have essentially served as the clinical pharmacist specialist in the outpatient GU oncology clinic of over four years. During that time, I have gotten to meet some really extraordinary individuals who have truly shaped me as a clinician and researcher. Matt Milowsky, I cannot adequately express just how much I gain from you as a mentor. You are a brilliant clinician, and an even better clinical researcher. Thank you for providing me with opportunities to collaborate with researchers at Memorial Sloan Kettering, and thank you for all you do for me (which I am sure is more than I know). Mary Dunn, what would I ever do without you? Thank you so very much for teaching me on all topics GU, thank you for making me laugh daily, for all the trips to Starbucks, and for being my friend. Young Whang, I have learned perhaps most from you. You are such an excellent teacher, and I have gained immensely from every single one of our conversations on medicine, medications, and science. Thank you for collaborating with me on the PRES-enzalutamide case report. Paul Godley and Ethan Basch, thank you both so much for supporting me, and for allowing me to be so integrated in the care of your patients. I know both of you have supported me extensively, and lobbied for me to work with the GU team post-dissertation. I cannot express to you just how touched I am by your confidence in me. And last but DEFINITELY not least, thank you Melissa Holt for being you- so supportive of me, and so central to the quality care of our patients. Literally, the best part of some of my work weeks was getting to see you, and how you would welcome me to clinic when I walked through the door. All of you are so talented and never cease to amaze me.

During my clinical experience, I was also fortunate to have my own “committee” of pharmacists that served as clinical advisors. I cannot say thank you enough times to Aimee Faso, Benyam Muluneh, and Meredith Keisler for being excellent role models of exemplary outpatient oncology pharmacists and CPPs. I want to be you guys when I a finally grow up and get the job of my dreams. I would also like to thank Benyam and Maurice Alexander for being my HOPS mentors. And, finally I would like to thank Lindsey Amerine, the members of the COG and all of the members of the CHIP pharmacy for helping navigate my clinical experience. It has been such a meaningful and positive adventure!

I am also extremely thankful for being able to go through this process in such a wonderful training environment at the UNC Eshelman School of Pharmacy. I would like to thank all of my mentors in DPET, most notably Herb Patterson, Heyward Hull, Craig Lee and Dhiren Thakker. At one point or another during my training, you all have provided me with excellent mentoring on myriad topics ranging from science to professional development. I would also like to thank our exemplary administrative staff, including: Arlo Brown, Kathy Maboll, Jessie Bishop, Aaron Todd, Anna Crollman, and members of the ITSOP staff. I would like to thank Kelly Scolaro for providing me with the opportunities to work with SHAC and SHAC Outreach. These have allowed me to hone mentoring skills and grow as a person. Thank you so much. Finally, I would like to convey my deepest appreciation to Brian Rybarczyk. You are so committed to professional development of graduate students, and really helped me to become a better person (not to mention a viable candidate for faculty positions).

I am so blessed to have been part of the UNC-Duke-Hamner Institute collaborative T32 training program in clinical pharmacology. I would like to first and foremost thank Kim

Brouwer as an important DPET mentor to me, but also as the person who brought me into the T32 training program. I would also like to thank Daniel Benjamin, Paul Watkins, Angela Kashuba, and Robert Noveck for this great opportunity to work with some great trainees in such a neat multi-disciplinary training environment. Thank you, as well, to the T32 administrators: Lisa Phillippe and Kirsten Leysieffer. It has been so wonderful, and I cannot thank all of you enough. I would also like to thank all of the organizations that helped fund my research and/or my stipend: the American Foundation for Pharmaceutical Education, GlaxoSmithKline, Khalid Ishaq, and the National Institutes of Health/National Institute for General Medicine Sciences.

I would like to thank all of the collaborators I was lucky enough to work with on the irinotecan pharmacogenetics replication project. Thank you so much to Jackie Ramirez and Mark Ratain from University of Chicago, to Gary Rosner from Johns Hopkins, to Wei Qiao from MD Anderson, and to Ron van Schaik, Ron Mathijssen, and Anne-Joy de Graan from Erasmus Medical Center in the Netherlands.

I would like to sincerely thank all of the mentors in my life outside of UNC that have helped shape and guide me through the years, and have all helped me complete my dissertation. James West, what can I say about you that gives you the proper credit for setting me on this path? Whatever words I use will fall desperately short. You are, without question, the most intelligent human being I have ever met. But it's not just that. You believed in me from day 1 and have continued to be one of my most ardent supporters and trusted mentors. You have impacted my life so positively in so many ways that I fear I can really never repay all of your support and kindness. I loved being your "mouse wrangler" back at CU, I appreciate you hooding me at my PharmD graduation, you were one of my groomsmen, and

now I am pleased to be able to call you one of my colleagues. You are truly one of the hoopiest froods I know. Jules Harral and Michelle Carr, you both helped provide me with the basics of molecular biology research, and helped ignite in me a passion for the work we do in the lab. Even though we don't still work together, I think you are two of the finest scientists I have ever met, and I am so privileged to call you my friends. If my experiences with you both at the Center for Genetic Lung Disease hadn't been so positive, I am not sure I would even be a scientist today. And someday, I hope we can get the lab back together…..I just need to figure out how to bridge oncology with PAH.

Howard McLeod and Chris Walko, I want to thank you both for being two of my biggest supporters while you were here at UNC, for continuing to advise me on all things science, oncology and pharmacogenetics, and for helping me to navigate the first steps in a career as a translational scientist. Even now that you have moved on to Moffitt Cancer Center in Tampa, you are still invested in my success, and I appreciate that deeply. Rob MacLaren, thank you so very much. I could honestly not have asked for a better PharmD Honors Project advisor, and I hope I lived up to your expectations as a student, and even as a babysitter for your kids. Thank you for your many letters of support, and for continuing to provide me with career guidance all these years later. Robert Page, you are certainly the most animated clinician scientist I know. Even though I am not interested in cardiology, you instilled in me a vigor and enthusiasm for direct patient care. Thank you for your never-ending support of me too. Steve Paugh, thank you for mentoring me during my summer in the POE program at St. Jude's Children's Research Hospital, and thank you for your continued help and guidance to this day. You certainly were indispensable in helping me design many of my assays I performed in my dissertation research. Finally, I would like to thank my ASCPT mentor,

Michelle Rudek, and my UNC Alumni mentor, Nader Moniri, for all of the hours they spent on the phone with me advising me on all facets of my science, professional development, and on planning for next steps in my career.

I could not have made it through this process without the love, kindness and support of the many friends and colleagues I have made since moving to North Carolina: Bob Schuck and Brittney Wright, thank you for being here in North Carolina and being our friends. We miss you every day. Colin Sheffield, thank you so much for being my friend and going to all the concerts with me. As you know, music fuels my soul. It has been so wonderful getting to know you, and I am so thankful you and Amanda Corbett have become such great friends. Kevin Watt, thank you for being the other "old man" in the DPET PhD program, and thanks for becoming one of my best friends here. I want to sincerely thank all of my other friends I've met while in North Carolina that I have not already mentioned, including: Jessica (Adams) and Ian Murphy, James Huckle and Ericka Mallow, Dylan Glatt (despite the fact that he went to Dakota Ridge), Paul Kim, Keith Carr, Dan and Taryn Golin (despite the fact she went to Chatfield), Jenna (Siskey) and VJ Nigro, Jai Patel, Bob Wittorf, Andy Madden, and Allison Schorzman. And a special shout-out goes to the best next-door neighbors and friends I could ask for: thank you so much Mike and Keiko Bury. An extra special shout-out goes out to my Russian friend, Olga Galeyeva. For just being you. You are very good. Stay weird.

I would also like to thank my friends back in Colorado for their support of me. Jon Bastone, you are my best man, and truly one of the best men I know. Dan Doyle, you are so intelligent and empathetic, and I really do appreciate the depth and breadth of our friendship. Lisa Yacko, you are undoubtedly one of my most cherished friends. Thank you for always

being there for me. Jill (Kennedy) Klem, you are one of the absolute funniest people I know and constantly make me laugh. My life is better with you in it. A special thank you goes out to my wonderful cousins Kelly Dimond and Erin Marvin (both of whom are more than just cousins, and have become two of my best friends). I love you both immensely and really appreciate all of your support over the years. And to any other friends I am forgetting, THANK YOU TOO!

Above all, I would like to thank my family for your unconditional love and unwavering support of me through all of the good times and bad. Julie and Jim Crona (aka Mom and Dad), the two of you set excellent examples for me as people that value education, and instilled in me a tenacious work ethic. It doesn't hurt that I also inherited some pretty great genes from you both! Meghan, Katie and Kevin (aka my siblings), I would like to thank you for being loving supporters of me throughout this process. Meghan, you provide me with an example of a person who is adventurous, and takes the risks I cannot. I am so jealous of your life in Ireland. Kevin, I am so proud of you and am quite honored that you chose pharmacy as a career (in part) because of me. That is quite a compliment. Katie, I want to give you an extra shout-out for being my running partner. I don't think I would have made it through this PhD without running becoming my stress reliever. I love that we have run marathons together! I love all three of you very, very much. A very special thank you also goes out to my godparents, Duane and Susie Bollig. I love you both very much. I would also like to thank my four-legged, furry family members for their unconditional love and support: Duke, Sweet Dee, Cooper and Danger- I love all of you and thank you so much.

Most of all, I want to thank my wife Lana. It has been a long road, but we made it. I could not have gotten through pharmacy school without you, and I certainly would have

never finished graduate school without your love, support, kindness, and humor. I appreciate you and everything you have done more me more than words can express, and more than you will ever know. You are the rose of my heart, and you are the love of my life. I love you more than anything, dear Nannybelle. Thank you so much for joining me on our adventure.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

%CV	Percent coefficient of variation (interpatient variability)
3C	Chromosome conformation capture
ABCB1	ATP-binding cassette, sub-family B (MDR/TAP), member 1
ABCC1	ATP-binding cassette, sub-family C (MRP), member 1
ABCC2	ATP-binding cassette, sub-family C (MRP), member 2
ADRs	Adverse drug reactions
AKT	Protein kinase B
ALL	Acute lymphoblastic leukemia
ALT	Alanine transaminase
ANC	Absolute neutrophil count
ANOVA	Analysis of variance
AST	Aspartate aminotransferase
AUC	Area under the curve
AXIS	Axitinib Versus Sorafenib in Advanced Renal Cell Carcinoma
BID	Twice per day
BRAF	v-Raf murine sarcoma viral oncogene homolog B
BSA	Body surface area
Cas-9	CRISPR-associated nuclease protein 9
CDK-1	Cyclin dependent kinase 1
cDNA	Complementary deoxyribonucleic acid
ChIP-Seq	Chromatin immunoprecipitation sequencing
CI	Confidence interval
c-KIT	v-Kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
C_{max}	Maximum or peak concentration

c-MET	MET proto-oncogene, receptor tyrosine kinase
CNV	Copy number variation
CPIC	The Clinical Pharmacogenetics Implementation Consortium
CRISPR	Clustered regularly interspaced short palindromic repeats
CT	Computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
CYP	Cytochrome P450
dbSNP	SNP Database
DLT	Dose limiting toxicity
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DPD/*DPYD*	Dihydropyrimidine dehydrogenase
EBM-2	Endothelial growth basal medium
EBP50	Ezrin-radixin-moesin-binding phosphoprotein-50
EC50	Half maximal effective concentration
EGFR	Epidermal growth factor receptor
ECM	Extra cellular matrix
ECOG	Eastern Cooperative Oncology Group
ELISA	Enzyme-linked immunosorbent assay
EMMA	Efficient mixed model association
EMSA	Electromobility shift assays
ENCODE	Encyclopedia of DNA Elements
eQTL	Expression quantitative trait locus/loci
ERK	Extracellular signal-regulated kinases

ERM	Ezrin-radixin-moesin
FBS	Fetal bovine serum
FDA	Food and Drug Administration
FDR	False discovery rate
FGFR-1	Fibroblast growth factor receptor-1
FLT-3	FMS-like tyrosine kinase-3
FLT-4	FMS-like tyrosine kinase-4
FOLFIRI	5-fluorouracil, leucovorin and irinotecan chemotherapy regimen
GWA	Genome wide association
GWAS	Genome wide association study/studies
HapMap	Haplotype Map
HCC	Hepatocellular carcinoma
HER-2	Human epidermal growth factor receptor 2
HIF1α	Hypoxia Inducible Factor-1α
HCS	High-content cell-based screening
HEATR7B1	Maestro heat-like repeat family member 2A (*MROH2A*)
HEK	Human embryonic kidney
HR	Hazard ratio
HUVEC	Human umbilical vein endothelial cells
HWE	Hardy–Weinberg equilibrium
IC_{50}	Half maximal inhibitory concentration
IFN	Interferon
IGFR	Insulin growth factor receptor
IL	Interleukin
INDEL	Insertion/Deletion

ITGAV	Integrin alpha-V
KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
LD	Linkage disequilibrium
LDH	Lactate dehydrogenase
LPEC	Liver parenchyma endothelial cells
LRT	Likelihood ratio test
MAF	Minor allele frequency
MAPK	Mitogen-activated protein kinases
MCL-1	Myeloid leukemia-1
MDP	Mouse Diversity Panel
MDR1	Multidrug resistance protein 1
MEFs	Mouse embryonic fibroblasts
MEK	Mitogen-activated protein kinase
MGI	The Mouse Genome Informatics Database
MRI	Magnetic resonance imaging
mRNA	Messenger ribonucleic acid
MSKCC	Memorial Sloan Kettering Cancer Center
mRCC	Metastatic renal cell carcinoma
MTD	Maximum tolerated dose
MTHFR	Methylene tetrahydrofolate reductase
mTOR	Mammalian target of rapamycin
MVD	Microvascular density
NCCN	National Comprehensive Cancer Network
NCI	National Cancer Institute
NIGMS	National Institute of General Medical Sciences

NIH	National Institutes of Health
NRP-1	Neuropilin-1
NSCLC	Non-small cell lung cancer
OBD	Optimal biologic dose
OCT	Organic cation transporter
OATP	Organic anion-transporting polypeptide
OR	Odds ratio
OS	Overall survival
pAKT	Phosphorylated protein kinase B
PANTHER	Protein Analysis Through Evolutionary Relationships
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDGF	Platelet derived growth factor
PDGFR	Platelet derived growth factor receptor
PFS	Progression free survival
Pgp	P-glycoprotein
PI3K	Phosphatidylinositol-4,5-bisphosphate 3-kinase
PGENI	Pharmacogenomics for Every Nation Initiative
PH	Proportional hazards
PharmGKB	Pharmacogenomics Knowledge Base
PK	Pharmacokinetics
PIM-1	Proviral integration site 1
PKA	Protein kinase A
PKB	Protein kinase B (also known as AKT)
PKC	Protein kinase C

PROVEAN	Protein Variation Effect Analyzer
PS	Performance status
QC	Quality control
QTL	Quantitative trait locus/loci
RAF	Rapidly accelerated fibrosarcoma
RANKL	Ligand of receptor activator of nuclear factor-κB
RCC	Renal cell carcinoma
RECIST	Response Evaluation Criteria in Solid Tumors
RFU	Relative fluorescent units
RNA	Ribonucleic acid
RR	Relative risk
SCAN	SNP and CNV Annotation
SE	Standard error
SEM	Standard error of the mean
SIFT	Sorting Intolerant From Tolerant
siRNA	Small interfering RNA
SLC22A1	Solute carrier family 22 (organic cation transporter), member 1
SLCO1B1	Solute carrier organic anion transporter family, member 1B1
SLCO1B3	Solute carrier organic anion transporter family, member 1B3
SN-38G	SN-38 glucuronide
SNP	Single nucleotide polymorphism
TALEN	Transcription activator-like effector nuclease
TARGET	Treatment Approaches in Renal Cancer Global Evaluation Trial
TF	Transcription factor
TGF-β	Transforming growth factor-β

TGN	Thioguanine nucleotides
TIME	Telomerase immortalized microvascular endothelial
TPMT	Thiopurine-S-methyltransferase
TTP	Time to progression
TYMS	Thymidylate synthetase
UCSC	University of California Santa Cruz
UGT	Uridine-5'-diphospho-glucuronosyltransferase
UGT1A9	Uridine 5'-diphospho-glucuronosyltransferase 1 family, polypeptide A9
ULN	Upper limit of normal
UTR	Untranslated region
VEGF	Vascular endothelial growth factor
VEGFA	Vascular endothelial growth factor A
VEGFC	Vascular endothelial growth factor C
VEGFR	Vascular endothelial growth factor receptor
VHL/*VHL*	von Hippel-Lindau
VOC	Valid object count
WT	Wild type
WWOX	WW domain-containing oxidoreductase
XRT	External radiation therapy

CHAPTER 1: RENAL CELL CARCINOMA: ANGIOGENESIS, VEGF-PATHWAY INHIBITORS AND BIOMARKERS

1.1 Overview

Cancers of the kidney and renal pelvis account for approximately 2-3% of all adult malignancies, and have increased in overall incidence over the past few decades. The most common subtype of kidney cancer arises from the renal parenchyma in the proximal tubules of the kidney and is classified as renal cell carcinoma (RCC). Approximately 30% of RCC patients will present initially with metastatic disease, and another 30% will relapse after surgical resection of their primary tumor. RCC responds poorly to standard cytotoxic chemotherapy and, prior to advent of targeted multikinase inhibitor therapies, interleukin-2 (IL-2) and interferon-α (INF-α) were the only systemic therapies commonly used for the treatment of advanced or metastatic RCC (mRCC).

However, over the past decade the treatment landscape for mRCC has changed dramatically due to the U.S. Food and Drug Administration (FDA) approval of multiple agents that target tumorigenic and angiogenic pathways. The approval of seven agents, which target angiogenic and/or oncogenic signaling pathways, has helped increase median survival time amongst mRCC patients. Nevertheless, despite these major advancements, most patients experience disease progression while on treatment and mRCC is eventually their cause of death.

Sorafenib is a multikinase inhibitor with potent activity against angiogenic, oncogenic, and stromal kinases, as well as the RAF/MEK/ERK signaling pathway, which

leads to inhibition of tumor proliferation and angiogenesis. Data from the pivotal phase III randomized, placebo-controlled, multicenter Treatment Approaches in Renal Cancer Global Evaluation Trial (TARGET) confirmed a significant overall survival (OS) and progression-free survival (PFS) benefit. These data led to its U.S. FDA approval in December 2005 for the treatment of patients with advanced or metastatic RCC.

Despite the recent U.S. FDA approval of several additional multikinase inhibitors for the treatment of mRCC, there is a clear unmet need to identify and validate prognostic and predictive biomarkers that associate with improved survival. Because anti-angiogenic multikinase inhibitors target, in addition to the tumor itself, non-malignant endothelial cells and tumor microenvironment, germline DNA variations likely affect the treatment efficacy and/or toxicity profiles of these drugs. In addition, because RCC is a highly vascularized tumor type and considerable interindividual variability in response to sorafenib is observed clinically, identification and validation of germline genetic variants that associate with sorafenib response may help determine which patients should be treated with sorafenib. In a crowded landscape of targeted agents for the treatment of mRCC, identification of predictive pharmacogenetic variants could certainly impact clinician treatment decisions and improve patient outcomes.

Additionally, the identification of novel prognostic markers may provide insight into RCC pathogenesis/prognosis, or identify patients who would benefit from more intensive therapies and/or monitoring. Germline single-nucleotide polymorphisms (SNPs) in angiogenesis pathway genes have associated with patient outcome in numerous tumor types, but the results are often inconsistent across studies and rarely validated. Furthermore, previous candidate gene pharmacogenetic studies of oral multikinase inhibitors have each

only interrogated a small number of candidate genes or SNPs. Moreover, for the variants identified in these studies, little information regarding their effects on angiogenesis, at the molecular and cellular level, is available.

This research stems from the hypothesis that germline genetic variants in mRCC patients will help explain the interindividual differences in sorafenib response and patient survival. This hypothesis will be addressed through three aims, described in detail below. The overall goal of this dissertation research is to identify and validate predictive germline genetic markers of sorafenib efficacy, and prognostic germline genetic markers that associate with RCC pathogenesis and/or prognosis.

1.2 Renal Cell Carcinoma

Cancers of the kidney and renal pelvis account for approximately 2-3% of all adult tumors, and the overall incidence has increased over the past few decades.[1,2] Renal cell carcinomas (RCC) arise from the epithelia that lines the renal tubules, and at least 85-90% of all malignancies arising in the kidney and renal pelvis can be classified as RCC.[3] The most common histological subtype of RCC is clear-cell RCC (70-80% of all cases of RCC).[4-6] Clear-cell RCC often presents as a single solid tumor located at the periphery of the renal parenchyma, and is defined by its optically clear cytoplasm, with nested clusters of cells surrounded by a dense endothelial network.[7,8]

Worldwide, RCC is the ninth most common type of cancer, with an estimated incidence of approximately 337,860 newly diagnosed cases and approximately 143,406 deaths in 2012.[9] In the U.S., RCC is the seventh most common cancer among men, and ninth most common cancer among women. It constitutes approximately 3.9% of all new cancers, with a median age of 64 years at diagnosis.[10] The American Cancer Society estimated that

approximately 63,920 new RCC cases were diagnosed, and 13,860 deaths (8,900 men and 4,960 women) occurred due to RCC in the U.S. in 2014.[1]

Median overall survival rates for RCC have improved over the past two decades, which could be attributed to improved screening and early detection of smaller tumors, the use of cytoreductive nephrectomy prior to the use of systemic therapy in advanced disease, and/or the U.S. FDA approval of multiple agents that target angiogenic and oncogenic signaling pathways. The 5-year and 10-year relative survival rates for kidney cancer are 72% and 62%, respectively. A majority of RCC cases are diagnosed at an early stage when disease is localized (64%), and the 5-year relative survival rate for these patients is 92%. Overall, the 5-year survival for all patients with RCC is 74%, and as high as 96% when patients present with stage I disease.[11,12] However, approximately 30% of RCC patients will present initially with metastatic disease, and an additional 30-50% of RCC patients, initially thought to be curable through nephrectomy, will relapse.[13-16] The median survival time for patients with metastatic disease is 10-12 months,[17] the 5-year survival rate for these patients is approximately 23%, and the 10-year survival is only 12.3%.[2,10,12]

1.3 Renal Cell Carcinoma and Angiogenesis

RCC arises from a series of mutations and selection events in cells of the proximal tubules of the nephron. These events ultimately result in the formation of cells that possess characteristics that are consistent with the hallmarks of cancer: unregulated cellular proliferation, growth in a hypoxic environment, recruitment of pro-angiogenic factors, evasion of anti-apoptotic signaling, invasion of basement membrane, and ultimately distant metastases.[18]

A seminal event in the pathogenesis of clear-cell RCC is loss of function of the von Hippel-Lindau (VHL) tumor suppressor gene. *VHL* was identified in 1993,[19] contains three exons, and is located on the short arm of chromosome 3 (3p25). Germline inheritance of mutated or deleted *VHL* alleles is the primary etiology for inherited clear-cell RCC. In addition, at least 75% of sporadic clear-cell RCC cases also occur as a consequence of aberrant *VHL* function.[18-21] Indeed, biallelic gene inactivation of *VHL* is a hallmark event that promotes clear-cell RCC tumor development, and it classically conforms to the Knudson 2-hit carcinogenesis model in cases of sporadic clear-cell RCC.[22] A deletion of one *VHL* allele results in a loss of heterozygosity in more than 90% of cases of sporadic clear-cell RCC.[23] Subsequently, the second allele can be inactivated through additional gene mutations,[24] or through gene silencing secondary to hypermethylation.[25,26] In contrast to inherited clear-cell RCC, both the first and second "hits" occur as a result of somatic mutations, rather than germline mutations.[20]

VHL is an upstream mediator of a family of transcription factors, known as hypoxia inducible factors. Under normoxic conditions, VHL marks hypoxia inducible factor-1α (HIF-1α) for ubiquitination and proteasomal degradation.[27] Ultimately, dysfunctional VHL protein prevents HIF-1α degradation, which allows HIF-1α to translocate into the nucleus to facilitate transcription of pro-angiogenic and mitogenic factors (including vascular endothelial growth factor [VEGF], transforming growth factor beta [TGF-β], and platelet derived growth factor [PDGF]). RCC is often associated with up-regulated activity of distal HIF-1α effectors (VEGFA, VEGFRs, TGF-β [and its receptor, EGFR], PDGF, and PDGFRs) and the RAF/MEK/ERK signaling pathway (Figure 1.1).[18,28,29] Because angiogenesis is so central to the pathophysiology of RCC, there is clear rationale for administering multikinase

inhibitors that target distal HIF-1α effectors, primarily pro-angiogenic effectors in the VEGF-pathway, to patients with mRCC.

1.4 The Treatment of Advanced or Metastatic Renal Cell Carcinoma

RCC responds poorly to traditional cytotoxic chemotherapy. While multiple agents (e.g. gemcitabine, vinblastine, and 5-fluorouracil) have been tested in patients with mRCC, response rates are extremely poor (4% to 6%).[3,30] One main mechanism of resistance to traditional chemotherapy could be related to the expression of MDR1, which encodes for the P-glycoprotein (Pgp) drug efflux transporter, in the proximal tubules of the kidney.[3] Prior to 2005, pharmacotherapeutic options for mRCC patients were limited to immunotherapies (IL-2 and INF-α). However, IL-2 and/or INF-α are highly toxic to patients, only a subset of RCC patients adequately respond to these therapies,[31] and prognosis for patients with mRCC receiving immunotherapy was poor with less than 10% achieving durable and complete remissions.[5]

Over the past decade the treatment landscape for mRCC has changed dramatically due to the U.S. FDA approval of multiple agents that target tumorigenic and angiogenic pathways. The approval of seven agents (Figure 1.2), which target angiogenic and/or oncogenic signaling pathways (notably, inhibitors of the VEGF-pathway and the mammalian target of rapamycin [mTOR] pathway) based on the pathophysiology of the disease (Figure 1.1), has helped increase median survival time amongst mRCC patients.[2,32-38]

Ultimately, 5-year survival rates for RCC patients with advanced disease have improved from 7.3% during 1992-1995 to 12.3% during 2004-2010,[2] which is likely to be at least partially attributable to the approval of multiple oral multikinase inhibitors that all have

a positive impact on patient overall survival. Nevertheless, despite these major advancements, most patients experience disease progression while on treatment and mRCC is eventually their cause of death.[7] In 2014, it was estimated that over 13,000 patients in the U.S. died from RCC,[2] and these statistics highlight the need for mRCC treatment optimization.

1.5 Sorafenib

Sorafenib tosylate (Nexavar; Bayer HealthCare Pharmaceuticals Corporation, Wayne, NJ; Onyx Pharmaceuticals, South San Francisco, CA) is an orally administered biaryl urea agent that is also a potent multikinase inhibitor. The chemical name of sorafenib is 4-[4-[[4-chloro-3-(trifluoromethyl)phenyl]carbamoylamino]phenoxy]-N-methyl-pyridine-2-carboxamide (Figure 1.3). It has a broad spectrum of activity in angiogenic, oncogenic, and stromal kinases, as well as the RAF/MEK/ERK signaling pathway. Sorafenib was originally developed as a RAF kinase inhibitor, but was subsequently shown to effectively inhibit VEGFR-1, -2 and -3, PDGFR-β, FMS-like tyrosine kinase-3 (FLT-3), fibroblast growth factor receptor-1 (FGFR-1), RAF-1, BRAF (wild-type and mutant $BRAF^{V600E}$), and c-KIT (cellular homolog of the feline sarcoma viral oncogene v-kit) receptor tyrosine kinases in multiple tumor cell lines (Table 1.1).[39-42] Sorafenib also exhibited broad-spectrum, dose-dependent inhibitory activity in multiple mouse xenograft models, including: breast, colon, lung, thyroid, and kidney tumors, as well as melanoma.[39,41] The anti-proliferative effects of sorafenib are largely dependent on the inhibition of oncogenic signaling pathways that regulate tumor proliferation.[41] Sorafenib has also been shown to induce apoptosis in numerous cell lines.[41] While the mechanisms underlying its pro-apoptotic effects are not well understood, one proposed hypothesis revolves around sorafenib's ability to inhibit

phosphorylation of initiation factor eIF4E combined with the loss of the anti-apoptotic myeloid leukemia-1 (MCL-1) protein.[43] Dynamic contrast-enhanced magnetic resonance imaging (MRI) in RCC patients revealed that sorafenib also significantly altered vascular permeability and tumor perfusion.[44]

Data from four dose-escalation phase I trials revealed that sorafenib was relatively safe at its maximum tolerated dose (MTD) of 400 mg twice daily (BID).[45-48] However, there was also a high degree of interpatient variability in the sorafenib pharmacokinetic profile for patients enrolled on these trials. The mean elimination half-life of sorafenib is approximately 25–48 hours. Multiple dosing at the MTD for seven days resulted in sorafenib accumulation levels 2.5- to 7-fold higher than when a single dose was administered.[40,47] Steady-state concentrations of sorafenib were reached after seven days of dosing, and no additional accumulation observed after steady-state was reached.[48] In the non-continuous trials, the mean peak plasma concentration (C_{max}) and area under the concentration–time curve (AUC) values were substantially greater on the last day than they were after sorafenib administration on the first day.[49] And, at 200 mg BID and at the MTD of 400 mg BID, the interpatient variability (%CV) in sorafenib exposure (measured by its AUC) ranged from 5 to 83%, and from 33 to 88% for sorafenib C_{max}.[49,50] Fortunately, even with this wide interpatient variability in sorafenib pharmacokinetics, there was not an observed association between increased sorafenib exposure and increased toxicities (notably: fatigue, diarrhea and dermatologic toxicities).[49]

Sorafenib metabolism is mediated through two parallel pathways: pyridine N-oxide oxidation catalyzed by cytochrome P450 (CYP) 3A4, and through glucuronide conjugation by uridine 5'-diphospho-glucuronosyltransferase (UGT) 1A9. At steady state, the most

prevalent circulating analyte detected in the plasma is parent sorafenib (70-85%); however, the main pyridine N-oxide is still detected at high levels, and has been shown to be as potent as the parent drug.[40,51] In patients with mild or moderate hepatic dysfunction (Child Pugh A and Child Pugh B) who received sorafenib twice daily at the MTD, AUC values for the N-oxide metabolite were 23-65% lower than for patients without hepatic impairment.[52]

As a target of glucuronide conjugation, sorafenib is believed to undergo extensive enterohepatic recirculation, as evidenced by occurrence of observable double peaks in the concentration–time profiles among patients treated with sorafenib. This is supported by population pharmacokinetic modeling, which adequately described sorafenib disposition when accounting for enterohepatic recirculation in the model.[50] Sorafenib is highly bound to plasma proteins (99.5%), and because of its lipophilic characteristics, it is widely distributed to tissues. However, recent studies have also shown that sorafenib also undergoes OCT1, OATP1B1, and OATP1B3-mediated active transport.[53,54]

Data from a phase II randomized discontinuation trial showed that sorafenib significantly improved PFS in mRCC patients.[55] Data from the pivotal phase III randomized, placebo-controlled, multicenter TARGET confirmed a significant PFS benefit, and showed a trend towards an OS benefit in patients with mRCC treated with sorafenib. Median PFS was significantly improved for patients treated with sorafenib,[33] and final survival analyses revealed improved OS for patients treated with sorafenib.[56] Based on these clinical trial data, the U.S. FDA approved sorafenib in December 2005 as the first anti-angiogenic multikinase inhibitor for the treatment of mRCC.

1.6 Renal Cell Carcinoma Biomarkers

The development and approval of oral multikinase inhibitors, such as sorafenib, that target the angiogenesis and the VEGF-pathway have improved overall survival for many patients with mRCC. However, there is a significant interindividual variability when it comes to the benefit of these medications. And, the overall response rate, defined generally as the proportion of patients with reduction in tumor burden of a predefined amount, only ranged from 10-44% in patients that received front-line VEGF-pathway inhibitor therapy.[33,36,37,57] The identification of prognostic and predictive biomarkers is an important next step in the evolution of mRCC treatment, and will help clinicians prioritize the use and sequence of the seven targeted agents approved over the past decade.

Prognostic biomarkers are used to evaluate phenotypes, which correlate with survival outcomes, independent of treatment.[12,58] Clinical prognostic biomarkers have been used extensively to estimate RCC prognosis. The Memorial Sloan Kettering Cancer Center (MSKCC) risk criteria score for estimating survival has been incorporated into routine clinical practice, and categorizes mRCC patients into low, intermediate and high risk categories. The MSKCC risk score examined five prognostic factors (serum hemoglobin levels, corrected serum calcium levels, serum lactate dehydrogenase (LDH) levels, interval between diagnosis and the start of treatment, and Karnofsky performance status) in mRCC patients.[59,60] More recently, newer prognostic models have been developed subsequent to the U.S. FDA approval of VEGF-targeting agents.[61,62] And, additional histological, molecular (e.g. circulating tumor cells, serum amyloid A protein, C-reactive protein, HIF-1α, phosphatidylinositol-4,5-bisphosphate 3-kinase [PI3K], etc.), and more recently genetic biomarkers have also been investigated as measures of prognosis.[12]

Predictive biomarkers are used to predict the clinical benefit and/or response to medications, and can be followed throughout the course of treatment.[12,58] No clinically validated biomarkers for RCC are utilized. However, types of predictive molecular biomarkers have been investigated, including: circulating biomarkers (e.g. VEGFA, sVEGFR2 and sVEGFR3), cytokine angiogenic factors (e.g. baseline IL-6 and elevated LDH), tissue-based biomarkers (e.g. *VHL* mutations), and factors in the mTOR pathway (e.g. elevated phosphor-S6 expression, and elevated phosphorylated protein kinase B [pAKT] expression).[58] In addition, single nucleotide polymorphisms (SNPs) that associate with differences in pharmacokinetics/pharmacodynamics, and that associate with differences in survival have been postulated to be predictive biomarkers of mRCC treatments.[63-68]

To date, mRCC remains incurable, despite the approval of several targeted therapies, and there is a clear unmet need to identify and validate markers that associate with improved survival. And, despite the U.S. FDA approval of multiple VEGF pathway inhibitors that have become the mainstay for pharmacotherapeutic treatment of mRCC, many unanswered questions remain regarding the choice of drug for an individual patient, the timing of and dose at treatment initiation, and the optimal sequencing of these agents for an individual patient. But, there are currently no validated molecular/genetic prognostic or predictive biomarkers that have been incorporated into routine clinical practice to help answer these questions and help clinician decision making. The identification, validation and clinical implementation of novel prognostic and predictive biomarkers are important towards optimizing mRCC therapies, and could certainly help ascertain which patients would receive the greatest clinical benefit from sorafenib.

1.7 Purpose of the Research

Despite the recent U.S. FDA approval of several multikinase inhibitors for the treatment of mRCC, there is a clear unmet need to identify and validate prognostic and predictive markers that associate with improved survival. Because anti-angiogenic multikinase inhibitors target, in addition to the tumor itself, non-malignant endothelial cells and tumor microenvironment, germline variations likely affect the treatment efficacy and/or toxicity profiles of these drugs. In addition, because RCC is a highly vascularized tumor type and considerable interindividual variability in response to sorafenib is observed clinically, identification and validation of germline genetic variants that associate with sorafenib response may help determine which patients should be treated with sorafenib. In a crowded landscape of targeted agents for the treatment of mRCC, identification of predictive pharmacogenetic variants will certainly impact clinician treatment decisions.

Additionally, the identification of novel prognostic markers may provide insight into RCC pathogenesis/prognosis, or identify patients who would benefit from more intensive therapies and/or monitoring. Germline single-nucleotide polymorphisms (SNPs) in angiogenesis pathway genes have associated with patient outcomes in numerous tumor types, but the results are often inconsistent across studies and results are rarely validated. Furthermore, previous candidate gene pharmacogenetic studies of oral multikinase inhibitors have interrogated small numbers of candidate genes or SNPs. Moreover, for the variants identified in these studies, little information regarding their effects on angiogenesis, at the molecular and cellular level, is available.

1.8 Specific Aims

The central hypothesis of this research is that identification and validation of germline genetic variants in mRCC patients will help explain the interindividual differences in sorafenib response and OS. This hypothesis will be addressed through three aims, described in detail below. The overall goal of these studies is to identify and validate predictive germline genetic markers of sorafenib efficacy and/or pharmacology, and prognostic germline genetic markers that associate with RCC pathogenesis and/or prognosis.

Aim 1. To genotype candidate SNPs from 56 candidate genes, using available genomic DNA from TARGET patients, and test associations with OS.

Hypothesis: Germline variants in genes related to RCC prognosis/pathogenesis, the angiogenesis pathway and/or sorafenib pharmacology will associate with OS in patients with mRCC enrolled on the phase III TARGET trial.

Significance: The oral multikinase inhibitor sorafenib helped revolutionize the treatment of mRCC, but mRCC remains incurable, even for patients with stage IV disease who have been treated with sorafenib.[69] In addition, wide interindividual variation in response to sorafenib clinically, coupled with the recent U.S. FDA approval of three additional oral anti-angiogenic multikinase inhibitors, has revealed an unmet need in the treatment of mRCC. Identification of predictive markers of sorafenib response may help identify a subpopulation of mRCC patients who will benefit most from sorafenib therapy, while identification of novel prognostic markers will provide clinicians and researchers with insights concerning the tumor biology of mRCC and identify patients that may require more intensive therapies and/or monitoring.

Rationale: Anti-angiogenic multikinase inhibitors target tumor cells, host endothelial cells, pericytes, and even the tumor microenvironment rather than simply targeting the tumor cell alone. Therefore, germline variation is likely an important determinant of drug response in multikinase inhibitors.[70] Identification and validation of germline genetic variants that significantly associate with survival will help identify patients who are optimal candidates for sorafenib therapy.

Aim 2. To validate functionality of germline variants (identified in Aim 1) that associate with OS in TARGET patients.

Hypothesis: Functional validation of germline variants that significantly associate with OS in TARGET patients can help elucidate the molecular effects of these variants on RCC pathogenesis/prognosis, angiogenesis and/or sorafenib pharmacology.

Significance: Findings from pharmacogenetic and pharmacogenomic studies (both candidate gene and genome wide association studies [GWAS]) continue to provide a plethora of information about genetic variation that underlies both disease pathology and responses to pharmacotherapy. However, a clear understanding of the molecular effects of candidate variants (selected from significant associations between genotype and clinical phenotypes) is often absent. Since a tagging SNP approach was employed to select SNPs for genotyping TARGET patient DNA, it is imperative that the causal variant(s) is identified. Therefore, it is important that a series of validation assays characterize the effect(s) of the variant on signaling pathways and/or drug response. Essential steps in the functional validation process, aside from genotyping and imputation, include: *in silico* analyses to identify and prioritize putatively functional variants, and *in vitro* validation of predicted molecular effects to

provide a basic explanation of the mechanistic processes that underlie the genotype-phenotype associations. Validation of these germline variants (in the absence of, or in conjunction with replicative genotyping in an independent, external cohort) is essential to their translation into potentially useful biomarkers that will inform treatment decisions concerning sorafenib therapy for mRCC patients.

Rationale: Many significant genotype-phenotype associations, derived from multikinase inhibitor pharmacogenetic studies, lack validation.[63-68] Information regarding the molecular effects, which underlie disease pathogenesis and/or response to therapy, is direly needed. Therefore, a sequential *in silico* → *in vitro* approach to validate germline variants of interest (garnered from Aim 1) will be employed. Functional validation of genetic variants that reveal significant associations with clinical phenotypes and drug response can increase the validity of the observed associations.[71,72] Elucidation of the molecular effects of variants will help translate the genotype-phenotype associations derived from pharmacogenetic studies into clinically useful biomarkers.

Aim 3. To use *in vitro* cell models to discover novel candidate genes and signaling pathways related to sorafenib cytotoxicity

Hypothesis: Differential cell health and response data (e.g. EC_{50} or IC_{50} values) from 32 MEF cells lines treated with sorafenib, can be used in GWAS to identify candidate genes associated with sorafenib response, which will ultimately lead to the discovery of novel genes for future pharmacogenetic testing in patients treated with sorafenib.

Significance: Aim 1 identified germline variants that significantly associate with OS through a candidate gene/candidate SNP approach. Since this approach leverages existing knowledge

about mRCC pathogenesis/prognosis, angiogenesis and/or sorafenib pharmacology, there is little chance that novel and previously unidentified signaling pathways or candidate genes will be discovered. This aim will use a cellular genetics approach, using high-content cellular imaging and genetic mapping, and will help discover novel genes and pathways involved with sorafenib cytotoxicity and provide a better understanding of the variability observed with this phenotype.

Rationale: Previous studies have shown that GWAS mapping can be successfully performed in a panel of diverse inbred strains of mice to identify genetic loci that contain candidate genes that modulate both single gene and polygenic traits.[73-76] But to date, there have been few examples of animal GWAS pharmacogenetics, and even those that have been published have not analyzed the contribution of genetics to multikinase inhibitor (e.g. sorafenib) cytotoxicity.[77,78] The use of genetically well-characterized inbred mouse strains provides a viable model system to analyze the genetic basis for cytotoxicity variability. Mouse embryonic fibroblasts (MEFs) from 32 inbred mouse strains are used for this high-throughput cellular genetics approach for four main reasons: they retain the exact genetic composition of the mouse strain from which they are isolated, selection of over 30 strains increases the likelihood of detecting genetic differences that underlie differences in sorafenib response, technological advances (e.g. high-content imaging) have allowed for better characterization of cellular phenotypes, and technological advances (e.g. siRNA loss of function and cDNA over-expression *in vitro* techniques) have allowed for functional validation of genetic loci.

TABLES

Table 1.1. Sorafenib inhibits angiogenic kinases and RAF/MEK/ERK pathway oncogenic kinases. Table adapted from Wilhelm, et al. *Cancer Res*. 2004;64:7099-7109. Abbreviations: BRAF, v-Raf murine sarcoma viral oncogene homolog B; c-KIT, v-Kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog; c-MET, MET proto-oncogene, receptor tyrosine kinase; CDK1, cyclin dependent kinase 1; EGFR, epidermal growth factor receptor, ERK, extracellular-signal-regulated kinase; FGFR-1, fibroblast growth factor receptor-1; FLT-3, fms-related tyrosine kinase 3; HER-2, human epidermal growth factor receptor 2; IC_{50}, half maximal inhibitory concentration; IGFR, insulin growth factor receptor ; MEK, ; mitogen-activated protein kinase kinase; PDGFR- β, platelet derived growth factor receptor β; PIM-1, proviral integration site 1; PKA, protein kinase A; PKB, protein kinase B; PKC, protein kinase C; RAF, rapidly accelerated fibrosarcoma; SD, standard deviation; VEGFR, vascular endothelial growth factor receptor.

Molecular Target Biochemical Activity	Sorafenib IC_{50} (mmol/L) ± SD
VEGFR-1	NA
VEGFR-2	90 ± 15
mVEGFR-2	15 ± 6
mVEGFR-3	20 ± 6
RAF-1	6 ± 3
BRAF WT	22 ± 6
$BRAF^{V600E}$	38 ± 9
FGFR-1	580 ± 100
mPDGFR-β	57 ± 20
c-KIT	68 ± 21
FLT-3	58 ± 20
IC50 >10,000 mmol/L: ERK-1, MEK-1, EGFR, HER-2, IGFR-1, c-MET, PKB, PKA, CDK1/CyclinB, PKC-α, PKC-γ, and PIM-1	

FIGURES

Figure 1.1. Overview of the RCC pathway with VEGF and therapeutic targets. Under normoxic conditions and with normal *VHL* function, the VHL protein is an integral part of the E3 ubiquitin ligase complex that marks HIF-*1*α for proteasomal degradation. Under hypoxic conditions and/or mutated VHL, HIF-1*α* is allowed to accumulate, which leads to the accumulation of HIF-1*α* transcription factors. HIF-1*α* can also accumulate secondary to activation of mTOR by PI3K/AKT signaling. Activated HIF-1*α* translocate to the nucleus and promotes transcription of pro-angiogenic genes, such as VEGF and PDGF. Transcriptional activation of these genes subsequently leads to the production of pro-angiogenic ligands that are released and able to bind to receptors present on the surface of the tumor cells, as well as on the surface host endothelial cells and/or pericytes. Activation of these receptors leads to increased migration, proliferation, and permeability of host vasculature. Bevacizumab is an inhibitor of the VEGFA ligand. Axitinib, pazopanib, sorafenib, and sunitinib are inhibitors of VEGFRs. Everolimus and temsirolimus are inhibitors of mTOR. Adapted from Rini BI, et al. *Lancet Oncol*. 2009;373:1119-1132. Abbreviations: AKT (also known as PKB), protein kinase B; HIF, hypoxia-inducible factor; mTOR, mammalian target of rapamycin; PDGF, platelet-derived growth factor; PDGFR, platelet-derived growth factor receptor; PI3K, phosphoinositide-3 kinase; TGF, transforming growth factor; VEGF, vascular endothelial growth factor; VEGFR, vascular endothelial growth factor receptor; VHL, von-Hippel Landau.

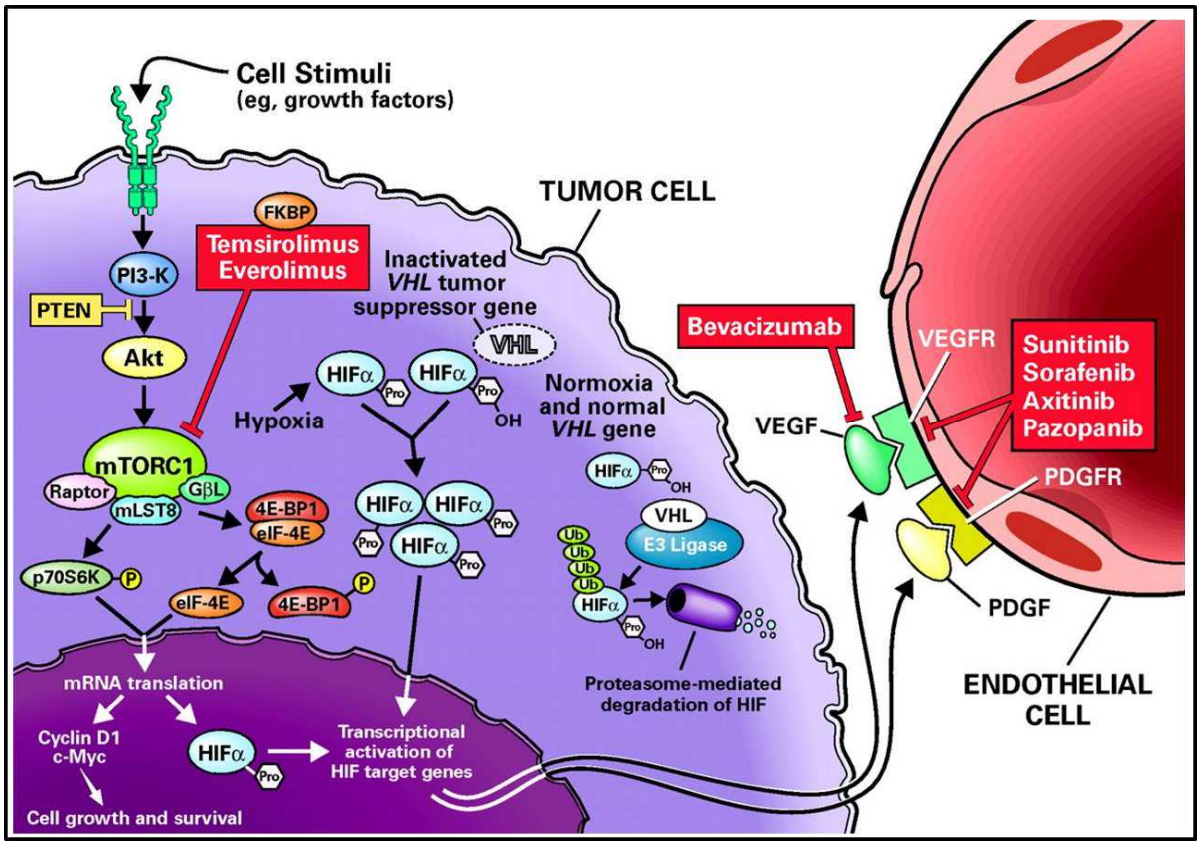

Cell Stimuli
(eg, growth factors)
TUMOR CELL
FKBP
Temsirolimus
Everolimus
PI3-K
PTEN
Akt
mTORC1
Raptor
GβL
mLST8
4E-BP1
eIF-4E
p70S6K
P
eIF-4E
4E-BP1
P
mRNA translation
Cyclin D1
c-Myc
HIFα
Pro
Transcriptional
activation of
HIF target genes
Cell growth and survival
Inactivated
VHL tumor
suppressor gene
VHL
HIFα
Pro
OH
Hypoxia
Normoxia
and normal
VHL gene
HIFα
Pro
OH
VHL
E3 Ligase
Ub
HIFα
Pro
OH
Proteasome-mediated
degradation of HIF
Bevacizumab
VEGF
VEGFR
Sunitinib
Sorafenib
Axitinib
Pazopanib
PDGFR
PDGF
ENDOTHELIAL
CELL

Figure 1.2. First and second line therapy recommendations for relapsed or Stage IV and surgically unresectable RCC. Per NCCN kidney cancer guidelines version 3.2015 (http://www.nccn.org/professionals/physician_gls/pdf/kidney.pdf), predictors of poor RCC prognosis include: lactate dehydrogenase >1.5 the upper limit of normal, hemoglobin level less than the lower limit of normal, corrected calcium >10.5 mg/dL (2.5 mmol/L), interval less than one year from original diagnosis until initiation of systemic therapy, Karnofsky performance score ≤70, ≥2 metastatic sites. Abbreviations: IFN, interferon; IL, interleukin; NCCN, National Comprehensive Cancer Network; PS, performance status RANKL, ligand of receptor activator of nuclear factor-κB; XRT, external radiation therapy.

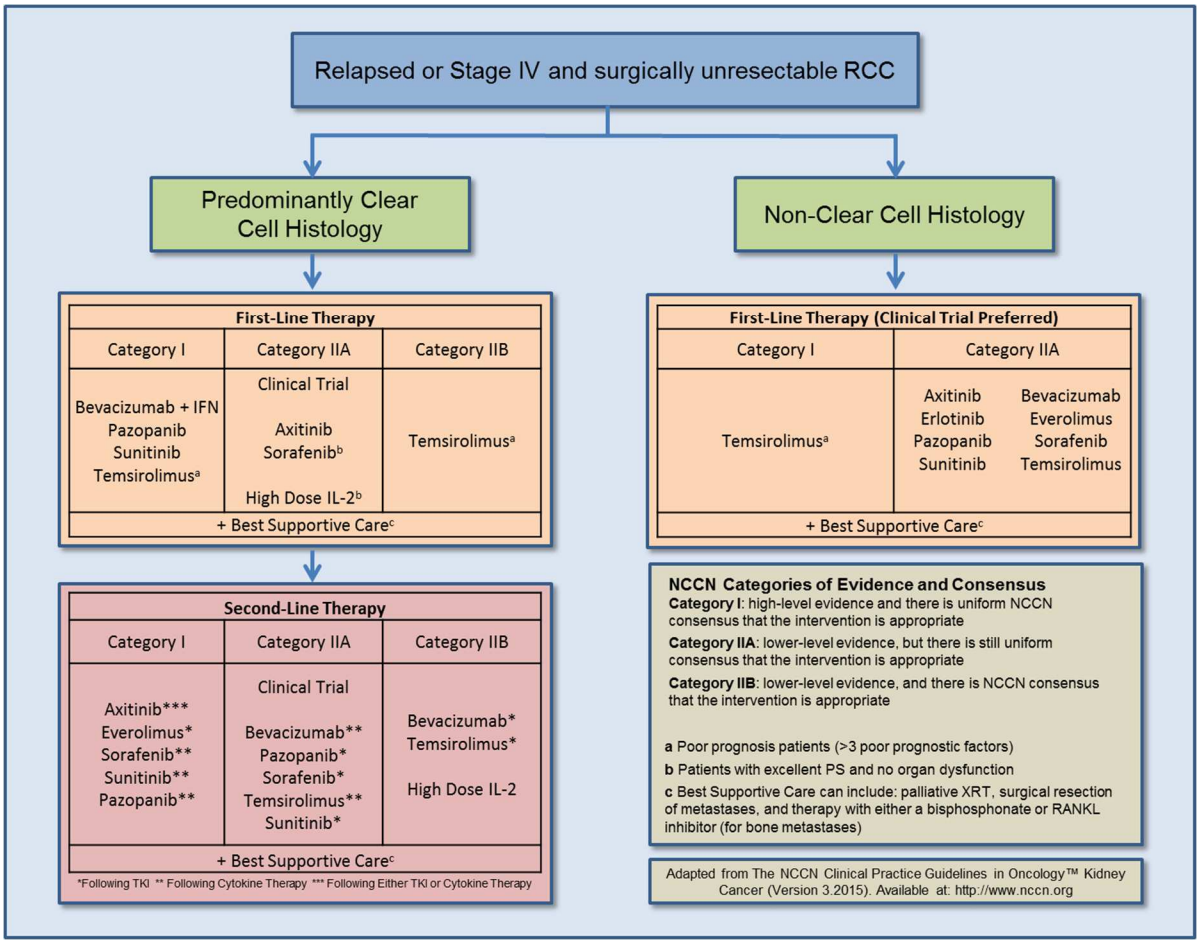
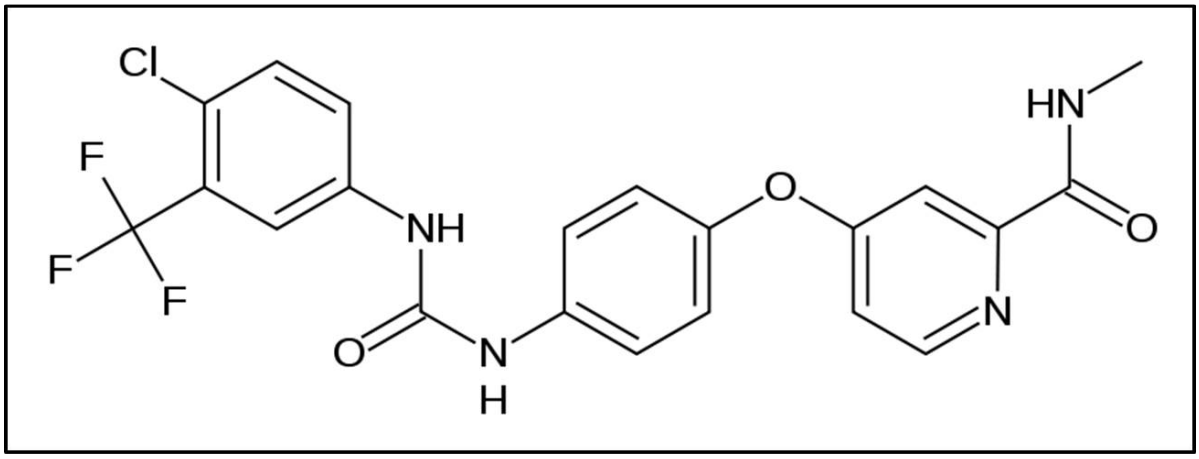


Figure 1.3. Chemical structure of sorafenib



REFERENCES

1. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA Cancer J Clin*. 2014;64:9-29.

2. Hackzell A, Uramoto H, Izumi H, Kohno K, Funa K. p73 independent of c-Myc represses transcription of platelet-derived growth factor beta-receptor through interaction with NF-Y. *J Biol Chem*. 2002;277:39769-39776.

3. Cohen HT, McGovern FJ. Renal-cell carcinoma. *N Engl J Med*. 2005;353:2477-2490.

4. Moch H, Gasser T, Amin MB, Torhorst J, Sauter G, Mihatsch MJ. Prognostic utility of the recently recommended histologic classification and revised TNM staging system of renal cell carcinoma: a Swiss experience with 588 tumors. *Cancer*. 2000;89:604-614.

5. Rini BI, Campbell SC, Escudier B. Renal cell carcinoma. *Lancet*. 2009;373:1119-1132.

6. Leibovich BC, Lohse CM, Crispen PL, et al. Histological subtype is an independent predictor of outcome for patients with renal cell carcinoma. *J Urol*. 2010;183:1309-1315.

7. Jonasch E, Gao J, Rathmell WK. Renal cell carcinoma. *BMJ*. 2014;349:g4797.

8. Sircar KaT, P. Pathologic Considerations. In: Lara Jr. PaJE, ed. Kidney Cancer: Principles and Practice. Heidelberg: Springer; 2012:17-28.

9. International Agency for Research on Cancer. GLOBOCAN. Kidney- estimated incidence, all ages: both sexes. (Accessed March 24, 2015. Available at: http://globocan.iarc.fr/ Pages/fact_sheets_population.aspx.)

10. Motzer RJ, Jonasch E, Agarwal N, et al. Kidney cancer, version 3.2015. *J Natl Compr Canc Netw*. 2015;13:151-159.

11. American Cancer Society. Cancer facts & figures. 2015. (Accessed March 20, 2015. Available at: http://www.cancer.org/acs/groups/content/@editorial/documents/document/acspc-044552.pdf.)

12. Li MaRW. Biomarkers for Renal Cell Carcinoma. In: Lara Jr. PaJE, ed. Kidney Cancer: Principles and Practice. Heidelberg: Springer; 2012:47-65.

13. Kane RC, Farrell AT, Saber H, et al. Sorafenib for the treatment of advanced renal cell carcinoma. *Clin Cancer Res*. 2006;12:7271-7278.

14. Frank I, Blute ML, Leibovich BC, Cheville JC, Lohse CM, Zincke H. Independent validation of the 2002 American Joint Committee on cancer primary tumor classification for renal cell carcinoma using a large, single institution cohort. *J Urol*. 2005;173:1889-1892.

15. Leibovich BC, Blute ML, Cheville JC, et al. Prediction of progression after radical nephrectomy for patients with clear cell renal cell carcinoma: a stratification tool for prospective clinical trials. *Cancer*. 2003;97:1663-1671.

16. Motzer RJ, Bander NH, Nanus DM. Renal-cell carcinoma. *N Engl J Med*. 1996;335:865-875.

17. Motzer RJ, Bacik J, Mazumdar M. Prognostic factors for survival of patients with stage IV renal cell carcinoma: memorial sloan-kettering cancer center experience. *Clin Cancer Res*. 2004;10:6302S-6303S.

18. Patel PH, Chadalavada RS, Chaganti RS, Motzer RJ. Targeting von Hippel-Lindau pathway in renal cell carcinoma. *Clin Cancer Res*. 2006;12:7215-7220.

19. Latif F, Tory K, Gnarra J, et al. Identification of the von Hippel-Lindau disease tumor suppressor gene. *Science*. 1993;260:1317-1320.

20. Kim WY, Kaelin WG. Role of VHL gene mutation in human cancer. *J Clin Oncol*. 2004;22:4991-5004.

21. Albiges L, Salem M, Rini B, Escudier B. Vascular endothelial growth factor-targeted therapies in advanced renal cell carcinoma. *Hematol Oncol Clin North Am*. 2011;25:813-33.

22. Knudson AG, Jr., Strong LC. Mutation and cancer: neuroblastoma and pheochromocytoma. *Am J Hum Genet*. 1972;24:514-532.

23. Gnarra JR, Tory K, Weng Y, et al. Mutations of the VHL tumour suppressor gene in renal carcinoma. *Nat Genet*. 1994;7:85-90.

24. Gallou C, Joly D, Mejean A, et al. Mutations of the VHL gene in sporadic renal cell carcinoma: definition of a risk factor for VHL patients to develop an RCC. *Hum Mutat*. 1999;13:464-475.

25. Schraml P, Struckmann K, Hatz F, et al. VHL mutations and their correlation with tumour cell proliferation, microvessel density, and patient prognosis in clear cell renal cell carcinoma. *J Pathol*. 2002;196:186-193.

26. Herman JG, Latif F, Weng Y, et al. Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc Natl Acad Sci U S A*. 1994;91:9700-9704.

27. Kamura T, Sato S, Iwai K, Czyzyk-Krzeska M, Conaway RC, Conaway JW. Activation of HIF1alpha ubiquitination by a reconstituted von Hippel-Lindau (VHL) tumor suppressor complex. *Proc Natl Acad Sci U S A*. 2000;97:10430-10435.

28. Oka H, Chatani Y, Hoshino R, et al. Constitutive activation of mitogen-activated protein (MAP) kinases in human renal cell carcinoma. *Cancer Res*. 1995;55:4182-4187.

29. Baldewijns MM, van Vlodrop IJ, Vermeulen PB, Soetekouw PM, van Engeland M, de Bruine AP. VHL and HIF signalling in renal cell carcinogenesis. *J Pathol*. 2010;221:125-138.

30. Yagoda A, Abi-Rached B, Petrylak D. Chemotherapy for advanced renal-cell carcinoma: 1983-1993. *Semin Oncol*. 1995;22:42-60.

31. Coppin C, Porzsolt F, Awa A, Kumpf J, Coldman A, Wilt T. Immunotherapy for advanced renal cell cancer. *Cochrane Database Syst Rev*. 2005:CD001425.

32. Hudes G, Carducci M, Tomczak P, et al. Temsirolimus, interferon alfa, or both for advanced renal-cell carcinoma. *N Engl J Med*. 2007;356:2271-2281.

33. Escudier B, Eisen T, Stadler WM, et al. Sorafenib in advanced clear-cell renal-cell carcinoma. *N Engl J Med*. 2007;356:125-134.

34. Escudier B, Pluzanska A, Koralewski P, et al. Bevacizumab plus interferon alfa-2a for treatment of metastatic renal cell carcinoma: a randomised, double-blind phase III trial. *Lancet*. 2007;370:2103-2111.

35. Motzer RJ, Escudier B, Oudard S, et al. Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomised, placebo-controlled phase III trial. *Lancet*. 2008;372:449-456.

36. Motzer RJ, Hutson TE, Tomczak P, et al. Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. *N Engl J Med*. 2007;356:115-124.

37. Sternberg CN, Davis ID, Mardiak J, et al. Pazopanib in locally advanced or metastatic renal cell carcinoma: results of a randomized phase III trial. *J Clin Oncol*. 2010;28:1061-1068.

38. Rini BI, Escudier B, Tomczak P, et al. Comparative effectiveness of axitinib versus sorafenib in advanced renal cell carcinoma (AXIS): a randomised phase 3 trial. *Lancet*. 2011;378:1931-1939.

39. Wilhelm SM, Carter C, Tang L, et al. BAY 43-9006 exhibits broad spectrum oral antitumor activity and targets the RAF/MEK/ERK pathway and receptor tyrosine kinases involved in tumor progression and angiogenesis. *Cancer Res*. 2004;64:7099-7109.

40. Product information. Nexavar (sorafenib). Wayne NBHP, Inc. June 2013. (Accessed January 29, 2015. Available at: http://berlex.bayerhealthcare.com/html/products/pi/Nexavar_PI.pdf.)

41. Wilhelm SM, Adnane L, Newell P, Villanueva A, Llovet JM, Lynch M. Preclinical overview of sorafenib, a multikinase inhibitor that targets both Raf and VEGF and PDGF receptor tyrosine kinase signaling. *Mol Cancer Ther*. 2008;7:3129-3140.

42. Wilhelm S, Chien DS. BAY 43-9006: preclinical data. *Curr Pharm Des*. 2002;8:2255-2257.

43. Yu C, Bruzek LM, Meng XW, et al. The role of Mcl-1 downregulation in the proapoptotic activity of the multikinase inhibitor BAY 43-9006. *Oncogene*. 2005;24:6861-6969.

44. Strumberg D. Sorafenib for the treatment of renal cancer. *Expert Opin Pharmacother*. 2012;13:407-419.

45. Clark JW, Eder JP, Ryan D, Lathia C, Lenz HJ. Safety and pharmacokinetics of the dual action Raf kinase and vascular endothelial growth factor receptor inhibitor, BAY 43-9006, in patients with advanced, refractory solid tumors. *Clin Cancer Res*. 2005;11:5472-5480.

46. Moore M, Hirte HW, Siu L, et al. Phase I study to determine the safety and pharmacokinetics of the novel Raf kinase and VEGFR inhibitor BAY 43-9006, administered for 28 days on/7 days off in patients with advanced, refractory solid tumors. *Ann Oncol*. 2005;16:1688-1694.

47. Strumberg D, Richly H, Hilger RA, et al. Phase I clinical and pharmacokinetic study of the Novel Raf kinase and vascular endothelial growth factor receptor inhibitor BAY 43-9006 in patients with advanced refractory solid tumors. *J Clin Oncol*. 2005;23:965-972.

48. Awada A, Hendlisz A, Gil T, et al. Phase I safety and pharmacokinetics of BAY 43-9006 administered for 21 days on/7 days off in patients with advanced, refractory solid tumours. *Br J Cancer*. 2005;92:1855-1861.

49. Strumberg D, Clark JW, Awada A, et al. Safety, pharmacokinetics, and preliminary antitumor activity of sorafenib: a review of four phase I trials in patients with advanced refractory solid tumors. *Oncologist*. 2007;12:426-437.

50. Jain L, Woo S, Gardner ER, et al. Population pharmacokinetic analysis of sorafenib in patients with solid tumours. *Br J Clin Pharmacol*. 2011;72:294-305.

51. Iyer R, Fetterly G, Lugade A, Thanavala Y. Sorafenib: a clinical and pharmacologic review. *Expert Opin Pharmacother*. 2010;11:1943-1955.

52. Miller AA, Murry DJ, Owzar K, et al. Phase I and pharmacokinetic study of sorafenib in patients with hepatic or renal dysfunction: CALGB 60301. *J Clin Oncol*. 2009;27:1800-1805.

53. Swift B, Nebot N, Lee JK, et al. Sorafenib hepatobiliary disposition: mechanisms of hepatic uptake and disposition of generated metabolites. *Drug Metab Dispos*. 2013;41:1179-1186.

54. Zimmerman EI, Hu S, Roberts JL, et al. Contribution of OATP1B1 and OATP1B3 to the disposition of sorafenib and sorafenib-glucuronide. *Clin Cancer Res*. 2013;19:1458-1466.

55. Ratain MJ, Eisen T, Stadler WM, et al. Phase II placebo-controlled randomized discontinuation trial of sorafenib in patients with metastatic renal cell carcinoma. *J Clin Oncol*. 2006;24:2505-2512.

56. Escudier B, Eisen T, Stadler WM, et al. Sorafenib for treatment of renal cell carcinoma: Final efficacy and safety results of the phase III treatment approaches in renal cancer global evaluation trial. *J Clin Oncol*. 2009;27:3312-3318.

57. Rixe O, Bukowski RM, Michaelson MD, et al. Axitinib treatment in patients with cytokine-refractory metastatic renal-cell cancer: a phase II study. *Lancet Oncol*. 2007;8:975-984.

58. Maroto P, Rini B. Molecular biomarkers in advanced renal cell carcinoma. *Clin Cancer Res*. 2014;20:2060-2071.

59. Motzer RJ, Bukowski RM, Figlin RA, et al. Prognostic nomogram for sunitinib in patients with metastatic renal cell carcinoma. *Cancer*. 2008;113:1552-1558.

60. Motzer RJ, Bacik J, Murphy BA, Russo P, Mazumdar M. Interferon-alfa as a comparative treatment for clinical trials of new therapies against advanced renal cell carcinoma. *J Clin Oncol*. 2002;20:289-296.

61. Heng DY, Xie W, Regan MM, et al. Prognostic factors for overall survival in patients with metastatic renal cell carcinoma treated with vascular endothelial growth factor-targeted agents: results from a large, multicenter study. *J Clin Oncol*. 2009;27:5794-5799.

62. Choueiri TK, Garcia JA, Elson P, et al. Clinical factors associated with outcome in patients with metastatic clear-cell renal cell carcinoma treated with vascular endothelial growth factor-targeted therapy. *Cancer*. 2007;110:543-550.

63. Xu CF, Bing NX, Ball HA, et al. Pazopanib efficacy in renal cell carcinoma: evidence for predictive genetic markers in angiogenesis-related and exposure-related genes. *J Clin Oncol*. 2011;29:2557-2564.

64. van Erp NP, Eechoute K, van der Veldt AA, et al. Pharmacogenetic pathway analysis for determination of sunitinib-induced toxicity. *J Clin Oncol*. 2009;27:4406-4412.

65. Peer CJ, Sissung TM, Kim A, et al. Sorafenib is an inhibitor of UGT1A1 but is metabolized by UGT1A9: implications of genetic variants on pharmacokinetics and hyperbilirubinemia. *Clin Cancer Res*. 2012;18:2099-2107.

66. Beuselinck B, Karadimou A, Lambrechts D, et al. Single-nucleotide polymorphisms associated with outcome in metastatic renal cell carcinoma treated with sunitinib. *Br J Cancer*. 2013;108:887-900.

67. Garcia-Donas J, Esteban E, Leandro-Garcia LJ, et al. Single nucleotide polymorphism associations with response and toxic effects in patients with advanced renal-

cell carcinoma treated with first-line sunitinib: a multicentre, observational, prospective study. *Lancet Oncol*. 2011;12:1143-1150.

68. van der Veldt AA, Eechoute K, Gelderblom H, et al. Genetic polymorphisms associated with a prolonged progression-free survival in patients with metastatic renal cell cancer treated with sunitinib. *Clin Cancer Res*. 2011;17:620-629.

69. Hiles JJ, Kolesar JM. Role of sunitinib and sorafenib in the treatment of metastatic renal cell carcinoma. *Am J Health Syst Pharm*. 2008;65:123-131.

70. Rodriguez-Antona C, Garcia-Donas J. Constitutional genetic variants as predictors of antiangiogenic therapy outcome in renal cell carcinoma. *Pharmacogenomics*. 2012;13:1621-1633.

71. Daly AK. Genome-wide association studies in pharmacogenomics. *Nat Rev Genet*. 2010;11:241-246.

72. Aslibekyan S, Claas SA, Arnett DK. To replicate or not to replicate: the case of pharmacogenetic studies: Establishing validity of pharmacogenomic findings: from replication to triangulation. *Circ Cardiovasc Genet*. 2013;6:409-412.

73. Grupe A, Germer S, Usuka J, et al. In silico mapping of complex disease-related traits in mice. *Science*. 2001;292:1915-1918.

74. Liang DY, Liao G, Wang J, et al. A genetic analysis of opioid-induced hyperalgesia in mice. *Anesthesiology*. 2006;104:1054-1062.

75. Liu P, Vikis H, Lu Y, Wang D, You M. Large-scale in silico mapping of complex quantitative traits in inbred mice. *PLoS One*. 2007;2:e651.

76. Bogue MA, Grubb SC. The Mouse Phenome Project. *Genetica*. 2004;122:71-74.

77. Harrill AH, Watkins PB, Su S, et al. Mouse population-guided resequencing reveals that variants in CD44 contribute to acetaminophen-induced liver injury in humans. *Genome Res*. 2009;19:1507-1515.

78. Guo Y, Weller P, Farrell E, et al. In silico pharmacogenetics of warfarin metabolism. *Nat Biotechnol*. 2006;24:531-536.

CHAPTER 2: TARGET TRIAL PATIENTS: IDENTIFICATION OF PREDICTIVE AND PROGNOSTIC GERMLINE VARIANTS ASSOCIATED WITH OVERALL SURVIVAL

2.1 Overview

Background:

Sorafenib is a potent inhibitor of multiple oncogenic, stromal and angiogenic receptor tyrosine kinases. Germline variants in VEGF-pathway genes and in sorafenib pharmacology genes might associate with prognosis and/or sorafenib efficacy in mRCC patients. The primary objective of this aim was to identify novel germline genetic markers associated with OS in mRCC patients from the phase III TARGET trial. The secondary objective of this aim was to determine if the variants associated with OS also associated with PFS.

Methods:

A total of 295 mRCC patients from the phase III TARGET trial were genotyped using 1,536 germline SNP variants from 56 candidate genes implicated in angiogenesis, sorafenib pharmacology and/or RCC prognosis/pathogenesis. Patients were either treated with sorafenib (n=155) or randomized to placebo (n=140). Directly genotyped variants were enriched through imputation, and imputation garnered an additional 10,097 germline variants (SNPs and insertions/deletions) for analyses. Germline variants were first tested for association with OS using the log rank test, and then in a multivariate analysis with clinical covariates using the Cox proportional hazards (PH) model. False discovery rate (FDR) was

used to correct for multiple comparisons for all variant-OS associations, from the multivariate model, with a p-value ≤0.05. Significant variants (p≤0.05 and q≥0.1) were then tested for associations with PFS using the log rank test and a multivariate Cox PH model. Germline variant effects are summarized using median OS, median PFS hazard ratios with 95% confidence intervals, p-values and FDR q-values.

Results:

Genomic DNA form 295 patients enrolled on the phase III TARGET trial were used in this study. A total of 11,117 germline variants (1,020 directly genotyped and 10,097 imputed variants) were used in the final analyses. The primary aim of the study was to determine if these variants significantly associated with OS. The secondary analysis of this study was to prospectively test whether variants that associated with OS also associated with PFS. This analysis identified five predictive variants that significantly associated with OS in mRCC patients treated with sorafenib (rs1885657 in *VEGFA*, rs3816375 in *ITGAV*, rs6719561 3' of *UGT1A9*, rs8047917 in *WWOX*, and rs200809375 3' of *NRP-1*), and an additional two possibly prognostic variants that associated with OS in a combined analysis of both treatment arms (rs307826 and rs3024987). PFS was correlated with OS among the genotyped patients, and five of the variants significantly associated with PFS as well.

Conclusions:

The identification of these seven variants is a key first step towards discovering and validating novel predictive and prognostic biomarkers to be used clinically in the treatment

of mRCC patients. While these associations are novel and encouraging findings, they require functional laboratory validation and/or replication.

2.2 Introduction

Sorafenib tosylate (sorafenib, Nexavar®) was the first small-molecule, oral, VEGF-pathway inhibitor approved by the U.S. FDA for the treatment of advanced and mRCC. Sorafenib is a potent multikinase inhibitor of numerous angiogenic, oncogenic, and stromal kinases. These include VEGFR2 and VEGFR3, PDGFR-β, FGFR-1, FLT-3, c-KIT, and genes that are part of the mitogen-activated protein kinase (MAPK) signaling cascade (including RAF-1 and BRAF).[1-4]

The approval of sorafenib in mRCC was based upon the results from a randomized discontinuation phase II trial, and the phase III TARGET trial. In the phase II trial, sorafenib significantly improved PFS in mRCC patients (24 vs. 6 weeks; p=0.0087).[5] The phase III TARGET trial was a double-blind randomized, placebo-controlled, multicenter study of 903 patients with advanced RCC who had failed previous systemic cytokine therapy (Figure 2.1). Data from an interim analysis of this pivotal phase III trial demonstrated a significant PFS benefit (5.5 vs. 2.8 months; HR=0.44, 95% CI 0.35-0.55; p<0.01). Based on these data, patients who progressed while receiving placebo were allowed to cross-over and receive open-label sorafenib. At the time of the interim analysis, data also revealed a trend toward improved OS for patients on sorafenib (14.7 months in the placebo arm, but had not been reached in the sorafenib arm at the time of interim analyses; HR=0.72, 95% CI 0.54-0.94; p=0.02); however, this result was not significant based on an *a priori* selected threshold for significance.[6] Final survival analyses, where cross-over patients censored, revealed improved

OS for patients treated with sorafenib (17.8 vs.14.3 months; HR=0.78, 95% CI 0.62-0.97; p=0.029).[7] Currently, sorafenib is used in the second-line setting after progression on an initial first-line therapy (Figure 2.2).[1,8-10] However, in select patients (e.g. patients with adequate organ function and performance status, or even the elderly), sorafenib can be utilized as a first-line option.[10] Several clinical trials focused on sorafenib have revealed a high degree of interindividual variability in sorafenib pharmacokinetics.[11-15] Considerable interindividual variability in response to sorafenib is also observed clinically.[5-7]

The MSKCC risk score (a score comprised of serum hemoglobin levels, corrected serum calcium levels, serum LDH levels, interval between diagnosis and the start of treatment, and Karnofsky performance status) was developed to help clinicians predict outcomes in patients with mRCC.[16-18] In the era of VEGF-targeting agents, additional prognostic scores have also been validated.[19,20] And, studies have analyzed how baseline soluble protein levels also associate with outcomes in these patients.[21] Studies have also found a correlation between circulating VEGF levels and mRCC prognosis, but results remain inconclusive because results have not been replicated. Plasma VEGF levels, as well as soluble VEGFR2 levels, have also been examined as predictive markers of treatment response. But again, the correlations were weak, and have not been validated.[22,23] Additional potential predictive biomarkers, such as cytokine angiogenic factors (e.g. baseline IL-6 and elevated LDH), tissue-based biomarkers (e.g. *VHL* mutations), and factors in the mTOR pathway (e.g. elevated phosphor-S6 expression, and elevated pAKT expression), have also been investigated, but have not been validated and are not used clinically.[24]

New recommendations have also begun to emerge, which are designed to help optimize sequencing of VEGF-pathway inhibitors, for patients likely to benefit from several

lines of treatment.[10] However, currently, there are still no validated molecular biomarkers to provide insights into mRCC prognosis, and no validated predictive biomarkers used routinely in clinic to help providers identify which patients will benefit most from a specific VEGF-pathway multikinase inhibitor (e.g. sorafenib). To date, only a small number of studies have been conducted that attempt to identify predictive pharmacogenetic variants, and none of them have been validated (e.g. in the laboratory and/or though replication). Finally, none of the previous pharmacogenetic studies have been conducted in a population of mRCC patients treated with sorafenib.[25-29]

RCC is a highly vascularized malignancy, due to molecular mechanisms abrogating the activity of the von *VHL* tumor suppressor gene, resulting in increased production of key growth factors integral to angiogenesis, including: VEGF, PDGF, TGF-α, and FGF.[30-34] Collectively, increased growth factor production results in epithelial cell proliferation in the tumor microenvironment, and dysregulated angiogenic signaling that results in induction of tumor angiogenesis and proliferation.[2,31,35,36] Sorafenib has been shown to inhibit angiogenic and non-angiogenic targets in multiple RCC models.[37-39]

Sorafenib affects tumor vascular endothelium,[40] the tumor microenvironment,[41] and as a VEGF-pathway inhibitor it also has effects on the host vascular endothelium and pericytes.[42,43] Because RCC is dependent on angiogenesis and the VEGF pathway, it is a pathway that is a viable target for pharmacotherapy.[44,45] And, because angiogenesis is primarily a host-mediated process,[46] there is excellent rationale for investigating germline variants as predictors of sorafenib efficacy and/or as markers with prognostic significance.

In this study, comprehensive assessments of the common germline DNA variants in genes related to the sorafenib pharmacology, angiogenesis and/or RCC

prognosis/pathogenesis were performed. Identification of germline variants that significantly associate with OS should help clarify which patients will benefit most from sorafenib. The primary objective of this Aim was to identify germline genetic markers that associate with OS. The secondary objective of the study was to prospectively analyze how the variants that associated with OS are also related to differences in PFS.

2.3 Patients, Materials and Methods

2.3.1 Patients and the TARGET Trial

TARGET was a multinational, double-blind, randomized, phase III trial that enrolled patients with unresectable or metastatic clear cell RCC, and who had received prior systemic cytokine therapy (n=903). Patients were randomized 1:1 to receive continuous treatment with either 400 mg sorafenib orally twice daily (n=451), or matched placebo (n=452). Patients remained on study until disease progression or discontinuation due to intolerable toxicity or death (Figure 1.1). Baseline characteristics for patients from the entire TARGET population were compared to those TARGET patients who consented for genotyping (Table 2.1).

The primary endpoint of TARGET was OS, defined as the time from the date of randomization until the date of death. The OS data used in these analyses were collected before patients in the placebo arm were permitted to cross over to receive open-label sorafenib. PFS was measured from the date of randomization until the date of progression. Disease progression was determined based on computed tomography (CT) or MRI, clinical progression, or death, with the use of the Response Evaluation Criteria in Solid Tumors (RECIST).[47] Investigators and independent radiologists who were unaware of the study-group assignments assessed PFS.[6]

Patient DNA collected from the peripheral blood from 295 patients (n=155 on sorafenib and n=140 on placebo) was used for genotyping. All patients provided written informed consent to participate in both the main TARGET trial and this pharmacogenetic study. The studies were conducted in accordance with the Declaration of Helsinki, and the study was approved by the institutional review board at each center.

2.3.2 Gene Selection

Fifty-six candidate genes were selected based on five main criteria. These candidate genes were selected because they are either important to angiogenesis, the VEGF pathway and the function of the host and tumor endothelium (*AKT1, CRK, EPO, FGF2, FGFR1, FLT1, FLT-4, FRS2, GRB2, ITGAV, ITGB5, KDR, KRAS, MAP2K4, MAP2K6, MAPK1, MAPK3, MAPK10, MAPK11, MAPK14, NOS3, NRAS, NRP-1, PGF, PIK3C2A, PIK3C2B, PIK3R5, PRKCA, PRKCE, PXLDC2, RAF1, VEGFA*, and *VEGFB*), are targets of sorafenib not directly linked with angiogenesis (*KIT* and *RET*), are associated with pericyte survival (*PDGFRα, PDGFRβ*), are related to sorafenib pharmacology and toxicity (*BGLAP, CDH13, CYP3A4, EXPH5, PMF1, STK39, UGT1A9*, and *WNK1*), have been associated with RCC pathogenesis (*EGFR, EPAS1, HIF1α, TGFα*, and VHL), or have been associated with prognosis (*CA9, IL8, IL17A, IL17F, STAT3*, and *WWOX*).

2.3.3 Variant Selection

From these candidate genes, 1,536 common germline SNP variants were selected for genotyping. Variant selection consisted of four step-wise procedures. First, SNP variants with minor allele frequency (MAF) >0.05 in European Americans, identified from data

obtained through a sequencing initiative of VEGF-pathway genes conducted by the National Heart, Lung and Blood Institute's DNA Resequencing and Genotyping Service (http://rsng.nhlbi.nih.gov/scripts/index.cfm),[48] and then from data from both 1000 Genomes (www.1000genomes.org) and HapMapIII,(www.hapmap.org) were selected. Second, variants not identified in Step 1 were selected based on dbSNP (http://www.ncbi.nlm.nih.gov/snp) annotation that predicted functionality (e.g. the variant creates or is located in: stop codon, missense, frameshift, insertion/deletion, 5' or 3' UTR, and/or a 5' or 3' splice site). Third, variants that were not identified in Steps 1-2 were selected if they were known to be expression quantitative trait loci (eQTL) in lymphoblastoid cell lines according to the SNP and CNV Annotation (SCAN) database.[49] Fourth, variants that were not previously identified in Steps 1-3 were selected based on already published variant-phenotype associations in the literature. Redundancy was minimized by excluding variants if they were in high linkage disequilibrium (LD) with ($r^2 \geq 0.80$) other identified variants.

2.3.4 Illumina GoldenGate Genotyping

The GoldenGate Assay (BeadArray technology) was selected as the platform on which to conduct the genotyping. Briefly, the GoldenGate Assay is designed for large scale genotyping of biallelic markers such as SNP variants, and was selected because it allows for a custom-built panel of 1,536 germline variants.[50,51] It consists of an initial allele-specific extension reaction followed by polymerase chain reaction (PCR) amplification, and allows for a high degree of multiplexing during the extension and amplification steps. In the GoldenGate Assay, allele-specific primers are hybridized to genomic DNA. The ligated

product then undergoes PCR amplification, and the amplified PCR products are then captured on beads carrying complementary target sequences for the SNP variant. Fluorescence signals from the extension and amplification steps are then read, and clustering for each individual variant is displayed in a scatter plot with the signal intensity on the y-axis and a signal intensity ratio on the x-axis (Figure 2.3).

Germline DNA was extracted from peripheral blood samples (FlexiGene DNA kit; Qiagen, Valencia, CA, USA). DNA concentrations were quantified by NanoDrop at the University of Chicago, and then Pico Green (Invitrogen, Carlsbad, CA) was used at the University of North Carolina Mammalian Genotyping Core to assess DNA quality and quantity. Genotypes were called using Illumina GenomeStudio software v2011. Variants were excluded from final analyses if there was evidence of a genotype call rate <97.5%, a MAF <1% across the cohort, or if they deviated from Hardy–Weinberg equilibrium (HWE; $p<0.0001$). GenTrain scores were derived from the Illumina GenomeStudio scatter plots to measure of variant detection reliability based on genotypic clustering distributions (Figure 2.3).[52] GenTrain scores <0.4 were removed and considered examples of poor clustering. For all significant variant-OS associations, clustering was manually inspected.

2.3.5 Imputation Procedures

Using the directly genotyped SNP variants genotypes at additional loci were imputed within a 5 Mb window of a gene center, and variants located within 25 kb of the gene transcription start and stop sites were retained for final analyses. Sample phasing and imputation were carried out using Impute2 version 2.30.[53,54] Phased reference haplotypes, based on 1,092 individuals from the 1000 Genomes Project (http://www.1000genomes.org;

2010 low-coverage whole-genome and 2011 high-coverage whole-exome sequence data freezes), were used. Only variants with information scores >0.90 and expected MAF >0.02 were retained. Genotypes were assigned their most likely genotype if their posterior probabilities ≥0.90, but were otherwise excluded to diminish the chance of type I error.

2.3.6 Polymerase Chain Reaction

Four SNP variants (rs1885657 and rs3024987 in *VEGFA*, rs8047917 in *WWOX* and rs307826 in *FLT-4*) were selected for follow-up confirmatory genotyping by TaqMan® SNP genotyping assays (Applied Biosystems, Foster City, CA, USA). TaqMan® assays were performed on two directly genotyped variants (rs1885657 and rs307826) to confirm Illumina Genome Studio genotyping calls, and were performed on two additional variants (rs3024987 and rs8047917) confirm Impute2 imputation calls. Each of the TaqMan® genotyping assays were carried out per the manufacturer's instruction using a CFX384 Real-Time System (Bio-Rad, Hercules, CA, USA) and allelic discrimination results were visualized on Bio-Rad CFX Manager software, version 1.6 (Bio-Rad).

For each assay, 120 ng of TARGET patient DNA were loaded onto 384-well plates with TaqMan® Genotyping Mastermix (Applied Biosystems, Foster City, CA, USA) containing VIC and FAM reporter dyes. Briefly, the VIC reporter dye is linked to the 5' end of the allele 1 probe, while the FAM reporter dye is linked to the 5' end of the allele 2 probe. Fluorescence signal of the two reporter dyes, generated during PCR amplification, distinguishes variant alleles. For rs1885657, sense 5'-GAGAGAAGCCCCTGTCACC-3' and antisense 5'-GCTGTGCTTTAGCTCTCGTG-3' primers were used. For rs307826, sense 5'-CCTGCCTGTATCCCTGACC-3' and antisense 5'-GGAGAGAGAGGCCATTACTGC-3'

primers were used. For rs3024987, sense 5'-GTCCCCTTTCCTCCTTGG-3' and antisense 5'-GGAGTTGGTGAGAGCTGGAG-3' were used. For rs8047917, sense 5'-CTGTGGGCTTGACTTGTCC-3' and antisense 5'-CCATCTCCATGCAGTTAAGC-3' were used. The cycling conditions were: an initial 10 minute step at 95°C for AmpliTaq Gold enzyme activation, followed by 40 cycles that alternated between a denaturation step for 15 seconds at 95°C and an annealing/extension step at 60°C for one minute.

Sanger-based DNA sequencing, using 3730xl Genetic Analyzers (Applied Biosystems) and performed at the UNC Mammalian Genotyping Core, was used to validate representative samples and determine thresholds for allelic discrimination. Analysis of the sequencing data, to confirm PCR thresholding prior to comparison against either Illumina Genome Studio or Impute2 genotype calls, was performed using Sequencher software, version 5.1 (Gene Codes Corporation, Ann Arbor, MI, USA). Comparison of PCR and sequencing to either Illumina Genome Studio or Impute2 calls was performed using MS Excel software (Microsoft, Redmond, WA, USA).

2.3.7 Statistical Analyses

The primary objective of this Aim was to identify germline genetic markers that associate with OS (Figure 2.1). To do so, univariate log rank tests and a multivariate Cox PH model were employed for each association. First, the effects of directly genotyped and imputed germline variants on OS were assessed using three separate log rank tests. The first was performed to analyze differences in OS by a germline variant among patients in the sorafenib arm only, the second was performed for the same reason but only among the

patients in the placebo group only, and the third was performed using both the placebo and sorafenib arms from TARGET.

Second, a multivariate Cox PH model was used to analyze the genotype-phenotype relationships for each variant while accounting for the effects of appropriate clinical covariates included in the model. Country, sex, age, race, Eastern Cooperative Oncology Group (ECOG) PS, time since RCC diagnosis, previous systemic IL-2 or INF administration, MSKCC prognostic score, number of metastatic sites, evidence of spread of distant metastasis to liver or lung, and treatment arm (sorafenib or placebo) were all covariates that were considered. First, a Chi-squared test of homogeneity was used for categorical variables when testing for differences in the covariate between treatment groups (country, race, ECOG PS, previous IL-2 or INF use, MSKCC prognostic score, and liver or lung metastases). Linear regression, with treatment as the independent variable, was used for quantitative variables (age, time since diagnosis, and number of metastatic sites). Associations between OS or PFS and potential covariates were then assessed using a likelihood ratio test applied to a Cox PH model. Third, potential covariates that remained significant ($p \leq 0.05$) after the previous likelihood ratio test were then tested, conditioned on one another, in a multivariate Cox PH model. Only variants that were conditionally significant ($p \leq 0.05$) were included in the final OS and PFS multivariate models. The results for these tests are provided in Tables 2.2 and 2.3.

All germline variant-OS associations were tested without *a priori* assuming any genetic model (i.e. did not assume an additive, dominant or recessive genetic model). FDR q-values were generated to account for multiple testing, and to ensure that type I error was limited so that spurious associations could be avoided.[55] FDR was employed in lieu of

correcting the family-wise error rate (e.g. Bonferroni correction) to account better for the correlation among tests induced by LD between germline variants. Variants were deemed significant if their multivariate model p-value was ≤0.05 and if their FDR q-value was ≤0.1.

Correlation analyses were conducted to ascertain if a significant relationship between OS and PFS existed among the TARGET genotyped patients. If the assumptions of normalized distributions were satisfied, a Pearson product moment correlation (Pearson's r) test was employed. If not, then a Spearman rank order correlation (Spearman's rho) test was used. For variants considered to be significantly associated with OS (p-value ≤0.05 and q-value ≤0.1), prospective testing was conducted to determine if there was also an association with PFS. As described above, univariate log rank tests were conducted followed by multivariate Cox PH models. Variant-PFS associations were considered significant if p-value ≤0.05. Because these associations were exploratory in nature, a correction for multiple testing was not applied.

All hazard ratios (HRs) and 95% confidence intervals (CIs) for each variant-OS and variant-PFS association were derived from the OS and PFS multivariate models, respectively. All analyses were conducted in R software,[56] and confirmed using SAS software, version 9.2 (SAS, Cary, NC, USA). GraphPad Prism® version 5.03 (GraphPad Software, Inc., La Jolla, CA, USA) software was also used in analyses and in the creation of all figures.

2.4 Results

2.4.1 Patient Characteristics

The genotyped patients used in this study were representative of the overall TARGET population. Baseline characteristics for those patients who consented for genotyping were similar to characteristics from the overall TARGET population (Table 2.1). In addition, OS was similar between the two groups (Figure 2.4). In the entire TARGET population, patients treated with sorafenib achieved a median OS of 175 days, compared with 162 days for the patients administered placebo. Among the genotyped TARGET patients, those treated with sorafenib achieved a median OS of 190 days, compared with 178 days for the patients administered placebo. PFS was also similar between the two groups (Figure 2.5). In the entire TARGET population, patients treated with sorafenib achieved a median PFS of 123 days, compared with 50 days for the patients administered placebo. Among the genotyped TARGET patients, those treated with sorafenib achieved a median PFS of 125 days, compared with 46 days for the patients administered placebo. These comparisons provided confidence that the subgroup of genotyped patients (n=295) were representative of the entire TARGET cohort (n=903).

2.4.2 Genotyping

The first round of GoldenGate genotyping, consisting of four total plates, was conducted at the University of Chicago. Among the TARGET patients who consented for genotyping and DNA was obtained (n=334), 68 failed the initial genotyping. Systematic assay error was assumed to be the reason underlying such a high failure rate because 56 of the 68 total failures came from plate 4 (with the remaining 12 failure divided evenly from

plates 1-3). As a result, a second round of GoldenGate genotyping was conducted at the University of North Carolina's Mammalian Genotyping Core. Prior to the second round of genotyping, Pico Green was performed on all failed DNA samples and on all positive control DNA samples to quantify the total amount and assess the quality of the DNA. A total of 29 additional patients (21 were form the original failed plate 4) had sufficient genotype calls and were included in the final analyses. Genotype calls for samples included as positive controls from plates 1-3 (n=7) in the second round of genotyping at the University of North Carolina were found to be in 100% concordance with the original genotype calls from the University of Chicago.

From the original 1,536 SNP variants that were directly genotyped, 416 were excluded. They were excluded if there was evidence of a genotype call rate below 97.5% (n=209), MAF <1% (n=123), HWE $p<0.0001$ (n=77), a combination of failing both MAF and HWE criteria (n=1), and poor clustering determined by an Illumina GenTrain score <0.4 (n=6). An additional 100 directly genotyped variants were excluded from the final analyses because they were not concordant between the two genotyping cores. Imputation was then conducted to enrich the directly genotyped variants. After quality control pruning, a total of 11,117 variants (1,020 directly genotyped SNP variants, and 10,097 imputed SNPs and insertion or deletions [indels]) were included to test associations between genotype and OS (Figure 2.6).

2.4.3 Confirmation of Genotyping and Imputation Calls

As an additional quality control measure, TaqMan® assays were performed on two directly genotyped variants (rs1885657 in *VEGFA* and rs307826 in *FLT-4*) and two imputed

variants (rs3024987 in *VEGFA* and rs8047917 in *WWOX*) that were significantly associated with differences in OS (Figure 2.7A-2.7D). These assays were performed to confirm the accuracy of the calls for both the directly genotyped variants performed on the Illumina GoldenGate Assay, and the imputed variants gained through Impute2 analyses. Results from these assays revealed a very high concordance between TaqMan® assays and either the Illumina GoldenGate Assay or Impute2 analyses. For rs1885657, there was greater than 96% concordance between TaqMan and Illumina GoldenGate genotyping. Among the 11 calls that were discordant, eight were due to failed genotyping on the TaqMan® assay (no call was returned). The remaining three discordant results were called homozygous wild type (WT) on the Illumina GoldenGate platform, but were called heterozygous by TaqMan®. For rs3024987, there was 99% concordance between the TaqMan® assay calls and results from Impute2. In this case, all three of the discordant calls were due to failed genotyping on the TaqMan® assay. For rs307826 and rs8047917, 100% concordance was observed. These data provided a high level of confidence in both the Illumina GoldenGate Assay genotyping calls and GenomeStudio clustering algorithm, and the Impute2 results.

2.4.4 Multivariate Model Covariates

Multiple clinical covariates that could conceivably influence OS and/or PFS were examined based on their potential affect variant-OS or variant-PFS associations. These included country, sex, age (years), race, ECOG PS, previous interleukin or interferon use, MSKCC prognostic score, number of metastatic sites, and evidence of spread of distant metastasis to liver or lung, and treatment arm (sorafenib or placebo). For OS, number of metastatic sites and the MSKCC prognostic risk score were included in its final multivariate

model (Table 2.2). For PFS, time since diagnosis and patient age were included in its final multivariate model (Table 2.3).

2.4.5 Germline variants associated with OS

Seven variants that associated with OS in either the patients solely from the sorafenib arm, or from a combined analysis of both treatment arms, were identified and deemed significant by an *a priori* established criteria (p-value ≤0.05 and FDR q≤0.1). Five germline variants within genes were identified. Two variants (rs1885657 and rs3024987) were in intronic regions of *VEGFA*, rs3816375 was in an intronic region in *ITGAV*, rs8047917 was in an intronic region of *WWOX*, and rs307826 was in a coding region of *FLT-4.* In addition, two intergenic variants (rs200809375 and rs6719561) were found to be in close proximity to *NRP-1* and *UGT1A9*, respectively (Table 2.4).

Three intragenic variants that associated with differences in OS in sorafenib-treated patients (rs1885657 in *VEGFA*, rs3816375 in *ITGAV*, and rs8047917 in *WWOX*) were identified. For rs1885657, patients with the TT or TC genotypes had a 17-fold decrease in risk of death, when compared to patients with the CC genotype (HR=17.32; 95% CI, 5.69 to 52.73; p=1.39x 10^{-4}; q=0.076; median OS 175 v 409 days) (Figure 2.8). For rs3816375, patients with the AA or AG genotype had a 6-fold decreased risk of death, when compared to patients with the GG genotype (HR =5.88; 95% CI, 2.11-16.39; p=4.87x10^{-4}; q=0.051; median OS 281 v 409 days) (Figure 2.9). And for rs8047917, patients with the TT genotype had a 4-fold decreased risk of death, when compared to patients with the TA genotype (HR=4.06; 95% CI, 1.88 to 8.78; p=3.27x10^{-4}; q=0.076) (Figure 2.10).

Two additional intergenic variants that associated with differences in OS for patients treated with sorafenib (SNP rs6719561 located 1.8 kB 3′ of *UGT1A9* and 606 bp 5′ of *HEATR7B1* and short indel rs200809375 located 7.5 kb 3′ of *NRP-1*) were identified. For rs6719561, patients with the CC or CT genotype had a nearly 4-fold decreased risk of death, when compared to those with the TT genotype (HR =3.82; 95% CI, 1.56 to 9.52; $p=3.27x10^{-4}$; q=0.076; median OS 195 v 216 days) (Figure 2.11). For rs200809375, patients without the insertion had a nearly 6-fold decreased risk of death, when compared to patients with at least one copy of the ATG insertion (HR=6.77; 95% CI, 2.62 to 17.5; $p=2.65x10^{-4}$; q=0.076; median OS 374 v 281 days) (Figure 2.12).

Finally, two intragenic variants, which associated with OS in a combined analysis of both treatment arms (SNP rs307826 in *FLT-4* and SNP rs3024987 in *VEGFA*), were identified. For rs307826, patients with the AA or AG genotypes had a nearly 14-fold decreased risk of death, when compared to patients with the GG genotype (HR=13.79; 95% CI 3.04 to 62.61; $p=1.24 \times 10^{-4}$; q=0.088; median OS 93 v 210 days) (Figure 2.13). For rs3024987, patients with the CC genotype had a 3-fold decreased risk of death when compared to patients with the CT genotype (HR=2.98; 95% CI, 1.66 to 5.37; $p=8.80x10^{-4}$; q=0.088; median OS 190 vs 210 days) (Figure 2.14). Decreased median OS was observed for patients with the CT genotype, when compared with those with the CC genotype, in patients from only the sorafenib (p=0.019) and in the patients from only the placebo arm (p=0.004), but these results were not deemed significant based q-values ≥0.1 (q=0.301 and q=0.371, respectively). This variant is in moderate LD with rs1885657 ($r^2=0.61$).

Unless specified above, none of the variant-OS associations observed in the sorafenib arm were also significant ($p\leq0.05$ and $q\leq0.1$) in analyses when the treatment arms were

combined. Similarly, neither of the significant variant-OS associations observed in the combined arm analyses was significant in the sorafenib arm alone. Finally, no associations between germline variants and OS among patients from the placebo arm met the criteria for significance ($p \leq 0.05$ and $q \leq 0.1$).

2.4.6 OS Variant Associations with PFS

PFS was correlated with OS in non-censored patients from the entire TARGET cohort (Figure 2.15A; Spearman $\rho=0.5083$, $p<0.0001$), genotyped TARGET patients (Figure 2.15B; Pearson $r=0.4681$, $p=0.0001$), genotyped TARGET patients treated with sorafenib (Figure 2.15C; Pearson $r = 0.4778$, $p=0.0076$), and genotyped TARGET patients treated with placebo (Figure 2.15D; Pearson $r=0.4701$, $p=0.0076$).

In this secondary endpoint, variants that significantly associated with OS were examined prospectively to test their associations with PFS. Among the five original variants that associated with OS in the sorafenib arm, rs1885657, rs6719561 and rs8047917 also significantly associated with PFS ($p \leq 0.05$). For rs1885657, patients from the sorafenib arm with the TT or TC genotypes had a four-fold decreased risk of disease progression, when compared to patients with the CC genotype (Figure 2.16; HR=4.00, 95% CI 1.57-10.21; $p=0.0037$). For rs6719561, patients from the sorafenib arm with the CC or CT genotypes had a nearly two-fold decreased risk of disease progression, when compared to patients with the TT genotype (Figure 2.17; HR=1.94, CI 1.00-3.76; $p=0.05$). For rs8047917, patients from the sorafenib arm with the TT genotypes had a 77% decreased risk of disease progression, when compared to patients with the TA genotype (Figure 2.17; HR=1.77, 95% CI 1.03-3.04; $p=0.0389$). Variant rs3816375 did not significantly associate with PFS.

Among the two variants that associated with OS in analyses when the treatment arms were combined, only rs307826 significantly associated with PFS ($p \leq 0.05$). For rs307826, patients with the AA and AG alleles decreased risk of disease progression, when compared to patients with the GG genotype (Figure 2.18; HR=2.92, 95% CI 1.19-7.19; p=0.0193). Also, rs307826 was associated with PFS among the patients from the sorafenib arm. Sorafenib-treated patients with the AA or AG alleles, had a nearly 9-fold decreased risk of disease progression (Figure 2.18; HR=8.62, 95% CI 1.98-37.04, p=0.0041). Variant rs3024987 did not significantly associate with PFS.

Two variants, rs200809375 and rs3816375, did not significantly associate with PFS in the sorafenib arm, nor in the combined arms. Similar to OS, none of the five variants significantly associated with PFS in the placebo arm.

2.5 Discussion

To date, this study represents the most comprehensive sorafenib pharmacogenetic study in mRCC patients, and the first to associate germline variants with OS and PFS. And, by analyzing 11,117 variants in 56 genes, it is the most comprehensive candidate gene pharmacogenetic study involving any of the four VEGF-pathway inhibitors approved by the U.S. FDA for the treatment of mRCC (axitinib, pazopanib, sorafenib and sunitinib). A candidate gene-candidate variant approach was selected in this case, instead of genome-wide studies or next generation sequencing approaches, based on how patients were consented for correlative studies during the TARGET trial.

The primary objective of this Aim was to identify germline genetic markers that associate with OS. The secondary objective of the study was to prospectively analyze how

the variants that associated with OS are also related to differences in PFS. Germline variants were examined in this study, and they are appropriate for pharmacogenetic studies because germline DNA has significant advantages over other nucleotide and protein biomarkers in that it is inherited, fixed, and relatively insensitive to time and environmental factors.[26] And, because angiogenesis is primarily a host-mediated process,[46] there is excellent rationale for investigating germline variants as predictors of sorafenib efficacy and/or as markers with prognostic significance.

In the TARGET study, OS was the primary endpoint and PFS was a secondary endpoint. In this study, OS was also used as the phenotype for the primary objective of this Aim. In the TARGET trial, sorafenib was used as monotherapy, and because there were no U.S. FDA approved VEGF-pathway multikinase inhibitors for the treatment of mRCC prior to sorafenib, OS results would not be confounded by concomitant and/or subsequent therapies. OS still remains the gold standard endpoint in clinical trials because OS is not a surrogate endpoint, and improved OS demonstrates a direct clinical benefit to the patient. Therefore, the identification of predictive and/or prognostic biomarkers that help provide a direct clinical benefit to mRCC patients, and help guide provider decision-making, is of the utmost importance. However, PFS was used as a phenotype for the secondary objective of this Aim. PFS has been shown to be a credible endpoint in oncology trials,[57] and can be a valuable surrogate endpoint for OS. With the recent U.S. FDA approval of multiple treatment options for patients with mRCC, a clear understanding of how each treatment impacts PFS (to ultimately extend OS for patients) will be crucial towards clarifying how these treatments are sequenced.

For the purposes of this Aim, it is important to distinguish between a prognostic and a predictive variant. Simply, a prognostic variant is one that is objectively measurable and is associated with clinical outcome in the absence of active therapy, or with the administration of standard of care or best supportive care medications. Essentially, this type of variant provides valuable information about the patient's overall cancer outcome.[58] The TARGET placebo patients provide the ideal scenario for evaluating the prognostic significance of germline variants in relation to OS and PFS. In this study, germline variants that exhibit similar OS and/or PFS effects in both treatment arms, where it is evident that there is not an association being driven by the effects of sorafenib, could be prognostic. Conversely, a predictive variant is one that is associated with response or lack of response, and provides information on the likely benefit of a particular therapy.[58] The sorafenib-treated TARGET patients are ideal for analyzing the interplay between germline variants, and sorafenib efficacy. In this study, germline variants that associate with the effects of OS and/or PFS, where the effects appear to be driven by sorafenib only, could be predictive.

VEGFA encodes for VEGFA (also known simply as VEGF), which has been shown to be a major driver of angiogenesis, in multiple preclinical models, through its interactions with VEGFR1-3.[59] Previously, baseline circulating VEGFA levels were shown to be prognostic for mRCC OS,[7] and circulating VEGF-165 (the predominant isoform) has also been shown to associate with tumor stage/grade and poorer prognosis.[60] More recently, VEGF-pathway multikinase inhibitor pharmacogenetic studies have shown that *VEGFA* SNP variants (e.g. rs699947, rs833061, rs2010963, and rs3025039) associated with differences in OS and/or PFS for patients treated with sunitinib or pazopanib.[26,61,62] In this study, rs1885657 (not in LD with the aforementioned variants) significantly associated with sorafenib efficacy

in TARGET patients. For sorafenib-treated patients with the CC genotype, this SNP associated with significantly worse OS and PFS and could be a predictive biomarker of sorafenib efficacy. A plausible hypothesis to explain these associations is that since rs1885657 is located in an intronic region of the *VEGFA*, two copies of the C allele resulted in altered transcription factor binding that ultimately caused *VEGFA* overexpression. Ultimately, increased *VEGFA* expression could have led to increased transcription and translation of VEGFA, which increased the pro-angiogenic capability of the ligand and overwhelmed sorafenib's ability to effectively inhibit angiogenesis. In turn, increased angiogenesis, as a result of *VEGFA* overexpression, could have led to a greater risk of disease progression and death in patients with the CC genotype. These results are certainly compelling; however, the dramatic effects on OS and PFS observed in the sorafenib-treated patients could be driven by a low number of uncensored patients with the CC genotype (n=5). Therefore, these results require validation in the laboratory and/or in an independent replication cohort of mRCC patients treated with sorafenib.

VEGFA variant rs3024987 was also found to significantly associate with poorer OS in TARGET patients. When both treatment arms were combined, patients with only one copy of the T allele experienced significantly shorter OS. This variant is also in an intronic region in *VEGFA*, is in moderate LD with rs1885657 (r^2=0.61), and also likely altered transcription factor binding, which led to increased *VEGFA* expression. This variant significantly associated with OS in an analysis where treatment arms were combined, and associations with OS were not significant (p-value ≤0.05 and q-value ≤0.1) in either the sorafenib or the placebo arm alone. However, there appears to be an effect in both treatment arms. The association between rs3024987 and OS trended towards significance in the sorafenib arm

(p=0.019) and the placebo arm (p=0.004), and each of the individual study arm Kaplan-Meier plots revealed a noticeable divergence in the OS curves for patients with the CC genotype versus those with the CT genotype. Therefore, rs3024987 could be prognostic of survival in mRCC patients.

FLT-4 encodes for VEGFR3, a transmembrane kinase receptor that has been traditionally linked to lymphangiogenesis.[63] But, VEGFR3 is also expressed in tumor vasculature.[64] Moreover, inhibition of VEGFR3 can suppress vascular network formation,[65] and preclinical models have suggested that VEGFR3 could possibly be more relevant than VEGFR2 (a primary driver of angiogenic signaling, and a main target for the VEGF-pathway inhibitors used to treat mRCC) for the development of metastases.[66] In this study, rs307826 was associated with decreased survival. In a combined analysis of both study arms, patients with the GG genotype were at higher risk for both disease progression and death. Interestingly, sorafenib-treated patients with the GG genotype were also at significantly higher risk for disease progression. Variant rs307826 results in a threonine to alanine substitution at position 494 (Thr494Ala), located in the fifth IgG-like domain of VEGFR3. Although the PolyPhen-2[67] and the Sorting Intolerant From Tolerant (SIFT)[68] algorithms predict this to be a benign and non-damaging amino acid substitution, several pharmacogenetic studies of mRCC patients treated with VEGF-pathway multikinase inhibitors showed that this variant is associated with significantly shorter OS and/or PFS.[27,28,69] A primary hypothesis to explain the effects of rs307826 on differences in OS and PFS is that rs307826 results in an amino acid in one of the signaling domains of the receptor, which is sufficient to allow VEGFR3 signaling to overcome sorafenib inhibition. Increased VEGFR3 signaling could perpetuate increased lymphangiogenesis and RCC spread to the

lymphatics, but also increase the pro-angiogenic capability of this receptor. Together, they could increase a patient's risk for disease progression and/or death.

It is unclear whether rs307826 is truly predictive of treatment response, or if it is a prognostic for mRCC. In each of the aforementioned studies, there were not a placebo control arms,[27,28] or the placebo arm was not genotyped.[69] And, in these studies, there were not a sufficient amount of GG patients to truly determine whether the variant's effect on reduced OS was driven by sorafenib treatment, and therefore it is important not to over-interpret these results. However, the results from the exploratory PFS analyses performed in this Aim may have helped clarify whether rs307826 could be a predictive variant. In the case of rs307826, replication is not as crucial since it has been associated with differences in OS and PFS in mRCC patients treated with oral VEGF-pathway inhibitors that inhibit many of the same targets as sorafenib (i.e. sunitinib and pazopanib).

Two additional intragenic variants, found in regulatory regions of their respective genes, significantly associated with OS (rs3816375 in *ITGAV* and rs8047917 in *WWOX*), and could be predictive variants of sorafenib efficacy for mRCC patients. *ITGAV* encodes for integrin alpha-V, which is a member of the integrin superfamily. Integrin alpha-V is an adhesion receptor that, when dimerized to a beta subunit, mediates cell-adhesion of the cytoskeleton to extra cellular matrix (ECM) proteins, including: vitronectin, fibronectin, fibrinogen, collagen, osteopontin and thrombospondin.[70] Aside from cellular adhesion, integrins have also been implicated in angiogenesis, proliferation and metastasis.[71-73] Although there have not been any studies linking variants in this gene to outcomes in mRCC or in patients treated with VEGF-pathway multikinase inhibitors, overexpression of *ITGAV* has been associated with progression, spread and poorer prognosis in patients with colorectal

cancer.[74,75] In this study, sorafenib-treated mRCC patients with GG genotype experienced significantly decreased OS, which could be directly related to overexpression of *ITGAV*. This variant is also located in an intronic region, and could have an effect on transcription factor binding and gene expression. Overexpression of *ITGAV* in these patients could result in increased stabilization of the cytoskeleton to ECM proteins, increased angiogenesis and metastasis, and ultimately poorer clinical outcomes.

WWOX encodes for WW domain-containing oxidoreductase, which has shown to be a tumor suppressor gene for multiple tumor types.[76] *WWOX* variants have been shown to associate with increased disease susceptibility in prostate cancer, clinical staging of lung cancer, and decreased survival in patients with multiple myeloma.[77-79] In this study, sorafenib-treated mRCC patients with one copy of an A allele at rs807917 experienced significantly decreased OS and PFS. A plausible hypothesis to explain these associations is that since rs8047917 is located in an intronic region of the gene, one copy of the A allele is sufficient to cause reduced *WWOX*. Therefore, sorafenib-treated patients with at least one A allele at rs8047917 experience less *WWOX* expression, and this could lead to their increased risk disease for progression and/or death.

Finally, two intragenic variants significantly associated with OS (SNP rs6719561 1.8 kb 3' of *UGT1A9* and indel rs200809375 7.5 kb 3' of *NRP-1*, and are potential predictive variants of sorafenib efficacy for mRCC patients. *UGT1A9* encodes for UGT1A9, which is integral to phase II metabolism and inactivation, through glucuronidation, of parent sorafenib and its N-oxide metabolite.[80,81] Previously, pharmacogenetic variants in *UGT1A9* have been associated with altered sorafenib pharmacokinetics and sorafenib-related toxicities.[82,83] Yet, this is the first study where a *UGT1A9* variant has been associated with differences in

survival in patients treated with sorafenib. In this study, mRCC patients with two copies of the T allele were shown to have decreased OS and PFS in these studies. One plausible hypothesis to explain the association could be that two copies of the T allele result in increased sorafenib glucuronidation, leading to decreased drug exposure. And, should patients be exposed to lower active drug concentrations over the duration of treatment, this could lead to increased risk of progression and death. However, the effects of this variant on OS and PFS may not be through *UGT1A9*. This intergenic variant is indeed in close proximity to *UGT1A9* (~1.8 kb 3' of the gene), it is also only approximately 600 bp 5' of *HEATR7B1* (also known as *MROH2A*). While the function of *HEATR7B1* is largely unknown, and there have not been any studies conducted on its relationship to sorafenib, it is conceivable that rs6719561's predictive effects on OS and PFS are through this gene. Further investigation into this hypothesis is warranted.

NRP-1 encodes for neuropilin-1 (one of two neuropilins), which is a membrane-bound receptor that binds several ligands; most notably the VEGF-165 isoform of VEGFA, and helps mediate several signaling pathways that control angiogenesis and cell migration.[84] It has also been shown that in aggressive types of RCC, *NRP-1* is overexpressed. And, in mouse xenograft models reduced *NRP-1* levels led to the implanted RCC cells having reduced tumor-forming ability.[85] One hypothesis for its association with OS in sorafenib-treated mRCC patients could be that only one copy of the small ATG insertion is sufficient enough to cause increased *NRP-1* expression. And, overexpressed *NRP-1* could increase the pro-angiogenic and pro-migratory capabilities of mRCC. However, similar to rs6719561, rs200809375's effects on OS may not be through *NRP-1*, and additional validation studies

should be conducted to determine how this germline variant influences survival in mRCC patients treated with sorafenib.

Data from the studies in this Aim have demonstrated that sorafenib-treated mRCC patients with deleterious genotypes in these variants may experience increased angiogenic and/or oncogenic signaling, or increased bioinactivation of the drug. This may result in decreased sorafenib efficacy, and ultimately shorter time to progression and/or survival. In addition, two germline variants were identified that have the potential to be prognostic variants of mRCC survival. Clearly, validation of these results is necessary before they are incorporated into routine clinical practice. The first necessary step, which will be addressed in Aim 2, is to validate these findings in a series of laboratory experiments designed to: 1) confirm the functionality of the variant, and 2) provide insight into the mechanistic underpinnings of sorafenib response and/or RCC prognosis/pathogenesis. The relationship between the predictive variants identified in this study and sorafenib-induced toxicities experienced by patients enrolled on TARGET should also be examined. Next, in future studies, these results should be replicated in an independent, external cohort of mRCC patients treated with sorafenib. Traditionally, replication of positive findings in an external, independent cohort of patients has served as one of the gold standard types of validation for genotype-phenotype associations.[86,87] And, replication studies will help to elucidate whether the variant-OS associations identified in this Aim, that seem to be driven by the effects of sorafenib, are truly predictive biomarkers. Finally, these potential biomarkers should be validated in a prospective clinical trial, which includes an interaction analysis design. This will provide the highest level of evidence for the validity of these germline variants as predictive and prognostic biomarkers for mRCC patients.[88]

In summary, amid a crowded landscape of targeted multikinase inhibitors recently for the treatment of mRCC, there is a clear unmet need to identify and validate predictive and prognostic pharmacogenetic germline variants as useful biomarkers to help guide provider treatment decisions. The identification of these seven variants is an important first step towards that pursuit.

TABLES

Table 2.1. **Patient characteristics for the entire TARGET population versus genotyped** TARGET patients. There were no significant differences between the two populations, nor were there within group differences between the sorafenib arm and the placebo arm at baseline. Abbreviations: ECOG, Eastern Cooperative Oncology Group; MSKCC, Memorial Sloan Kettering Cancer Center

All TARGET Patients (n = 903)			Genotyped TARGET Patients (n = 295)		
Variable	Sorafenib (n=451)	Placebo (n=452)	Variable	Sorafenib (n=155)	Placebo (n=140)
Male Sex - no. (%)	315 (70)	340 (75)	Male Sex - no. (%)	114 (74)	108 (77)
Median Age - year (range)	58 (19-86)	59 (29-84)	Median Age - year (range)	59 (19-80)	58 (31-82)
ECOG Performance Status - no. (%)			ECOG Performance Status - no. (%)		
0	219 (49)	210 (46)	0	83 (54)	81 (58)
1	223 (49)	236 (52)	1	72 (46)	58 (42)
2	7 (2)	4 (1)	2	0	0
Number of Metastatic Sites - no. (%)			Number of Metastatic Sites - no. (%)		
1	62 (14)	63 (14)	1	26 (17)	26 (19)
2	131 (29)	129 (29)	2	42 (27)	40 (29)
>2	256 (57)	258 (57)	>2	87 (56)	74 (52)
Previous cytokine use - no. (%)	374 (83)	368 (81)	Previous cytokine use - no. (%)	141 (91)	122 (88)
Median Duration of Disease - year (range)	2 (<1-19)	2 (<1-20)	Median Duration of Disease - year (range)	1.7 (<1-19.5)	1.5 (<1-16)
MSKCC Prognostic Risk - no. (%)			MSKCC Prognostic Risk - no. (%)		
Low	233 (52)	228 (50)	Low	73 (47)	68 (49)
Intermediate	218 (48)	223 (49)	Intermediate	82 (53)	72 (51)

Table 2.2. Covariate selection for the OS multivariate model. For the univariate and multivariate Cox PH models, a likelihood ratio test was used. Abbreviations: ECOG, Eastern Cooperative Oncology Group; LRT = likelihood ratio test; MSKCC, Memorial Sloan Kettering Cancer Center; P, p-value; PH, proportional hazards; PS, performance status.

* denotes covariates included in the final multivariate model

Covariate	Placebo-Sorafenib P	Univariate Cox PH P	Multivariate Cox PH P
Sex	0.56	0.52	
Country	0.91	0.54	
Age	0.07	0.58	
Race	0.10	0.40	
ECOG PS	0.42	0.06	
Time since diagnosis	0.25	0.05	0.06
Previous cytokine use	0.39	0.96	
MSKCC score	0.89	<0.01	0.05*
No. of metastatic sites	0.88	<0.01	<0.01*
Lung or liver metastases	1.00	0.14	

Table 2.3. Covariate selection for the PFS multivariate model. For the univariate and multivariate Cox PH models, a likelihood ratio test was used. Abbreviations: ECOG, Eastern Cooperative Oncology Group; MSKCC, Memorial Sloan Kettering Cancer Center; P, p-value; PH, proportional hazards; PS, performance status.

>45% of the cells had expected counts less than 5 (46% for country and 60% for race). In these cases, the Fisher's Exact test was used; * denotes covariates included in the final multivariate model.

Covariate	Placebo-Sorafenib P	Univariate Cox PH P	Multivariate Cox PH P
Sex	0.48	0.95	
Country	0.75#	0.12	
Age	0.07	0.02	<0.01*
Race	0.05#	0.06	
ECOG PS	0.42	0.77	
Time since diagnosis	0.24	<0.01	0.06
Previous cytokine use	0.37	0.67	
MSKCC score	0.80	0.54	
No. of metastatic sites	0.85	0.09	
Lung or liver metastases	0.98	0.03	0.02*

Table 2.4. Significant variants associated with OS in TARGET patients. Seven total variants (five variants in four genes and two intergenic variants) significantly associated with OS by both p-value ($p<0.05$) and FDR q-value ($q<0.1$). Five variants associated with OS in patients from the sorafenib arm, while two variants associated with OS when both treatment arms were combined. Variants are listed in descending order, from smallest to largest, by p-value. No variants were significant by p-value and after FDR correction in patients from the placebo arm. Alt, Alternate; Chr, chromosome; FDR, false discovery rate; HR, hazard ratio; HWE, Hardy Weinberg equilibrium; MAF, minor allele frequency; Pos, position; Ref, reference; * denotes imputed variants

Variant ID	Chr	Gene	Ref/Alt Allele	Feature	MAF	HR (95% CI)	P-value	FDR Q-value
Sorafenib Arm (n=155)								
rs1885657	6	*VEGFA*	T/C	Intron	0.17	17.32 (5.69-52.73)	$1.39x10^{-04}$	0.076
rs200809375*	10	NA	A/ATG	7.5 kB 3' of *NRP-1*	0.22	6.77 (2.62-17.5)	$2.65x10^{-04}$	0.076
rs6719561*	2	NA	C/T	1.8 kB 3' of *UGT1A9*	0.34	3.82 (1.56-9.52)	$3.27x10^{-04}$	0.076
rs8047917	16	*WWOX*	T/A	Intron	0.08	4.06 (1.88-8.78)	$3.27x10^{-04}$	0.076
rs3816375	2	*ITGAV*	A/G	Intron	0.38	5.88 (2.11-16.39)	$4.87x10^{-04}$	0.051
All Patients (n=295)								
rs307826	5	*FLT-4*	A/G	Missense (Thr494Ala)	0.10	13.79 (3.04-62.61)	$1.24x10^{-04}$	0.088
rs3024987*	6	*VEGFA*	C/T	Intron	0.11	2.98 (1.66-5.37)	$8.80x10^{-04}$	0.088

FIGURES

Figure 2.1. TARGET trial design

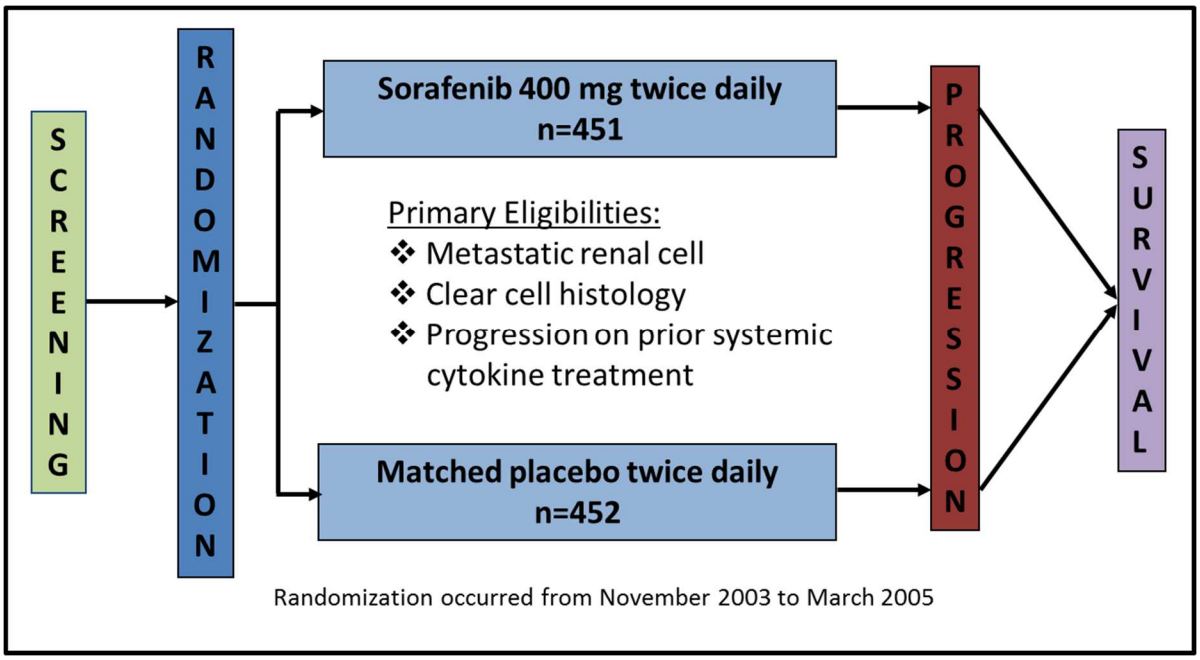


Figure 2.2. First and second line therapy recommendations for relapsed or Stage IV and surgically unresectable RCC. Per NCCN kidney cancer guidelines version 3.2015 (http://www.nccn.org/professionals/physician_gls/pdf/kidney.pdf), predictors of poor RCC prognosis include: lactate dehydrogenase >1.5 the upper limit of normal, hemoglobin level less than the lower limit of normal, corrected calcium >10.5 mg/dL (2.5 mmol/L), interval less than one year from original diagnosis until initiation of systemic therapy, Karnofsky performance score ≤70, and ≥2 metastatic sites. Abbreviations: IFN, interferon; IL, interleukin; NCCN, National Comprehensive Cancer Network; PS, performance status RANKL, ligand of receptor activator of nuclear factor-κB; RCC, renal cell carcinoma; XRT, external radiation therapy.

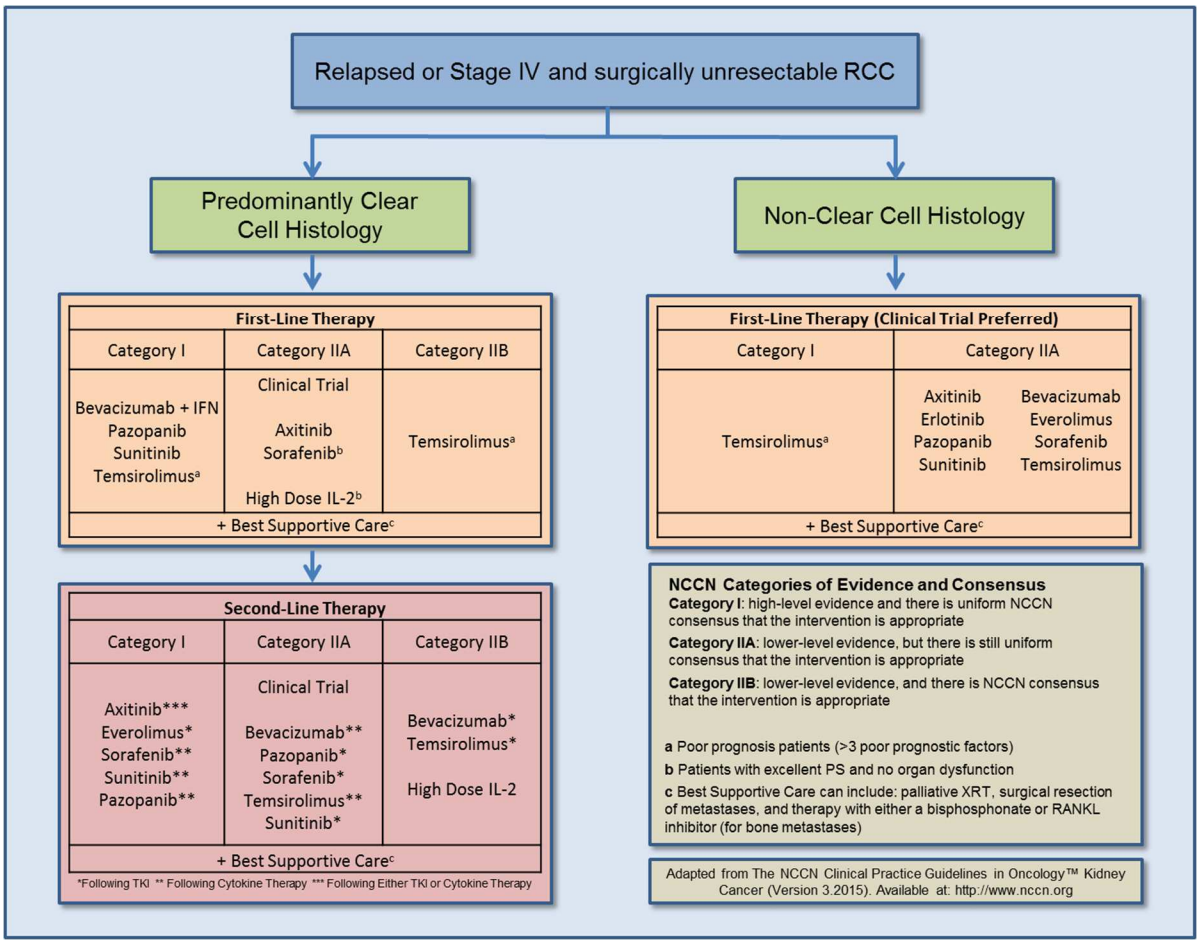


Figure 2.3. Illumina GenomeStudio output from GoldenGate assay. Fluorescence signals from the extension and amplification steps were read, and clustering for each individual variant was displayed in a scatter plot with the signal intensity on the y-axis and a signal intensity ratio on the x-axis. GenTrain scores were derived from the Illumina GenomeStudio scatter plots to measure detection reliability based on genotypic clustering distribution. Clustering distributions for three variants are included here as an example.

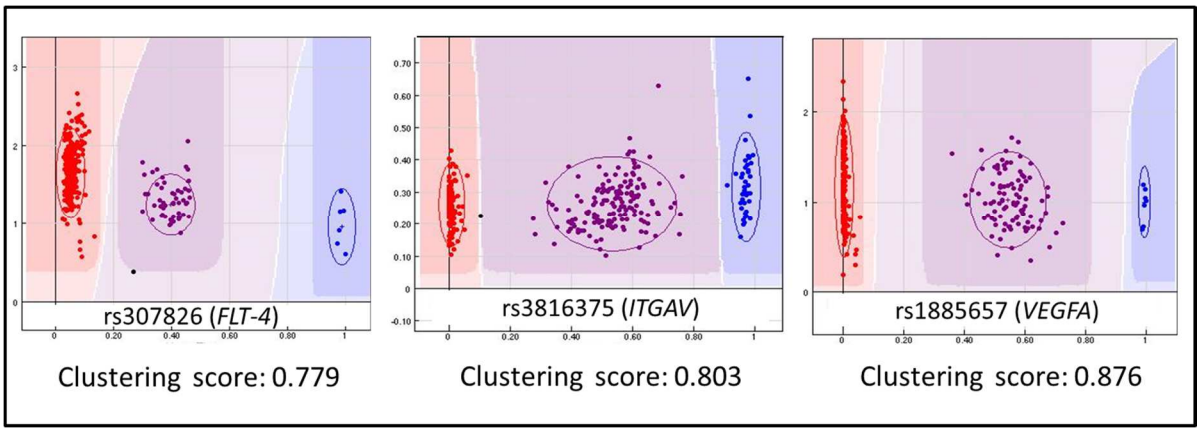


Figure 2.4. Kaplan-Meier analysis of OS in the entire TARGET population versus the genotyped TARGET patients. The curves show the percent of patients in each treatment group who survived while on trial (y-axis) versus time in days (x-axis) for all patients from the entire TARGET population (left panel) and the genotyped TARGET patients (right panel). Vertical bars on the survival curves indicate censored observations.

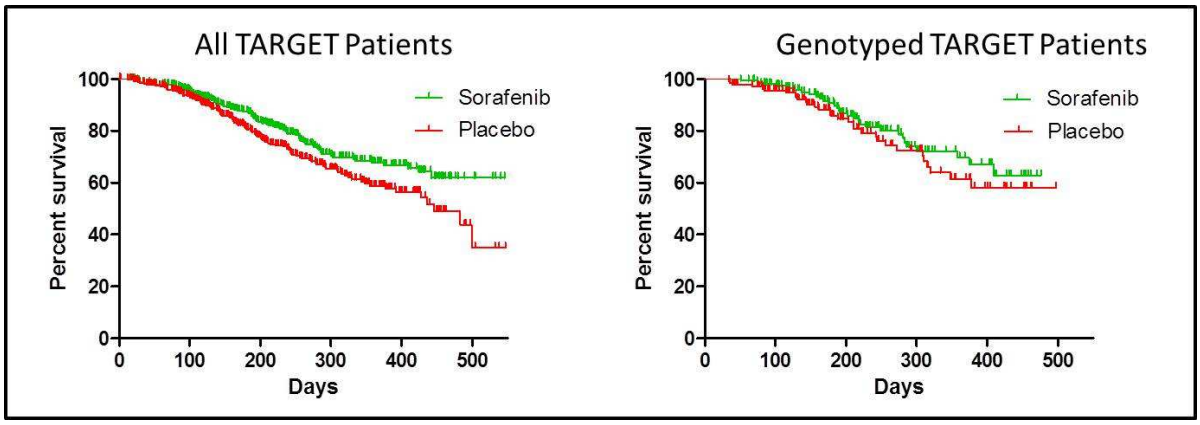


Figure 2.5. Kaplan-Meier analysis of PFS in the entire TARGET population versus the genotyped TARGET patients. The curves show the percent of patients in each treatment group who survived without progression while on trial (y-axis) versus time in days (x-axis) for all patients from the entire TARGET population (left panel) and the genotyped TARGET patients (right panel). Vertical bars on the survival curves indicate censored observations.

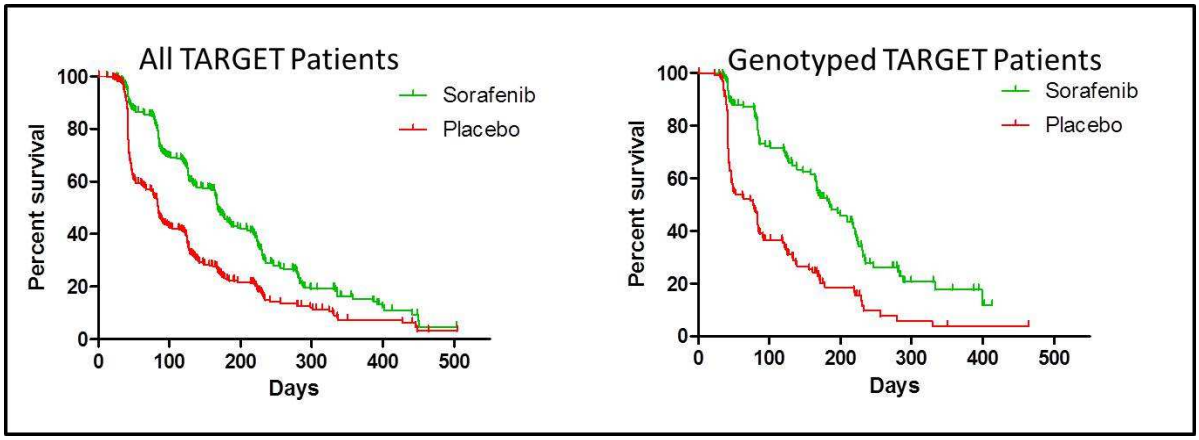


Figure 2.6. TARGET genotyping study schematics. Genomic DNA from 295 patients from the phase III TARGET trial was used in this genotyping study that sought to find associations between 11,117 germline variants (1,020 directly genotyped SNPs and 10,097 imputed SNPs and indels) and clinical phenotypes (OS and PFS). Abbreviations: HWE, Hardy Weinberg equilibrium; Indels, insertions/deletions; MAF, minor allele frequency; SNP, single nucleotide polymorphism; TARGET, Treatment Approaches in Renal Cancer Global Evaluation Trial.

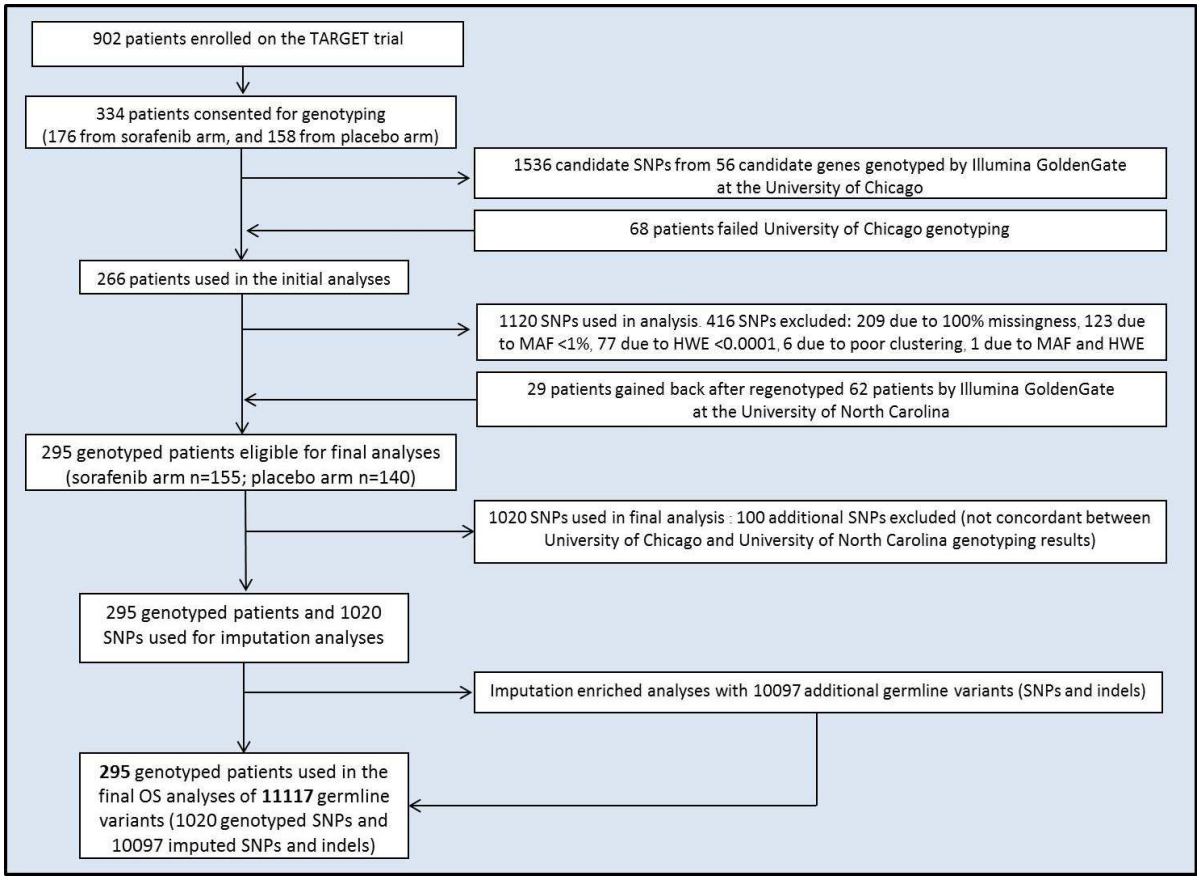


Figure 2.7. Results from confirmatory PCR assays. To corroborate genotype calls from both the Illumina GoldenGate Assay and Impute2, Taqman® assays were conducted on two directly genotyped variants: rs1885657 (panel A) and rs307826 (panel B), and two imputed variants: rs3024987 (panel C) and rs8047917 (panel D).

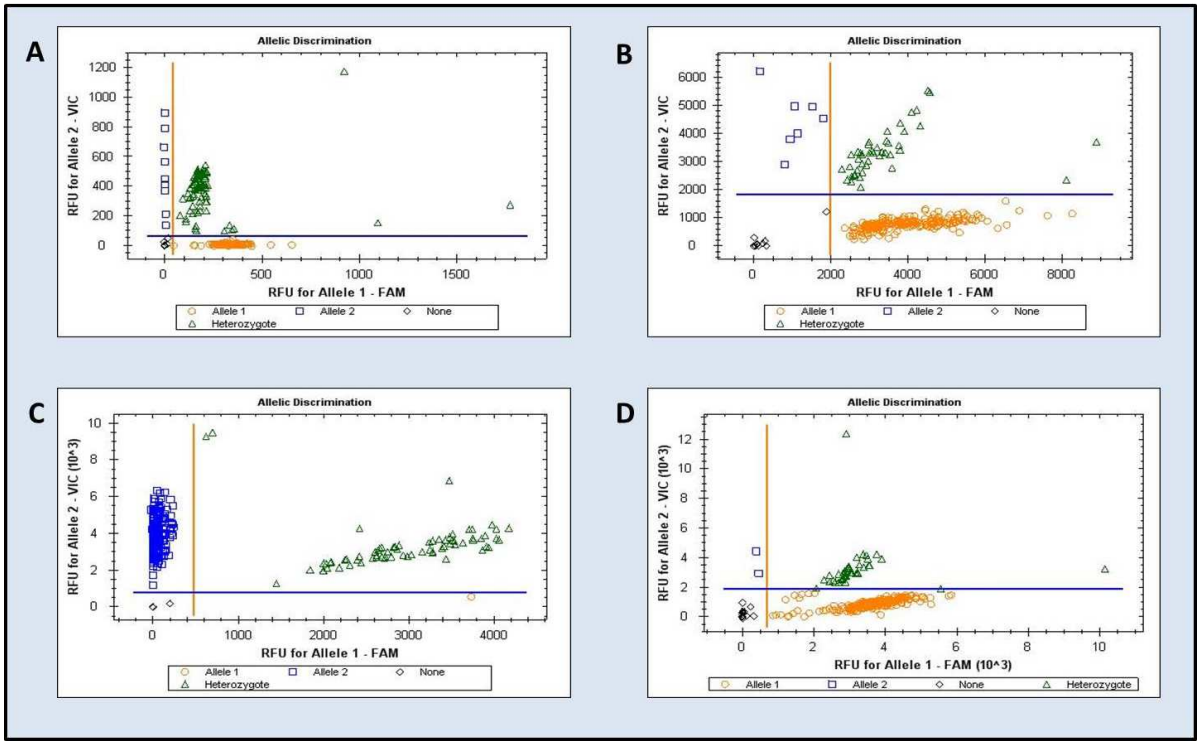


Figure 2.8. Kaplan-Meier analysis of OS for rs1885657 (*VEGFA*). The large panel shows that sorafenib-treated patients with the TT genotype live significantly shorter than patients with CC or TC genotypes. The other two panels are provided, for reference, to show the effects of the variant on OS in the patients from the placebo arm, and when both arms are combined. The curves show the percent of patients in each treatment group who survived without progression while on trial (y-axis) versus time in days (x-axis). Vertical bars on the survival curves indicate censored observations.

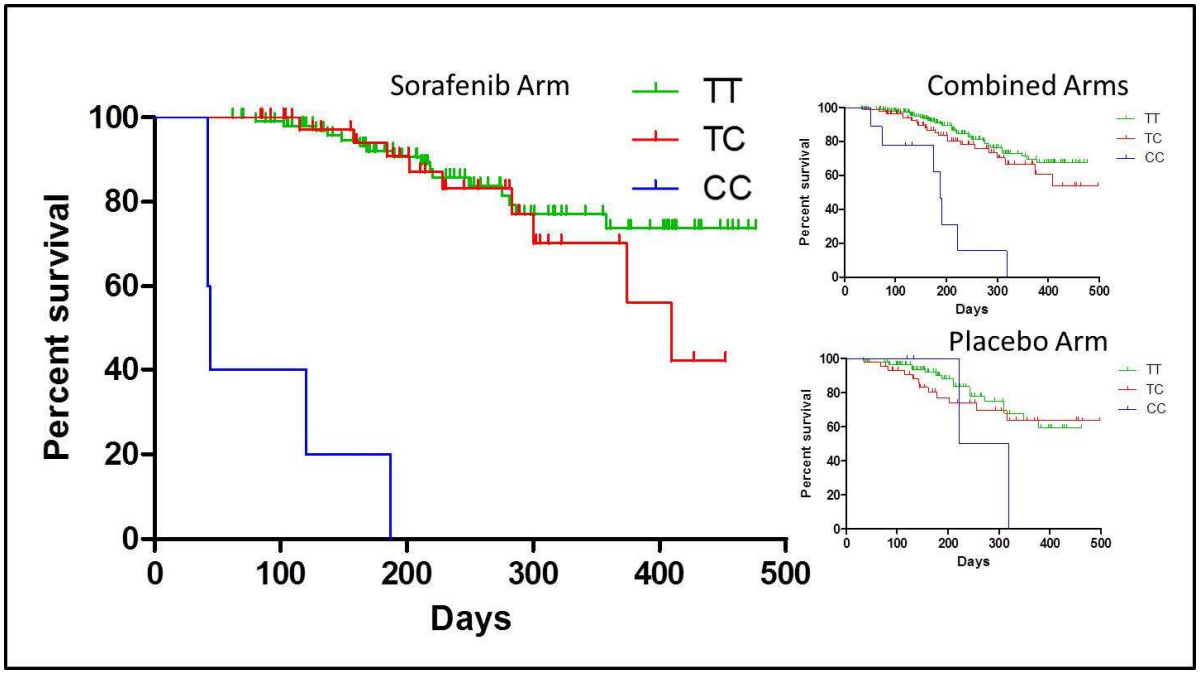


Figure 2.9. Kaplan-Meier analysis of OS for rs3816375 (*ITGAV*). The large panel shows that sorafenib-treated patients with the GG genotype live significantly shorter than patients with AA or AG genotypes. The other two panels are provided, for reference, to show the effects of the variant on OS in the patients from the placebo arm, and when both arms are combined. The curves show the percent of patients in each treatment group who survived without progression while on trial (y-axis) versus time in days (x-axis). Vertical bars on the survival curves indicate censored observations.

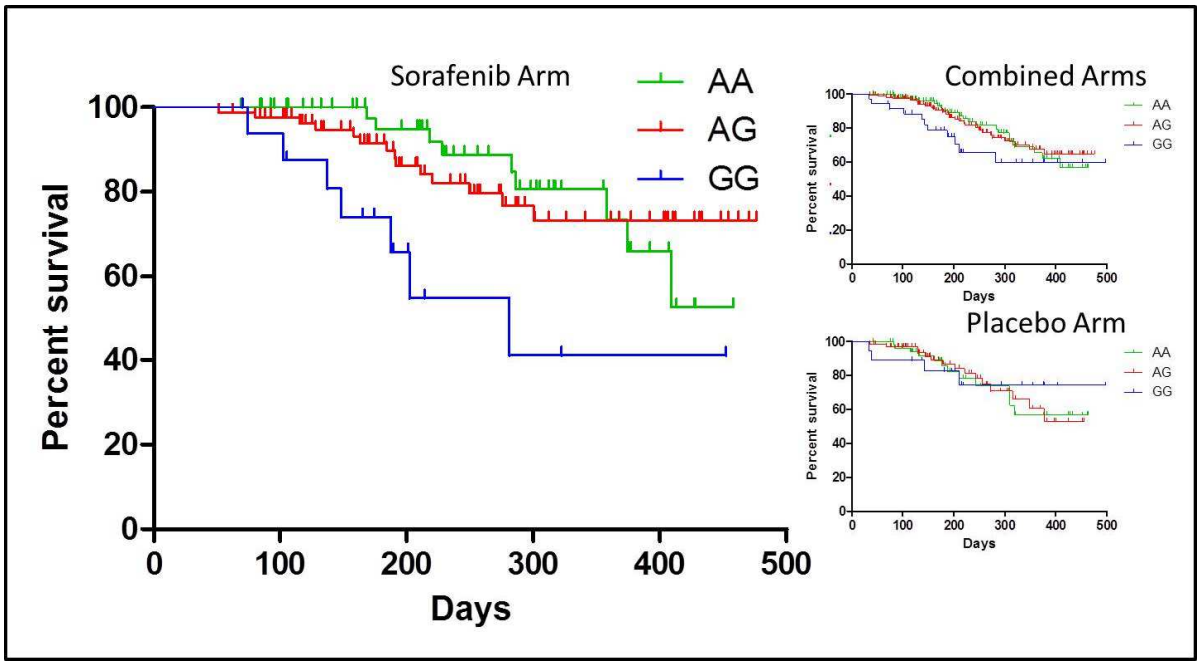


Figure 2.10. Kaplan-Meier analysis of OS for rs8047917 (*WWOX*). The large panel shows that sorafenib-treated patients with the AT genotype live significantly shorter than patients with TT genotype. The other two panels are provided, for reference, to show the effects of the variant on OS in the patients from the placebo arm, and when both arms are combined. The curves show the percent of patients in each treatment group who survived without progression while on trial (y-axis) versus time in days (x-axis). Vertical bars on the survival curves indicate censored observations.

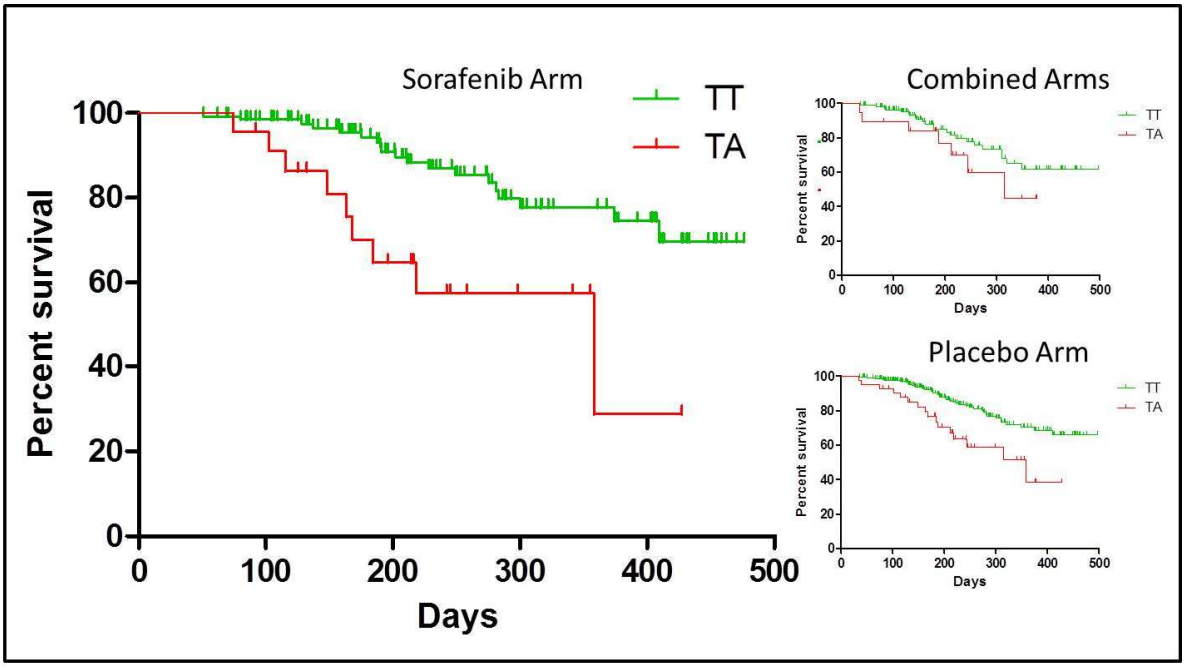


Figure 2.11. Kaplan-Meier analysis of OS for rs6719561 (3' of *UGT1A9*). The large panel shows that sorafenib-treated patients with the TT genotype live significantly shorter than patients with CC or CT genotypes. The other two panels are provided, for reference, to show the effects of the variant on OS in the patients from the placebo arm, and when both arms are combined. The curves show the percent of patients in each treatment group who survived without progression while on trial (y-axis) versus time in days (x-axis). Vertical bars on the survival curves indicate censored observations.

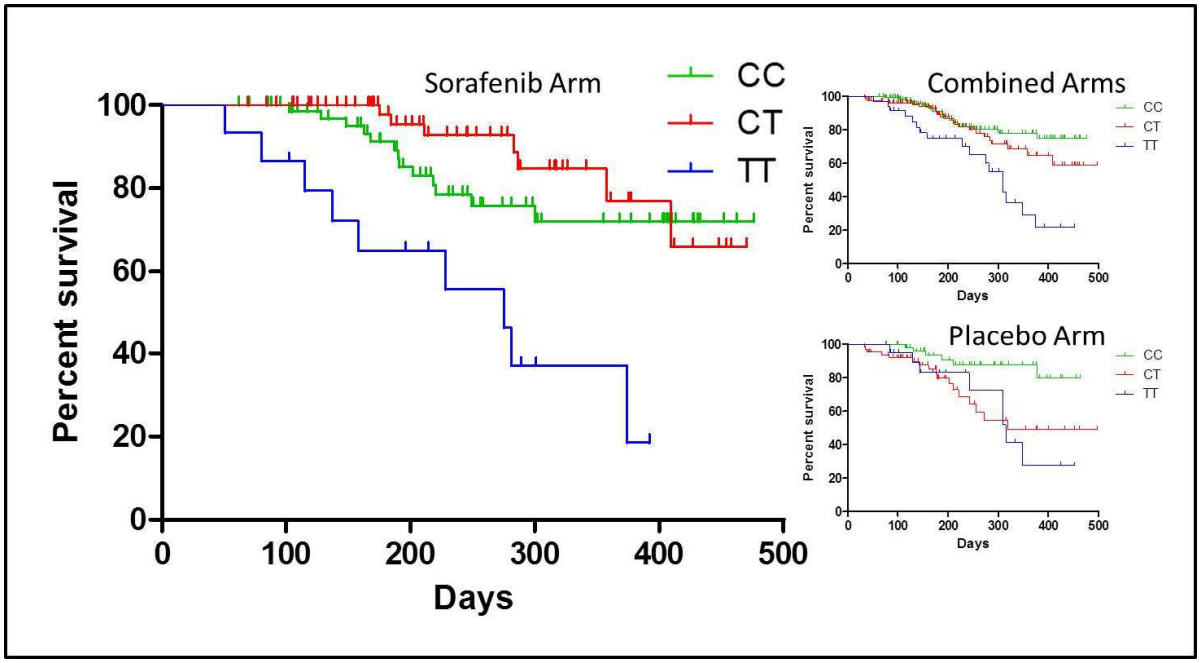


Figure 2.12. Kaplan-Meier analysis of OS for rs200809375 (3' of *NRP-1*). The large panel shows that sorafenib-treated patients with the at least one ATG insertion genotype live significantly shorter than patients with the AA genotype. The other two panels are provided, for reference, to show the effects of the indel on OS in the patients from the placebo arm, and when both arms are combined. The curves show the percent of patients in each treatment group who survived without progression while on trial (y-axis) versus time in days (x-axis). Vertical bars on the survival curves indicate censored observations.

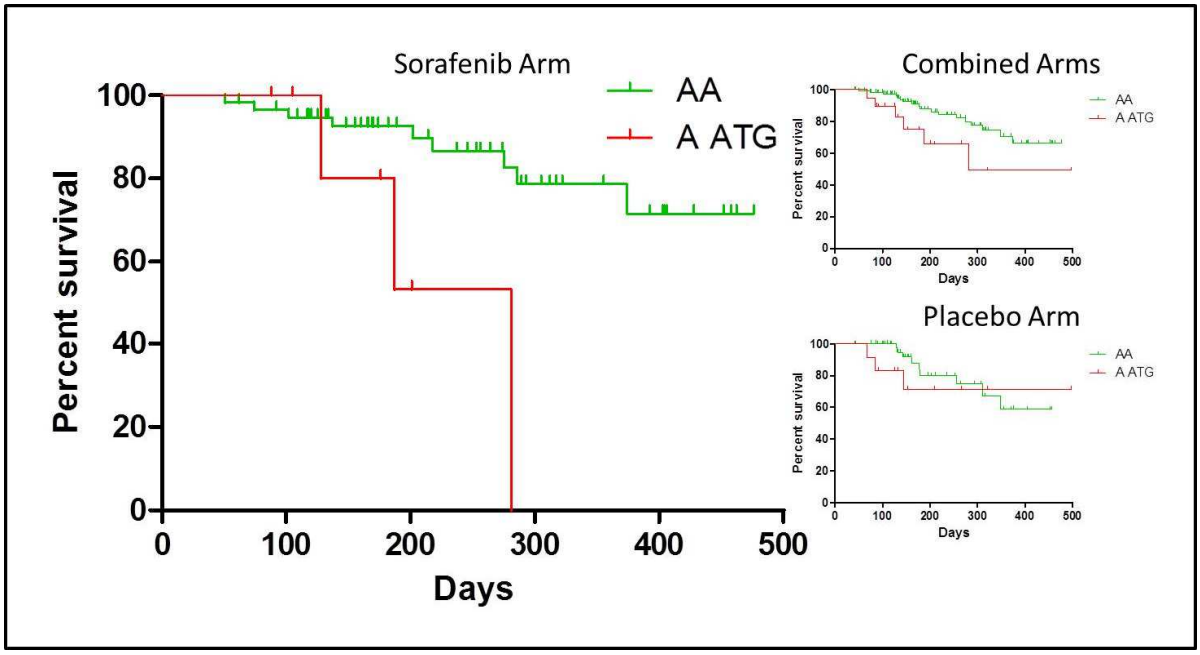


Figure 2.13. Kaplan-Meier analysis of OS for rs307826 (*FLT-4*). The large panel shows that in a combined group of patients, those with the GG genotype live significantly shorter than patients with AA or AG genotypes. The other two panels are provided, for reference, to show the effects of the variant on OS in the patients from the placebo and sorafenib arms. The curves show the percent of patients in each treatment group who survived without progression while on trial (y-axis) versus time in days (x-axis). Vertical bars on the survival curves indicate censored observations.

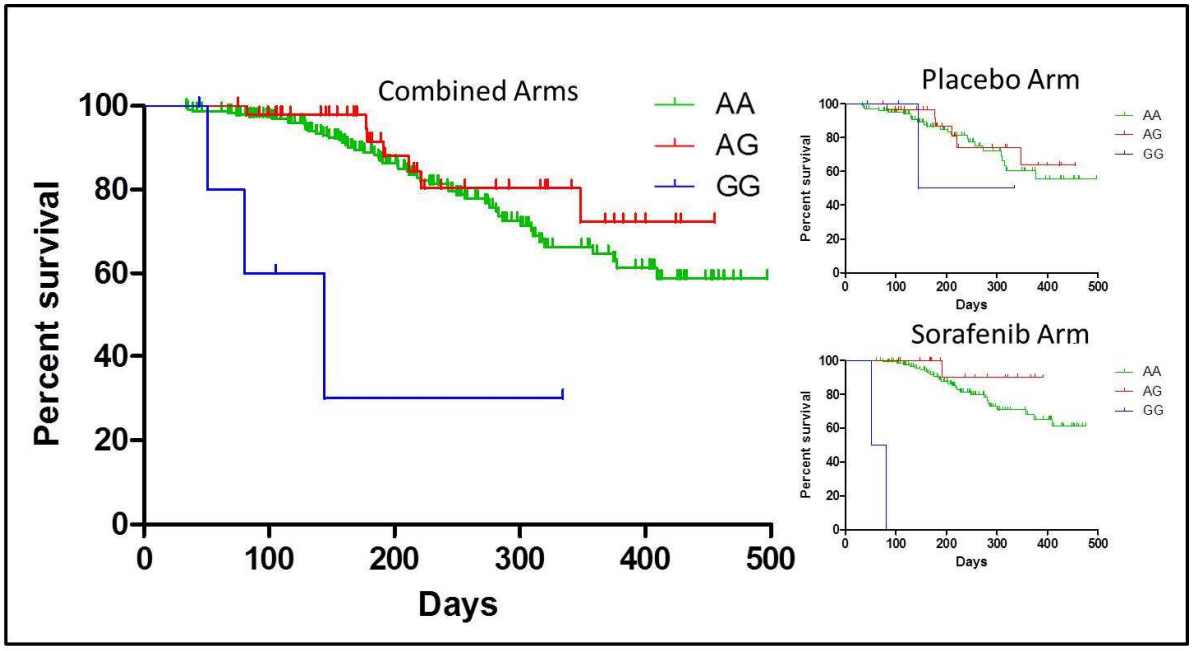


Figure 2.14. Kaplan-Meier analysis of OS for rs3024987 (*VEGFA*). The large panel shows that in a combined group of patients, those with the TC genotype live significantly shorter than patients with CC genotype. The other two panels are provided, for reference, to show the effects of the variant on OS in the patients from the placebo and sorafenib arms. The curves show the percent of patients in each treatment group who survived without progression while on trial (y-axis) versus time in days (x-axis). Vertical bars on the survival curves indicate censored observations.

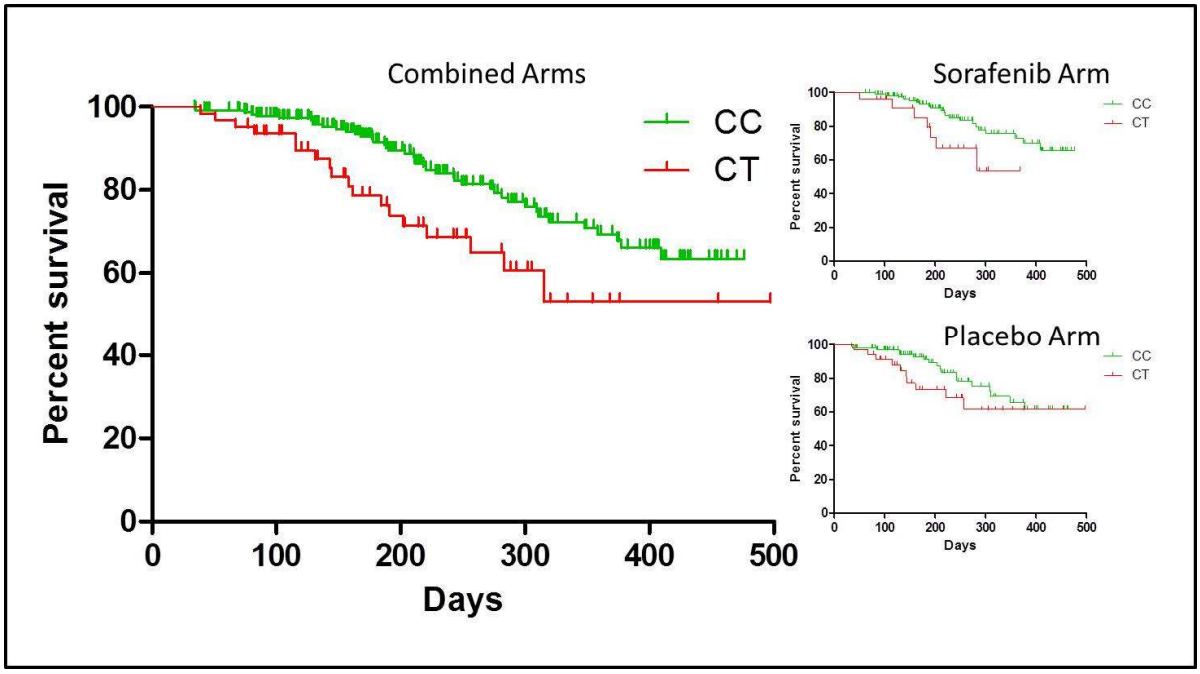


Figure 2.15. Correlation analyses between OS and PFS. Correlation analyses between OS and PFS are provided for **A)** all TARGET patients, **B)** among all genotyped TARGET patients, **C)** among genotyped TARGET patients treated with sorafenib, and **D)** among TARGET patients administered placebo. For each of the correlation analyses, patients that were censored in the survival analyses were removed. For the correlation test between OS and PFS for all TARGET patients, a Spearman rank correlation test was used. For all others, a Pearson product moment correlation was used.

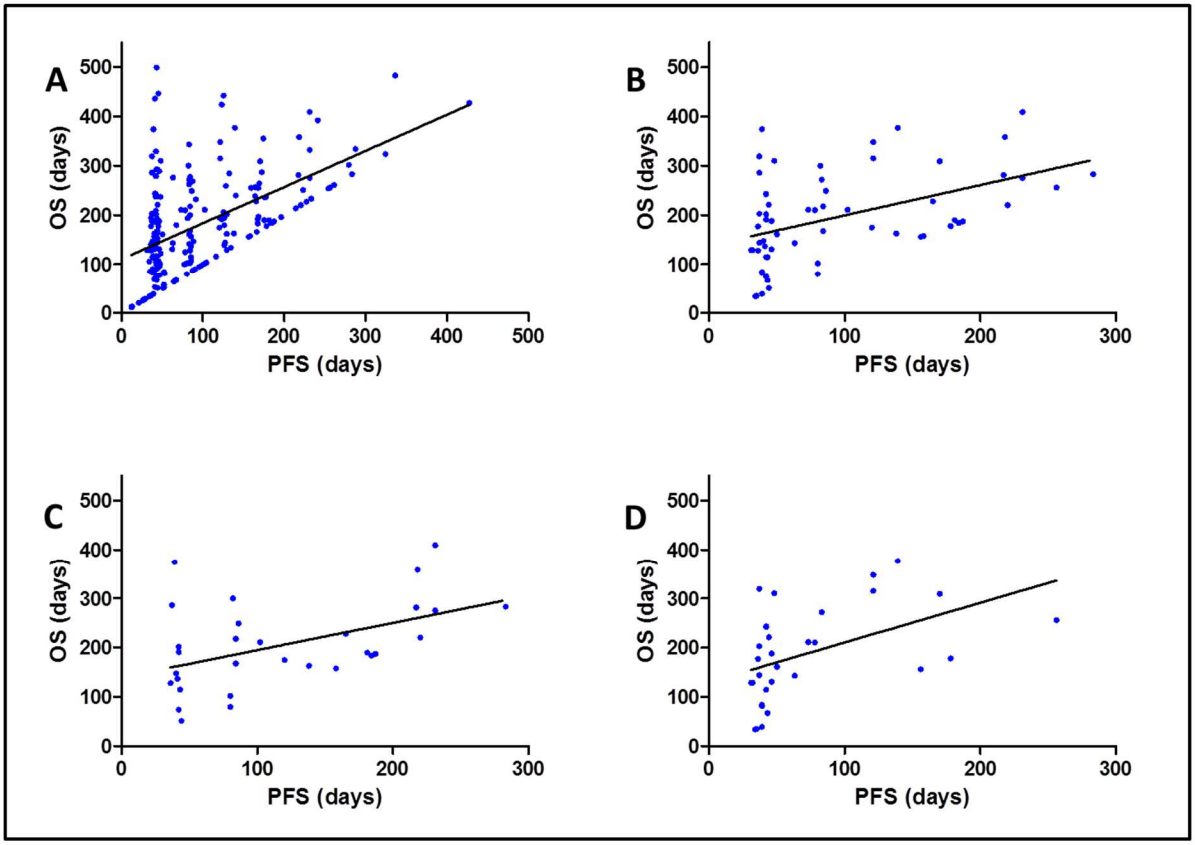


Figure 2.16. Kaplan-Meier analysis of PFS for rs1885657 (*VEGFA*). Variant rs1885657, which associated with OS, also significantly associated with PFS ($p \leq 0.05$) in **A)** sorafenib-treated patients, and **B)** when both treatment arms were combined. The curves show the percent of patients in each treatment group who survived without progression while on trial (y-axis) versus time in days (x-axis). Vertical bars on the survival curves indicate censored observations. Abbreviations: PFS, progression free survival; *VEGFA*, vascular endothelial growth factor A.

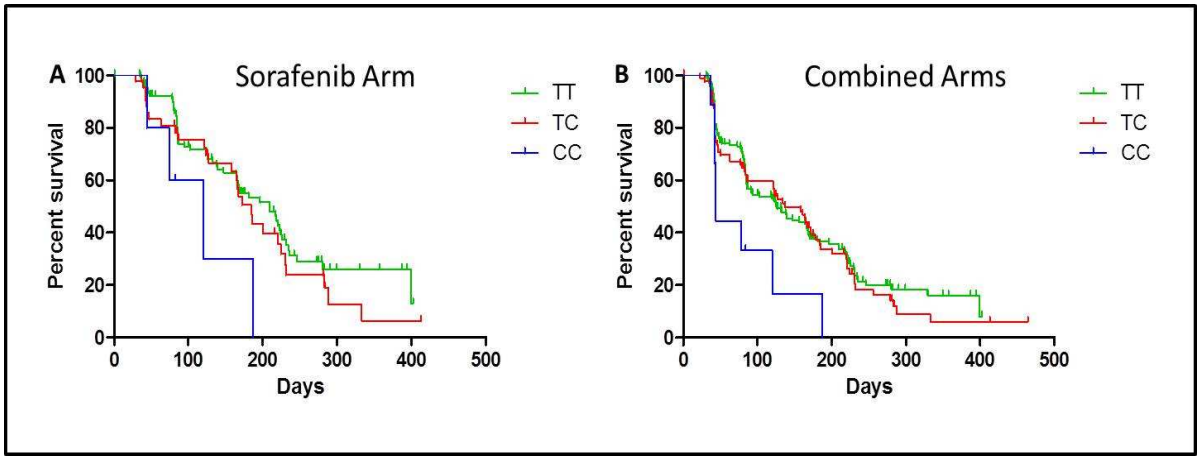


Figure 2.17. Kaplan-Meier analysis of PFS for three variants previously with OS. Three variants, associated with OS, were also associated with PFS: **A)** rs8047917 in *WWOX*, **B)** rs3024987 in *VEGFA*, and **C)** rs6719561, which is 3' of *UGT1A9,* also significantly associated with PFS ($p \leq 0.05$) in sorafenib-treated patients. The curves show the percent of patients in each treatment group who survived without progression while on trial (y-axis) versus time in days (x-axis). Vertical bars on the survival curves indicate censored observations. Abbreviations: PFS, progression free survival; *UGT1A9*, Uridine 5'-diphospho-glucuronosyltransferase 1 family, polypeptide A9; *VEGFA*, vascular endothelial growth factor A; *WWOX*, WW domain-containing oxidoreductase.

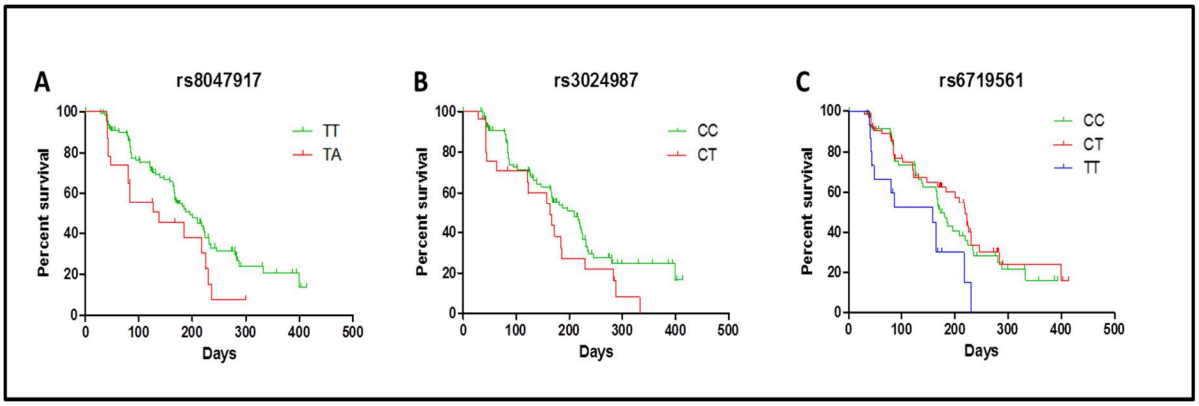
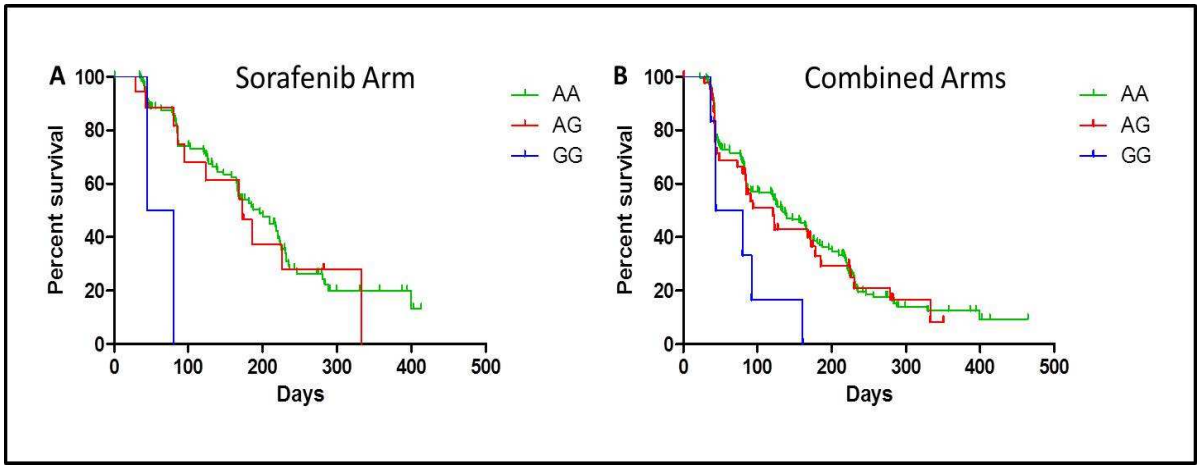


Figure 2.18. Kaplan-Meier analysis of PFS for rs307826 (*FLT-4*). Variant rs307826, associated with OS, also significantly associated with PFS ($p \leq 0.05$) in **A)** sorafenib-treated patients, and **B)** when both treatment arms were combined. The curves show the percent of patients in each treatment group who survived without progression while on trial (y-axis) versus time in days (x-axis). Vertical bars on the survival curves indicate censored observations. Abbreviations: *FLT-4*, fms-related tyrosine kinase 4; PFS, progression free survival.



REFERENCES

1. Chen AM, Jahan TM, Jablons DM, Garcia J, Larson DA. Risk of cerebral metastases and neurological death after pathological complete response to neoadjuvant therapy for locally advanced nonsmall-cell lung cancer: clinical implications for the subsequent management of the brain. *Cancer*. 2007;109:1668-1675.

2. Wilhelm SM, Carter C, Tang L, et al. BAY 43-9006 exhibits broad spectrum oral antitumor activity and targets the RAF/MEK/ERK pathway and receptor tyrosine kinases involved in tumor progression and angiogenesis. *Cancer Res*. 2004;64:7099-7109.

3. Wilhelm S, Chien DS. BAY 43-9006: preclinical data. *Curr Pharm Des*. 2002;8:2255-2257.

4. Hilger RA, Scheulen ME, Strumberg D. The Ras-Raf-MEK-ERK pathway in the treatment of cancer. *Onkologie*. 2002;25:511-518.

5. Ratain MJ, Eisen T, Stadler WM, et al. Phase II placebo-controlled randomized discontinuation trial of sorafenib in patients with metastatic renal cell carcinoma. *J Clin Oncol*. 2006;24:2505-2512.

6. Escudier B, Eisen T, Stadler WM, et al. Sorafenib in advanced clear-cell renal-cell carcinoma. *N Engl J Med*. 2007;356:125-134.

7. Escudier B, Eisen T, Stadler WM, et al. Sorafenib for treatment of renal cell carcinoma: Final efficacy and safety results of the phase III treatment approaches in renal cancer global evaluation trial. *J Clin Oncol*. 2009;27:3312-3318.

8. Tokes AM, Szasz AM, Geszti F, et al. Expression of proliferation markers Ki67, cyclin A, geminin and aurora-kinase A in primary breast carcinomas and corresponding distant metastases. *J Clin Pathol*. 2015;68:274-282.

9. Ravaud A, Gross-Goupil M. Overcoming resistance to tyrosine kinase inhibitors in renal cell carcinoma. *Cancer Treat Rev*. 2012;38:996-1003.

10. Escudier B, Szczylik C, Porta C, Gore M. Treatment selection in metastatic renal cell carcinoma: expert consensus. *Nat Rev Clin Oncol*. 2012;9:327-337.

11. Awada A, Hendlisz A, Gil T, et al. Phase I safety and pharmacokinetics of BAY 43-9006 administered for 21 days on/7 days off in patients with advanced, refractory solid tumours. *Br J Cancer*. 2005;92:1855-1861.

12. Clark JW, Eder JP, Ryan D, Lathia C, Lenz HJ. Safety and pharmacokinetics of the dual action Raf kinase and vascular endothelial growth factor receptor inhibitor, BAY 43-9006, in patients with advanced, refractory solid tumors. *Clin Cancer Res*. 2005;11:5472-5480.

13. Strumberg D, Richly H, Hilger RA, et al. Phase I clinical and pharmacokinetic study of the Novel Raf kinase and vascular endothelial growth factor receptor inhibitor BAY 43-9006 in patients with advanced refractory solid tumors. *J Clin Oncol*. 2005;23:965-972.

14. Strumberg D, Clark JW, Awada A, et al. Safety, pharmacokinetics, and preliminary antitumor activity of sorafenib: a review of four phase I trials in patients with advanced refractory solid tumors. *Oncologist*. 2007;12:426-437.

15. Moore M, Hirte HW, Siu L, et al. Phase I study to determine the safety and pharmacokinetics of the novel Raf kinase and VEGFR inhibitor BAY 43-9006, administered for 28 days on/7 days off in patients with advanced, refractory solid tumors. *Ann Oncol*. 2005;16:1688-1694.

16. Motzer RJ, Mazumdar M, Bacik J, Berg W, Amsterdam A, Ferrara J. Survival and prognostic stratification of 670 patients with advanced renal cell carcinoma. *J Clin Oncol*. 1999;17:2530-2540.

17. Motzer RJ, Bukowski RM, Figlin RA, et al. Prognostic nomogram for sunitinib in patients with metastatic renal cell carcinoma. *Cancer*. 2008;113:1552-1558.

18. Motzer RJ, Bacik J, Murphy BA, Russo P, Mazumdar M. Interferon-alfa as a comparative treatment for clinical trials of new therapies against advanced renal cell carcinoma. *J Clin Oncol*. 2002;20:289-296.

19. Choueiri TK, Garcia JA, Elson P, et al. Clinical factors associated with outcome in patients with metastatic clear-cell renal cell carcinoma treated with vascular endothelial growth factor-targeted therapy. *Cancer*. 2007;110:543-550.

20. Heng DY, Xie W, Regan MM, et al. Prognostic factors for overall survival in patients with metastatic renal cell carcinoma treated with vascular endothelial growth factor-targeted agents: results from a large, multicenter study. *J Clin Oncol*. 2009;27:5794-5799.

21. Zurita AJ, Jonasch E, Wang X, et al. A cytokine and angiogenic factor (CAF) analysis in plasma for selection of sorafenib therapy in patients with metastatic renal cell carcinoma. *Ann Oncol*. 2012;23:46-52.

22. Yang JC. Bevacizumab for patients with metastatic renal cancer: an update. *Clin Cancer Res*. 2004;10:6367S-6370S.

23. Deprimo SE BC, Smeraglia J, Shalinsky DR, Freddo J, Baum CM, Rini BI, Michaelson D, Motzer RJ, Spinella DG. Soluble protein biomarkers of pharmacodynamic activity of the multi-targeted kinase inhibitor SU11248 in patients with metastatic renal cell cancer. *Proc Am Assoc Cancer Res*. 2005:108, abstract 464.

24. Maroto P, Rini B. Molecular biomarkers in advanced renal cell carcinoma. *Clin Cancer Res*. 2014;20:2060-2071.

25. van der Veldt AA, Eechoute K, Gelderblom H, et al. Genetic polymorphisms associated with a prolonged progression-free survival in patients with metastatic renal cell cancer treated with sunitinib. *Clin Cancer Res*. 2011;17:620-629.

26. Xu CF, Bing NX, Ball HA, et al. Pazopanib efficacy in renal cell carcinoma: evidence for predictive genetic markers in angiogenesis-related and exposure-related genes. *J Clin Oncol*. 2011;29:2557-2564.

27. Beuselinck B, Karadimou A, Lambrechts D, et al. Single-nucleotide polymorphisms associated with outcome in metastatic renal cell carcinoma treated with sunitinib. *Br J Cancer*. 2013;108:887-900.

28. Garcia-Donas J, Esteban E, Leandro-Garcia LJ, et al. Single nucleotide polymorphism associations with response and toxic effects in patients with advanced renal-cell carcinoma treated with first-line sunitinib: a multicentre, observational, prospective study. *Lancet Oncol*. 2011;12:1143-1150.

29. van Erp NP, Eechoute K, van der Veldt AA, et al. Pharmacogenetic pathway analysis for determination of sunitinib-induced toxicity. *J Clin Oncol*. 2009;27:4406-4412.

30. Oka H, Chatani Y, Hoshino R, et al. Constitutive activation of mitogen-activated protein (MAP) kinases in human renal cell carcinoma. *Cancer Res*. 1995;55:4182-4187.

31. Cohen HT, McGovern FJ. Renal-cell carcinoma. *N Engl J Med*. 2005;353:2477-2490.

32. Sabo E, Boltenko A, Sova Y, Stein A, Kleinhaus S, Resnick MB. Microscopic analysis and significance of vascular architectural complexity in renal cell carcinoma. *Clin Cancer Res*. 2001;7:533-537.

33. Ferda J, Hora M, Hes O, Ferdova E, Kreuzberg B. Assessment of the kidney tumor vascular supply by two-phase MDCT-angiography. *Eur J Radiol*. 2007;62:295-301.

34. Anderson H, Yap JT, Wells P, et al. Measurement of renal tumour and normal tissue perfusion using positron emission tomography in a phase II clinical trial of razoxane. *Br J Cancer*. 2003;89:262-267.

35. Chang YS, Adnane J, Trail PA, et al. Sorafenib (BAY 43-9006) inhibits tumor growth and vascularization and induces tumor apoptosis and hypoxia in RCC xenograft models. *Cancer Chemother Pharmacol*. 2007;59:561-574.

36. Grothey A, Galanis E. Targeting angiogenesis: progress with anti-VEGF treatment with large molecules. *Nat Rev Clin Oncol*. 2009;6:507-518.

37. Liu L, Cao Y, Chen C, et al. Sorafenib blocks the RAF/MEK/ERK pathway, inhibits tumor angiogenesis, and induces tumor cell apoptosis in hepatocellular carcinoma model PLC/PRF/5. *Cancer Res*. 2006;66:11851-11858.

38. Kim S, Yazici YD, Calzada G, et al. Sorafenib inhibits the angiogenesis and growth of orthotopic anaplastic thyroid carcinoma xenografts in nude mice. *Mol Cancer Ther*. 2007;6:1785-1792.

39. Yuen JS, Sim MY, Siml HG, et al. Inhibition of angiogenic and non-angiogenic targets by sorafenib in renal cell carcinoma (RCC) in a RCC xenograft model. *Br J Cancer*. 2011;104:941-947.

40. Murphy DA, Makonnen S, Lassoued W, Feldman MD, Carter C, Lee WM. Inhibition of tumor endothelial ERK activation, angiogenesis, and tumor growth by sorafenib (BAY43-9006). *Am J Pathol*. 2006;169:1875-1885.

41. Murakami M, Zhao S, Zhao Y, et al. Evaluation of changes in the tumor microenvironment after sorafenib therapy by sequential histology and 18F-fluoromisonidazole hypoxia imaging in renal cell carcinoma. *Int J Oncol*. 2012;41:1593-1600.

42. Kamba T, McDonald DM. Mechanisms of adverse effects of anti-VEGF therapy for cancer. *Br J Cancer*. 2007;96:1788-1795.

43. Pignochino Y, Grignani G, Cavalloni G, et al. Sorafenib blocks tumour growth, angiogenesis and metastatic potential in preclinical models of osteosarcoma through a mechanism potentially involving the inhibition of ERK1/2, MCL-1 and ezrin pathways. *Mol Cancer*. 2009;8:118.

44. Rini BI, Small EJ. Biology and clinical development of vascular endothelial growth factor-targeted therapy in renal cell carcinoma. *J Clin Oncol*. 2005;23:1028-1043.

45. Rini BI, Sosman JA, Motzer RJ. Therapy targeted at vascular endothelial growth factor in metastatic renal cell carcinoma: biology, clinical results and future development. *BJU Int*. 2005;96:286-290.

46. Azam F, Mehta S, Harris AL. Mechanisms of resistance to antiangiogenesis therapy. *Eur J Cancer*. 2010;46:1323-1332.

47. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst*. 2000;92:205-216.

48. Pare-Brunet L, Glubb D, Evans P, et al. Discovery and functional assessment of gene variants in the vascular endothelial growth factor pathway. *Hum Mutat*. 2014;35:227-235.

49. Gamazon ER, Zhang W, Konkashbaev A, et al. SCAN: SNP and copy number annotation. *Bioinformatics*. 2010;26:259-262.

50. Lin CH, Yeakley JM, McDaniel TK, Shen R. Medium- to high-throughput SNP genotyping using VeraCode microbeads. *Methods Mol Biol*. 2009;496:129-142.

51. Shen R, Fan JB, Campbell D, et al. High-throughput SNP genotyping on universal bead arrays. *Mutat Res*. 2005;573:70-82.

52. Fan JB, Oliphant A, Shen R, et al. Highly parallel SNP genotyping. *Cold Spring Harb Symp Quant Biol*. 2003;68:69-78.

53. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5:e1000529.

54. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*. 2007;39:906-913.

55. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res*. 2001;125:279-284.

56. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, 2012. (Accessed at http://www.R-project.org.).

57. Johnson JR, Williams G, Pazdur R. End points and United States Food and Drug Administration approval of oncology drugs. *J Clin Oncol*. 2003;21:1404-1411.

58. Oldenhuis CN, Oosting SF, Gietema JA, de Vries EG. Prognostic versus predictive value of biomarkers in oncology. *Eur J Cancer*. 2008;44:946-953.

59. Rini BI. Vascular endothelial growth factor-targeted therapy in metastatic renal cell carcinoma. *Cancer*. 2009;115:2306-2312.

60. Bukowski RM. Prognostic factors for survival in metastatic renal cell carcinoma: update 2008. *Cancer*. 2009;115:2273-2281.

61. Scartozzi M, Bianconi M, Faloppi L, et al. VEGF and VEGFR polymorphisms affect clinical outcome in advanced renal cell carcinoma patients receiving first-line sunitinib. *Br J Cancer*. 2013;108:1126-1132.

62. Kim JJ, Vaziri SA, Rini BI, et al. Association of VEGF and VEGFR2 single nucleotide polymorphisms with hypertension and clinical outcome in metastatic clear cell renal cell carcinoma patients treated with sunitinib. *Cancer*. 2012;118:1946-1954.

63. Kaipainen A, Korhonen J, Mustonen T, et al. Expression of the fms-like tyrosine kinase 4 gene becomes restricted to lymphatic endothelium during development. *Proc Natl Acad Sci U S A*. 1995;92:3566-3570.

64. Smith NR, Baker D, James NH, et al. Vascular endothelial growth factor receptors VEGFR-2 and VEGFR-3 are localized primarily to the vasculature in human primary solid cancers. *Clin Cancer Res*. 2010;16:3548-3561.

65. Tammela T, Zarkada G, Wallgard E, et al. Blocking VEGFR-3 suppresses angiogenic sprouting and vascular network formation. *Nature*. 2008;454:656-660.

66. Roberts N, Kloos B, Cassella M, et al. Inhibition of VEGFR-3 activation with the antagonistic antibody more potently suppresses lymph node and distant metastases than inactivation of VEGFR-2. *Cancer Res*. 2006;66:2650-2657.

67. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7:248-249.

68. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4:1073-1081.

69. Xu CF BH, Bing N, Sternberg C, Xue Z, McCann L, King K, Spraggs C, Mooser V, Pandite LN. . Association of genetic markers in angiogenesis- or exposure-related genes with overall survival in pazopanib-treated patients with advanced renal cell carcinoma. *J Clin Oncol*. 2011;29(Supp 7):abstract 303.

70. Tucker GC. Inhibitors of integrins. *Curr Opin Pharmacol*. 2002;2:394-402.

71. Tsuchiya Y, Sawada S, Tsukada K, Saiki I. A new pseudo-peptide of Arg-Gly-Asp (RGD) inhibits intrahepatic metastasis of orthotopically implanted murine hepatocellular carcinoma. *Int J Oncol*. 2002;20:319-324.

72. Cui R, Takahashi F, Ohashi R, et al. Abrogation of the interaction between osteopontin and alphavbeta3 integrin reduces tumor growth of human lung cancer cells in mice. *Lung Cancer*. 2007;57:302-310.

73. Mettouchi A, Meneguzzi G. Distinct roles of beta1 integrins during angiogenesis. *Eur J Cell Biol*. 2006;85:243-247.

74. Waisberg J, De Souza Viana L, Affonso Junior RJ, et al. Overexpression of the ITGAV gene is associated with progression and spread of colorectal cancer. *Anticancer Res*. 2014;34:5599-5607.

75. Ha SY, Shin J, Kim JH, et al. Overexpression of integrin alphav correlates with poor prognosis in colorectal cancer. *J Clin Pathol*. 2014;67:576-581.

76. Del Mare S, Salah Z, Aqeilan RI. WWOX: its genomics, partners, and functions. *J Cell Biochem*. 2009;108:737-745.

77. Lange EM, Beebe-Dimmer JL, Ray AM, et al. Genome-wide linkage scan for prostate cancer susceptibility from the University of Michigan Prostate Cancer Genetics Project: suggestive evidence for linkage at 16q23. *Prostate*. 2009;69:385-391.

78. Jenner MW, Leone PE, Walker BA, et al. Gene mapping and expression analysis of 16q loss of heterozygosity identifies WWOX and CYLD as being important in determining clinical outcome in multiple myeloma. *Blood*. 2007;110:3291-3300.

79. Frullanti E, Galvan A, Falvella FS, et al. Multiple genetic loci modulate lung adenocarcinoma clinical staging. *Clin Cancer Res*. 2011;17:2410-2416.

80. Product information. Nexavar (sorafenib). Wayne NBHP, Inc. . June 2013. (Accessed January 29, 2015. Available at: http://berlex.bayerhealthcare.com/html/products/pi/Nexavar_PI.pdf.)

81. Iyer R, Fetterly G, Lugade A, Thanavala Y. Sorafenib: a clinical and pharmacologic review. *Expert Opin Pharmacother*. 2010;11:1943-1955.

82. Boudou-Rouquette P, Narjoz C, Golmard JL, et al. Early sorafenib-induced toxicity is associated with drug exposure and UGTIA9 genetic polymorphism in patients with solid tumors: a preliminary study. *PLoS One*. 2012;7:e42875.

83. Jain L, Woo S, Gardner ER, et al. Population pharmacokinetic analysis of sorafenib in patients with solid tumours. *Br J Clin Pharmacol*. 2011;72:294-305.

84. Soker S, Takashima S, Miao HQ, Neufeld G, Klagsbrun M. Neuropilin-1 is expressed by endothelial and tumor cells as an isoform-specific receptor for vascular endothelial growth factor. *Cell*. 1998;92:735-745.

85. Cao Y, Wang L, Nandy D, et al. Neuropilin-1 upholds dedifferentiation and propagation phenotypes of renal cell carcinoma cells by activating Akt and sonic hedgehog axes. *Cancer Res*. 2008;68:8667-8672.

86. Chanock SJ, Manolio T, Boehnke M, et al. Replicating genotype-phenotype associations. *Nature*. 2007;447:655-660.

87. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat Genet*. 2001;29:306-309.

88. Buyse M, Sargent DJ, Grothey A, Matheson A, de Gramont A. Biomarkers and surrogate end points--the challenge of statistical validation. *Nat Rev Clin Oncol*. 2010;7:309-317.

CHAPTER 3: VALIDATION OF GERMLINE VARIANTS THAT ASSOCIATE WITH OVERALL SURVIVAL IN TARGET TRIAL PATIENTS

3.1 Overview

Background:

Statistical associations between genetic variants and outcomes in cancer studies should be supported with molecular mechanistic evidence of variant function to aid in biomarker validation. The primary objective of this Aim was to validate the functionality of germline variants were significantly associated with OS in TARGET patients, through a sequential *in silico* → *in vitro* approach, and to test the hypothesis that functional validation will elucidate the molecular effects of these variants on mRCC biology and sorafenib pharmacology.

Methods:

Germline variants, identified in Aim 1, that significantly associated with OS (multivariate model $p<0.05$ and FDR $q<0.1$) were analyzed using a selection of *in silico* bioinformatic tools to prioritize which would be validated using laboratory *in vitro* assays. Several resources were used to provide information on variant allele frequencies and LD patterns (e.g. 1,000 Genomes Project data, Ensembl and dbSNP). Others helped to predict the myriad ways variants influence gene regulation by leveraging ENCODE data (e.g. Ensembl, the UCSC genome browser, HaploReg, and Regulome DB). One non-synonymous germline

variant in a coding region of *FLT-4* (rs307826) was assessed for function using cell viability assays. Dual reporter gene luciferase assays were used to examine effects of variants on gene expression. Only intragenic variants were selected for functional validation; therefore, no experiments were in the two intergenic variants identified in Aim 1 (rs6719561 and rs200809375).

Results:

In silico tools helped predict functionality for five intragenic variants which were associated with OS in Aim 1. These tools prioritized intronic SNPs in *VEGFA* (rs1885657 and rs3024987) and rs307826 for functional validation. Cell viability assays revealed that rs307826 (G allele) results in a more resistant phenotype for HEK-293 cells treated with sorafenib. Dual reporter gene luciferase assays validated functionality of the two *VEGFA* variants (as well as two variants in perfect LD with rs1885657) in three different cell lines.

Conclusions:

In silico tools were useful for prioritizing variants for functional, *in vitro* validation. *In vitro* experiments validated the functionality of intronic *VEGFA* variants (rs1885657 and rs3024987), and the coding variant in *FLT-4* (rs307826). Additional experiments should be conducted to further elucidate the mechanisms underlying molecular effects that these variants have on OS in mRCC patients.

3.2 Introduction

Findings from pharmacogenetic and pharmacogenomic studies (e.g. candidate gene, GWAS, and now next generation sequencing) continue to provide a plethora of information about genetic variation that underlies both disease pathology and responses to pharmacotherapy. Statistical associations between genetic variants and outcomes (e.g. OS and/or PFS) in cancer studies should be properly supported with molecular mechanistic evidence of variant function, and the lack of mechanistically-based effects underlying clinical phenotypes is a major limitation for biomarker validation.[1] And, a clear understanding of the molecular effects of candidate variants, selected from significant associations between genotype and clinical phenotypes, is often absent. In fact, none of the previous pharmacogenetic studies of oral VEGF-pathway inhibitors used for the treatment of mRCC have conducted studies to validate their observed clinical associations.[2-8]

Replication of positive findings in an external, independent cohort of patients has traditionally served as a gold standard for validation of genotype-phenotype relationships.[9,10] However, validation of clinical variant-phenotype associations through replication can be problematic because independent, external cohorts of patients that match the discovery cohort (e.g. baseline patient characteristics, clinical presentation, and pharmacotherapies used in the treatment of their cancers) are not always easily obtainable, and do not provide insights into the mechanisms underlying disease pathology and/or drug response. Currently, a suitable replication cohort of mRCC patients treated with sorafenib is not available. Therefore, functional validation of these clinical associations becomes even more important to the overall goal of identifying and validating clinically useful biomarkers to aid clinicians in the treatment of mRCC.

The primary objective of this Aim was to validate the functionality of germline variants, identified in Aim 1, through a sequential *in silico* → *in vitro* approach, and to test the hypothesis that functional validation will elucidate the molecular effects of these variants on mRCC pathogenesis, angiogenesis and/or sorafenib pharmacology.

3.3 Materials and Methods

3.3.1 In Silico Functional Prediction

In silico bioinformatic analyses will help to predict the functionality of the significant variants identified in Aim 1, and will help prioritize which variants will be pursued for *in vitro* functional validation. Genotype data from Caucasian patients of the 1,000 Genomes Project (http://www.1000genomes.org) was downloaded and visualized in HaploView to investigate LD and haplotype structure.[11] The Broad Institute's HaploReg version 2 (http://www.broadinstitute.org/mammals/haploreg/haploreg_v2.php) and version 3 (http://www.broadinstitute.org/mammals/haploreg/haploreg_v3.php), the dbSNP website (http://www.ncbi.nlm.nih.gov/projects/SNP), and the Ensembl website (www.ensembl.org) were also be used to assess LD for variants of interest and to help analyze haplotype structure.[12,13] To help prioritize variants of interest for *in vitro* validation, only those in high LD ($r^2 > 0.8$; based on Caucasians from the 1000 Genomes Project) with a significant variant identified in Aim 1 were considered relevant to the haplotype structure.

Additionally, data from the Encyclopedia of DNA Elements (ENCODE) Consortium was extensively queried, through the use of online resources (e.g. the HaploReg, the University of California Santa Cruz [UCSC] Genome Browser [https://genome.ucsc.edu], and the Center for Genomics and Personalized Medicine at Stanford University's Regulome

DB version 1.1 website [http://regulomedb.org]).[14-16] Output from these *in silico* bioinformatic tools that provided evidence of variant functionality helped prioritize candidate variants for *in vitro* validation based on the potential of the variant to alter regulatory pathways and/or gene function. Variants from Aim 1 that were associated with OS, and additional variants in high LD with the variants identified from Aim 1, that had predicted evidence of histone modifications (promoter or enhancer histone marks), open chromatin (DNaseI hypersensitivity peaks), or changes to transcription factor binding motifs were prioritized for *in vitro* validation in the laboratory.

Finally, the SIFT and PolyPhen-2 algorithms were used to better understand the effects of amino acids substitutions caused by non-synonymous variants in gene coding regions.[17,18]

3.3.2 Laboratory Functional Validation

Cell Culture

HEK293 (ATCC) cells were maintained-in DMEM with 10% FBS and 1% penicillin/streptomycin, and Caki-1 human clear-cell metastatic renal cell carcinoma cells were cultured in McCoy's 5A (Iwakata and Grace Mod.) with L-glutamine media (ATCC, Manassas, VA, USA) with 10% fetal bovine serum (FBS) (Thermo Fisher Scientific, Waltham, MA, USA) and 1% and penicillin/streptomycin (Mediatech, Manassas, VA, USA). Human telomerase-immortalized microvascular endothelial (TIME) cells and human liver parenchyma endothelial cells (LPEC) were cultured in endothelial basal media (EBM-2) with 5% FBS and supplemented with the EGM®-2 MV bullet kit (Lonza Inc., Allendale, NJ,

USA), per the manufacturer instructions. All cells were maintained in 5% CO_2 at 37 ^oC and grown to desired confluency (80-90%).

Dual Reporter Gene Luciferase Assays

Dual reporter gene luciferase assays were performed by inserting a pGL4.26 [luc2/minP/Hygro] plasmid (Promega, Madison, WI, USA), with a minimal promoter and *Firefly* luciferase gene into all cell lines that were assayed. This pGL4.26 plasmid was utilized in all reporter gene assays. The pGL4.26 plasmids with cloned sequences of interest were synthesized by GeneWiz (Research Triangle Park, NC, USA). Restriction sites were identified using CLC Sequence Viewer 6.6.2 (CLC Bio, Cambridge, MA), and mutagenesis primers were designed using Agilent Technology's QuikChange Primer Design program (http://www.genomics.agilent.com/primerDesignProgram) Site-directed mutagenesis was carried out using a QuikChange II XL Site-Directed Mutagenesis Kit (Agilent, Santa Clara, CA), per manufacturer instructions. Sanger-based DNA sequencing, using 3730xl Genetic Analyzers (Applied Biosystems) and performed at the UNC Mammalian Genotyping Core, was used to confirm site-directed mutagenesis.

Two *VEGFA* variants that were prioritized for functional validation (rs1885657 and rs3024987) were tested for their functional effects on the transcriptional activity of *VEGFA*. An additional two variants that are in perfect LD (r^2=1.0) with rs1885657 (rs58159269, and rs943070) were also selected. First, a 1,320 bp *VEGFA* fragment (chr6:43739302-43740622) was cloned upstream of a minimal promoter in pGL4.26 using KpnI and NheI for the 5' and 3' restriction sites, respectively. DNA clones containing the reference sequence with each reference allele (rs1885657 T, rs943070 C, and rs58159269 T) were generated. DNA clones

containing each variant allele (rs1885657 C, rs943070 G, and rs58159269 C) were all generated through site-directed mutagenesis. A separate DNA clone containing all three variant alleles (rs1885657 C/rs943070 G/rs58159269 C) was also generated through site-directed mutagenesis.

For rs3024987, a 463 bp *VEGFA* fragment (chr6:43741162-43741625) was cloned upstream of a minimal promoter in pGL4.26 using NheI and XhoI for the 5' and 3' restriction sites, respectively. DNA clones containing the reference sequence with the reference allele (rs304987 C) were generated. Again, allelic variation (rs3024987 T) was introduced through site-directed mutagenesis, and DNA clones containing the variant allele were produced.

Once cells were grown to desired confluency (80-90%), they were seeded in 24-well plates at a density of $5x10^4$ cells per well. Caki-1 cells lines were transfected using lipofectamine (Invitrogen), using the reporter gene plasmid construct of interest and a Renilla HSV-TK plasmid (Promega), in 24-well plates. TIME and LPEC cell lines were transfected in 24-well plates by TransIT®-2020 reagent (Mirus Bio LLC, Madison, WI, USA) and used a reporter gene plasmid of interest with a Renilla plasmid containing an SV40 promoter (Promega). Additionally, LPEC cells were treated with CombiMag (Oz Biosciences, San Diego, CA, USA) and were transfected under magnetoporation. All cells were lysed 40 h after transfection, and the luciferase assays were then conducted according to the manufacturer instructions (Promega). Firefly luciferase and Renilla were loaded into 96-well plates and read on a Beckman Coulter DTX 880 Multimode Detector (Beckman Coulter, Inc., Brea, CA, USA).

Each mutant of the three constructs for SNPs in the *VEGFA* haplotype (rs1885657, rs58159269, and rs943070), a "triple variant" construct that contained all three mutations, a reference construct, and an empty pGL4.26 plasmid were each transfected in four independent experiments, using triplicate wells. For rs3024987, a construct containing the mutant, a reference construct, and an empty pGL4.26 plasmid were each transfected in four independent experiments, using triplicate wells. Luciferase activity was defined as a ratio of Firefly to Renilla luciferase. Luciferase activity for each of the variant constructs was normalized to the luciferase activity of the empty pGL4.26 vector. For rs304987, differences in luciferase activity were tested between the variant and reference, using a two-sided t-test. For rs1885657, a one-way ANOVA was conducted, and then pairwise comparisons to the reference were conducted for each variant, and also for the "triple variant." A Dunnett's correction was used to correct for multiple tests for rs1885657. All calculations were performed and figures created using GraphPad Prism® version 5.03 (GraphPad Software, Inc., La Jolla, CA, USA) software. All statistical analyses were confirmed using SAS software, version 9.2 (SAS, Cary, NC, USA).

Cell Viability Assays

Sorafenib tosylate (LC Laboratories, Woburn, MA) was reconstituted in 100% DMSO (Thermo), and stock concentrations (0.5–100 mM) were prepared. Working concentrations were then prepared in DMEM before addition to cells. The final concentrations of sorafenib used in cell viability assays ranged from 0.5–30 μM. The final concentration of DMSO introduced to cells was 0.2% in all experimental wells. The VEGFR3 Thr494Ala amino acid substitution, caused by rs307826, was introduced into the

pCMV6-XL5 expression vector containing *FLT-4* cDNA (Origene Technologies, Rockville, MD, USA) through the use of the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA, USA), as per the manufacturer's instructions. Sanger-based sequencing was again performed at the UNC Mammalian Genotyping Core to confirm site-directed mutagenesis.

Once cells were grown to desired confluency (80-90%), they were seeded in 24-well plates at a density of $5x10^4$ cells per well. After 24 h, reference and variant vectors were transfected into HEK-293 cells using Lipofectamine® 2000 (Invitrogen). *In vivo*, VEGFC is the ligand that is required for VEGFR3 dimerization and activation. So, after a 24 h transfection, cells were stimulated with VEGFC diluted in DMEM to a final 200 ng/mL concentration, or with a matched volume of DMEM media for unstimulated cells, for 1 h prior to the administration of sorafenib. Twelve concentrations of sorafenib (final concentration: 0 μM, 0.5 μM, 1 μM, 2 μM, 5 μM, 6 μM, 8 μM, 9 μM, 10 μM, 15 μM, 18 μM, 20 μM, and 30 μM) were then added to cells, in triplicate, and incubated for 72 h. The concentrations of sorafenib selected for these assays overlap the estimated pharmacologically relevant concentration range for sorafenib (approximately 6-15 μM).[19,20] Triplicate wells were also treated with 100% DMSO (final concentration of 20% in DMEM media), and 1% DMSO (final concentration of 0.2% in DMEM media) as controls. After a 72 h sorafenib incubation, alamarBlue® cell viability reagent (Life Technologies, Grand Island, NY, USA) was added to cells at an amount equal to 10% of the culture volume and incubated at 37 ºC for 2 h.

Fluorescence, with excitation wavelength at 530-560nm and emission wavelength at 590 nm, was measured using a Beckman Coulter DTX 880 Multimode plate reader

(Beckman Coulter, Inc.), and relative fluorescent units (RFU) were reported. Using the RFU output, the following equation was used to calculate the percent viability for each sorafenib concentration: $((PV_{conc} - PV_{D20})/(PV_{D0.2} - PV_{D20}) * 100)$, where PV_{conc} is the percent viability of cells for a given concentration in experimental wells, PV_{D20} is the percent viability of cells in control wells treated with 20% DMSO, and $PV_{D0.2}$ is the percent viability of cells in control wells treated with 0.2% DMSO. Experiments were performed, on different days in duplicate, and an average percent viability was used for all analyses. For all analyses, sorafenib concentrations were log_{10} transformed and a four-parameter non-linear regression model was used to assess log_{10} concentration versus percent viability, and to obtain IC_{50} values. Two-sided t-tests were used to assess differences between IC_{50} values for VEGFC-stimulated mutant-transfected HEK cells and VEGFC-stimulated reference-transfected HEK cells, and between IC_{50} values for VEGFC-stimulated mutant-transfected HEK cells and unstimulated mutant-transfected HEK cells. All calculations were performed and figures created using GraphPad Prism® version 5.03, and all statistical analyses were again confirmed using SAS software, version 9.2.

3.4 Results

3.4.1 In silico predictions

To understand more about the variants identified and their LD structure, as well as to prioritize variants for *in vitro* validation based on their predicted functionality, a series of *in silico* analyses were conducted. Variants from Aim 1 with evidence of promoter or enhancer histone modifications (e.g. H3K4me1, H3K4me3 and H3K27ac marks), evidence of DNase sensitivity, evidence that they influence transcription factor binding through changes to

transcription factor binding motifs and/or evidence of being an expression quantitative trait locus (eQTL), were prioritized for *in vitro* functional validation.

For the *VEGFA* variants identified in Aim 1 (rs1885657 and rs3024987) and two variants in perfect LD with rs1885657 (rs58159269 and rs943070), there was substantial evidence that they exist in a regulatory region. In multiple cell lines, there was evidence of both promoter and enhancer histone marks (H3K4me1, H3K4me3 and H3K27ac marks) (Table 3.1 and Figure 3.1). Briefly, methylation or acetylation modifications, or marks, to histones can influence how accessible the chromatin is to transcription, thereby influencing gene expression. H3K4me1 histone mark is a mono-methylation of lysine 4 of the H3 histone protein and is associated with enhancers and DNA regions downstream of transcription start sites. An H3K4me3 histone mark indicates tri-methylation of the same lysine residue and is associated with promoters that are active or poised to be activated. And, a H3K27ac histone mark indicates acetylation of lysine 27 of H3, and like H3K4me1, a H3K27ac mark is also often indicative of a transcription enhancer.[21]

These four *VEGFA* variants also revealed evidence of altered transcription factor binding and/or altered transcription factor motifs, and areas of DNase hypersensitivity (Table 3.1 and Figure 3.1). Variants in active regulatory regions (e.g. promoters and enhancers) often reveal evidence of DNase hypersensitivity. Areas of DNase hypersensitivity indicate loss of condensed chromatin structure, which exposes DNA and makes it accessible for transcription.[21,22]

No renal endothelial cell lines were analyzed during the ENCODE project. This was a potential limitation to leveraging ENCODE resources because RCC is highly dependent on angiogenesis and host vascular endothelium. In addition, analyzing data from an endothelial

cell type was essential because sorafenib, in part, targets host vascular endothelium. Therefore, examining data from human umbilical vein endothelial cells (HUVEC) was important because it provided insight into regulatory elements within the cell type closest to renal endothelial cells. In HUVECs, there was evidence of H3K4me1, H3K4me3, and H3K27ac marks at rs1885657. For rs58159269, there was evidence of H3K4me3, and H3K27ac marks. For rs943070 and rs3024987, there was evidence of H3K27ac (Figure 3.2).

By Regulome DB, the variant predicted to have the most influence on regulation of *VEGFA* was rs59159269. It had a Regulome DB score of 2b (Table 3.2), which means there was evidence of altered transcription factor binding sites, any transcription factor motif, a DNase footprint, and a DNase peak. Variants rs1885657, rs943070, and rs3024987 all had a Regulome DB score of 4, which means there was evidence of a DNase peak and altered transcription factor binding. Collectively, there was sufficient ENCODE prediction data to prioritize these variants for functional validation.

There was only minimal evidence in ENCODE that rs3816375 in *ITGAV* and rs8047917 in *WWOX* were in regions that would affect gene regulation (Tables 3.3 and 3.4, respectively). There is no data to provide a Regulome DB score for rs3816375, and the Regulome DB score for rs8047917 was 6. Based on minimal ENCODE evidence in the UCSC Browser, HaploReg and Regulome DB, these two variants were not prioritized for functional validation.

Variant rs30726 was not predicted to be deleterious or cause a harmful amino acid substitution. While a threonine to alanine amino acid substitution results in a change from a medium-sized and polar amino acid to a small-sized and hydrophobic amino acid,[23] rs307826 received a score of 0.53 from SIFT, which means the algorithm predicted that the amino acid

substitution would be tolerated.[18] Similarly, rs307826 received a score of 0.005 from PolyPhen-2, which means the algorithm predicted that the amino acid substitution would be benign.[17]

3.4.2 Variant Effects on Cell Viability

To validate the functionality of rs307826 and to gain insight into whether this non-synonymous variant in a coding region of *FLT-4* has an effect on VEGFR3 signaling, vectors containing reference and mutant cDNA were transfected into HEK-293 cells and subsequently treated with sorafenib. Cell viability assays were conducted to generate IC_{50} values for VEGFC-stimulated and unstimulated reference cells, as well as VEGFC-stimulated and unstimulated mutant cells.

Cells were treated in triplicate wells with 12 concentrations of sorafenib (0–30 μM), and cell viability relative to untreated controls (0.2% DMSO only) was determined through the alamarBlue® cell viability assay. First, a statistically significant difference in sorafenib cytotoxicity was observed between reference-transfected and mutant-transfected HEK-293 cells that were not stimulated by VEGFC ($p<0.0001$). The IC_{50} for reference-transfected cells was 7.58 μM, while the IC_{50} for mutant-transfected cells was 15.45 μM (Figure 3.3). Second, a statistically significant difference in sorafenib cytotoxicity was observed between reference-transfected and mutant-transfected HEK-293 cells that were both stimulated by VEGFC prior to sorafenib administration ($p<0.0001$). The IC_{50} for reference-transfected cells was 2.02 μM, while the IC_{50} for mutant-transfected cells was 7.67 μM (Figure 3.3). Finally, a statistically significant difference in sorafenib cytotoxicity was observed between VEGF-stimulated and unstimulated mutant-transfected HEK-293 cells ($p<0.0001$) (Figure 3.3).

3.4.3 Variant Effects on Transcriptional Activity

To provide insight into the mechanisms underlying the variant-OS associations observed among intronic variants identified in Aim 1, variants were assessed for their function as potential effectors of regulatory activity using dual reporter gene luciferase assays. Genomic regions containing the reference sequence, or sequence with the variants (or those in perfect LD with the identified variant) were cloned into a pGL4.26 plasmid construct containing a minimal promoter and downstream of the *Firefly* luciferase gene. Allelic variation (rs1885657 C allele, rs58159269 C allele, rs943070 G allele, or rs3024987 T allele) were introduced by site-directed mutagenesis. These assays were performed to determine how each variant (or the construct with three variants) affected the transcriptional activity of the minimal promoter.

For the *VEGFA* variant that associated with shorter OS in patients treated with sorafenib (rs1885657), and the two variants in perfect LD (rs943070 and rs58159269), significant increases in luciferase activity in LPEC cells were observed individually for all three variants, as well as for the "triple variant." For rs1885657 (T>C), the C allele increased luciferase activity by an average of 48% ($p<0.05$). For rs58159269 (T>C), the C allele increased luciferase by an average of 70% ($p<0.001$). For rs943070 (C>G), the G allele increased luciferase by an average of 40% ($p<0.05$). And, for the triple variant, luciferase was increased by an average of 98% ($p<0.01$), when compared to the reference alleles (Figure 3.4).

In the TIME human endothelial cell line, increased luciferase activity was again observed. For rs1885657 (T>C), the C allele increased luciferase activity by an average of 57% ($p<0.001$). For rs58159269 (T>C), the C allele increased luciferase by an average of

80% ($p<0.001$). And, for the triple variant, luciferase was increased by an average of 99% ($p<0.001$), when compared to the reference alleles. Luciferase activity was not significantly different between rs943070 and the reference allele in TIME cells (Figure 3.4).

Finally, in the clear-cell mRCC Caki-1 cell line significant increases in luciferase activity were observed individually for rs1885657, rs58159269 and rs943070, as well as for a haplotype of all three SNPs. For rs1885657 (T>C), the C allele increased luciferase activity by an average of 30% ($p<0.05$). For rs58159269 (T>C), the C allele increased luciferase by an average of 56% ($p<0.001$). For rs943070 (C>G), the G allele increased luciferase by an average of 35% ($p<0.01$). And, for the triple variant, luciferase was increased by an average of 70% ($p<0.001$), when compared to reference (Figure 3.4).

For *VEGFA* rs3024987 (C>T), which was associated with shortened OS in a combined analysis of both study arms, significant increases in luciferase activity were observed in all three cell lines. In LPEC cells, the T allele increased luciferase activity by an average of 34% ($p=0.0032$), when compared to reference. In the TIME cells, the T allele increased luciferase activity by an average of 38% ($p=0.0002$). Finally, in Caki-1 cells, the T allele increased luciferase activity by an average of 32% ($p=0.0001$), when compared to reference (Figure 3.5).

3.5 Discussion

The primary objective of this Aim was to validate the functionality of germline variants, identified in Aim 1, through a sequential *in silico* → *in vitro* approach, and to test the hypothesis that functional validation will elucidate the molecular effects of these variants on mRCC pathogenesis, angiogenesis and/or sorafenib pharmacology. For candidate gene-

candidate variant studies, where a tagging SNP variant approach has been applied or where the functionality of the variant is unknown, relying on clinical associations severely limits the ability to translate candidate variants into clinically useful prognostic or predictive biomarkers.

The initial step towards predicting if the candidate variants identified in Aim 1 were functional was to understand the LD structure of the other variants associated with the candidate. Sequencing data from the 1,000 Genomes Project was important for identifying other variants in high LD with the candidate variant that could also be functional. After gaining an appreciation of the haplotype and LD structure of each candidate variant, *in silico* bioinformatic tools were leveraged to understand more about whether the identified variants were in regulatory regions.[24,25] Additional *in silico* resources, such as HaploReg, and RegulomeDB, that provided an interpretation of ENCODE data (e.g. histone modifications, transcription factor binding, eQTLs, and chromatin structure) helped predict the functionality of intronic variants and prioritized them for validation,[16,21,22] The *in silico* predictions for variant functionality were consistent between HaploReg and Regulome DB (except for rs3816375, for which there was no prediction information in Regulome DB). Ensembl Variant Effect Predictor (http://uswest.ensembl.org/info/docs/tools/vep/index.html?redirect=no) incorporated resources like SIFT[18] and PolyPhen-2[17] to help predict the effects of non-synonymous variants. Table 3.6 has been provided to summarize key results from Aims 1 and 2, including; the clinical association for each variant, *in silico* predictions concerning function as regulatory elements or as a deleterious amino acid substitution, and *in vitro* experiment results.

The *in silico* bioinformatic tools predicted that the four *VEGFA* variants identified in Aim 1 would be functional, and thus they were prioritized for laboratory validation. In Aim 1, patients in the sorafenib arm with the CC genotype for rs1885657achieved significantly worse OS and PFS. This association could be caused by differences in regulatory elements (e.g. enhancer activity that results in altered transcription factor binding) at rs1885657, which could result in increased *VEGFA* expression. In addition, analyses of combined treatment arms revealed that patients with at least one copy of the T allele for rs3024987 also achieved poorer OS. This intronic variant is also likely to alter transcription factor binding and increase *VEGFA* expression. In both instances, increased *VEGFA* expression could conceivably lead to increased angiogenic capability of mRCC.

Data from functional assays performed in in Aim 2 support the hypothesis that rs1885657 and rs3024987 are functionally active variants. Luciferase assay data validated the *in silico* predictions, revealing increased luciferase activity for rs1885657 rs943070, rs58159269, and rs3024987. Luciferase assays were chosen to help validate these variants because they test the transcriptional effects of genetic variants in potential regulatory regions, and provide an initial understanding of the mechanisms underlying the variant-OS associations observed in Aim 1.[25]

The luciferase assay data also provided evidence that these variants could all be functional and acting independently to affect *VEGFA* expression through multiple different transcriptional regulatory elements. Because a tagging variant approach was used for many of the variants selected for Golden Gate genotyping in Aim 1, the original hypothesis tested in these experiments was that rs1885657 might only be a proxy for another functional variant. However, these data have shown that rs1885657 and two other variants in perfect LD

(rs58159269 and rs943070) are all likely to be biologically relevant, and that the entire three-SNP haplotype is indeed functional. These data support a new hypothesis that combinational effects of multiple enhancer variants, in high LD with one another, will together increase *VEGFA* expression, and confer differences in OS to mRCC patients.[26] While previous pharmacogenetic studies of oral VEGF-pathway inhibitors have identified SNPs that associate with OS and/or PFS in mRCC patients, none have attempted to functionally validate of their findings.[5,27,28] Prior to these studies, rs1885657 and rs3024987 were of unknown significance. But, increased cellular transcription of *VEGFA* by the C allele of rs1885657, the C allele of rs58259269, the G allele of rs94307 and/or the T allele of rs3024987, combined with their respective associations with OS, align well with the hypothesis that increased *VEGFA* expression increase the angiogenic potential of mRCC.

In Aim 1, analyses when treatment arms were combined revealed that patients with the GG genotyped for rs307826 achieved significantly worse OS and PFS. The SNP variant rs307826 causes a threonine to alanine substitution at position 494 (Thr494Ala), located in the fifth IgG-like domain of VEGFR3. While the PolyPhen-2[17] and SIFT[18] algorithms, did not predict this variant to be deleterious and cause a damaging amino acid substitution, several pharmacogenetic studies of mRCC patients treated with oral VEGF-pathway inhibitors have associated rs307826 with significantly shorter OS and/or PFS.[2,3,8]

FLT-4 encodes for VEGFR3, a transmembrane kinase receptor that has been traditionally linked to lymphangiogenesis.[29] But, VEGFR3 is also expressed in tumor vasculature,[30] and inhibition of VEGFR3 can suppress vascular network formation.[31] Preclinical models have also suggested that VEGFR3 could possibly be more relevant than VEGFR2 (a primary driver of angiogenic signaling, and a main target for the VEGF-pathway

inhibitors used to treat mRCC) for the development of metastases.[32] One notable study recently characterized the structure and function of VEGFR3, and revealed that the fifth IgG-like receptor domain is critical to receptor signaling.[33] Cell viability assays, which showed the cells overexpressing the variant G allele, were more resistant to sorafenib. These data support the primary hypothesis that this amino acid substitution results in increased VEGFR3 signaling, resistance to sorafenib, and thus an increased risk of disease progression and death.

While more common for non-synonymous variants to result in amino acid substitutions that alter protein function, evidence supports the alternate hypothesis that this non-synonymous variant can also modify *FLT-4* expression.[34-36] Although the *in silico* predictive evidence is not as compelling as for the *VEGFA* variants, HaploReg does reveal evidence of altered enhancer marks and changes to transcription factor binding motifs (Table 4.5). Therefore, investigating rs307826 as an effector of gene expression should also be considered.

Although not prioritized for validation by the *in silico* resources that use ENCODE data, rs3816375 in *ITGAV* and rs8047917 in *WWOX* could still potentially be functional. As previously mentioned, ENCODE experiments were not performed in a kidney endothelial cell line, nor in an mRCC line. These two variants were significantly associated with OS after a rigorous statistical design, which sought to avoid Type I error and detection of spurious variant-phenotype associations, was employed in Aim 1. Therefore, it would be premature to dismiss these two variants completely. For the two intergenic variants that associated with OS (rs6719561 and rs200809375), validation is still required. Luciferase assays could potentially provide insight into their activity as enhancers; however, because they are not located within genes it would be difficult to ascertain on which gene they act. In

addition, the physiologic relevance on a clinical phenotype of variants in intergenic regions could be directly related to currently unknown non-coding RNAs, (e.g. microRNAs, or macro long non-coding RNAs).[37] Certainly, additional future validation experiments involving these two variants should be conducted.

Functional evidence by one validation method should not be relied upon because it is rarely sufficient to provide a mechanistic foundation to support clinical associations.[25,38] Additional laboratory validation experiments should be conducted on the variants identified in Aim 1 to ascertain if they influence gene expression, and to characterize the mechanistic underpinnings that result in altered gene expression. For example, variants with evidence of luciferase activity should be prioritized for additional *in vitro* functional validation by electromobility shift assays (EMSA). And, phosphorylation assays assessed through traditional Western blotting, enzyme-linked immunosorbent assays (ELISA), or by leveraging a mass spectrometry-based proteomic approach could provide necessary confirmation that rs307826 influences VEGFR3 signaling.

Ultimately, variant constructs used in reporter gene assays are analyzed out of their natural context, and state. And, for cell viability and/or phosphorylation assays, cDNA vectors likely overexpress the gene more than what occurs *in vivo*.[25] Therefore, the creation of isogenic endothelial cell lines, using clustered regularly interspaced short palindromic repeats (CRISPR)/cas9 or transcription activator-like effector nuclease (TALEN) technologies, could for the first time isolate how these variants influence gene expression, receptor signaling, even angiogenesis.

In summary, the findings from Aim 2 provide evidence that the variants identified in Aim 1, which significantly associated with OS in TARGET patients treated with sorafenib,

are functional. These data lay provide validation of these germline variants, and also lay the groundwork for future studies that will further elucidate the molecular mechanisms underlying these variants. These data provide the next essential step towards conducting prospective trials that will then confirm them as predictive and prognostic biomarkers.

TABLES

Table 3.1. HaploReg output for rs1885657 and rs3024987. *VEGFA* variants in high LD ($r^2 \geq 0.8$) with rs1885657 are shown. Additionally, rs3024987, which is in moderate LD with rs1885657 ($r^2 \geq 0.6$), is included. The Regulome DB score for rs1885657 is 4, for rs58159269 is 2b, for rs943070 is 4, and for rs3024987 is 4. In TARGET, the CC genotype for rs1885657 was associated with significantly shorter OS in patients from the sorafenib arm ($p=1.39 \times 10^{-04}$, $q=0.076$). The CT genotype was associated with significantly shorter OS in a combined analysis of both patient arms ($p=8.80 \times 10^{-04}$, $q=0.088$). Abbreviations: LD, linkage disequilibrium; SNP, single nucleotide polymorphism.

Position	LD (r^2)	Variant	Ref	Alt	EUR freq	Promoter histone marks	Enhancer histone marks	DNAse	Proteins bound	Motifs changed	GENCODE genes	dbSNP functional annotation
chr6:43739446	1	rs58159269	T	C	0.20	24 organs	6 organs	30 organs	13 bound proteins	14 altered motifs	*VEGFA*	intronic
chr6:43740094	1	rs1885657	T	C	0.20	22 organs	12 organs	18 organs	4 bound proteins	5 altered motifs	*VEGFA*	intronic
chr6:43740451	0.99	rs943070	C	G	0.20	21 organs	13 organs	28 organs	7 bound proteins	ATF3, BDP1	*VEGFA*	intronic
chr6:43740840	0.61	rs3024987	C	T	0.13	19 organs	15 organs	GI, MUS		5 altered motifs	*VEGFA*	intronic

Table 3.2. Regulome DB scoring system. This table is adapted from one found on the Regulome DB website (http://regulomedb.org/). Abbreviations: eQTL, expression quantitative trait locus/loci; TF, transcription factor.

Score	Supporting Data
1a	eQTL + TF binding + matched TF motif + matched DNase Footprint + DNase peak
1b	eQTL + TF binding + any motif + DNase Footprint + DNase peak
1c	eQTL + TF binding + matched TF motif + DNase peak
1d	eQTL + TF binding + any motif + DNase peak
1e	eQTL + TF binding + matched TF motif
1f	eQTL + TF binding / DNase peak
2a	TF binding + matched TF motif + matched DNase Footprint + DNase peak
2b	TF binding + any motif + DNase Footprint + DNase peak
2c	TF binding + matched TF motif + DNase peak
3a	TF binding + any motif + DNase peak
3b	TF binding + matched TF motif
4	TF binding + DNase peak
5	TF binding or DNase peak
6	Other

Table 3.3. HaploReg output for rs3816375. *ITGAV* variants in high LD with rs3816375 ($r^2 \geq 0.8$) are included. There was no data in Regulome DB to provide a score for rs3816375. In TARGET, the GG genotype for rs3816375 was associated with significantly shorter OS in patients from the sorafenib arm ($p=4.87 \times 10^{-04}$, q=0.051). Abbreviations: LD, linkage disequilibrium; SNP, single nucleotide polymorphism.

Position	LD (r^2)	Variant	Ref	Alt	EUR freq	Promoter histone marks	Enhancer histone marks	DNAse	Proteins bound	Motifs changed	GENCODE genes	dbSNP functional annotation
chr2:187503226	1	rs4667108	T	C	0.40		BLD, SKIN			Cdx, TATA	*ITGAV*	intronic
chr2:187505486	1	rs3816375	A	G	0.40		BLD, BONE	LNG		4 altered motifs	*ITGAV*	intronic
chr2:187513708	0.87	rs202185248	23-mer	T	0.58					Mef2	*ITGAV*	intronic
chr2:187517655	0.91	rs4667109	T	C	0.39					CEBPG, Foxp1, PLZF	*ITGAV*	intronic
chr2:187532010	0.92	rs41265951	G	T	0.39					TATA	*ITGAV*	intronic
chr2:187533741	0.93	rs2290083	T	C	0.39					5 altered motifs	*ITGAV*	intronic

Table 3.4. HaploReg output for rs8047917. *WWOX* variants in high LD with rs8047917 ($r^2 \geq 0.8$) are included. Regulome DB score for rs8047917 is 5, for rs77533819 is 6, and for rs7190335 is 6. There is evidence of siPhycons for rs7190335. In TARGET, the TA genotype for rs8047917 was associated with significantly shorter OS in patients from the sorafenib arm ($p=3.27 \times 10^{-04}$, q=0.076). Abbreviations: LD, linkage disequilibrium; SNP, single nucleotide polymorphism.

Position	LD (r^2)	Variant	Ref	Alt	EUR freq	Promoter histone marks	Enhancer histone marks	DNAse	Proteins bound	Motifs changed	GENCODE genes	dbSNP functional annotation
chr16:78184378	0.83	rs8052567	C	G	0.07					Ets	*WWOX*	intronic
chr16:78187380	0.98	rs16947192	T	C	0.07					ATF3	*WWOX*	intronic
chr16:78188868	1	rs77533819	T	C	0.07		4 organs			5 altered motifs	*WWOX*	intronic
chr16:78189414	1	rs8047917	T	A	0.07		6 organs	LNG		E4F1,HMG-IY, XBP-1	*WWOX*	intronic
chr16:78190218	0.96	rs7190335	T	C	0.07		BRST, BRN, SKIN	SKIN, LNG		5 altered motifs	*WWOX*	intronic

Table 3.5. HaploReg output for rs307826. There are no variants in high LD with rs307826 ($r^2 \geq 0.8$); however, this SNP is in low LD ($r^2=0.38$) with rs307821, which is another non-synonymous variant previously associated with survival in patients treated with pazopanib and sunitinib. There is evidence of SiPhycons with rs307826 (not shown in table). The Regulome DB score for rs307826 is 5. In TARGET, the GG genotype was associated with significantly shorter OS in a combined analysis of both patient arms ($p=1.24 \times 10^{-04}$, q=0.088). Abbreviations: LD, linkage disequilibrium; SNP, single nucleotide polymorphism.

Position	LD (r^2)	variant	Ref	Alt	EUR freq	Promoter histone marks	Enhancer histone marks	DNAse	Proteins bound	Motifs changed	GENCODE genes	dbSNP functional annotation
chr5:180051003	1	rs307826	A	G	0.11		10 organs			Egr-1 ,Irf, NRSF	*FLT-4*	missense

Table 3.6. Summary of key results. For the seven variants identified in Aim 1, this table summarizes the key results from Aims 1-2: the clinical association, *in silico* predictions, and *in vitro* validation results. Abbreviations: HEK, human embryonic kidney; IC_{50}, half maximal inhibitory concentration.

Variant	**Gene**	**Clinical Association**	**HaploReg Evidence**	**Regulome DB Score**	**SIFT/ PolyPhen-2**	***In Vitro* Validation Results**
rs307826	*FLT-4*	GG associated with decreased OS in **combined analysis:** • $p=1.24x10^{-04}$ • q=0.088	Moderate: • Enhancer histone marks in 10 organs • 3 altered motifs	5	SIFT: • 0.53 • Tolerated PolyPhen-2: • 0.005 • Benign	Resistant IC_{50} in HEK cells transfected with rs307826 and treated with sorafenib
rs1885657	*VEGFA*	CC associated with decreased OS in **sorafenib arm:** • $p=1.39x10^{-04}$ • q=0.076	High: • Promoter histone marks in 22 organs • Enhancer histone marks in 12 organs • DNase hypersensitivity in 18 organs • 4 proteins bound • 5 altered motifs	4	N/A	**Increased** luciferase activity (whole element and variant allele) in: • Caki-1 • LPEC • TIME
rs3024987	*VEGFA*	CT associated with decreased OS in **combined analysis:** • $p=8.80X10^{-04}$ • q=0.088	High: • Promoter histone marks in 19 organs, • Enhancer histone marks in 15 organs • DNase hypersensitivity in 2 organs • 5 altered motifs	4	N/A	**Increased** luciferase activity (variant allele) in: • Caki-1 • LPEC • TIME
rs3816375	*ITGAV*	GG associated with decreased OS in **sorafenib arm:** • $p=4.87x10^{-04}$ • q=0.051	Moderate: • Enhancer histone marks in 2 organs • DNase hypersensitivity in 1 organ • 4 altered motifs	N/A	N/A	N/A
rs8047917	*WWOX*	TA associated with decreased OS in **sorafenib arm:** • $p=3.27x10^{-04}$ • q=0.076	Moderate: • Enhancer histone marks in 6 organs • DNase hypersensitivity in 1 organ • 3 altered motifs	5	N/A	N/A

rs200809375	3' of *NRP-1*	ATG associated with decreased OS **in sorafenib arm:** • $p=2.65x10^{-04}$ • q=0.076	Minimal: • 2 altered motifs	6	N/A	N/A
rs6719561	3' of *UGT1A9*	Associated with OS in **sorafenib arm:** • $p=3.27x10^{-04}$ • q=0.076	Minimal: • 2 altered motifs	5	N/A	N/A

FIGURES

Figure 3.1. UCSC Browser's ENCODE output for *VEGFA* variants. ENCODE output from the University of Santa Cruz (UCSC) Genome Browser for variants rs1885657, rs3024987, rs58159269 and rs943070 is represented in this figure under a region on chromosome 6 (43,739,300-43,741,000). **A)** Regions where there is evidence of histone modifications are shown here (H3K4me1, H3K3me3, and H3K27ac marks). Individual colors represent evidence in individual cell lines. **B)** Evidence of DNase hypersensitivity is shown here. Black represents a region with more hypersensitivity. **C)** ChIP-Seq evidence of altered binding motifs for different transcription factors. Black represents evidence of altered transcription factor binding, while the lighter the shade of grey represents fewer cell lines that demonstrated altered binding for an given transcription factor. Abbreviations: ChIP-Seq, chromatin immunoprecipitation-sequencing.

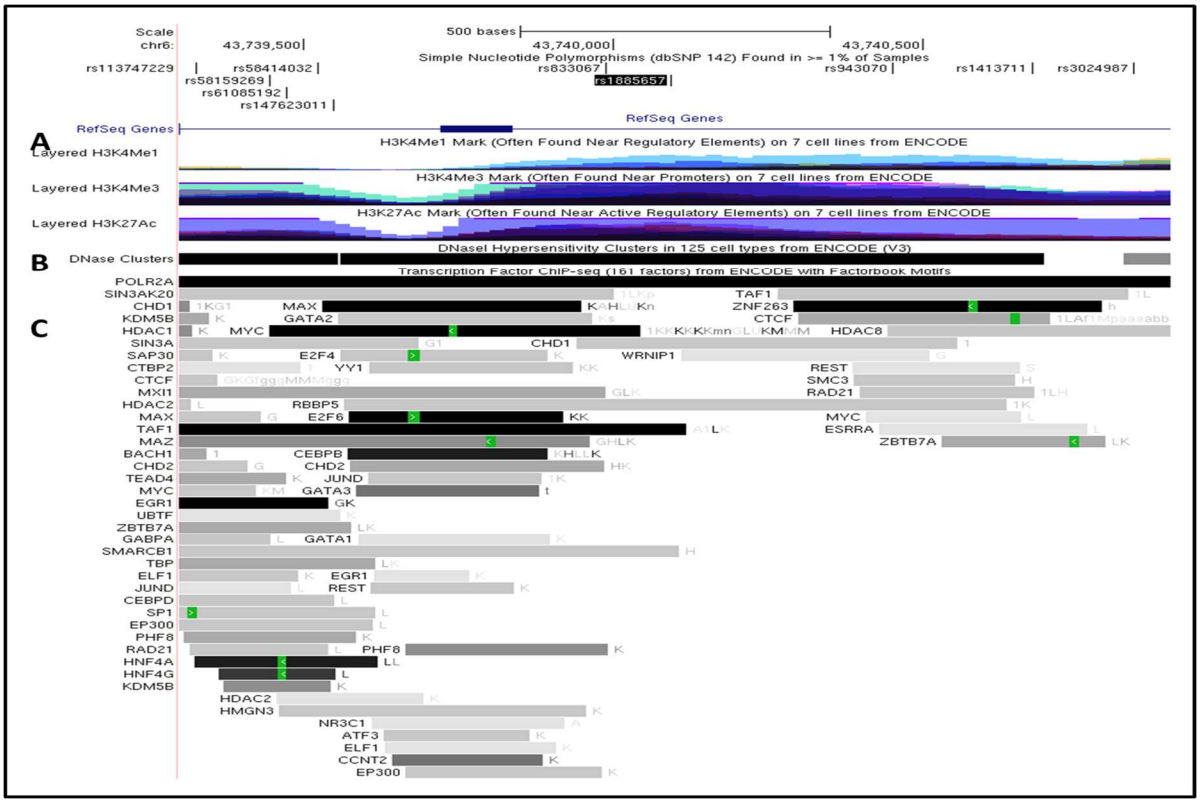


Figure 3.2. Histone marks in HUVEC cells for *VEGFA* variants. ENCODE output from the UCSC Genome Browser for variants rs1885657, rs3024987, rs58159269 and rs943070 is represented in this figure under a region on chromosome 6 (43,739,300-43,741,000). Regulatory regions where marks are present that are evident of histone modifications (H3K4me1, H3K3me3, and H3K27ac marks) are shown for HUVEC cells. Abbreviations: HUVEC, human umbilical vein endothelial cells; UCSC, University of Santa Cruz.

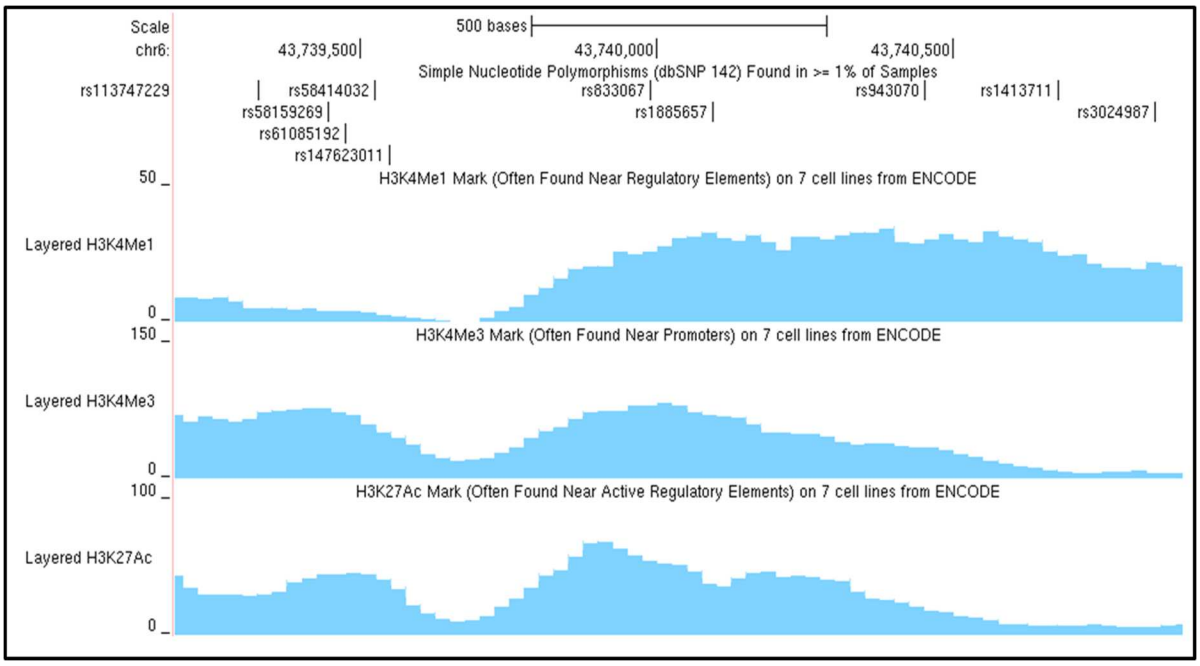


Figure 3.3. Effects of rs307826 on cell viability in reference-transfected, and mutant-transfected HEK-293 cells. Transfected cells were treated 12 concentrations of sorafenib (0–30 μM) and then with alamarBlue® to assess cell viability and generate IC_{50} values. Stimulated cells were treated with VEGFC (200 ng/mL) prior to sorafenib administration. Relative fluorescence units (RFUs) were generated, and RFUs were used to determine the percent of viable cells present for a given concentration. Concentrations were compared to untreated controls with 0.2% DMSO. The IC_{50} for unstimulated reference-transfected cells was 7.58, while the IC_{50} for VEGFC-stimulated reference-transfected cells was 2.02. The IC_{50} for unstimulated mutant-transfected cells was 15.4, while the IC_{50} for VEGFC-stimulated mutant-transfected cells was 7.67. For the purpose of this figure, reference constructs with the A allele were referred to as reference, and mutant constructs with the rs307826 G allele were referred to as Mut.

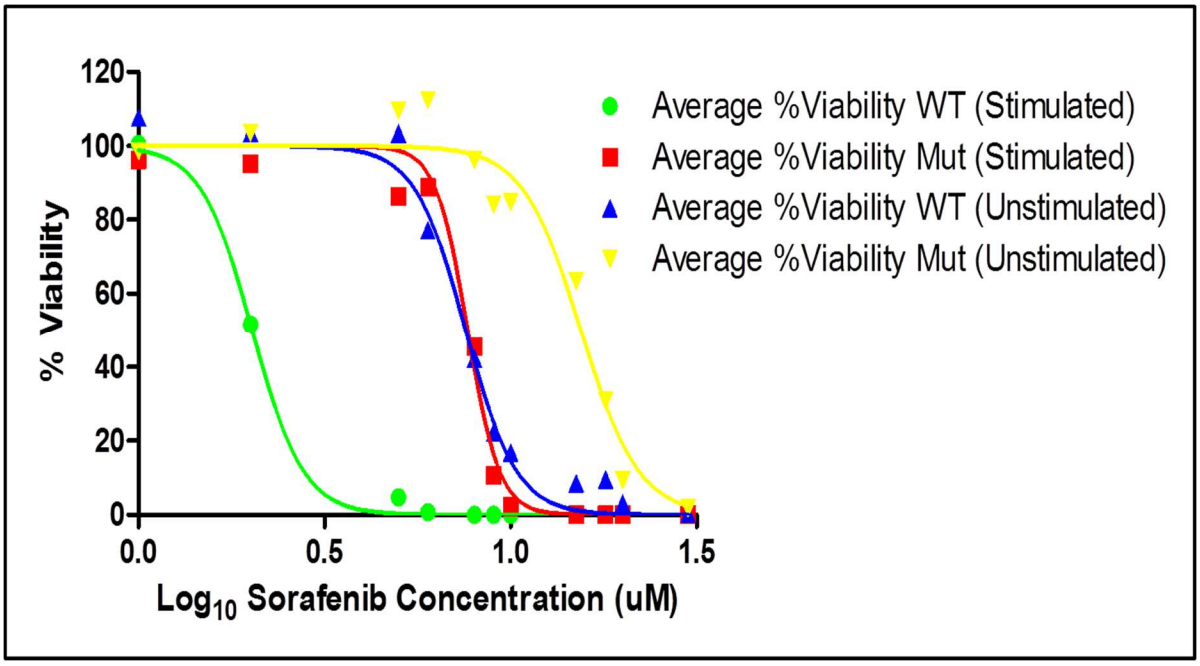


Figure 3.4. Dual reporter gene assay results for luciferase activity of SNPs in *VEGFA*. Relative luciferase activity of SNPs is represented in Caki-1, TIME and LPEC cell lines (from left to right). SNP luciferase activity is normalized to empty pGL4.26 plasmid luciferase activity. The mean ± SEM of the transfection experiments in quadruplicate is shown. In this figure, WT refers to the reference allele. Abbreviations: LPEC, liver parenchyma endothelial cells; SEM, standard error of the mean ; SNP, single nucleotide polymorphism; TIME, telomerase-immortalized microvascular endothelial; *VEGFA*, vascular endothelial growth factor A; WT, wild type. *p≤0.05, **p≤0.01, ***p≤0.001.

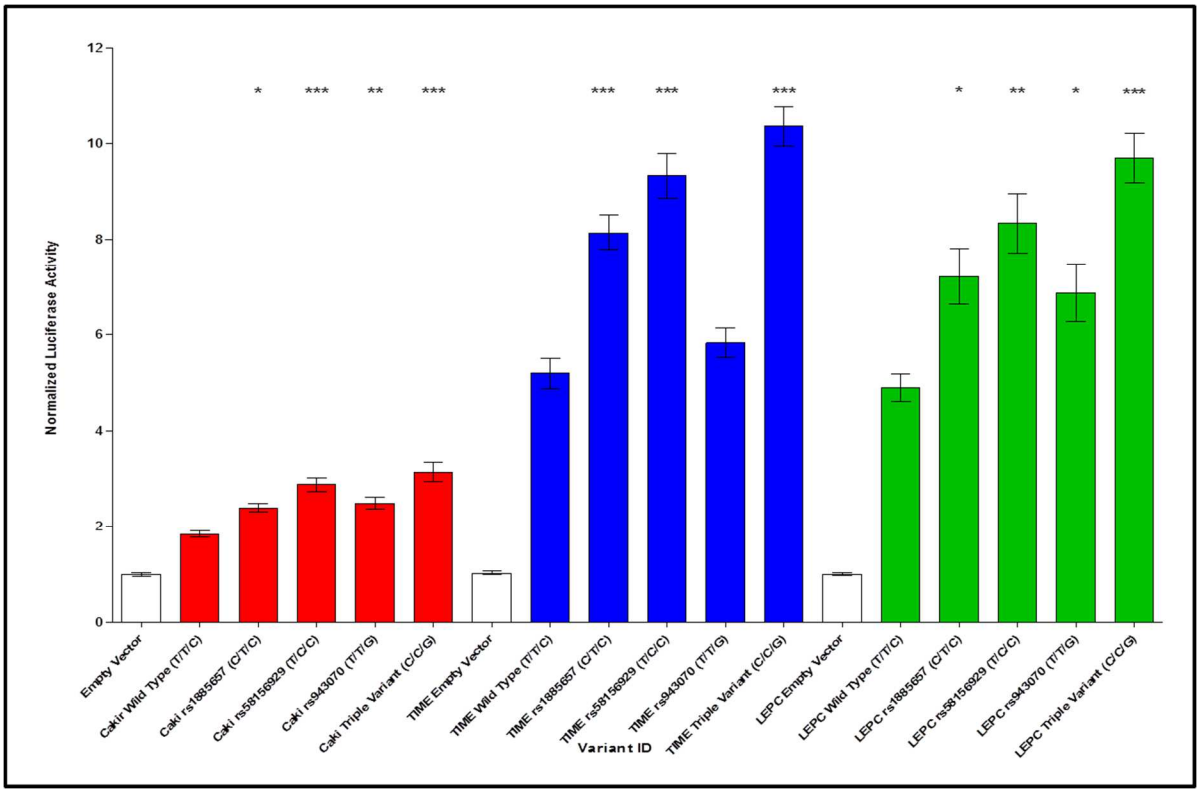
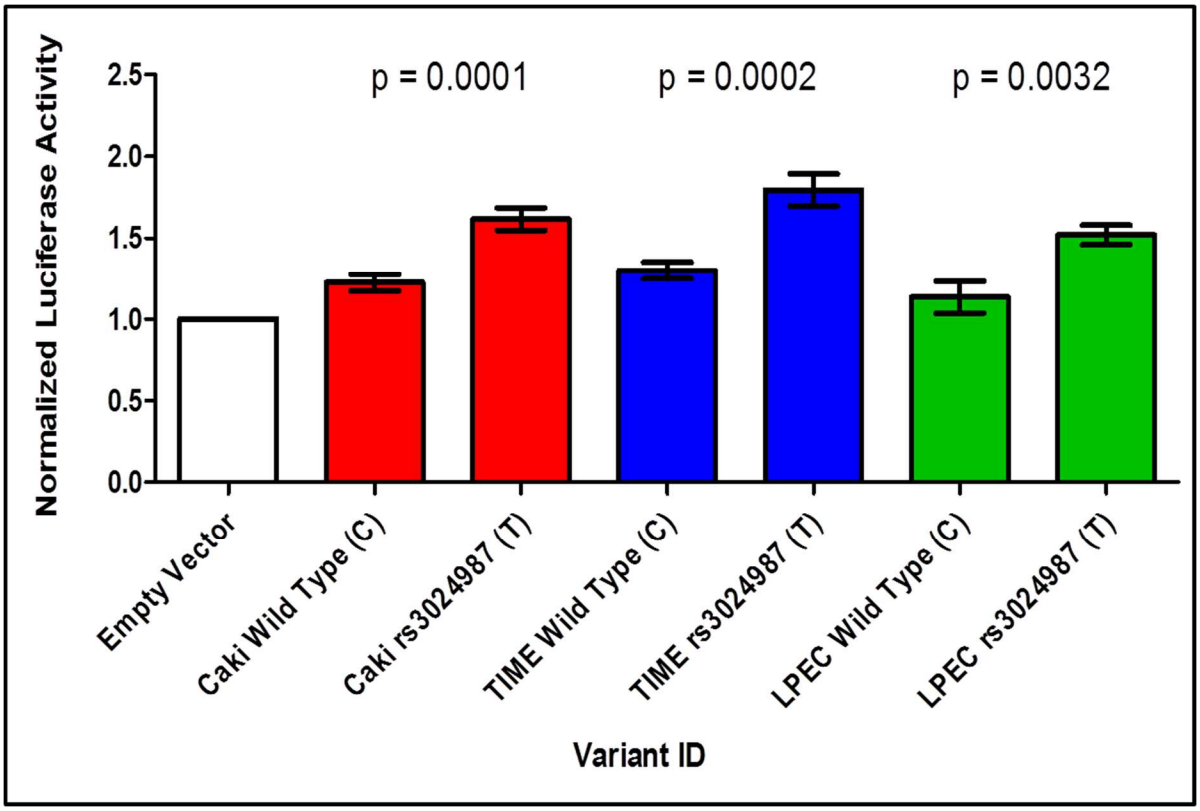


Figure 3.5. Dual reporter gene assay results for luciferase activity of rs3024987. Relative luciferase activity of SNPs is represented in Caki-1, TIME and LPEC cell lines (from left to right). SNP luciferase activity is normalized to empty pGL4.26 plasmid luciferase activity. The mean ± SEM of the transfection experiments in quadruplicate is shown. Abbreviations: LPEC, liver parenchyma endothelial cells; SEM, standard error of the mean ; SNP, single nucleotide polymorphism; TIME, telomerase-immortalized microvascular endothelial; *VEGFA*, vascular endothelial growth factor A; WT, wild type.



REFERENCES

1. Glubb DM, Cerri E, Giese A, et al. Novel functional germline variants in the VEGF receptor 2 gene and their effect on gene expression and microvessel density in lung cancer. *Clin Cancer Res*. 2011;17:5257-67.

2. Garcia-Donas J, Esteban E, Leandro-Garcia LJ, et al. Single nucleotide polymorphism associations with response and toxic effects in patients with advanced renal-cell carcinoma treated with first-line sunitinib: a multicentre, observational, prospective study. *Lancet Oncol*. 2011;12:1143-50.

3. Beuselinck B, Karadimou A, Lambrechts D, et al. Single-nucleotide polymorphisms associated with outcome in metastatic renal cell carcinoma treated with sunitinib. *Br J Cancer*. 2013;108:887-900.

4. Peer CJ, Sissung TM, Kim A, et al. Sorafenib is an inhibitor of UGT1A1 but is metabolized by UGT1A9: implications of genetic variants on pharmacokinetics and hyperbilirubinemia. *Clin Cancer Res*. 2012;18:2099-107.

5. Scartozzi M, Bianconi M, Faloppi L, et al. VEGF and VEGFR polymorphisms affect clinical outcome in advanced renal cell carcinoma patients receiving first-line sunitinib. *Br J Cancer*. 2013;108:1126-32.

6. van der Veldt AA, Eechoute K, Gelderblom H, et al. Genetic polymorphisms associated with a prolonged progression-free survival in patients with metastatic renal cell cancer treated with sunitinib. *Clin Cancer Res*. 2011;17:620-9.

7. van Erp NP, Eechoute K, van der Veldt AA, et al. Pharmacogenetic pathway analysis for determination of sunitinib-induced toxicity. *J Clin Oncol*. 2009;27:4406-12.

8. Xu CF BH, Bing N, Sternberg C, Xue Z, McCann L, King K, Spraggs C, Mooser V, Pandite LN. Association of genetic markers in angiogenesis- or exposure-related genes with overall survival in pazopanib-treated patients with advanced renal cell carcinoma. *J Clin Oncol*. 2011;29(Supp 7):abstract 303.

9. Chanock SJ, Manolio T, Boehnke M, et al. Replicating genotype-phenotype associations. *Nature*. 2007;447:655-60.

10. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat Genet*. 2001;29:306-9.

11. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005;21:263-5.

12. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res*. 2012;40:D930-4.

13. Cunningham F, Amode MR, Barrell D, et al. Ensembl 2015. *Nucleic Acids Res*. 2015;43:D662-9.

14. Rosenbloom KR, Armstrong J, Barber GP, et al. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res*. 2015;43:D670-81.

15. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22:1790-7.

16. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*.2004;306:636-40.

17. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7:248-9.

18. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4:1073-81.

19. Wilhelm SM, Adnane L, Newell P, Villanueva A, Llovet JM, Lynch M. Preclinical overview of sorafenib, a multikinase inhibitor that targets both Raf and VEGF and PDGF receptor tyrosine kinase signaling. *Mol Cancer Ther*. 2008;7:3129-40.

20. Rahmani M, Nguyen TK, Dent P, Grant S. The multikinase inhibitor sorafenib induces apoptosis in highly imatinib mesylate-resistant bcr/abl+ human leukemia cells in association with signal transducer and activator of transcription 5 inhibition and myeloid cell leukemia-1 down-regulation. *Mol Pharmacol*. 2007;72:788-95.

21. Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? *Mol Cell*. 2013;49:825-37.

22. Heintzman ND, Stuart RK, Hon G, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet*. 2007;39:311-8.

23. Artimo P, Jonnalagedda M, Arnold K, et al. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res*. 2012;40:W597-603.

24. Glubb DM, Paugh SW, van Schaik RH, Innocenti F. A guide to the current Web-based resources in pharmacogenomics. *Methods Mol Bio*l. 2013;1015:293-310.

25. Glubb DM, Innocenti F. Architecture of pharmacogenomic associations: structures with functional foundations or castles made of sand? *Pharmacogenomics*. 2013;14:1-4.

26. Corradin O, Saiakhova A, Akhtar-Zaidi B, et al. Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res*. 2014;24:1-13.

27. Xu CF, Bing NX, Ball HA, et al. Pazopanib efficacy in renal cell carcinoma: evidence for predictive genetic markers in angiogenesis-related and exposure-related genes. *J Clin Oncol*. 2011;29:2557-64.

28. Kim JJ, Vaziri SA, Rini BI, et al. Association of VEGF and VEGFR2 single nucleotide polymorphisms with hypertension and clinical outcome in metastatic clear cell renal cell carcinoma patients treated with sunitinib. *Cancer*. 2012;118:1946-54.

29. Kaipainen A, Korhonen J, Mustonen T, et al. Expression of the fms-like tyrosine kinase 4 gene becomes restricted to lymphatic endothelium during development. *Proc Natl Acad Sci USA*. 1995;92:3566-70.

30. Smith NR, Baker D, James NH, et al. Vascular endothelial growth factor receptors VEGFR-2 and VEGFR-3 are localized primarily to the vasculature in human primary solid cancers. *Clin Cancer Res*. 2010;16:3548-61.

31. Tammela T, Zarkada G, Wallgard E, et al. Blocking VEGFR-3 suppresses angiogenic sprouting and vascular network formation. *Nature*. 2008;454:656-60.

32. Roberts N, Kloos B, Cassella M, et al. Inhibition of VEGFR-3 activation with the antagonistic antibody more potently suppresses lymph node and distant metastases than inactivation of VEGFR-2. *Cancer Res*. 2006;66:2650-7.

33. Leppanen VM, Tvorogov D, Kisko K, et al. Structural and mechanistic insights into VEGF receptor 3 ligand binding and activation. *Proc Natl Acad Sci USA*. 2013;110:12960-5.

34. Zhang W, He T, Wang Q, et al. Interleukin-1 receptor-associated kinase-2 genetic variant rs708035 increases NF-kappaB activity through promoting TRAF6 ubiquitination. *J Biol Chem*. 2014;289:12507-19.

35. Duellman T, Warren C, Yang J. Single nucleotide polymorphism-specific regulation of matrix metalloproteinase-9 by multiple miRNAs targeting the coding exon. *Nucleic Acids Res*. 2014;42:5518-31.

36. Reamon-Buettner SM, Sattlegger E, Ciribilli Y, Inga A, Wessel A, Borlak J. Transcriptional defect of an inherited NKX2-5 haplotype comprising a SNP, a nonsynonymous and a synonymous mutation, associated with human congenital heart disease. *PLoS One*. 2013;8:e83295.

37. Kohtz JD, Berghoff EG. Regulatory long non-coding RNAs and neuronal disorders. *Physiol Behav*. 2010;100:250-4.

38. Wang D, Chen H, Momary KM, Cavallari LH, Johnson JA, Sadee W. Regulatory polymorphism in vitamin K epoxide reductase complex subunit 1 (VKORC1) affects gene expression and warfarin dose requirement. *Blood*. 2008;112:1013-21.

CHAPTER 4: IDENTIFYING GENETIC MARKERS FOR CYTOTOXIC RESPONSE TO SORAFENIB IN MOUSE EMBRYONIC FIBROBLAST CELLS

4.1 Overview

Background:

Sorafenib is an oral multikinase inhibitor that decreases tumor angiogenesis and proliferation. The antitumor efficacy and toxicity profiles of sorafenib vary among patients. Novel pathways and targets of sorafenib activity remain to be identified, and no predictive biomarkers of sorafenib activity exist to help guide clinicians. This aim sought to identify novel genes associated with sorafenib activity by using an *in vitro* methodology based upon mouse genomics and high-throughput screening of multiple cell health parameter phenotypes.

Methods:

Primary mouse embryonic fibroblasts (MEFs) from 32 inbred strains were profiled for sorafenib cytotoxicity utilizing high content imaging and simultaneous evaluation of cell health parameters. The 32 strains were genomically characterized previously. MEFs were treated with ten concentrations of sorafenib (0–300 μM), incubated for 24 and 72 h, then fixed and stained. Cell viability, MEF cell membrane permeability, mitochondrial membrane potential, and cytochrome C release were assessed (Table 4.1). Image analysis software assessed the effects of sorafenib on each phenotype. Dose response curves were generated from data using a Brain-Cousens model, and EC_{50} or IC_{50} values for each strain were

identified. Genome-wide association (GWA) mapping, using the efficient mixed-model associations (EMMA) and SNPster algorithms, was performed to identify quantitative trait loci (QTLs) associated with sorafenib activity. Approximately 277,000 single nucleotide polymorphisms were tested, and genomic loci discovered by EMMA, with p-values $<1.0x10^{-7}$, were identified as potential candidate genes. These genes were then examined using stringent, multi-faceted criteria before being selected for future laboratory validation.

Results:

Interstrain IC_{50} variability among the 32 MEF strains was observed after 72 h sorafenib incubations (17.2-44.5 μM). One peak (chromosome 9 from 51-52 Mb; $p=1.0x10^{-8}$), which reached genome-wide significance and significantly associated with cytochrome C release, was identified. From this peak, candidate genes that may underlie variability in sorafenib-induced cytochrome C release from mitochondria have been identified. A total of nine genes, expressed in MEF cells at mRNA level, are present in this QTL. Interstrain IC_{50} variability, which associated with VOC, was also observed after 72 h sorafenib incubations (17-32 μM). One peak potentially associated with cell viability (chromosome 4 from 119,500,000-120,750,000; $p=2.2x10^{-5}$). From this peak, candidate genes that may underlie variability in sorafenib cytotoxicity and cell viability have been identified. A total of 13 candidate genes, expressed in MEF cells at mRNA level, are present in this QTL.

Conclusions:

This innovative high content cellular genetics approach has detected robust interstrain cellular differences in sorafenib activity. One QTL region, which reached genome-wide significance and potentially associates with sorafenib-induced cytochrome C release from mitochondria, was identified. An additional QTL was identified that potentially associates with sorafenib cytotoxicity and cell viability. Candidate genes for functional validation have been prioritized through a multi-faceted set of criteria. Future steps for this work include functional validation of candidate genes, using knockdown and overexpression approaches, in MEF and human cell lines. Ultimately, variants in candidate genes that are successfully validated will be genotyped in TARGET patients and tested for associations with OS, PFS, and sorafenib-induced toxicities.

4.2 Introduction

Since the approval of sorafenib in 2005, major advancements have been made in the treatment of metastatic clear-cell RCC. These advancements, defined as improved OS and PFS for a majority of patients, are largely due to molecular knowledge underlying disease pathology. Subsequent to sorafenib, three additional angiogenesis inhibitors (axitinib, pazopanib, and sunitinib), and two inhibitors of mTOR (everolimus and temsirolimus) were approved by the U.S. FDA. While axitinib, pazopanib, sorafenib, and sunitinib belong to the same class of oral multikinase inhibitors that target the VEGF pathway, they all differ in the exact kinases they inhibit, the potency of kinase inhibition, clinical pharmacology, dosing and dose intensity, and ultimately clinical efficacy.

Sorafenib is an exceptionally promiscuous oral multikinase inhibitor used in the treatment of RCC, hepatocellular carcinoma and thyroid cancer.[1,2] Originally, sorafenib was developed as a RAF inhibitor based on *in vitro* screen showing its potent inhibition of RAS serine/threonine isoforms.[3,4] Similarly, it was also shown to be a potent inhibitor of the several other targets in the RAS/MEK/ERK signal transduction cascade, the c-KIT receptor, initiation factor eIF4E and the anti-apoptotic protein MCL-1.[2,5-7] Most notably though, sorafenib has been shown to be an effective VEGF-pathway and angiogenesis inhibitor by targeting several kinases, including: VEGFR-1-3, PDGFR-β, FLT-3, and FGF-1.[2,6,7]

However, recent data has begun to emerge revealing novel and previously unknown targets of sorafenib. More recently, sorafenib has been shown to be a ligand for multiple serotonin receptors,[8] an inhibitor of additional anti-apoptotic proteins not directly tied to angiogenesis (e.g. cyclin B1, cyclin D1 and survivin),[9] and has been found to also affect the immune system (e.g. on the function of dendritic cells, and by modulating effector CD4+ cell and regulatory T cell function).[10,11] There are very likely many novel and still undiscovered targets of sorafenib, and many of these targets are likely susceptible to germline genetic variation.

Previous studies have shown that fine genome mapping, through genome wide association studies (GWAS), can be successfully performed in a panel of diverse inbred strains of mice to identify genetic loci that contain candidate genes that modulate both single gene and polygenic traits.[12-15] But to date, there have been few examples of animal GWAS pharmacogenetics, and even those that have been published have not analyzed the contribution of genetics to multikinase inhibitor (e.g. sorafenib) activity.[16-19] Moreover, previous pharmacogenetic studies of VEGF-pathway inhibitors of the same class as sorafenib

(e.g. pazopanib and sunitinib) have focused on associations between variants in drug metabolizing/transporter genes or genes involved with angiogenesis, and clinical phenotypes (e.g. OS, PFS and drug-induced toxicities).[20-24] And, the only study in humans to test genotype-clinical phenotype associations in mRCC patients was not conducted in patients treated with sorafenib.[25]

In this Aim, MEFs from 32 inbred isogenic strains were treated with sorafenib in a concentration-response format. Phenotypic measurements of cell health were obtained through the use of high-content imaging, and GWAS was performed to identify potential candidate quantitative trait loci QTLs and candidate genes that underlie sorafenib activity (Figure 4.1). The use of genetically well-characterized inbred mouse strains provides a viable model system to analyze the genetic basis for drug activity and cytotoxicity. These MEF strains can be used in a high-throughput cellular genetics approach for four main reasons: retention of the exact genetic composition of the mouse strain from which they are isolated, selection of these 32 strains increases the likelihood of detecting genetic differences that could underlie differences in sorafenib response and/or identify novel nodes in sorafenib signaling pathways, technological advances (e.g. high-content imaging) have allowed for better characterization of cellular phenotypes, and technological advances (e.g. siRNA loss of function and cDNA over-expression *in vitro* validation techniques) have allowed for functional validation of QTLs. This Aim will test the hypothesis that differential cell health and response data (e.g. EC_{50} or IC_{50} values) from 32 MEF cells lines treated with sorafenib, can be used in GWAS to identify candidate genes associated with sorafenib response, which will ultimately lead to the discovery of novel genes for future pharmacogenetic testing in patients treated with sorafenib.

4.3 Materials and Methods

4.3.1 Animals

Primary mouse embryonic fibroblast (MEF) cells derived from 32 inbred strains were obtained from The Jackson Laboratory (Bar Harbor, ME) were used in this screen: 129S1/SvImJ, A/J, AKR/J, BALB/cByJ, BTBR T^{+} *tf*/J, BUB/BnJ, C3H/HeJ, C57BL/6J, C57BR/cdJ, CBA/J, CE/J, CZECHII/EiJ, DBA/2J, FVB/NJ, I/LnJ, KK/HlJ, LP/J, MA/MyJ, MOLF/EiJ, MRL/MpJ, NOD/ShiLtJ, NON/ShiLtJ, P/J, PERA/EiJ, PL/J, PWK/PhJ, RIIIS/J, SEA/GnJ, SJL/J, SM/J, SWR/J, and WSB/EiJ. Cells were harvested from male mice aged 10-12 weeks. The isogenic strains chosen for this screen were a subset of strains from the Mouse Phenome Project.[12]

4.3.2 Cell Culture

Each of the 32 MEF cell lines was expanded from passage 0 to passage 2 using DMEM (Cellgro, Manassas, VA) supplemented with 10% FBS (Cellgro), 1% non-essential amino acid solution (Sigma Aldrich, Milwaukee, WI), and 1% penicillin/streptomycin (Sigma Aldrich). MEF cells were maintained at 37^{o}C and 5% CO_2. When MEF cells reached 90% confluence they were harvested. MEF cells were seeded into 384-well, PDL-coated Aurora 200 micron COP polymer plates (Aurora Biosciences/Brooks Life Science Systems, Fremont, CA) using Multidrop 384 dispensers (Thermo Scientific, Waltham, MA) to a final concentration of $5x10^4$ cells/mL per well. Seeding densities were confirmed by an Invitrogen Countess automated cell counter (Invitrogen, Carlsbad, CA). For each 384-well microtiter plate (n=4 replicates), 12 wells were assigned to each of the 32 strains (Figure 4.2A). Cells were plated with two different densities: 1500 cells/well for the 24 h time point and 1000

cells/well for the 72 h time point. A final volume of 5 µL of cells in supplemented DMEM was administered to each well. Plates were incubated overnight at 37°C and 5% CO_2.

4.3.3 High Content Screening

The cytotoxic effects of sorafenib on MEF cells were screened using a nine-point logarithmic dose response. Sorafenib response was assessed 24 and 72 h post administration. Stock solutions were prepared in DMSO to create a master microtiter plate for the screens: 75 mM, 25 mM, 8.33 mM, 2.78 mM, 0.93 mM, 0.31 mM, 0.10 mM, 0.03 mM, and 0.01 mM. Three-fold serial dilutions were prepared using DMSO (Sigma Aldrich) and a Biomek 2000 (Beckman Coulter, Brea, CA), and the final sorafenib concentrations used in the screens were: 300 µM, 100 µM, 33 µM, 11 µM, 3.7 µM, 1.2 µM, 0.41 µM, 0.14 µM, and 0.045 µM. Two separate positive control microplates, containing valinomycin in DMSO, were prepared. For the 24 h time point, valinomycin with a concentration of 8.33 mM (final diluted concentration of 33 µM) was prepared. For the 72 h time point, valinomycin with a concentration of 5 µM (final diluted concentration of 20 nM) was prepared. The MEF cells were dosed with 200 nL of each sorafenib concentration, in quadruplicate, from the stock solution in the master microtiter plate using a slot pin tool (V&P Scientific, San Diego, CA) on a Biomek FX liquid handling automation system (Beckman Coulter). For plate reference control compounds, one well for each cell line was treated with valinomycin in DMSO, one well was treated with vehicle only (DMSO 0.4% final concentration) and one well for each cell line was treated with media only (Figure 4.2A).

After 24 h or 72 h incubation with the compounds, cells were labeled and processed using the Thermo Cellomics Multiparameter Cytotoxicity 3 Kit protocol (Thermo Scientific,

Rockford, IL) and HCA Image Amp protocol (Perkin Elmer, Waltham, MA). First, MEF cells were stained with Mitotracker Orange dye (Invitrogen, Carlsbad, CA) to assess mitochondrial membrane potential, and YOYO-1 cell permeability dye (included in the Multiparameter Cytotoxicity 3 Kit) and incubated at 37^{o}C. Next, MEF cells were fixed with 16% paraformaldehyde, washed, labeled with anti-Cytochrome C antibody (Thermo), and finally counterstained with Hoechst 33342 nuclear stain (included with the Multiparameter Cytotoxicity 3 Kit). All dyes, buffer, and antibodies were administered using Multidrop 384 well dispensers. All wash steps were completed using a BioTek ELx405 Select microtiter plate washer (BioTek, Winooski, VT).

4.3.4 Image Acquisition and Analysis

Plates were sealed once staining was completed, and subsequently imaged on the Cellomics ArrayScan VTi™ high content imager (Thermo), using 10X/0.45NA objective lens (Figure 4.2B). Cells expressing fluorescent bioprobes for Hoechst 33342 (386 nm), YOYO-1 (485 nm), Mitotracker Orange (549 nm), and Cytochrome C–Alexafluor 647 (650 nm) were excited with LED light engine (L μM encor, Beaverton, OR), and emitted light through a Quadband filter set (Semrock, VT) to collect fluorescence signal using a 12-bit Orca-ER II CCD Camera (Hamamatsu, Japan). Exposure times were <0.05 seconds, and four fields per well were captured. Next, a compartmental analysis algorithm identified: valid object counts based on Hoechst nuclear staining (cell loss), mean average intensity of YOYO-1 positive cells in the nucleus (membrane permeability), mean average intensity of Mitotracker Orange in the cytoplasm (mitochondrial function), and mean average intensity of

Cytochrome C–Alexafluor 647 (apoptosis) in the cytoplasm (Table 4.1, Figure 4.2C). All generated data was exported using the vHCS View client (Thermo).

Captured images were manually inspected for data integrity. Wells were rejected for replicate analyses from the final analysis if an image was not properly focused, contained a fluorescent artifact, or if debris occupied >20 % the field area. Outliers that failed the QC criteria were filtered and removed from the final data analyses.

4.3.5 Response Analyses

Values generated by the image analysis algorithm for each MEF strain were normalized to the DMSO control well on each plate. For the cell viability phenotype, hereafter referred to as Valid Object Count (VOC), values were normalized against the DMSO average among the quadruplicate plates for each individual MEF line). Dose response curves were created in F-Curve software, for each MEF cell line by fitting the normalized quadruplicates to a Brain-Cousens model implemented by the drc package (version 1.8-1) for R (Figure 4.2D).[26] Brain-Cousens was chosen to account for biphasic response and biological shifts sometimes observed from toxic compound dose responses (e.g. hormesis).[27] The Brain-Cousens equation is: $f(x,(b,c,d,e)) = c+((d+fx-c)/(1+\exp(b(\log(x)-\log(e)))))$.

All curves were visually inspected for outliers. In cases where outliers were removed, a new fit was performed in F-curve. Dose-response curves, generated using F-curve and the Brain Cousens model, were used to calculate the individual effective concentration (EC) or inhibitory concentration (IC) values of each MEF line. For the phenotypic endpoints where response decreased (VOC and cytochrome C release), IC_{50} through to EC_{80} values in six stepwise increments of 5% were interpolated from the fitted function (Figure 4.3).

4.3.6 Quantitative Trait Loci Mapping

GWA mapping was performed using the respective half-maximal effective concentration (EC_{50}) and half-maximal inhibitory concentration (IC_{50}) values for each cell health phenotype using the EMMA and SNPster algorithms.[28,29] Only phenotypes with EC_{50} or IC_{50} values for ≥27 strains were used in GWAS. The SNPster algorithm uses a bootstrap model to calculate the significance of association analysis from an inferred haplotype structure determined by overlapping three-SNP windows for each strain. A one-way ANOVA F-statistic from associations between the haplotypes and either EC_{50} or IC_{50} values for each phenotype. Then, p-values were generated after the phenotype values were bootstrapped ($1x10^6$ times), and thus the maximum –log(P) value was 6.[28] Conversely, EMMA used single SNPs for GWA mapping, where the ANOVA F-statistic and p-values were generated from associations between the individual marker and phenotypic value.[29] EMMA SNPs were filtered to remove non-informative markers and redundant SNPs, where adjacent polymorphisms have the same allelic distribution pattern across the inbred mouse strains. P-values generated from these associations were corrected for multiple comparison testing using a Bonferroni correction, and therefore the threshold for genome-wide significance using EMMA was a p-value of $1.8x10^{-7}$, or a –log(P) value of ≥6.74.

Although EC_{50} and IC_{50} values were chosen *a priori* to represent the phenotype for each cell health parameter, a QC step was performed next to assess the robustness of genotype-phenotype associations. The identification of QTLs for each of the four phenotypes verified selecting the SNPs with the highest association scores (top 2% of association scores by –logP value) across EC_{50-80} and IC_{50-80} ranges, and the top 2% association scores were averaged at each SNP position (Figure 4.4A and 4.4B). Genomic regions with mean –logP

≥4.0 (including a window of 100 kb on each side) were selected for further analysis using EC_{50} or IC_{50} association data (Figure 4.4C). Genes that completely or partially overlap the regions associated with drug response were selected for pathway and functional enrichment analysis. Candidate genes were selected from the University of California Santa Cruz (UCSC) mouse RefSeq database (assembly mm9, http://genome.ucsc.edu/) (Figure 4.5).

The SNP genotypes used for association mapping were obtained from the Mouse Diversity Array set, available from the Comparative Genomics Database website (CGD; http://cgd.jax.org/cgdsnpdb/).[30] SNPs were trimmed from the final set based on redundancy (identical haplotype pattern at a particular locus), missingness (few calls for a specific SNP among the 32 strains), and if they demonstrated a lack of genetic variability among the 32 strains. After pruning, approximately 277,000,000 SNPs were used for the analyses. Manhattan plots were visualized using R version 3.1.0 and the UCSC Mouse Genome Browser on the Mouse July 2007 (NCBI37/mm9) Assembly (https://genome.ucsc.edu).[31]

4.3.7 Candidate Gene Selection

Candidate gene selection was conducted using a multi-step, multi-faceted approach (Figure 4.6). Genomic regions with mean –logP ≥4.0 for a given phenotype were moved forward for further analysis. Next, to increase efficiency of all downstream analyses, the first step in candidate gene selection was to examine genes with likely cellular responses in MEFs, so only genes with known expression levels in MEFs were selected. Subsequently, expression levels of genes were examined in kidney cells from six strains of inbred mice (*A/J, AKR/J, C3H/HeJ, C57BL/6J, CBA/J,* and *DBA/2J*). Kidney cell expression was prioritized over other tissue types because of its relevance to the patients and tumor type

described extensively in Aims 1-2. Expression levels were measured in MEFs and using the Affymetrix Mouse Genome 430 2.0 Array (Santa Clara, CA). Genes were considered expressed if their expression level was greater than 50 for at least one of the strains following data processing with the gcRMA algorithm.

Next, QTL regions that overlapped using both SNPster and EMMA algorithms were prioritized, and genes in these regions were still considered candidates. However, it is important to note that these QTL regions need not have a $-\log P \geq 6.74$ to still be considered. Candidate genes were further prioritized based on: literature evidence biology that underlies sorafenib response or cancer pathogenesis, differential haplotype structure between MEF strains with low EC_{50} or IC_{50} values versus those with high values, evidence of functional annotation using the UCSC Genome Browser and/or the Ensembl website (www.ensembl.org), evidence of cis-acting gene regulation among candidate genes in a QTL region, if there was presence of potentially deleterious non-synonymous coding SNPs within a QTL, or if there were direct connections between a candidate gene and sorafenib or known targets of sorafenib (e.g. receptors, transcription factors, kinases, etc.).[32-35]

The Mouse Genome Informatics Database (MGI; http://www.informatics.jax.org) and the CGD resources were used to examine non-synonymous coding SNPs in candidate genes within 30 MEF strains used for GWAS. Build 137 on the dbSNP website (http://www.ncbi.nlm.nih.gov/projects/SNP) was used to confirm the SNP position (m38) and amino acid substitution for a given non-synonymous SNP. PROVEAN (Protein Variation Effect Analyzer) version 1.1.322, PANTHER (Protein Analysis Through Evolutionary Relationships) version 9.0 and SIFT (Sorting Intolerant From Tolerant) version 1.03 were utilized to assess that candidate non-synonymous SNPs would cause deleterious

effects to protein structure or function.[36-39] For PROVEAN, a score of -2.5 indicates that the SNP results in alterations to the functional effect on the protein. For the PANTHER, a subSPEC (substitution position-specific evolutionary conservation) score of -3 indicates that there is a 50% probability that a score is deleterious ($P_{deleterious}$= 0.5), MSA indicates the number of multiple sequence alignments, NIC (number of independent counts) is an estimate of observations used to calculate the amino acid probabilities. P_{wt} and $P_{substituted}$ refer to the respective probabilities of the wild-type (WT) and substituted amino acids. For the SIFT algorithm, a score <0.05 indicates that the SNP would likely cause deleterious effects on the protein. Protein data for a given SNP was considered significant if PROVEAN predicted the protein to likely be deleterious, if PANTHER returned a score ≤-3, and/or if SIFT returned a score <0.05. Mouse Phylogeny Viewer (https://msub.csbio.unc.edu/) was used to examine haplotype structure among the MEF strains for a given QTL or candidate gene.[40] Finally, Ingenuity® Pathway Analysis was used to determine if direct connections between a candidate gene and sorafenib or known targets of sorafenib (e.g. receptors, transcription factors, kinases, etc.) could be predicted (http://www.ingenuity.com).[35]

4.4 Results

4.4.1 High Content Imaging Screen

A high content imaging screening approach based on mouse genetics was selected to simultaneously evaluate multiple cell health parameters involved with early and late apoptosis. The study design allowed MEF cells from 32 inbred mouse lines of the Mouse Phenome Project to be plated on a single 384-well plate and then treated with sorafenib in a 10-point concentration-response curve manner (and with positive and negative controls) for

each strain. A total of four 384-well plates were successfully stained, imaged and analyzed during the experiments conducted for this Aim. Approximately 1.5% of the wells from all four plates were discarded due to image quality issues after visual inspection of each of the wells in question. The WSB/EiJ required re-analysis across the four plates due to issues related to the analysis algorithms versus the strain's small-sized nucleus. Ultimately, issues to do with nucleus size also confounded analyses in F-Curve, and the WSB/EiJ strain was removed. In addition, due to problems with plating during the screen, the SWR/J strain was removed. The final SNPster and EMMA GWA mapping studies included 30 MEF lines for four cell health parameter phenotypes and both 24 h and 72 h sorafenib incubation time points.

4.4.2 Quantitative Trait Loci Mapping

Using GWA mapping through the SNPster and EMMA, only two of the four phenotypes (Cytochrome C release and VOC) at the 72 hour time point revealed QTL regions with –logP value ≥4.0, and QTL regions that overlapped on both algorithms were identified. No QTL regions with –logP value ≥4.0 were identified for any of the four phenotypes after 24 h incubation with sorafenib. Similarly, no QTL regions with –logP value ≥4.0 were identified for the cell permeability or mitochondrial membrane potential phenotypes after 72 h incubation with sorafenib.

IC_{50} variability across the strains was observed for Cytochrome C release (Figure 4.7) and VOC (Figure 4.8) after a 72 h incubation with sorafenib, which indicated significant genetic variability among the strains for these two phenotypes. One genome-wide significant QTL ($p=1.0x10^{-8}$, –logP=8), which overlapped using both SNPster and EMMA algorithms,

was identified for the Cytochrome C release phenotype: chromosome 9, position 50,500,000-52,500,000 (Figure 4.9). A total of seven genes (Figure 4.10) were identified under this QTL (*Arhgap20, Btg4, Fdx1, Layn, Pou2af1, Rdx*, and *Zc3h12c*). The gene directly under the genome-wide significant peak, *Arhgap20*, was not expressed in MEF or mouse kidney cells. Among the genes under this genome-wide significant peak, only *Fdx1, Layn, Rdx*, and *Zc3h12c* were expressed in both MEF cells (Table 4.2) and in mouse kidney cells. *Fdx1* expression in kidney cells was 12,679.96, *Layn* expression was 64.23, *Rdx* expression was 1,829.55, and *Zc3h12c* was 196.79. These four genes associated with Cytochrome C release were prioritized for candidate gene selection.

A second notable QTL ($p=2.2x10^{-5}$, $-logP=4.66$), which overlapped using both SNPster and EMMA algorithms, was identified for the VOC phenotype: chromosome 4, position 119,000,000-121,500,000 (Figure 4.11). A total of 13 genes (Figure 4.12) were identified under this QTL (*Ctps, Col9a2, Exo5, Edn2, Foxo6, Hivep3, Kcnq4, Nfyc, Rims3, Scmh1, Slfnl1, Smap2,* and *Zpf69*). Among the genes under this genome-wide significant peak, only *Ctps, Exo5, Nfyc, Scmh1, Smap2, and Zpf69* were expressed in both MEF cells (Table 4.3) and in mouse kidney cells. Notably, *Scmh1* was the gene closest to being directly under the peak. No probe sets were available to assess expression in *Foxo6*. *Ctps* expression in kidney cells was 615.64, *Exo5* expression was 534.41, *Nfyc* expression was 498.13, *Scmh1*expression was 87.54, *Smap2* expression was 222.71 in one probe set, and *Zpf69* expression was 125.74. These four genes associated with Cytochrome C release were prioritized for candidate gene selection

4.4.3 Candidate Gene Selection

Four genes (*Fdx1, Layn, Rdx*, and *Zc3h12c)* within the genome-wide significant QTL associated with Cytochrome C release (chromosome 9, position 50,500,000-52,500,000) were expressed in MEFs and mouse kidney tissue. For this phenotype, seven strains (SM/J, RIIIS/J, C57BL/6J, C57BR/cdJ, PL/J, MA/MyJ, and AKR/J) were more sensitive to the cytotoxic effects of sorafenib after a 72 h incubation. Differences in haplotype structure among these four genes were noted for the seven sensitive strains, when compared to the 23 strains more insensitive to sorafenib in terms of cytochrome C release (Figure 4.13). The only exception involved *Layn*, where the second most sensitive strain (RIIIS/J) had the same haplotype structure as the insensitive strains. For this reason, *Layn* was not considered a viable candidate gene. Among genes within this QTL, all were included in pathways directly related to sorafenib response and/or known signaling targets of sorafenib; however, one was found to have no direct connections to sorafenib or known targets of sorafenib (*Zc3h12c*) and was excluded. Only *Rdx* was shown to have a direct connection to a known target of sorafenib, *FLT-1* (Figure 4.14). Only three potentially deleterious non-synonymous SNPs, in two genes, were identified among these potential candidates (Table 4.4). Only two non-synonymous SNPs in *Layn* (rs33773426 and rs32764902) were predicted to be potentially deleterious. Finally, none of the six candidate genes within the QTL region associated with VOC revealed evidence of cis-acting gene regulation (data not shown). Considering data from all criteria, *Rdx* will be prioritized as a candidate gene for future validation studies.

Six genes (*Ctps, Exo5, Nfyc, Scmh1, Smap2, and Zpf69)* within the QTL that associated with VOC/cell viability (chromosome 4, position 119,000,000-121,500,000) were expressed in MEFs and mouse kidney tissue. For this phenotype, 13 strains (C57BL/6J,

CZECHII/EiJ, RIIIS/J, FVB/NJ, SEA/GnJ, NON/LtJ, AKR/J, C57BR/cdJ, PWK/PhJ, BUB/BnJ, MA/MyJ, KK/HIJ, and A/J) were considered more sensitive to the cytotoxic effects of sorafenib after a 72 h incubation. Differences in haplotype structure among these four genes were noted for the 13 sensitive strains, when compared to the 17 strains more insensitive to sorafenib in terms of cell viability (Figure 4.15). Among genes within this QTL, all were included in pathways directly related to sorafenib response and/or known signaling targets of sorafenib; however, two were found to have no direct connections to sorafenib or known targets of sorafenib (*Slfnl1* and *Zfp69*) and were excluded. Only *Nfyc* was shown to have a direct connection to a known target of sorafenib (PDGFR-β). *Nfyc* was also shown to have a direct connection to the oncogenic protein MYC (Figure 4.16). Only nine potentially deleterious non-synonymous SNPs, in five genes, were identified among these potential candidates (Table 4.5). Only one non-synonymous SNP in *Schmh1* (rs28256862) and two SNPs in *Zpf69* (rs27485619 and rs32769909) were predicted to be potentially deleterious. Finally, none of the six candidate genes within the QTL region associated with VOC revealed evidence of cis-acting gene regulation (data not shown). Considering data from all criteria, *Nfyc* and *Scmh1* will be prioritized as a candidate gene for future validation studies.

4.5 Discussion

The primary objective of this Aim was to test the hypothesis that that differential cell health and response data (e.g. EC_{50} or IC_{50}) from 32 MEF cells lines treated with sorafenib, can be used in GWAS to identify candidate genes associated with sorafenib response, which

will ultimately lead to the discovery of novel genes for future pharmacogenetic testing in patients treated with sorafenib.

There are many advantages of conducting a high content *in vitro* screen of drugs based on mouse cellular genetics. For instance, this type of high-throughput technology allows for multiplexed measurements of several endpoints, and the evaluation of subtle cytological changes in response to drug can be observed.[18] In addition, inbred mouse lines allow for greater reproducibility, when compared to primary human cell lines, due to genetic stability of the strains, which allows for robust GWA mapping studies based on the known and static nature of inbred and immortalized mouse lines. These strains are genetically and phenotypically diverse, and have a substantial number of recombinations, which improves GWA mapping resolution.[18,28] Therefore, a relatively small number of inbred mouse strains can be used, like had been done in this Aim, to discover new genotype-phenotype associations that are applicable to human populations. This also enables this type of high-throughput screen to capture a broad range of drug response variance, and identify genes that underlie drug response at a cellular level.[18,41]

Aim 1 identified germline variants that significantly associate with OS and PFS through a candidate gene/candidate SNP approach. Since this approach leverages existing knowledge about mRCC pathogenesis/prognosis, angiogenesis and/or sorafenib pharmacology, there is virtually no chance that novel and previously unidentified signaling pathways and/or candidate genes will be discovered. This cellular genetics approach, using high-content cellular imaging and genetic mapping, has helped discover novel genes and pathways involved with sorafenib cytotoxicity and provide a better understanding of the variability observed with this phenotype.

A total of seven genes expressed in MEF cells at mRNA level were present in the QTL that associated with the Cytochrome C release phenotype after a 72 h incubation with sorafenib. After a multi-step and multi-faceted schema (Figure 4.6), one candidate gene was selected for future *in vitro* validation studies. *Rdx* encodes for radixin, which is a component of the ezrin-radixin-moesin (ERM)-binding phosphoprotein-50 (EBP50) complex, and is conserved between humans and mice. In humans, radixin, as part of the EBP50 complex (also known as NHERF1), belongs to the family of PDZ scaffolding proteins, has been shown to act as a tumor suppressor in multiple tumor types.[42-46] Additionally, EBP50 has been shown to promote apoptosis in hepatocellular carcinoma (HCC) by modulating β-catenin/E-cadherin, and to potentially enhance the metastatic potential of renal cell carcinoma.[47,48] This is important because sorafenib has been approved by the U.S. FDA for the treatment of both mRCC and unresectable HCC.[2] More recently, in an osteosarcoma model, sorafenib was shown to inhibit phosphorylation of the ERM complex.[49]

A total of 13 genes expressed in MEF cells at mRNA level were present in the QTL that associated with VOC/cell viability after a 72 h incubation with sorafenib. After the same criteria were applied as before (Figure 4.6), two candidate genes were selected for future *in vitro* validation studies. *Nfyc* encodes for the C subunit of the nuclear factor-gamma (NF-γ) transcription factor. This transcription factor complex is conserved between humans and mice, and binds to CCAAT motifs in the promoter regions. In humans, NF-γC is integral to the trimerization of NF-γA, NF-γB and NF- γC, and is a target for regulatory proteins MYC and p53.[50] NF-γC has also been shown to regulate DNA-dependent transcription of one of sorafenib's primary targets, PDGFR-β, which is for pericyte formation.[51] *Scmh1* encodes for sex comb on midleg homolog 1 (Drosophila) protein, which forms multiprotein complexes

with other polycomb group proteins to help retain the transcriptionally repressed state of other genes. Traditionally associated with male spermatogenesis, *Scmh1*is also conserved between humans and mice.[52] Nothing is currently known about whether *Scmh1* directly affects cancer risk, pathogenesis, or prognosis; however, *Scmh1* has recently been shown to directly and indirectly regulate geminin stability.[53] Geminin is a key protein involved in DNA replication, has been linked to cancer, and therefore *Scmh1* variants could be important in cancer and in sorafenib response.[54-57]

Future studies to validate these candidate genes will certainly be conducted. First, the prioritized candidate genes (e.g. *Rdx* and *Fdx1* for Cytochrome C release, as well as *Nfyc* and *Scmh1* for VOC) will be validated in MEF cells using *in vitro* knockdown and overexpression approaches to confirm the capability of this high content imaging screening approach as a viable method for gene selection. *In vitro* methods, using and siRNA approach for knockdown and a pCMV-SPORT6 vector-based approach for overexpression, have been previously described.[18] Validation of candidate genes identified in this study will follow a similar experimental design. Second, for genes validated by *in vitro* knockdown and/or overexpression in MEF cells and are conserved in humans, a second round of knockdown and overexpression experiments will be conducted. Orthologous human siRNA and overexpression vectors will be introduced to a human endothelial cell line (e.g. TIME cells) and a human mRCC cell line (e.g. Caki-1 cells) to confirm the putative effects of these genes on sorafenib response. Although translating results from an *in vitro* mouse screen to human *in vivo* studies may seem like a daunting task, it has been successfully accomplished previously.[16,17,58] Therefore, for genes that are validated in both MEFs and human cell lines,

candidate SNPs will be identified, and genotyped in TARGET patients from the sorafenib arm to test associations with clinical phenotypes (OS, PFS, and sorafenib-induced toxicities).

In summary, this innovative high content cellular genetics approach has detected robust interstrain cellular differences in sorafenib activity. One QTL region, which reached genome-wide significance, which potentially associates with sorafenib-induced cytochrome C release from mitochondria, was identified. An additional QTL was identified that potentially associates with sorafenib cytotoxicity and cell viability. Candidate genes for functional validation have been prioritized through a multi-faceted set of criteria. Future steps for this work include functional validation of candidate genes, using knockdown and overexpression approaches, in MEF and human cell lines.

TABLES

Table 4.1. **Description of the high content imaging output data features**. This description includes HCS feature, the corresponding probes used in assay, biological indication measured, and IC_{50} curve fitting trends projected following normalization of data to DMSO control. All generated data was exported using the vHCS View client from Thermo. Abbreviations: DMSO, dimethyl sulfoxide; HCS, high content screening.

No.	HCS Feature	IC_{50} Trend	Probe Reagent	Biological Indicator
1	Valid Object Count	Down	Hoechst 33342 Stain	Cell count / cell loss
2	Mean Size Intensity of YoYo-1 positive cells in nucleus	Up	YOYO-1 Dye	Comprised nuclear membrane
3	Mean Average intensity of Mitotracker in the cytoplasm	Down	Mitotracker Orange Dye	Mitochondria function impaired
4	Mean Intensity of the Cytochrome C–Alexafluor 647	Down	Anti-Cytochrome C Antibody	Apoptosis

Table 4.2. MEF cell expression of candidate genes within the QTL associated with Cytochrome C release. A total of seven genes were identified under that QTL region, and their respective expression levels, for all available expression probe sets, are presented in this table. Expression level values for six MEF lines and MEF expression levels reported in the BioGPS website (www.biogps.org) are included. Abbreviations: MEF, mouse embryonic fibroblast; QTL, quantitative trait locus/loci.

Gene	Full Name	Probeset ID	BioGPS	DBA/2J	C57BL/6J	A/J	AKR/J	C3H/HeJ	CBA/J
Arhgap20	Rho GTPase activating protein 20	1427522_at	9.27	7.308	7.161	7.29	7.732	7.296	7.331
		1429918_at	4.64	5.779	5.802	5.771	5.985	5.766	5.804
Btg4	B cell translocation gene 4	1426520_at	5.37	4.447	4.467	4.44	4.544	4.446	4.519
Fdx1	Ferredoxin 1	1449108_at	427.69	450.403	602.196	407.316	519.105	455.606	458.031
Layn	Layilin	1442608_at	37.78	26.79	50.201	44.105	49.095	45.52	44.256
		1444165_at	190.09	78.682	90.642	95.183	114.413	92.013	72.725
Pou2af1	POU domain, class 2, associating factor 1	1416957_at	4.64	6.061	6.094	6.088	6.193	6.089	6.329
Rdx	Radixin	1416179_a_at	5282.61	3173.472	3026.9	3018.068	3043.828	2336.223	2598.385
		1416180_a_at	841.34	558.965	465.09	480.598	400.75	389.205	424.604
		1448236_at	1394.17	1069.25	1001.683	956.406	931.784	870.561	933.532
Zc3h12c	Zinc finger CCCH type containing 12C	1437111_at	29.09	12.706	10.79	10.788	10.886	10.775	9.29
		1441787_at	13.25	7.611	7.66	9.071	7.789	7.651	7.689
		1444402_at	87.68	69.316	67.363	60.562	79.447	51.012	59.074

Table 4.3. MEF cell expression of candidate genes within the QTL associated with Valid Object Count. A total of 13 genes were identified under that QTL region, and their respective expression levels, for all available expression probe sets, are presented in this table. Expression level values for six MEF lines and MEF expression levels reported in the BioGPS website (www.biogps.org) are included. Abbreviations: MEF, mouse embryonic fibroblast; QTL, quantitative trait locus/loci.

Gene	Full Name	Probeset ID	BioGPS	DBA/2J	C57BL/6J	A/J	AKR/J	C3H/HeJ	CBA/J
Ctps	cytidine 5'-triphosphate synthase	1416563_at	2799.2	3289.424	2690.554	2226.492	3148.194	2547.071	2913.795
Col9a2	collagen, type IX, alpha 2	1450673_at	4.64	5.723	5.72	5.686	5.821	5.713	5.821
Exo5	exonuclease 5	1428903_at	187.44	147.003	175.026	172.678	156.161	160.716	146.567
Edn2	endothelin 2	1449161_at	4.64	5.599	5.61	5.712	5.685	5.613	5.601
Foxo6	forkhead box O6	none	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Hivep3	human immunodeficiency virus type I enhancer binding protein 3	1439660_at	8.68	4.062	4.064	4.05	4.111	4.05	4.072
		1450132_at	4.64	6.208	6.168	6.204	6.307	6.195	6.221
		1421150_at	4.64	4.219	4.23	4.213	4.266	4.212	4.253
		1429134_at	5.02	5.899	5.369	5.383	5.498	5.604	5.431
		1458802_at	8.21	56.808	15.764	15.702	20.173	15.884	16.002
Kcnq4	potassium channel, voltage gated KQT-like subfamily Q, member 4	1435721_at	5.39	3.96	4.206	4.188	4.24	4.19	4.377
Nfyc	nuclear transcription factor Y, gamma	1448963_at	523.45	300.231	247.408	283.674	237.023	242.022	267.626
Rims3	regulating synaptic membrane exocytosis 3	1459042_at	4.64	5.843	5.841	5.833	5.98	5.852	5.897
		1435971_at	4.68	4.799	4.78	4.75	4.838	4.778	4.808
Scmh1	sex comb on midleg homolog 1	1426241_a_at	281.63	57.533	100.8	53.924	114.368	91.96	90.253
		1439554_at	8.57	5.244	5.279	5.25	5.35	5.249	5.275
		1441573_at	7.28	6.3	6.338	6.29	6.407	6.304	6.333
Slfnl1	schlafen like 1	1455838_at	7.13	4.385	4.396	4.368	4.501	4.376	4.51
Smap2	small ArfGAP 2	1450675_at	164.3	174.581	104.31	112.917	110.652	110.834	110.447
Zfp69	zinc finger protein 69	1458274_at	4.64	9.08	8.936	9.11	9.499	9.368	9.19
		1425209_at	14.79	11.118	16.559	12.406	11.115	9.893	10.529
		1435916_at	79.19	61.218	67.965	61.141	54.539	47.654	46.759
		1425210_s_at	55.95	17.165	18.55	17.514	16.922	17.508	14.864

Table 4.4. Non-synonymous coding SNPs and deleterious protein effects for genes associated with Cytochrome C release. The Mouse Genome Informatics Database (MGI; http://www.informatics.jax.org) and the Center for Genome Dynamics website (CGD;; http://cgd.jax.org/cgdsnpdb) were used to examine non-synonymous coding SNPs in candidate genes. Build 137 on the dbSNP website (http://www.ncbi.nlm.nih.gov/ projects/SNP) was used to confirm SNP position (m38) and amino acid substitution for each non-synonymous SNP. PROVEAN, PANTHER Classification System and SIFT (via Ensembl; http://www.ensembl.org) were used to assess that candidate non-synonymous SNPs would cause deleterious effects to protein structure or function. For PROVEAN, a score of -2.5 predicts alterations to the functional effect on the protein. For the PANTHER, a subSPEC (substitution position-specific evolutionary conservation) score of -3 predicts a 50% probability that a score is deleterious ($P_{deleterious}$=0.5), MSA is the number of multiple sequence alignments, NIC (number of independent counts) is an estimate of observations used to calculate the amino acid probabilities. P_{wt} and $P_{substituted}$ (P_{subs}) refer to the respective probabilities of the WT and substituted amino acids. For SIFT, a score <0.05 predicts the SNP to likely cause deleterious effects on the protein. Protein data was considered significant if PROVEAN predicted the protein to likely be deleterious, if PANTHER score ≤ -3, and/or if SIFT score <0.05. Protein data was highlighted in red if the protein was likely deleterious using PROVEAN and/or SIFT, and/or had a PANTHER score ≤-3. Protein data was highlighted in yellow if the PANTHER score was ≤-2 but >-3. Protein data was highlighted in yellow if the PANTHER score was ≤-2 but >-3. No non-synonymous variants, among the 30 selected MEF strains, were identified in *Fdx1* and *Pou2af1* using MGI and CGD resources.

Gene	SNP ID (dbSNP Build 137)	Position (m38)	Substitution	Protein Change	PROVEAN Score	Prediction	SIFT	Prediction	Sub SPEC	Pdel	MSA Position	Pwt	P subs	NIC
Arhgap20	rs46602632	9:51831708	G>A	Arg282Lys	0.867	Neutral	1	Tolerated	-0.62	0.08	216	0.18	0.36	1.005
	rs47224784	9:51848572	A>G	Ile574Met	-0.727	Neutral	0.07	Tolerated	2.38	0.35	501	0.36	0.03	1.128
	rs51861578	9:51849843	G>T	Arg998Leu	-0.059	Neutral	0.71	Tolerated	-0.36	0.07	916	0.12	0.09	1.128
Btg4	rs13461391	9:51116640	A>G	Ile37Val	-0.346	Neutral	0.34	Tolerated	-0.65	0.09	47	0.34	0.35	1.952
	rs13461390	9:51117988	A>G	Ile158Val	0.29	Neutral	1	Tolerated	-0.89	0.11	174	0.17	0.23	1.952
	rs32767530	9:51119206	G>A	Cys202Tyr	2.231	Neutral	1	Tolerated	-0.81	0.10	218	0.10	0.15	1.601
Layn	rs33773426	9:51057643	A>C	Thr267Pro	-1.316	Neutral	0.63	Tolerated	-2.60	0.40	272	0.03	0.62	0.858
	rs32764902	9:51063251	C>A	Ser196Tyr	-2.179	Neutral	0.01	Deleterious	-1.92	0.25	201	0.11	0.01	1.049
Rdx	rs49859267	9:52081129	T>A	Ser401Thr	-0.202	Neutral	0.39	Tolerated	N/A- no hit					
Zc3h12c	rs51475643	9:52144261	G>A	Val83Ile	-0.16	Neutral	0.26	Tolerated	N/A- no hit					

Table 4.5. Non-synonymous coding SNPs and deleterious protein effects for genes associated with Valid Object Count. The Mouse Genome Informatics Database (MGI; http://www.informatics.jax.org) and the Center for Genome Dynamics website (CGD;; http://cgd.jax.org/cgdsnpdb) were used to examine non-synonymous coding SNPs in candidate genes. Build 137 on the dbSNP website (http://www.ncbi.nlm.nih.gov/ projects/SNP) was used to confirm SNP position (m38) and amino acid substitution for each non-synonymous SNP. PROVEAN, PANTHER Classification System and SIFT (via Ensembl; http://www.ensembl.org) were used to assess that candidate non-synonymous SNPs would cause deleterious effects to protein structure or function. For PROVEAN, a score of -2.5 predicts alterations to the functional effect on the protein. For the PANTHER, a subSPEC (substitution position-specific evolutionary conservation) score of -3 predicts a 50% probability that a score is deleterious ($P_{deleterious}$=0.5), MSA is the number of multiple sequence alignments, NIC (number of independent counts) is an estimate of observations used to calculate the amino acid probabilities. P_{wt} and $P_{substituted}$ (P_{subs}) refer to the respective probabilities of the WT and substituted amino acids. For SIFT, a score <0.05 predicts the SNP to likely cause deleterious effects on the protein. Protein data was considered significant if PROVEAN predicted the protein to likely be deleterious, if PANTHER score ≤ -3, and/or if SIFT score <0.05. Protein data was highlighted in red if the protein was likely deleterious using PROVEAN and/or SIFT, and/or had a PANTHER score ≤-3. Protein data was highlighted in yellow if the PANTHER score was ≤-2 but >-3. Protein data was highlighted in yellow if the PANTHER score was ≤-2 but >-3. No non-synonymous variants, among the 30 selected MEF strains, were identified in *Fdx1* and *Pou2af1* using MGI and CGD resources

Gene	SNP ID (dbSNP Build 137)	Position (m38)	Substitution	Protein Change	PROVEAN Score	Prediction	SIFT	Prediction	Sub SPEC	Pdel	MSA Position	Pwt	Psubs	NIC
Col9a2	rs27518053	4:121039754	C>T	Leu14Phe	-0.946	Neutral	0.05	Tolerated	-2.30	0.33	13	0.55	0.04	0.919
	rs27517992	4:121049711	C>T	Thr298Ile	-0.979	Neutral	0.05	Tolerated	-0.59	0.08	431	0.12	0.18	1.327
	rs48849764	4:121053103	G>A	Ala482Thr	0.71	Neutral	0.49	Tolerated	-1.27	0.15	603	0.21	0.05	0.919
	rs45643752	4:121053104	C>T	Ala482Val	-1.15	Neutral	0.29	Tolerated	-0.36	0.07	603	0.21	0.13	0.919
	rs27500263	4:121054292	G>A	Arg610His	-2.382	Neutral	0.02	Deleterious	-0.76	0.10	871	0.17	0.10	1.327
Edn2	rs27502234	4:120161517	C>G	Ser3Cys	-0.874	Neutral	0.1	Tolerated	-2.32	0.34	43	0.11	0.01	1.397
	rs27502233	4:120161519	G>A	Ala4Thr	-0.341	Neutral	0.1	Tolerated	-1.28	0.15	44	0.28	0.10	1.397
Foxo6	rs46653639	4:120269032	C/T	Ala189Thr	0.529	Neutral	1	Tolerated	-3.75	0.68	84	0.05	0.02	19.67
Hivep3	rs27517176	4:120095806	G>T	Asp4420Tyr	-1.452	Neutral	No SIFT Score		N/A- no hit					
	rs27517175	4:120096389	C>T	Ala634Val	0.571	Neutral	No SIFT Score		N/A- no hit					
	rs27517174	4:120096692	G>A	Gly735Glu	0.719	Neutral	No SIFT Score		N/A- no hit					
	rs27517173	4:120097642	A>G	Ser1052Gly	-1.547	Neutral	No SIFT Score		N/A- no hit					
	rs27517171	4:120098851	G>A	Gly1455Ser	0.742	Neutral	No SIFT Score		N/A- no hit					
	rs27517170	4:120099034	A>C	Met15116Leu	-0.111	Neutral	No SIFT Score		N/A- no hit					
	rs31783440	4:120132367	G>A	Gly2005Glu	-1.095	Neutral	No SIFT Score		N/A- no hit					
Scmh1	rs28256862	4:120508090	A>G	Lys317Arg	-1.668	Neutral	0.03	Deleterious	-2.53	0.39	315	0.43	0.06	2.28
	rs28256861	4:120508182	G>A	Val348Ile	-0.061	Neutral	1	Tolerated	-1.18	0.14	345	0.09	0.13	2.28
Slfnl1	rs28256783	4:120533203	T>C	Val17Ala	-0.595	Neutral	0.3	Tolerated	N/A- no hit					
	rs28256782	4:120533307	A/G	Ser52Gly	1.988	Neutral	1	Tolerated	N/A- no hit					
	rs28256775	4:120535130	T>G	His207Gln	1.465	Neutral	1	Tolerated	N/A- no hit					
Zfp69	rs27485635	4:120930397	A>G	Thr574Ala	-0.329	Neutral	0.82	Tolerated	N/A- no hit					
	rs27485631	4:120931234	C>G	His295Asp	-2.404	Neutral	0.65	Tolerated	-1.54182	0.18875	163	0.18199	0.03284	1.001
	rs27485619	4:120934354	G>A	Gly197Ser	-0.401	Neutral	0.87	Tolerated	-2.77593	0.44422	64	0.71378	0.03472	1.091
	rs32769909	4:120935120	A>G	Gln153Arg	2.012	Neutral	1	Tolerated	-2.00436	0.2698	22	0.47005	0.04965	0.996
	rs27485601	4:120947398	A>C	Glu92Ala	-0.193	Neutral	1	Tolerated	N/A- no hit					
	rs32810951	4:120947503	C>T	Thr57Ile	-0.081	Neutral	1	Tolerated	N/A- no hit					

FIGURES

Figure 4.1. A high-throughput cellular genetics approach to identify QTLs and candidate genes that associate with sorafenib response.

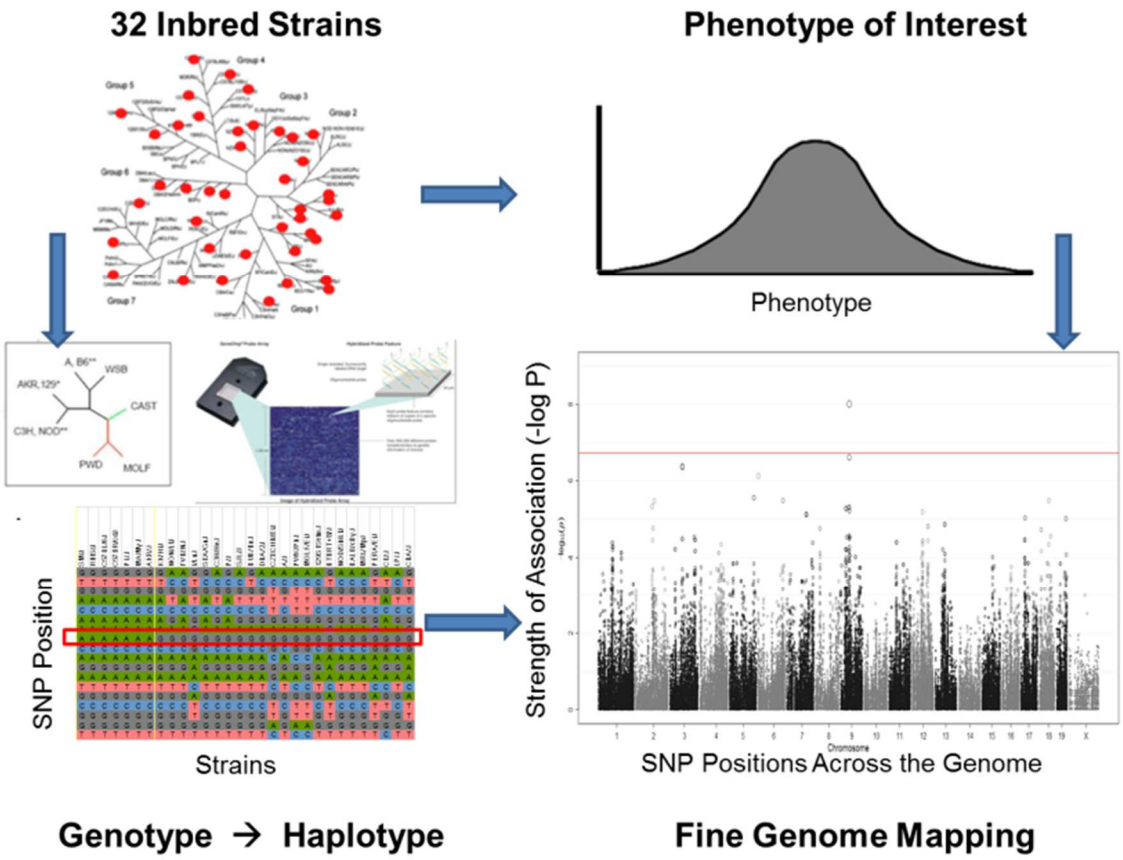


Figure 4.2. Experimental workflow for high-throughput screening of MEF cell lines in a concentration-response format after administration of sorafenib. A) Layout of 32 MEF strains on a single 384-well microtiter plate. 1500 cells/well for 24 h time point plates, and 1000 cells/well for 72 h time point plates in a total of 50 μL of media were dispensed into each well. A total of 200 nL of ten sorafenib concentrations (0-300 μM) was added to each of ten wells. **B)** Cell staining reagents were added, incubated and then imaged via high-content imaging. From left to right: image 1 was nuclei staining representing cell viability, image 2 was membrane permeability, image 3 was mitochondrial membrane potential, and image 4 was cytochrome C antibody staining. **C)** Images were analyzed using Cellomics vHCS Toolbox Compartmental Analysis software. **D)** Dose response curves for each of the 32 MEF strains are generated by F-curve software for each of the four cell health phenotypes. IC_{50} values from the dose response curves are then used for GWAS. This figure was adapted, with permission, from Suzuki, et al. *Front Genet*. 2014;5:272. Abbreviations: GWAS, genome wide association studies; HCS, high content screening; IC_{50}, half maximal inhibitory concentration; MEF, mouse embryonic fibroblast;

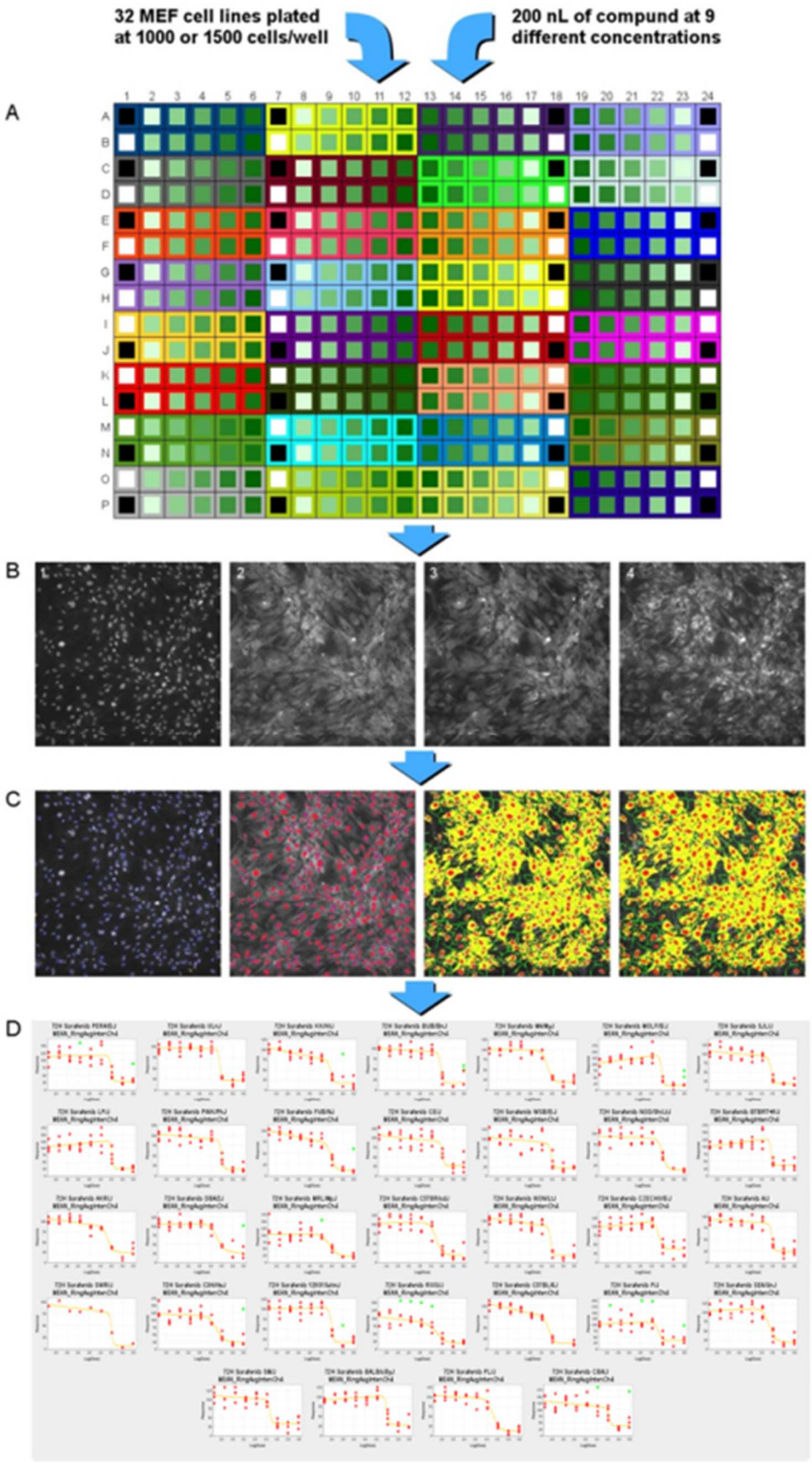

32 MEF cell lines plated
at 1000 or 1500 cells/well
200 nL of compund at 9
different concentrations
A
B
C
D

Figure 4.3. Dose response curves and IC_{50} values generated using a Brain-Cousens model. Endpoints with a downward trend produced IC_{80} to IC_{50} values, decreasing in 5%-effect steps. Cytochrome C release from MA/MyJ MEF cells after 72 hour sorafenib incubation is displayed here. Abbreviations: IC_{50}, half maximal inhibitory concentration; IC_{80}, concentration to achieve 80% inhibition; MEF, mouse embryonic fibroblasts.

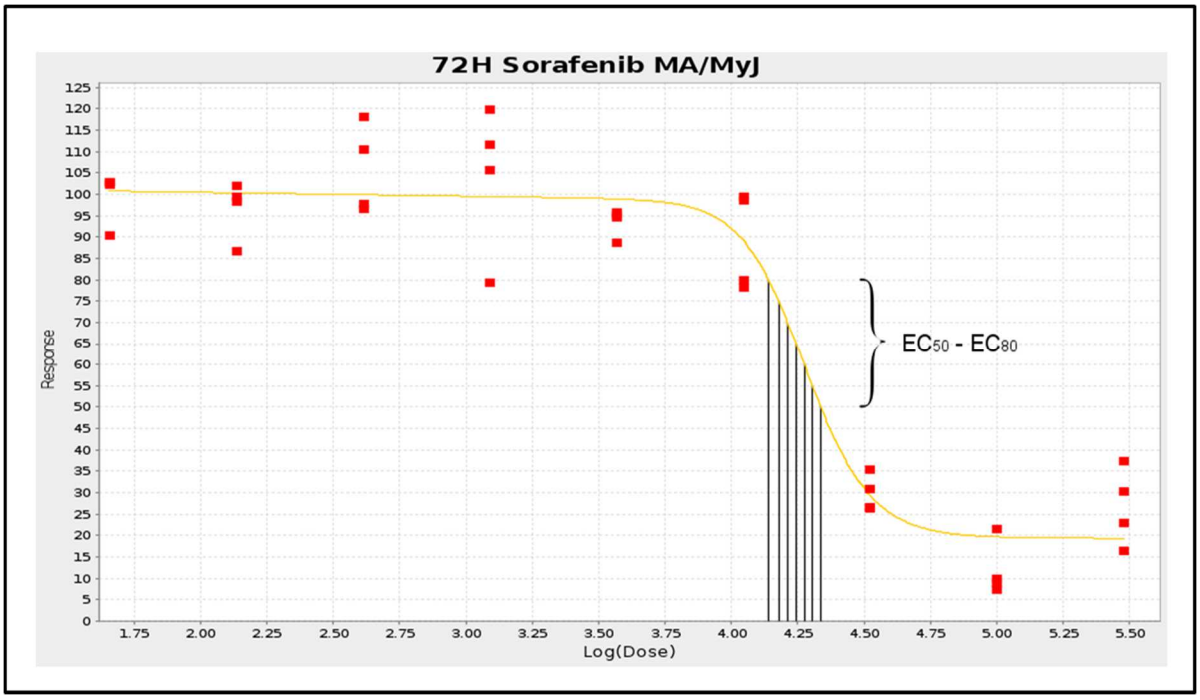


Figure 4.4. Example of QTL interval selection. A) Up to seven SNPster GWA results, derived from IC_{50} through IC_{80}, were used to account for slight differences in dose-response curve slopes. **B)** The 2% SNPs with highest –logP scores were averaged between the SNPster results to detect the SNPs that have consistently higher association scores. **C)** The region surrounding SNPs with an average –logP greater than 4.0 was determined as the QTL. If SNPs were located within 1 Mb of each other, the entire interval between them was included. Additionally, 100 kb on each side of the outermost SNPs were also included. The resulting interval, in this diagram, is indicated by the blue bar. Abbreviations: GWA, genome wide association; IC_{50}, half maximal inhibitory concentration; IC_{80}, concentration to achieve 80% inhibition; QTL, quantitative trait locus/loci; SNP, single nucleotide polymorphism.

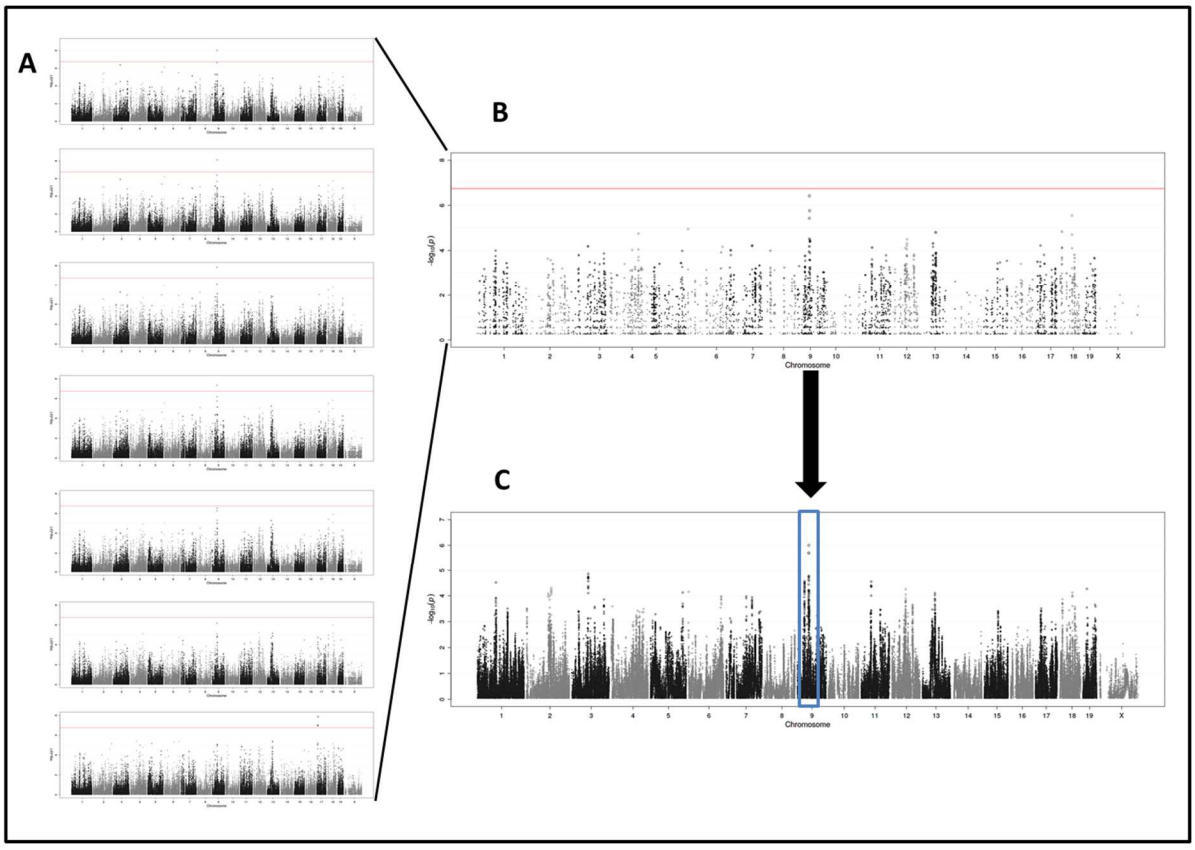


Figure 4.5. Example of candidate gene selection from GWAS. A) GWA Manhattan plot (here Cytochrome C release after 72 h sorafenib incubation is presented). The region with the highest –logP value is selected for further analysis. **B)** Detailed region under the identified QTL with candidate genes. **C)** Haplotype structure of the inbred MEF strains in the QTL region. Abbreviations: Genome wide association; GWAS, genome wide association studies, MEF, mouse embryonic fibroblasts; QTL, quantitative trait locus/loci.

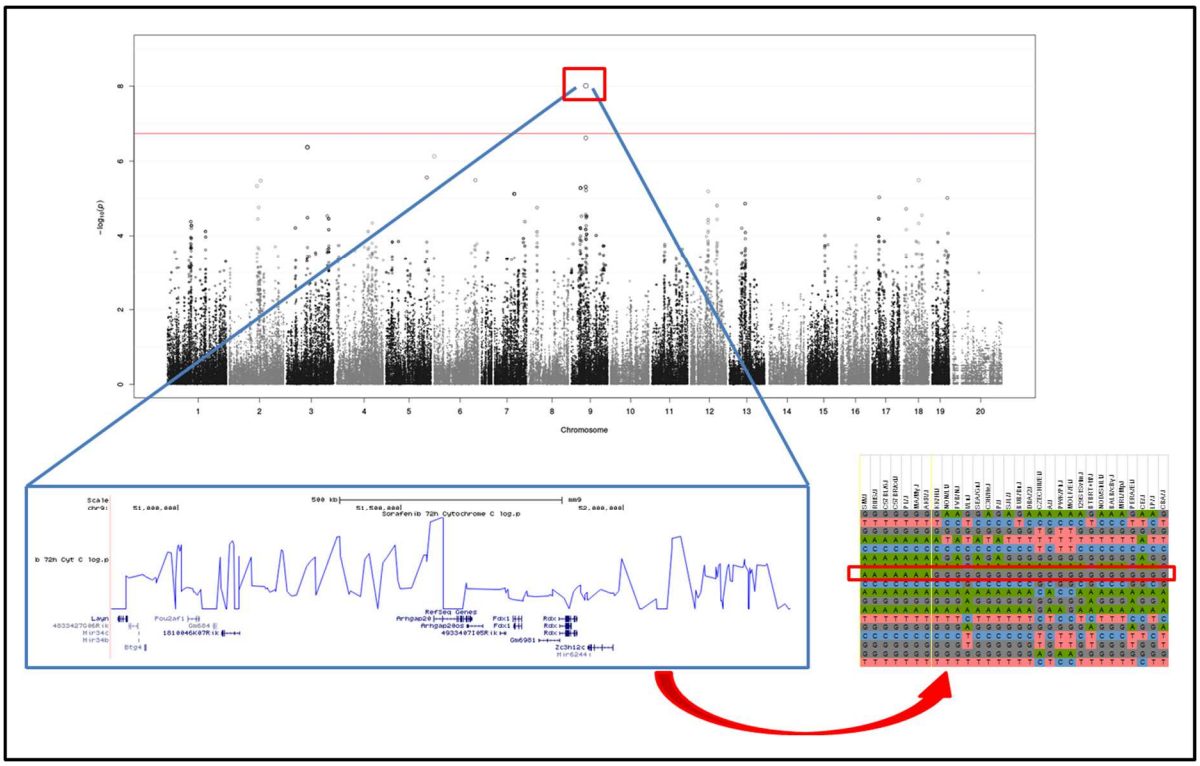


Figure 4.6. A multi-step and multi-faceted schematic for candidate gene selection.

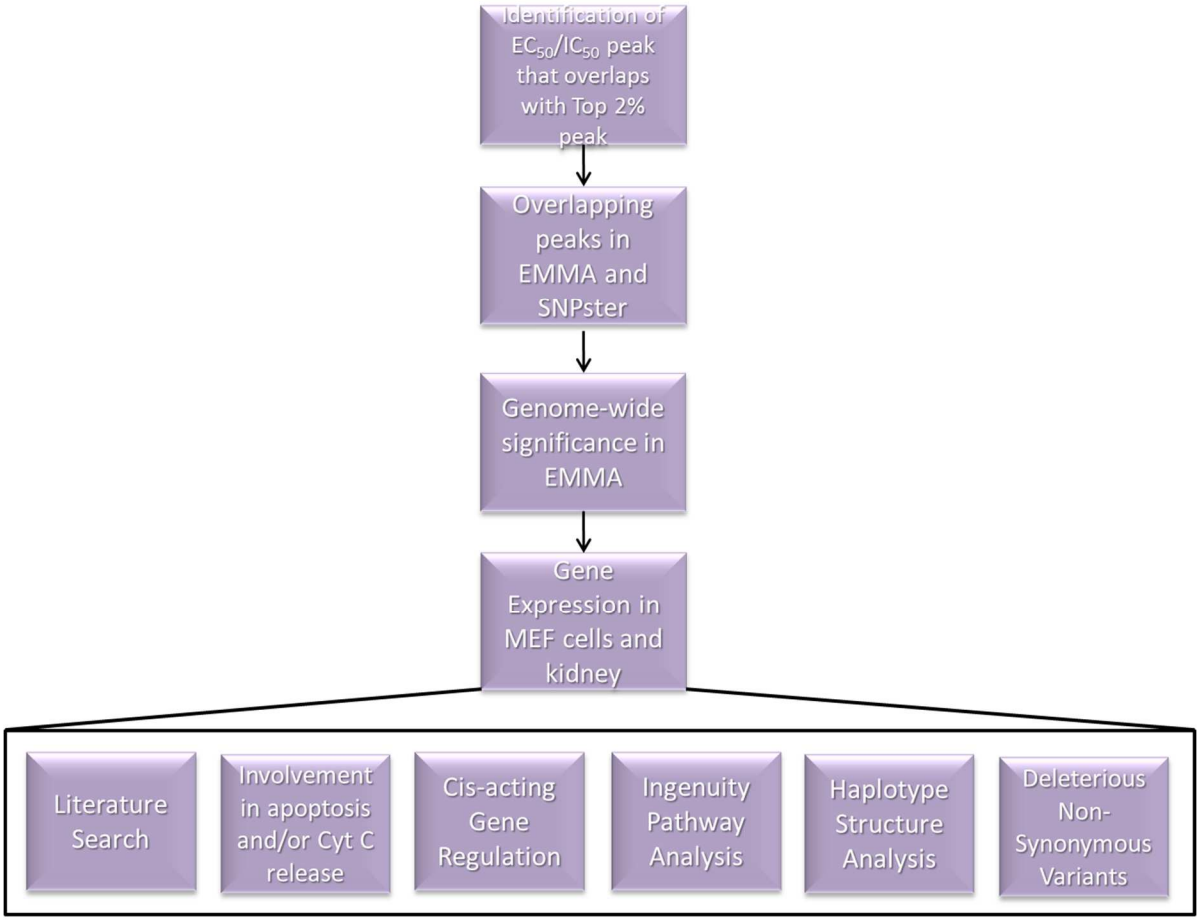


Figure 4.7. Distribution of IC_{50} values associated with Cytochrome C release across 32 MEF strains. The distribution of Cytochrome C release IC_{50} values shows variability (17-45μM) across 32 MEF strains. These data were then used as the basis for GWAS. Thirty of the 32 strains represented here were used for GWAS, with only SWR/J, and WSB/EiJ excluded from final analyses. Abbreviations: GWAS, genome wide association studies; IC_{50}, half maximal inhibitory concentration; MEF, mouse embryonic fibroblasts.

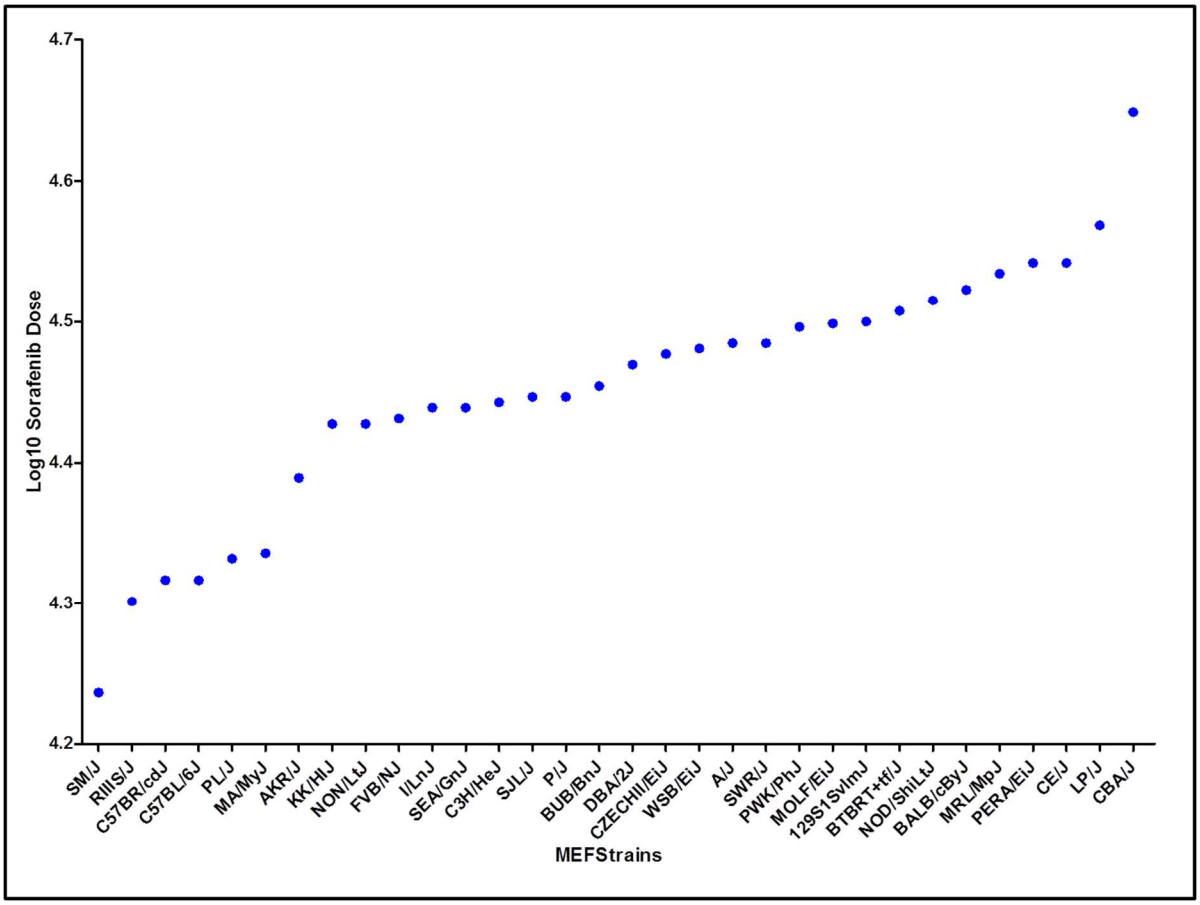


Figure 4.8. Distribution of IC_{50} values associated with Valid Object Count release across 32 MEF strains. The distribution of Valid Object Count IC_{50} values shows variability (17-32μM) across 32 MEF strains. These data were then used as the basis for GWAS. Thirty of the 32 strains represented here were used for GWAS, with only SWR/J, and WSB/EiJ excluded from final analyses. Abbreviations: GWAS, genome wide association studies; IC_{50}, half maximal inhibitory concentration; MEF, mouse embryonic fibroblasts.

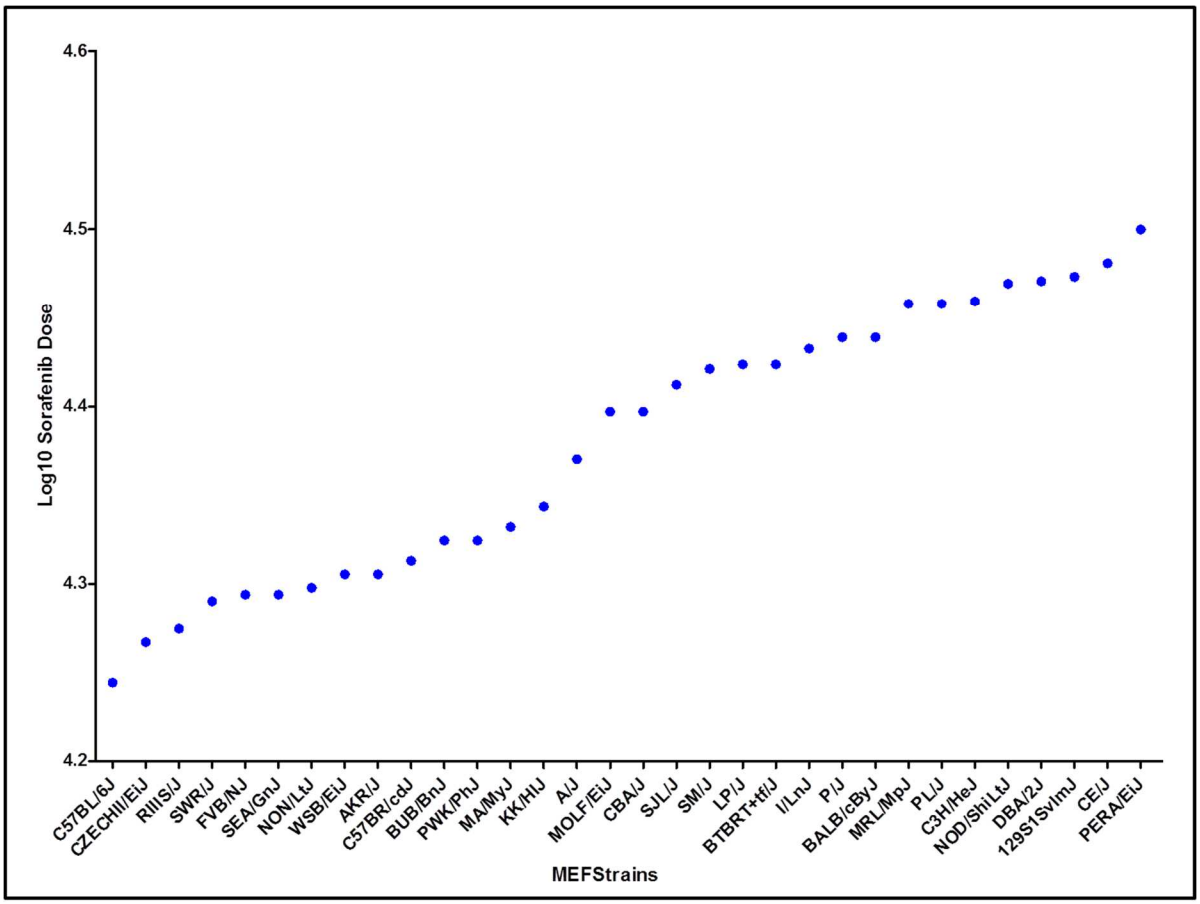


Figure 4.9. Manhattan plots for Cytochrome C release. GWA Manhattan plots were generated for MEF cells exposed to sorafenib for 72 hours and associated with Cytochrome C release. Panel A shows results generated from the EMMA algorithm and panel B shows results generated from SNPster. One notable QTL with a –logP value ≥4.0 was identified (chromosome 9, position 50,500,000-52,500,000). This QTL reached genome-wide significance by the EMMA algorithm with a –logP value ≥6.74. Abbreviations: EMMA, efficient mixed model association; GWA, genome wide association; MEF, mouse embryonic fibroblasts, QTL, quantitative trait locus/loci.

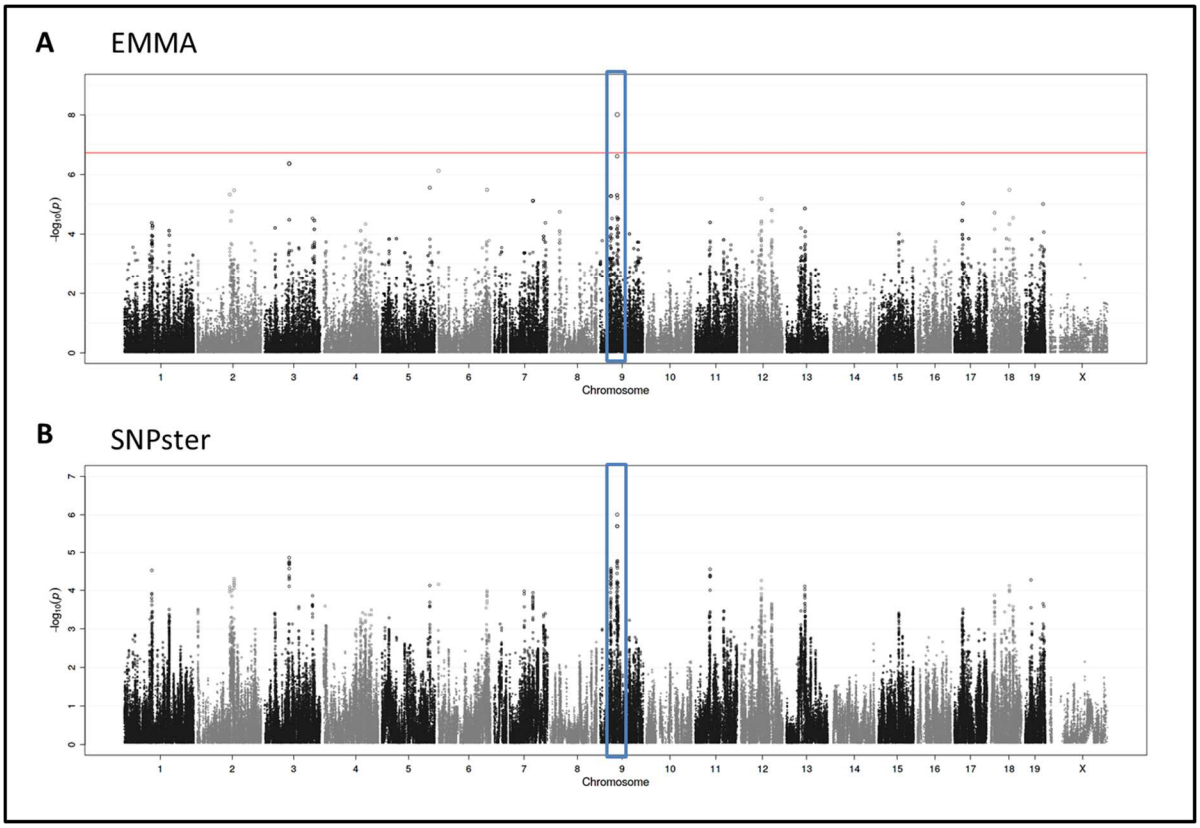


Figure 4.10. Genes under the QTL associated with Cytochrome C release. A total of seven genes were identified under a QTL on chromosome 9 (50,500,000-52,500,000). Abbreviations: QTL, quantitative trait locus/loci.

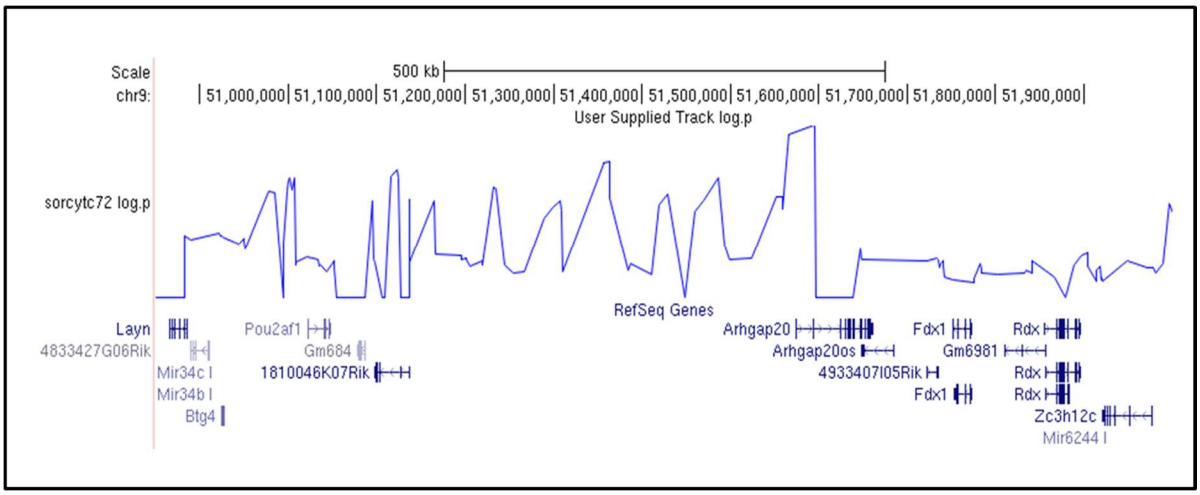


Figure 4.11. Manhattan plots for Valid Object Count. GWA Manhattan plots were generated for MEF cells exposed to sorafenib for 72 hours and associated with VOC. Panel A shows results generated from the EMMA algorithm and panel B shows results generated from SNPster. One notable QTL with a –logP ≥4.0 was identified (chromosome 4, position 119,000,000-121,500,000). This QTL was notable, albeit not genome-wide significant, by the EMMA algorithm with a –logP value ≥4.66. Abbreviations: EMMA, efficient mixed model association; GWA, genome wide association; MEF, mouse embryonic fibroblasts, QTL, quantitative trait locus/loci; VOC, valid object count.

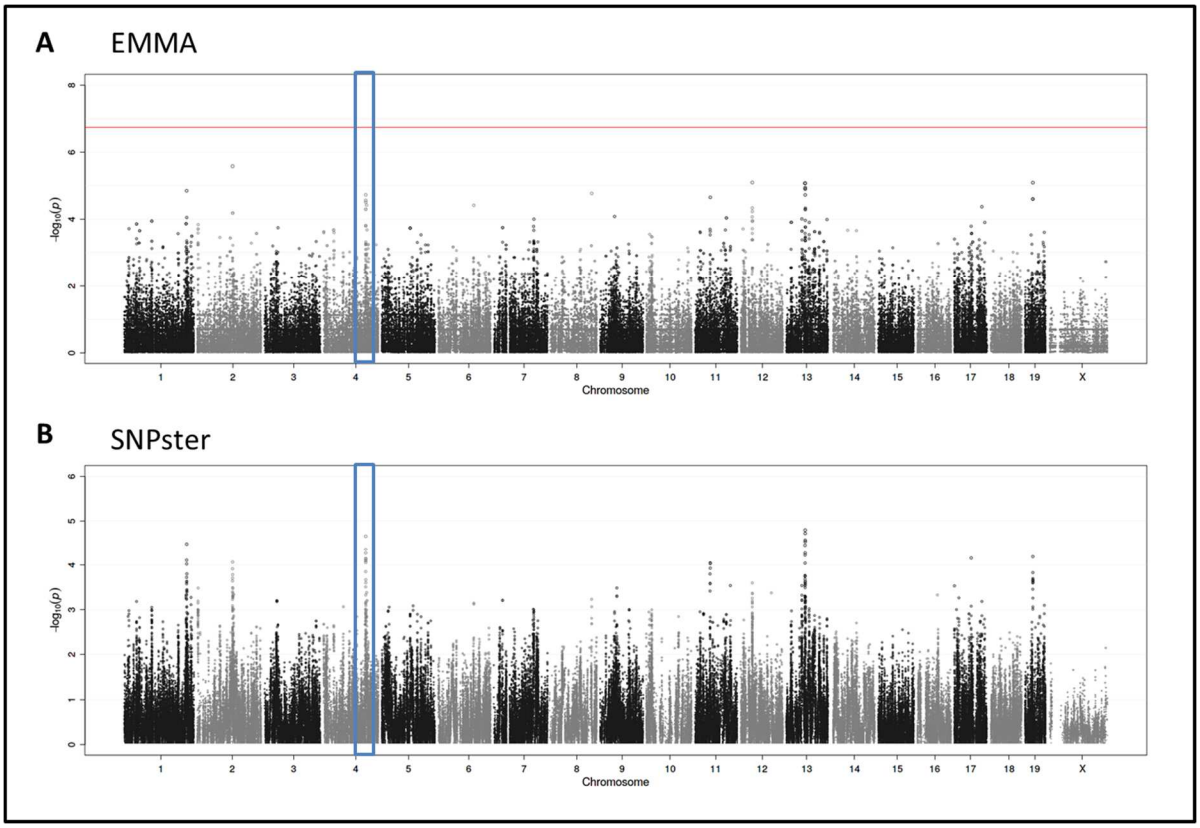


Figure 4.12. Genes under the QTL associated with Valid Object Count. A total of 13 genes were identified under a QTL on chromosome 4 (position 119,000,000-121,500,000). Abbreviations: QTL, quantitative trait locus/loci.

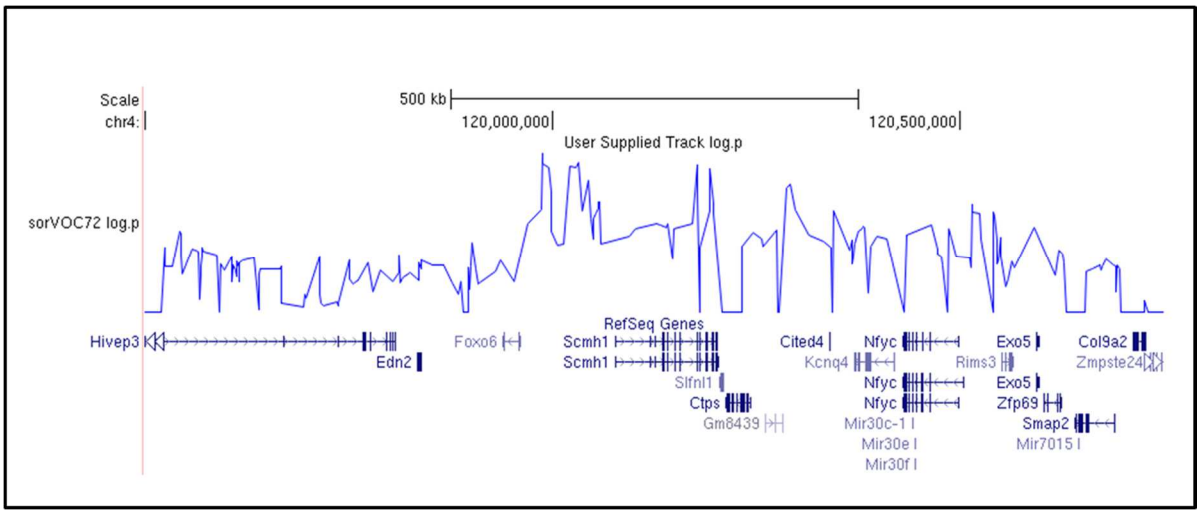


Figure 4.13. Haplotype structure for genes associated with the Cytochrome C release QTL. The haplotype structure of select inbred mouse strains within chromosome 9 (position 50,500,000-52,500,000), which represents the areas directly under and surrounding the genome-wide significant peak observed for this phenotype after 72 h incubation of sorafenib. Strains are shown in descending order from most sensitive to least sensitive. The blue arrow represents the location directly under the peak (position 51595846), the red arrow represents the location of *Arhgap20* (position 51573457-51661164), and the green arrow represents the location of *Rdx* (position 51855255-51896843). The haplotype structure was visualized with the Mouse Phylogeny Viewer (https://msub.csbio.unc.edu/). Abbreviations: QTL, quantitative trait locus/loci.

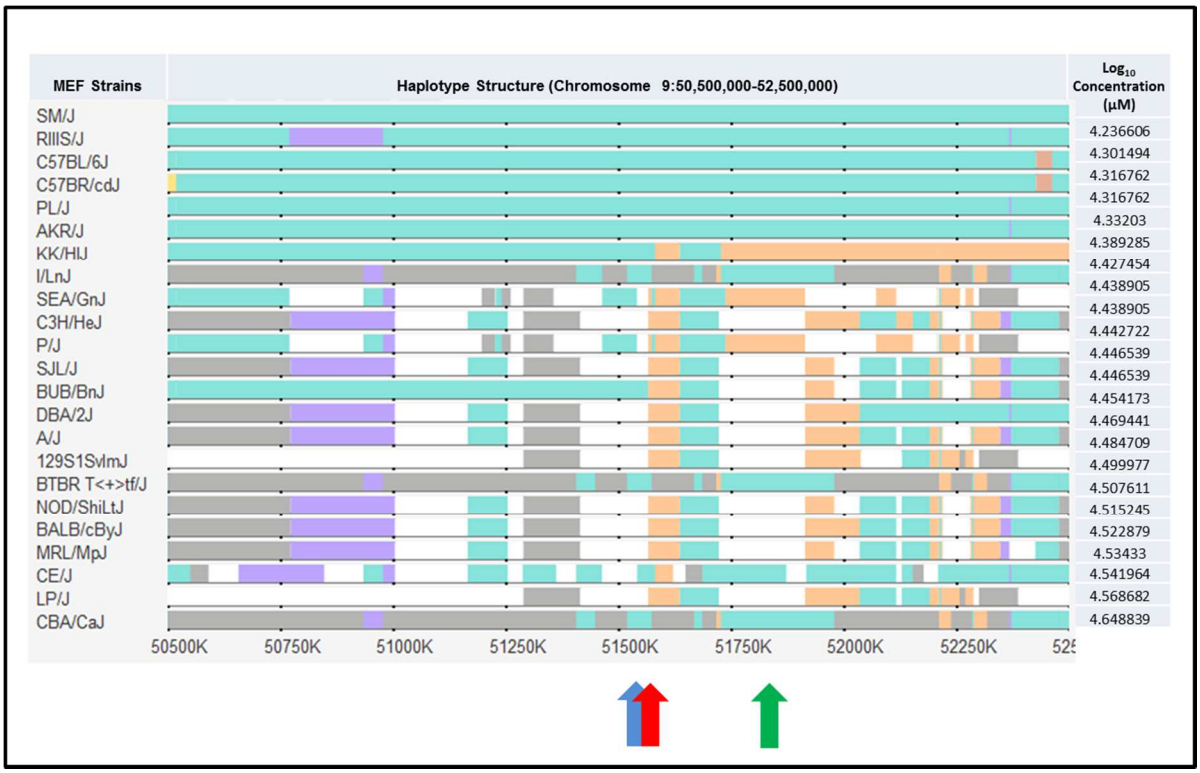


Figure 4.14. Ingenuity pathway analysis for genes with the QTL associated with Cytochrome C release. Seven genes within the QTL associated with Cytochrome C release were input into an Ingenuity Pathway Analysis with sorafenib and known targets of sorafenib (*BRAF, FLT-1, FLT-3, FLT-5, FGFR1, KIT, KDR, PDGFR-β, RET,* and *RAF-1*). Only direct connections were included. *Zc3h12c* is not represented in this analysis because no direct connections to sorafenib and/or sorafenib targets were identified. A direct connection between a candidate gene and sorafenib or a sorafenib target is denoted in red. Abbreviations: QTL, quantitative trait locus/loci.

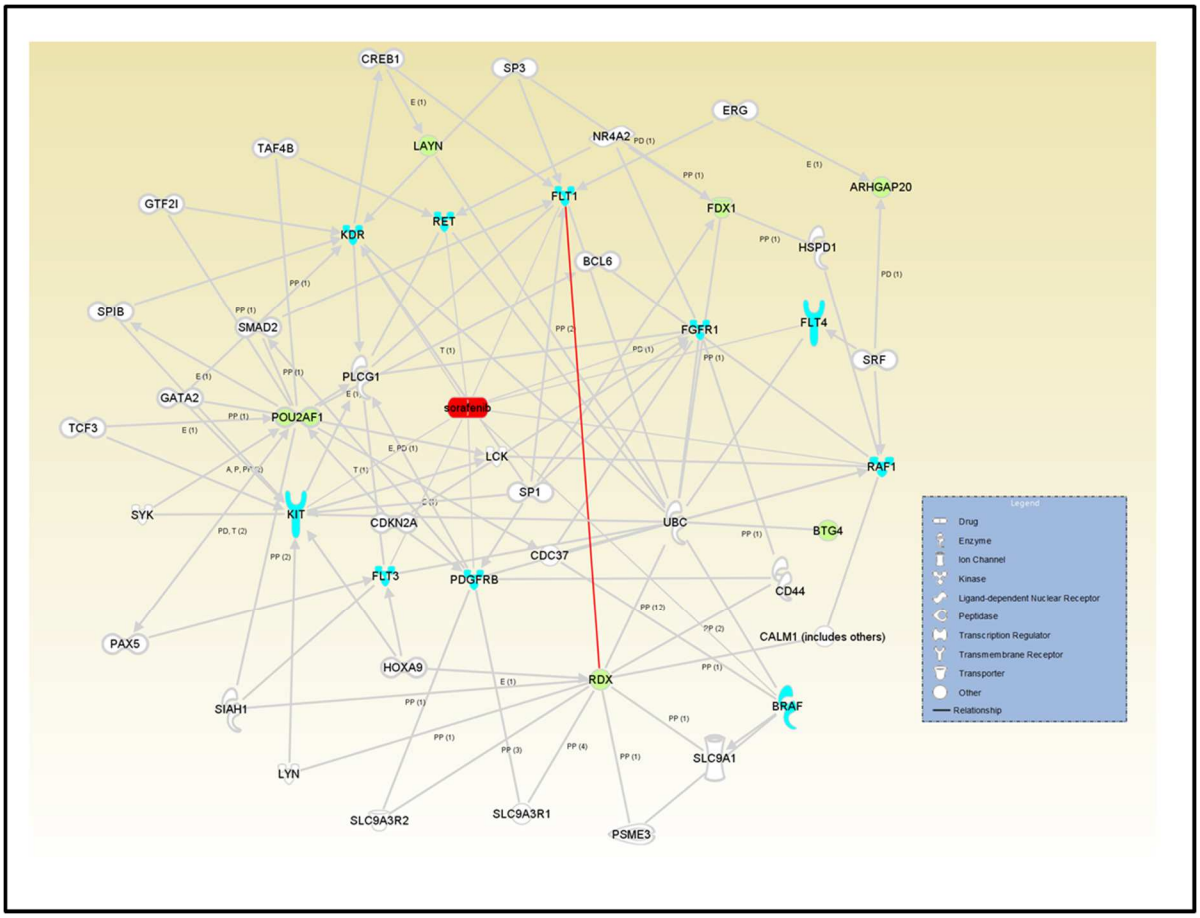


Figure 4.15. Haplotype structure for genes associated with the Valid Object Count QTL. The haplotype structure of the inbred mouse strains within chromosome 4 (position 119,000,000-121,500,000), which represents the areas directly under and surrounding the highest peak observed for this phenotype after 72 h incubation of sorafenib. Strains are shown in descending order from most sensitive to least sensitive. The blue arrow represents the location directly under the peak (position 119987827), the red arrow represents the location of *Scmh1* (position 120077886-120202804), and the green arrow represents the location of *Nfyc* (position 120430040-120498346). The haplotype structure was visualized with the Mouse Phylogeny Viewer (https://msub.csbio.unc.edu/). Abbreviations: QTL, quantitative trait locus/loci.

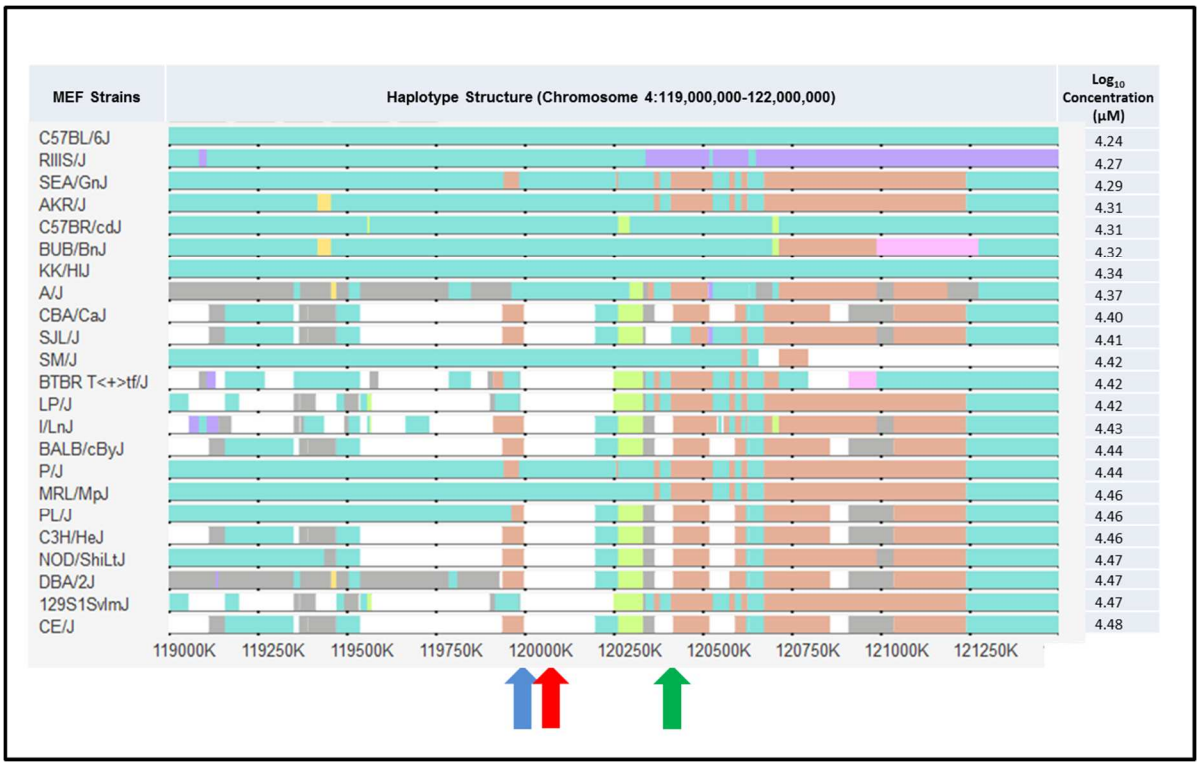
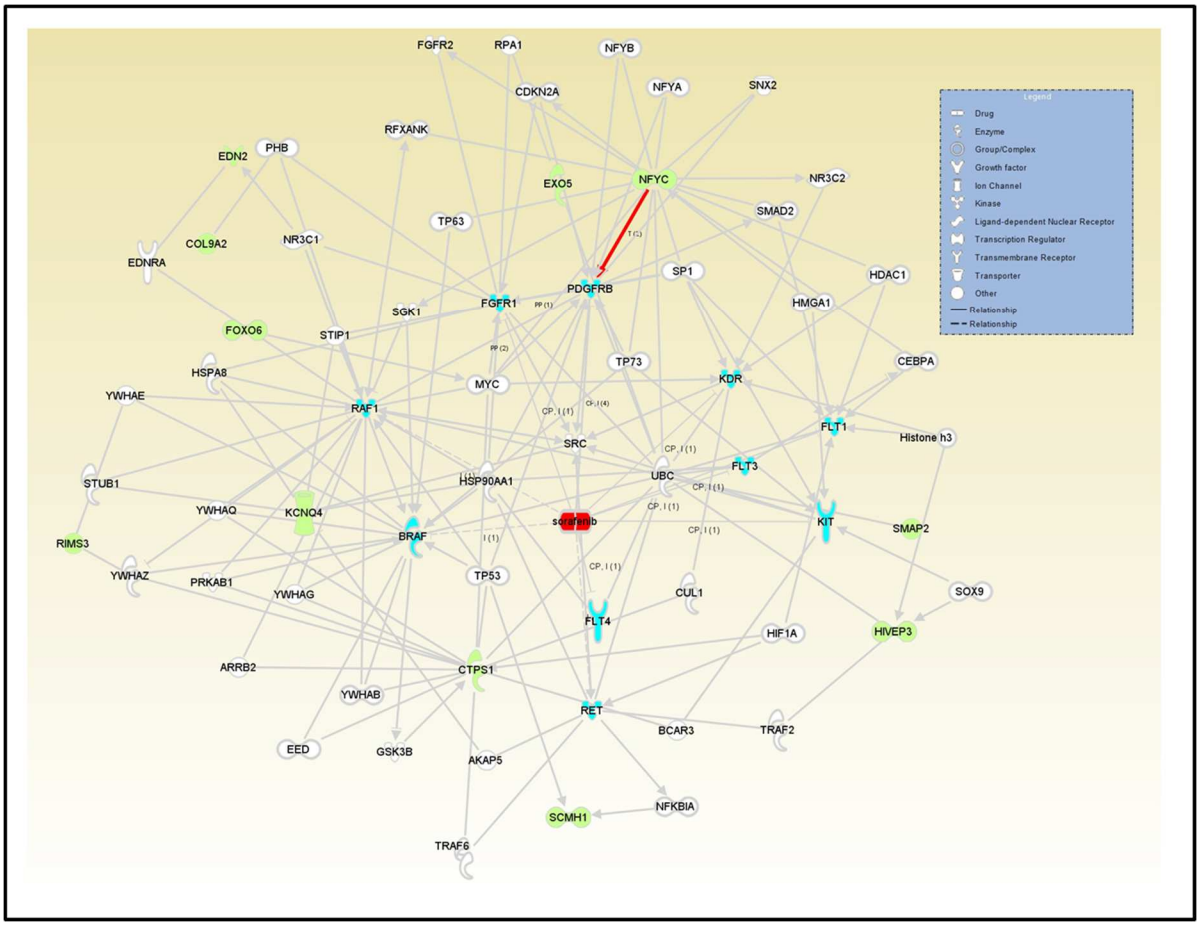


Figure 4.16. Ingenuity pathway analysis for genes with the QTL associated with Valid Object Count. Thirteen genes within the QTL associated with VOC were input into an Ingenuity Pathway Analysis with sorafenib and known targets of sorafenib (*BRAF, FLT-1, FLT-3, FLT-5, FGFR1, KIT, KDR, PDGFR-β, RET,* and *RAF-1*). Only direct connections were included. *Zfp69* and *Slfnl1* are not represented in this analysis because no direct connections to sorafenib and/or sorafenib targets were identified. A direct connection between a candidate gene and sorafenib or a sorafenib target is denoted in red. Abbreviations: QTL, quantitative trait locus/loci.



REFERENCES

1. Hampton T. "Promiscuous" anticancer drugs that hit multiple targets may thwart resistance. *JAMA*. 2004;292:419-22.

2. Product information. Nexavar (sorafenib). Wayne NBHP, Inc. June 2013. Accessed January 29, 2015. Available at: http://berlex.bayerhealthcare.com/html/products/pi/Nexavar_PI.pdf.

3. Wan PT, Garnett MJ, Roe SM, et al. Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. *Cell*. 2004;116:855-67.

4. Wilhelm SM, Carter C, Tang L, et al. BAY 43-9006 exhibits broad spectrum oral antitumor activity and targets the RAF/MEK/ERK pathway and receptor tyrosine kinases involved in tumor progression and angiogenesis. *Cancer Res*. 2004;64:7099-109.

5. Yu C, Bruzek LM, Meng XW, et al. The role of Mcl-1 downregulation in the proapoptotic activity of the multikinase inhibitor BAY 43-9006. *Oncogene* .2005;24:6861-9.

6. Wilhelm S, Chien DS. BAY 43-9006: preclinical data. *Curr Pharm Des*. 2002;8:2255-7.

7. Wilhelm SM, Adnane L, Newell P, Villanueva A, Llovet JM, Lynch M. Preclinical overview of sorafenib, a multikinase inhibitor that targets both Raf and VEGF and PDGF receptor tyrosine kinase signaling. Mol Cancer Ther. 2008;7:3129-40.

8. Lin X, Huang XP, Chen G, et al. Life beyond kinases: structure-based discovery of sorafenib as nanomolar antagonist of 5-HT receptors. *J Med Chem* .2012;55:5749-59.

9. Yuen JS, Sim MY, Siml HG, et al. Inhibition of angiogenic and non-angiogenic targets by sorafenib in renal cell carcinoma (RCC) in a RCC xenograft model. *Br J Cancer*. 2011;104:941-7.

10. Cabrera R, Ararat M, Xu Y, et al. Immune modulation of effector CD4+ and regulatory T cell function by sorafenib in patients with hepatocellular carcinoma. Cancer Immunol Immunother. 2013;62:737-46.

11. Hipp MM, Hilf N, Walter S, et al. Sorafenib, but not sunitinib, affects function of dendritic cells and induction of primary immune responses. *Blood* .2008;111:5610-20.

12. Bogue MA, Grubb SC. The Mouse Phenome Project. *Genetica*. 2004;122:71-4.

13. Grupe A, Germer S, Usuka J, et al. In silico mapping of complex disease-related traits in mice. *Science*. 2001;292:1915-8.

14. Liang DY, Liao G, Wang J, et al. A genetic analysis of opioid-induced hyperalgesia in mice. *Anesthesiology*. 2006;104:1054-62.

15. Liu P, Vikis H, Lu Y, Wang D, You M. Large-scale in silico mapping of complex quantitative traits in inbred mice. *PLoS One*. 2007;2:e651.

16. Harrill AH, Watkins PB, Su S, et al. Mouse population-guided resequencing reveals that variants in CD44 contribute to acetaminophen-induced liver injury in humans. *Genome Res* .2009;19:1507-15.

17. Guo Y, Weller P, Farrell E, et al. In silico pharmacogenetics of warfarin metabolism. *Nat Biotechnol*. 2006;24:531-6.

18. Suzuki OT, Frick A, Parks BB, et al. A cellular genetics approach identifies gene-drug interactions and pinpoints drug toxicity pathway nodes. *Front Genet*. 2014;5:272.

19. Frick A FY, Richards K, Damania B, Parks B, Suzuki O, Benton CS, Chan E, Thomas RS, Wiltshire T. Immune cell-based sreening assay for response to anticancer agents: applications in pharmacogenomics. . *Pharmacogenomics and Personalized Medicine*. 2015;8:81-98.

20. van der Veldt AA, Eechoute K, Gelderblom H, et al. Genetic polymorphisms associated with a prolonged progression-free survival in patients with metastatic renal cell cancer treated with sunitinib. *Clin Cancer Res*. 2011;17:620-9.

21. van Erp NP, Eechoute K, van der Veldt AA, et al. Pharmacogenetic pathway analysis for determination of sunitinib-induced toxicity. *J Clin Oncol*. 2009;27:4406-12.

22. Scartozzi M, Bianconi M, Faloppi L, et al. VEGF and VEGFR polymorphisms affect clinical outcome in advanced renal cell carcinoma patients receiving first-line sunitinib. *Br J Cancer*. 2013;108:1126-32.

23. Beuselinck B, Karadimou A, Lambrechts D, et al. Single-nucleotide polymorphisms associated with outcome in metastatic renal cell carcinoma treated with sunitinib. *Br J Cancer*. 2013;108:887-900.

24. Garcia-Donas J, Esteban E, Leandro-Garcia LJ, et al. Single nucleotide polymorphism associations with response and toxic effects in patients with advanced renal-cell carcinoma treated with first-line sunitinib: a multicentre, observational, prospective study. *Lancet Oncol*. 2011;12:1143-50.

25. Xu CF, Bing NX, Ball HA, et al. Pazopanib efficacy in renal cell carcinoma: evidence for predictive genetic markers in angiogenesis-related and exposure-related genes. *J Clin Oncol.* 2011;29:2557-64.

26. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, 2012. Available at: http://www.R-project.org..

27. Brain P, Cousens R. An equation to describe dose responses where there is stimulation of growth at low doses. *Weed Research*. 1989;29:285-91.

28. McClurg P, Janes J, Wu C, et al. Genomewide association analysis in diverse inbred mice: power and population structure. *Genetics*. 2007;176:675-83.

29. Kang HM, Zaitlen NA, Wade CM, et al. Efficient control of population structure in model organism association mapping. *Genetics*. 2008;178:1709-23.

30. Yang H, Wang JR, Didion JP, et al. Subspecific origin and haplotype diversity in the laboratory mouse. *Nat Genet*. 2011;43:648-55.

31. Waterston RH, Lindblad-Toh K, Birney E, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature*. 2002;420:520-62.

32. Moreau Y, Tranchevent LC. Computational tools for prioritizing candidate genes: boosting disease gene discovery. *Nat Rev Genet.* 2012;13:523-36.

33. Flicek P, Ahmed I, Amode MR, et al. Ensembl 2013. *Nucleic Acids Res*. 2013;41:D48-55.

34. Dreszer TR, Karolchik D, Zweig AS, et al. The UCSC Genome Browser database: extensions and updates 2011. *Nucleic Acids Res*. 2012;40:D918-23.

35. Thomas S, Bonchev D. A survey of current software for network analysis in molecular biology. *Hum Genomics*. 2010;4:353-60.

36. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4:1073-81.

37. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One*. 2012;7:e46688.

38. Thomas PD, Campbell MJ, Kejariwal A, et al. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res*. 2003;13:2129-41.

39. Mi H, Lazareva-Ulitsky B, Loo R, et al. The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res*. 2005;33:D284-8.

40. Wang JR, de Villena FP, McMillan L. Comparative analysis and visualization of multiple collinear genomes. *BMC Bioinformatics*. 2012;13 Suppl 3:S13.

41. Paigen K, Eppig JT. A mouse phenome project. *Mamm Genome.* 2000;11:715-7.

42. Zheng JF, Sun LC, Liu H, Huang Y, Li Y, He J. EBP50 exerts tumor suppressor activity by promoting cell apoptosis and retarding extracellular signal-regulated kinase activity. *Amino Acids*. 2010;38:1261-8.

43. Hayashi Y, Molina JR, Hamilton SR, Georgescu MM. NHERF1/EBP50 is a new marker in colorectal cancer. *Neoplasia*. 2010;12:1013-22.

44. Bartholow TL, Becich MJ, Chandran UR, Parwani AV. Immunohistochemical analysis of ezrin-radixin-moesin-binding phosphoprotein 50 in prostatic adenocarcinoma. *BMC Urol*. 2011;11:12.

45. Lin YY, Hsu YH, Huang HY, et al. Aberrant nuclear localization of EBP50 promotes colorectal carcinogenesis in xenotransplanted mice by modulating TCF-1 and beta-catenin interactions. *J Clin Invest.* 2012;122:1881-94.

46. Claperon A, Guedj N, Mergey M, et al. Loss of EBP50 stimulates EGFR activity to induce EMT phenotypic features in biliary cancer cells. *Oncogene*. 2012;31:1376-88.

47. Peng XL, Ji MY, Yang ZR, Song J, Dong WG. Tumor suppressor function of ezrin-radixin-moesin-binding phosphoprotein-50 through beta-catenin/E-cadherin pathway in human hepatocellular cancer. *World J Gastroenterol.* 2013;19:1306-13.

48. Hsu YH, Lin WL, Hou YT, et al. Podocalyxin EBP50 ezrin molecular complex enhances the metastatic potential of renal cell carcinoma through recruiting Rac1 guanine nucleotide exchange factor ARHGEF7. *Am J Pathol.* 2010;176:3050-61.

49. Pignochino Y, Grignani G, Cavalloni G, et al. Sorafenib blocks tumour growth, angiogenesis and metastatic potential in preclinical models of osteosarcoma through a mechanism potentially involving the inhibition of ERK1/2, MCL-1 and ezrin pathways. *Mol Cancer.* 2009;8:118.

50. Romier C, Cocchiarella F, Mantovani R, Moras D. The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y. *J Biol Chem*. 2003;278:1336-45.

51. Hackzell A, Uramoto H, Izumi H, Kohno K, Funa K. p73 independent of c-Myc represses transcription of platelet-derived growth factor beta-receptor through interaction with NF-Y. J Biol Chem 2002;277:39769-76.

52. Wu HH, Su B. Adaptive evolution of SCML1 in primates, a gene involved in male reproduction. *BMC Evol Biol*. 2008;8:192.

53. Yasunaga S, Ohtsubo M, Ohno Y, et al. Scmh1 has E3 ubiquitin ligase activity for geminin and histone H2A and regulates geminin stability directly or indirectly via transcriptional repression of Hoxa9 and Hoxb4. *Mol Cell Biol*. 2013;33:644-60.

54. Montanari M, Macaluso M, Cittadini A, Giordano A. Role of geminin: from normal control of DNA replication to cancer formation and progression? Cell Death Differ.. 2006;13:1052-6.

55. Zhu W, Depamphilis ML. Selective killing of cancer cells by suppression of geminin activity. *Cancer Re*s. 2009;69:4870-7.

56. Yagi T, Inoue N, Yanai A, et al. Prognostic significance of geminin expression levels in Ki67-high subset of estrogen receptor-positive and HER2-negative breast cancers. *Breast Cancer*. 2014 Aug [Epub ahead of print].

57. Tokes AM, Szasz AM, Geszti F, et al. Expression of proliferation markers Ki67, cyclin A, geminin and aurora-kinase A in primary breast carcinomas and corresponding distant metastases. *J Clin Pathol*. 2015;68:274-82.

58. Zhang X, Liu HH, Weller P, et al. In silico and in vitro pharmacogenetics: aldehyde oxidase rapidly metabolizes a p38 kinase inhibitor. *Pharmacogenomics J*. 2011;11:15-24.

CHAPTER 5: DISCUSSION, PERSPECTIVE AND FUTURE DIRECTIONS

5.1 Summary and Scope

Over the past decade the treatment landscape for mRCC has changed dramatically due to the U.S. FDA approval of multiple agents that target tumorigenic and angiogenic pathways. The approval of seven agents, which target the VEGF-pathway (axitinib, pazopanib, sorafenib and sunitinib), and mTOR (everolimus and temsirolimus) have helped increase median survival time amongst mRCC patients.[1-8] Nevertheless, despite these major advancements, many mRCC patients experience disease progression while on treatment, and eventually die from their cancer. There is currently a shortage of clinically relevant prognostic and predictive pharmacogenetic markers that have been translated for routine incorporation into clinical practice.[9-11]

A "one-size-fits-all" paradigm might apply well to the Frank Zappa album of the same name, or might be appropriate for clothing like sweatpants and socks; however, it is a poor way to approach the treatment of advanced and metastatic cancer patients. It has been well established that, for cancer patients, responses to treatment vary greatly,[12-14] and it is understood that genetic variations in therapeutic targets of medications can lead to altered treatment efficacy.[15] Pharmacogenetics examines interindividual genetic variability that influences the course of drug action so that medication regimens may be optimized to maximize response, while minimizing drug-induced toxicity. Personalized medicine through genetics and genomics is the overarching goal for the field of pharmacogenetics and vital to

the evolution of effective treatment paradigms (Appendix 1).[16] Currently, there is an unmet need, yet an opportunity, to identify and validate predictive and prognostic pharmacogenetic biomarkers that can help clinicians individualize treatment strategies for mRCC patients.

Sorafenib is a multikinase inhibitor with potent activity against angiogenic, oncogenic, and stromal kinases, as well as the RAF/MEK/ERK signaling pathway, which leads to inhibition of tumor proliferation and angiogenesis. Data from the pivotal phase III randomized, placebo-controlled, TARGET trial confirmed a significant OS and PFS benefit. Several clinical trials focused on sorafenib have revealed a high degree of interindividual variability in sorafenib pharmacokinetics.[17-21] Considerable interindividual variability in response to sorafenib is also observed clinically.[2,22,23]

This dissertation research stemmed from the hypothesis that germline genetic variants in mRCC patients will help explain the interindividual differences in sorafenib response and patient survival. Thus, the primary objective of this dissertation was to identify and validate germline genetic variants that are predictive of sorafenib response, as well as variants that are prognostic markers of survival in mRCC patients. The key findings of this work include: 1) the identification of five germline variants associated with OS in patients from the sorafenib and two variants associated with OS in a combined analysis of both treatment arms, 2) the validation of four intronic variants (rs3816375 in *ITGAV*, rs1885657 and rs3024987 in *VEGFA*, and rs8047917 in *WWOX*) in dual reporter gene luciferase assays, as well as the validation of one exonic variant (rs307826 in *FLT-4*) in a cell viability assay, and 3) the identification of a genome-wide significant QTL on mouse chromosome 9 associated with cytochrome C release, and a second QTL on mouse chromosome 4 associated with cell viability, in 32 MEF strains treated with sorafenib for 72 h. The major translational

implication of this dissertation research is that it provides a foundation for future laboratory and replication studies to confirm the analytical validity of these findings, followed by prospective clinical trials that will evaluate both the clinical validity and clinical utility of these variants as predictive biomarkers of sorafenib efficacy and/or prognostic biomarkers of survival for patients with mRCC.

5.2 Key Findings

Sorafenib affects tumor vascular endothelium,[24] the tumor microenvironment,[25] and as a VEGF-pathway inhibitor it also has effects on the host vascular endothelium and pericytes.[26,27] Because RCC is dependent on angiogenesis and the VEGF pathway, it is a pathway that is a viable target for pharmacotherapy.[28,29] And, because angiogenesis is primarily a host-mediated process,[30] there is excellent rationale for investigating germline variants as predictors of sorafenib efficacy and/or as markers with prognostic significance.

The primary objective of Aim 1 was to identify germline genetic markers that associate with OS. The secondary objective of the study was to prospectively analyze how the variants that associated with OS are also related to differences in PFS. Five germline variants within genes were found to significantly associate with OS (multivariate $p \leq 0.05$ and $q \leq 0.1$): two variants (rs1885657 and rs3024987) were in intronic regions of *VEGFA*, rs3816375 was in an intronic region in *ITGAV*, rs8047917 was in an intronic region of *WWOX*, and rs307826 was in a coding region of *FLT-4.* In addition, two intergenic variants (rs200809375 and rs6719561) were found to be in close proximity to *NRP-1* and *UGT1A9*, respectively. Five of the seven variants associated with OS in the sorafenib-treated TARGET patients (rs1885657, rs3816375, rs6719561, rs8047917, and rs200809375), while two of the

variants only associated with OS in a combined analysis of both TARGET treatment arms (rs30786 and rs3024987). FDR was employed in lieu of correcting the family-wise error rate (e.g. Bonferroni correction). Variants were deemed significant if their multivariate model p-value was ≤0.05 and also if their FDR q-value was ≤0.1. A p-value threshold of ≤0.05 was certainly appropriate for these analyses because discovery pharmacogenetic cohorts commonly utilize this same p-value threshold, and reserve corrections for multiple comparisons for subsequent replication pharmacogenetic studies that test associations in independent cohorts of patients. However, a replication cohort was not identified to validate these findings, and therefore an FDR correction was employed in this discovery cohort to reduce the chances of obtaining spurious findings due to type I error. FDR was selected over a Bonferroni correction because Bonferroni can be overly conservative because all tested associations are considered independent observations, which is clearly not the case in genetic association studies where variants are often in linkage with one another (and therefore not truly independent observations). While there is currently no consensus on the appropriate FDR threshold in the literature, FDR q-value ≤0.1 is an appropriate threshold because it was chosen *a priori* with the assumption that the variants significantly associated with OS would be subsequently validated through a combination of *in silico* prediction tools and *in vitro* experiments.

PFS was correlated with OS, so these seven variants were also tested for associations with PFS. Among the five original variants that associated with OS in the sorafenib arm, rs1885657, rs6719561 and rs8047917 also significantly associated with PFS (p≤0.05). Among the two variants that associated with OS in analyses when the treatment arms were combined, only rs307826 significantly associated with PFS (p≤0.05). Interestingly, rs307826

was also associated with PFS among the patients from the sorafenib arm. Because these analyses were a secondary endpoint, and exploratory in nature, they were not corrected for multiple comparisons.

The primary objective of Aim 2 was to functionally validate germline variants that associated with OS (identified through Aim 1) to elucidate the molecular drivers underlying differences in sorafenib efficacy, as well as mRCC pathogenesis, angiogenesis and/or sorafenib pharmacology. First, *in silico* analyses, which leveraged information from ENCODE, predicted which variants were likely functional, and helped prioritize laboratory validation of variants based on their potential to alter regulatory pathways and/or gene function. The Broad Institute's HaploReg website (http://www.broadinstitute.org/mammals/haploreg/haploreg_v3.php), the UCSC Genome Browser (https://genome.ucsc.edu), and Stanford's Center for Genomics and Personalized Medicine Regulome DB website (http://regulomedb.org) revealed that the intronic SNP variants, identified in Aim 1, in *VEGFA* (rs1885657 and rs3024987) were likely to be located in functionally active regions. Data from *in vitro* dual reporter gene luciferase assays confirmed the *in silico* predictions, and showed increased luciferase activity in the rs1885657 variant C allele (and in a "triple variant" construct that contained variant alleles for two SNPs in perfect LD with rs1885657: the rs943070 variant G allele, and the rs58159269 variant T allele), and the rs3024987 variant T allele. These data are consistent with the observed clinical association that patients with two copies of the variant C allele at rs1885657 experience significantly shorter OS and PFS, and support the hypothesis that the haplotype, which includes rs1885657 C allele, the rs943070 G allele, and the rs58159269 C allele, causes increased *VEGFA* expression, which results in increased angiogenesis, and ultimately shorter patient survival. Cell viability assays

revealed that the non-synonymous SNP in *FLT-4* (rs307826), which causes threonine to alanine amino acid substitution in the fifth IgG-like domain of VEGFR3, results in a more resistant phenotype when HEK-293 cells were treated with sorafenib. These data are consistent with the observed clinical association that patients with two copies of the variant G allele at this locus experience significantly shorter OS and PFS, and support the hypothesis that the amino acid substitution results in increased VEGFR3 signaling and resistance to sorafenib. The findings from Aim 2 provide evidence that the variants identified in Aim 1 are functional, and it lays the groundwork for future studies that will aim to better understand the molecular mechanism underlying these variants, and trials to validate them as predictive and prognostic biomarkers.

The primary objective of Aim 3 was to test the hypothesis that that differential cell health and response data (e.g. EC_{50} or IC_{50}) from 32 MEF cells lines treated with sorafenib, can be used in GWAS to identify candidate genes associated with sorafenib response, which will ultimately lead to the discovery of novel genes for future pharmacogenetic testing in patients treated with sorafenib. A total of seven genes were present in the QTL that associated with the cytochrome C release phenotype after a 72 h incubation with sorafenib. One candidate gene (*Rdx*) was selected for future *in vitro* validation studies. *Rdx* encodes for radixin, which is a component of the ezrin-radixin-moesin-binding phosphoprotein-50 (EBP50) complex, has been shown to act as a tumor suppressor in multiple tumor types.[31-35] A total of 13 genes were present in the QTL that associated with cell viability after a 72 h incubation with sorafenib. Two candidate genes (*Nfyc* and *Scmh1*) were selected for future *in vitro* validation studies. *Nfyc* is a transcription factor that is a key regulator of MYC and p53,[36] and has been shown to regulate DNA-dependent transcription of *PDGFR-β*.[37] *Scmh1*

encodes for a protein that helps repress transcription, and has been shown to regulate geminin. Interestingly, geminin has been shown to regulate DNA replication and has been linked to cancer.[38-41]

5.3 Future Directions

Although the results described by this dissertation research have laid the foundation for these variants as biomarkers, there are still many future studies that should be conducted. First, this dissertation has presented only the results of variants that satisfied extremely stringent statistical criteria, which included an FDR multiple comparisons correction. However, variants that did not pass the FDR correction ($q \leq 0.1$), but were significant by p-value ($p \leq 0.05$) should be evaluated, and potentially prioritized (along with the variants described in Aims 1-2) for replication pharmacogenetic studies in an independent cohort of sorafenib-treated patients.

Replication of positive findings in an external, independent cohort of patients has served as the gold standard for validation of genotype-phenotype relationships.[42,43] Pharmacogenetic studies have identified a vast set of genetic variants as predictors of chemotherapy efficacy and toxicity. The majority of these proposed variants have failed to produce similar results across different studies, which has limited the clinical utility of pharmacogenetics.[9,10] Therefore, prospective replication of pharmacogenetic findings in independent and external cohorts of patients is essential to hasten the implementation of pharmacogenetics into routine clinical practice (Appendix 2).[44] Currently, a suitable replication cohort of mRCC patients treated with sorafenib is not available; however, collaborations with investigators on the Axitinib Versus Sorafenib in Advanced Renal Cell

Carcinoma (AXIS) trial should be pursued.[8] Briefly, AXIS was a randomized phase III trial compared axitinib to sorafenib as second-line therapy in patients with mRCC. This cohort would not only have the potential to replicate findings from this work in their sorafenib-treated patients, patients from the axitinib arm could potentially help identify variants that have a class effect among VEGF-pathway inhibitors used in the treatment of mRCC. Additionally, collaboration with the investigators on the ASSURE (E2805) trial (ClinicalTrials.gov NLM Identifier NCT00326898) should be pursued. This placebo-controlled phase III trial examined the effectiveness, using PFS as the primary endpoint, of either sorafenib or sunitinib in the treatment of mRCC patients in the adjuvant treatment setting. While this cohort might not be optimal since ASSURE patients have not relapsed post-nephrectomy and do not have metastatic disease, this cohort could still be advantageous for two reasons: 1) it is a placebo-controlled trial, and thus comparisons to the TARGET patients from the placebo arm could be made, and 2) the sunitinib-treated patients could potentially help identify variants that have a class effect among VEGF-pathway inhibitors used in the treatment of mRCC.

Next, all directly genotyped and imputed variants used in the OS analyses should be used to test associations with PFS, and also with common sorafenib-induced toxicities (e.g. hypertension, diarrhea, rash, and hand foot skin reaction). Replication of those results in an independent cohort of sorafenib-treated patients is also important. Additionally, there is accumulating evidence that sorafenib pharmacokinetics are at least partially mediated by drug transporters.[45,46] Candidate variants in *SLC22A1*, *SLCO1B1* and *SLCO1B3*, which encode for OCT1, OATP1B1 and OATP1B3, respectively, could be selected for the next round of genotyping using DNA from TARGET patients. Ultimately, a prospective clinical

trial, analyzing the effects of these variants on OS and PFS in sorafenib-treated patients, should be conducted to confirm clinical validity of the associations and their clinical utility as mRCC biomarkers.

Additional functional validation of variants that associated with OS, identified in Aim 1, is required. Results from the cell viability assays provided proof of concept that rs307826 alters receptor signaling and confers increased resistance to sorafenib. However, the hypothesis that rs307826 results in altered receptor signaling requires confirmation. This can be accomplished through phosphorylation assays and traditional Western blotting. Alternate approaches, which would quantitatively compare the how the rs307826 G allele affects VEGFR3 phosphorylation when compared to the reference A allele, would be to leverage either ELISA assays or a proteomic mass spectrometry approach. While the non-synonymous variant likely causes a change in protein function, an alternative hypothesis is that the rs307826 G allele could alter *FLT-4* expression. Therefore, dual reporter gene luciferase assays could be conducted to assess variant effects on gene expression.

Additional laboratory validation experiments should be conducted on the variants identified in Aim 1 to ascertain if they influence gene expression, and to characterize the mechanistic underpinnings that result in altered gene expression. Variants that are predicted by ENCODE to influence transcription factor binding, based ChIP-Seq data and evidence of changes to transcription factor binding motifs, will be prioritized for additional *in vitro* functional validation by EMSA. Additional cell viability assays could be conducted to associate differences in gene expression with sorafenib response. Finally, using CRISPR)/cas9 or TALEN technologies, the creation of isogenic endothelial cell lines could for the first time isolate how variants (e.g. SNPs) directly affect cellular phenoypes (e.g.

angiogenesis and/or VEGF-pathway inhibitor responses). TIME cell lines will be utilized because they retain many of the features of primary endothelial cells,[47] without issues related to senescence and loss of endothelial phenotypes after multiple cell passages.[48]

For the two intergenic variants that associated with OS (rs6719561 and rs200809375), validation is still required. Dual reporter gene luciferase assays would have indeed provided insight into their activity as enhancers; however, because they are not located within genes it would be difficult to ascertain on which gene the enhancer is active. One possible validation method that could be employed is the chromosome conformation capture (3C) technology. Intergenic regulatory elements can influence gene regulation through *cis*-interactions (i.e. on a gene in close proximity to the variant) or through *trans*-interactions (i.e. on a gene a great distance away from the variant) by engaging in direct physical interactions with target genes or with other elements.[49] Therefore, the 3C technology could provide valuable validation that these variants help to regulate *UGT1A9* and *NRP-1*, respectively. Variant rs6719561 is in close proximity to two genes (approximately 1.5 kb 3' of *UGT1A9*, and 600 bp 5' of *HEATR7B1*), and thus provides some ambiguity about which gene is associated with differences in OS. After 3C is performed, the promoter of *UGT1A9* and *HEATR7B1*could be cloned into a vector with a construct containing rs6719561, and dual reporter gene luciferase assays could be conducted to detect differences in activity.

Future studies to validate the candidate genes identified in Aim 3 will certainly be conducted in MEF cells using *in vitro* knockdown and overexpression approaches. The prioritized candidate genes (*Rdx* and *Fdx1* for Cytochrome C release, and *Nfyc* and *Scmh1* for VOC) will be validated using and the same siRNA knockdown and pCMV-SPORT6 vector-based overexpression approaches that have been previously described.[50] Any of the

genes validated through these siRNA and/or overexpression experiments in MEF cells will be repeated in orthologous human siRNA and overexpression vectors, and tested in human cell lines (e.g. the human endothelial TIME cells, and the human mRCC Caki-1 cells) to confirm the effects of these genes on sorafenib response. Finally, for genes that validate in human and MEF cell lines, a second candidate gene-candidate SNP genotyping study will be conducted using TARGET patient DNA.

5.4 Conclusions

In conclusion, the primary objective of this dissertation was to identify and validate germline genetic variants that are predictive of sorafenib response, as well as variants that are prognostic markers of survival in mRCC patients. The identification and validation of predictive pharmacogenetic variants may help determine which patients should be treated with sorafenib. In a crowded landscape of targeted agents for the treatment of mRCC, identification, validation, and then ultimately clinical incorporation of these predictive variants into routine clinical practice could help guide clinician treatment decisions. Additionally, the identification of novel prognostic markers may provide insight into RCC pathogenesis/prognosis, or identify patients who would benefit from more intensive therapies and/or monitoring. The studies from Aims 1-2 of this dissertation have identified and validated seven pharmacogenetic variants that associate with OS among mRCC patients originally enrolled on the phase III TARGET trial. Among the seven identified germline variants, five (rs3816375 in *ITGAV*, rs1885657 in *VEGFA*, rs8047917 in *WWOX*, as well as rs6719561 in a region 3' of *UGT1A9*, and rs200809375 in a region 3' of *NRP-1*) have the potential to be variants that are predictive of sorafenib efficacy. It is unclear if the remaining

two variants that associated with OS (rs307826 in *FLT-4* and rs3024987 in *VEGFA*) are purely prognostic, or if they also possess predictive effects. Replication of these associations in an independent and external cohort of mRCC patients treated with sorafenib is of utmost importance before incorporating them into routine clinical practice. Additional laboratory validation of these variants, to elucidate the molecular underpinnings of their predictive and/or prognostic effects, should also be conducted. In Aim 3, innovative high content cellular genetics approach has detected robust interstrain cellular differences in sorafenib activity. One QTL region, which reached genome-wide significance, which potentially associates with sorafenib-induced cytochrome C release from mitochondria, was identified. An additional QTL was identified that potentially associates with sorafenib cytotoxicity and cell viability. Candidate genes for functional validation have been prioritized through a multi-faceted set of criteria. Future steps for this work include functional validation of candidate genes, using knockdown and overexpression approaches, in MEF and human cell lines. Collectively, this dissertation research has identified variants in genes known to be central to angiogenesis and/or sorafenib pharmacology, as well as QTLs that influence sorafenib's pro-apoptotic capabilities. This work lays a critical foundation for future studies that will help validate novel biomarkers that will aid clinicians in the treatment of patients with mRCC.

REFERENCES

1. Hudes G, Carducci M, Tomczak P, et al. Temsirolimus, interferon alfa, or both for advanced renal-cell carcinoma. *N Engl J Med*. 2007;356:2271-81.

2. Escudier B, Eisen T, Stadler WM, et al. Sorafenib in advanced clear-cell renal-cell carcinoma. *N Engl J Med*. 2007;356:125-34.

3. SEER Stat Fact Sheets: Kidney and Renal Pelvis Cancer. Accessed March 1, 2015. Available at: http://seer.cancer.gov/statfacts/html/kidrp.html.

4. Escudier B, Pluzanska A, Koralewski P, et al. Bevacizumab plus interferon alfa-2a for treatment of metastatic renal cell carcinoma: a randomised, double-blind phase III trial. *Lancet*. 2007;370:2103-11.

5. Motzer RJ, Escudier B, Oudard S, et al. Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomised, placebo-controlled phase III trial. *Lancet*. 2008;372:449-56.

6. Motzer RJ, Hutson TE, Tomczak P, et al. Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. *N Engl J Med*. 2007;356:115-24.

7. Sternberg CN, Davis ID, Mardiak J, et al. Pazopanib in locally advanced or metastatic renal cell carcinoma: results of a randomized phase III trial. *J Clin Oncol*. 2010;28:1061-8.

8. Rini BI, Escudier B, Tomczak P, et al. Comparative effectiveness of axitinib versus sorafenib in advanced renal cell carcinoma (AXIS): a randomised phase 3 trial. *Lancet*. 2011;378:1931-9.

9. Coate L, Cuffe S, Horgan A, Hung RJ, Christiani D, Liu G. Germline genetic variation, cancer outcome, and pharmacogenetics. *J Clin Oncol*. 2010;28:4029-37.

10. Innocenti F, Schilsky RL. Translating the cancer genome into clinically useful tools and strategies. *Dis Model Mech*. 2009;2:426-9.

11. Marsh S, McLeod HL. Pharmacogenomics: from bedside to clinical practice. *Hum Mol Genet*. 2006;15 Spec No 1:R89-93.

12. Hoskins JM, Carey LA, McLeod HL. CYP2D6 and tamoxifen: DNA matters in breast cancer. *Nat Rev Cancer*. 2009;9:576-86.

13. Evans WE, Relling MV. Moving towards individualized medicine with pharmacogenomics. *Nature*. 2004;429:464-8.

14. Evans WE, McLeod HL. Pharmacogenomics--drug disposition, drug targets, and side effects. *N Engl J Med*. 2003;348:538-49.

15. Chin L, Andersen JN, Futreal PA. Cancer genomics: from discovery science to personalized medicine. *Nat Med*. 2011;17:297-303.

16. Crona D, Innocenti F. Can knowledge of germline markers of toxicity optimize dosing and efficacy of cancer therapy? *Biomark Med*. 2012;6:349-62.

17. Awada A, Hendlisz A, Gil T, et al. Phase I safety and pharmacokinetics of BAY 43-9006 administered for 21 days on/7 days off in patients with advanced, refractory solid tumours. *Br J Cancer*. 2005;92:1855-61.

18. Clark JW, Eder JP, Ryan D, Lathia C, Lenz HJ. Safety and pharmacokinetics of the dual action Raf kinase and vascular endothelial growth factor receptor inhibitor, BAY 43-9006, in patients with advanced, refractory solid tumors. *Clin Cancer Res*. 2005;11:5472-80.

19. Strumberg D, Richly H, Hilger RA, et al. Phase I clinical and pharmacokinetic study of the Novel Raf kinase and vascular endothelial growth factor receptor inhibitor BAY 43-9006 in patients with advanced refractory solid tumors. *J Clin Oncol*. 2005;23:965-72.

20. Strumberg D, Clark JW, Awada A, et al. Safety, pharmacokinetics, and preliminary antitumor activity of sorafenib: a review of four phase I trials in patients with advanced refractory solid tumors. *Oncologist*. 2007;12:426-37.

21. Moore M, Hirte HW, Siu L, et al. Phase I study to determine the safety and pharmacokinetics of the novel Raf kinase and VEGFR inhibitor BAY 43-9006, administered for 28 days on/7 days off in patients with advanced, refractory solid tumors. *Ann Oncol.* 2005;16:1688-94.

22. Ratain MJ, Eisen T, Stadler WM, et al. Phase II placebo-controlled randomized discontinuation trial of sorafenib in patients with metastatic renal cell carcinoma. *J Clin Oncol*. 2006;24:2505-12.

23. Escudier B, Eisen T, Stadler WM, et al. Sorafenib for treatment of renal cell carcinoma: Final efficacy and safety results of the phase III treatment approaches in renal cancer global evaluation trial. *J Clin Oncol.* 2009;27:3312-8.

24. Murphy DA, Makonnen S, Lassoued W, Feldman MD, Carter C, Lee WM. Inhibition of tumor endothelial ERK activation, angiogenesis, and tumor growth by sorafenib (BAY43-9006). *Am J Pathol.* 2006;169:1875-85.

25. Murakami M, Zhao S, Zhao Y, et al. Evaluation of changes in the tumor microenvironment after sorafenib therapy by sequential histology and 18F-fluoromisonidazole hypoxia imaging in renal cell carcinoma. *Int J Oncol*. 2012;41:1593-600.

26. Kamba T, McDonald DM. Mechanisms of adverse effects of anti-VEGF therapy for cancer. *Br J Cancer*. 2007;96:1788-95.

27. Pignochino Y, Grignani G, Cavalloni G, et al. Sorafenib blocks tumour growth, angiogenesis and metastatic potential in preclinical models of osteosarcoma through a

mechanism potentially involving the inhibition of ERK1/2, MCL-1 and ezrin pathways. *Mol Cancer*. 2009;8:118.

28. Rini BI, Small EJ. Biology and clinical development of vascular endothelial growth factor-targeted therapy in renal cell carcinoma. *J Clin Oncol.* 2005;23:1028-43.

29. Rini BI, Sosman JA, Motzer RJ. Therapy targeted at vascular endothelial growth factor in metastatic renal cell carcinoma: biology, clinical results and future development. *BJU Int*. 2005;96:286-90.

30. Azam F, Mehta S, Harris AL. Mechanisms of resistance to antiangiogenesis therapy. *Eur J Cancer*. 2010;46:1323-32.

31. Zheng JF, Sun LC, Liu H, Huang Y, Li Y, He J. EBP50 exerts tumor suppressor activity by promoting cell apoptosis and retarding extracellular signal-regulated kinase activity. *Amino Acids*. 2010;38:1261-8.

32. Hayashi Y, Molina JR, Hamilton SR, Georgescu MM. NHERF1/EBP50 is a new marker in colorectal cancer. *Neoplasia*. 2010;12:1013-22.

33. Bartholow TL, Becich MJ, Chandran UR, Parwani AV. Immunohistochemical analysis of ezrin-radixin-moesin-binding phosphoprotein 50 in prostatic adenocarcinoma. *BMC Urol*. 2011;11:12.

34. Lin YY, Hsu YH, Huang HY, et al. Aberrant nuclear localization of EBP50 promotes colorectal carcinogenesis in xenotransplanted mice by modulating TCF-1 and beta-catenin interactions. *J Clin Invest*. 2012;122:1881-94.

35. Claperon A, Guedj N, Mergey M, et al. Loss of EBP50 stimulates EGFR activity to induce EMT phenotypic features in biliary cancer cells. Oncogene 2012;31:1376-88.

36. Romier C, Cocchiarella F, Mantovani R, Moras D. The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y. *J Biol Chem*. 2003;278:1336-45.

37. Hackzell A, Uramoto H, Izumi H, Kohno K, Funa K. p73 independent of c-Myc represses transcription of platelet-derived growth factor beta-receptor through interaction with NF-Y. *J Biol Chem*. 2002;277:39769-76.

38. Montanari M, Macaluso M, Cittadini A, Giordano A. Role of geminin: from normal control of DNA replication to cancer formation and progression? *Cell Death Differ. 2006*;13:1052-6.

39. Zhu W, Depamphilis ML. Selective killing of cancer cells by suppression of geminin activity. *Cancer Res*. 2009;69:4870-7.

40. Yagi T, Inoue N, Yanai A, et al. Prognostic significance of geminin expression levels in Ki67-high subset of estrogen receptor-positive and HER2-negative breast cancers. *Breast Canc*er. 2014 Aug [Epub ahead of print]..

41. Tokes AM, Szasz AM, Geszti F, et al. Expression of proliferation markers Ki67, cyclin A, geminin and aurora-kinase A in primary breast carcinomas and corresponding distant metastases. *J Clin Pathol.* 2015;68:274-82.

42. Chanock SJ, Manolio T, Boehnke M, et al. Replicating genotype-phenotype associations. *Nature*. 2007;447:655-60.

43. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat Genet*. 2001;29:306-9.

44. Crona DJ, Ramirez J, Qiao W, et al. Clinical validity of new genetic biomarkers of irinotecan neutropenia: an independent replication study. *Pharmacogenomics J*. 2015 April 14 [Epub ahead of print].

45. Swift B, Nebot N, Lee JK, et al. Sorafenib hepatobiliary disposition: mechanisms of hepatic uptake and disposition of generated metabolites. *Drug Metab Dispos*. 2013;41:1179-86.

46. Zimmerman EI, Hu S, Roberts JL, et al. Contribution of OATP1B1 and OATP1B3 to the disposition of sorafenib and sorafenib-glucuronide. *Clin Cancer Res*. 2013;19:1458-66.

47. Venetsanakos E, Mirza A, Fanton C, Romanov SR, Tlsty T, McMahon M. Induction of tubulogenesis in telomerase-immortalized human microvascular endothelial cells by glioblastoma cells. *Exp Cell Res*. 2002;273:21-33.

48. Freedman DA. Senescence and its bypass in the vascular endothelium. *Front Biosci*. 2005;10:940-50.

49. Dostie J, Richmond TA, Arnaout RA, et al. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res*. 2006;16:1299-309.

50. Suzuki OT, Frick A, Parks BB, et al. A cellular genetics approach identifies gene-drug interactions and pinpoints drug toxicity pathway nodes. *Front Genet*. 2014;5:272.

APPENDIX 1[1]: CAN KNOWLEDGE OF GERMLINE MARKERS OF TOXICITY OPTIMIZE DOSING AND EFFICACY OF CANCER THERAPY?

A1.1 Overview

The systemic treatment of cancer with traditional cytotoxic chemotherapeutic agents and more targeted agents is often complicated by the onset of adverse drug reactions (ADRs). Pharmacogenetic prediction of ADRs might have consequences for dosing and efficacy. This review discusses relevant examples where the germline variant-toxicity relationship has been validated as an initial step in developing clinically useful pharmacogenetic markers, and provides examples where germline variants have influenced dosing strategies and/or survival or other outcomes of efficacy. This review will also provide insight into the reasons why more pharmacogenetic markers have not been routinely integrated into clinical practice.

A1.2 Introduction

It is well established that when cancer patients are administered chemotherapy, their clinical response will vary greatly.[1-3] While many patients experience a complete or partial response, for a sizable proportion of these patients the chemotherapy will be largely ineffective despite the onset of toxicities, which might be severe, and even life threatening. Often these adverse drug reactions (ADRs) will lead to hospitalizations.[4] Numerous factors

[1] This chapter was published in *Biomarkers in Medicine* in 2012 (Crona D and Innocenti F. *Biomark Med*. 2012;6(3):349-62).

influence whether a patient will experience an ADR, including baseline clinical and demographic characteristics and genetics. Among available efficacious regimens, the selection of the safest regimen for a given patient, is not guided by molecular biomarkers, but, rather, by considerations related to the patient's age, comorbidities, and performance status.

Traditional cytotoxic chemotherapy and more novel molecularly-targeted agents (some of them also called "biologics") are the two core pharmacotherapeutic paradigms at the center of modern cancer care. Phase I trials identify the dose-limiting toxicities (DLTs) of a new chemotherapeutic agent, and [5] doses are advanced until the onset of DLTs to define its maximum tolerated dose (MTD) to be used in phase II trials. For most approved chemotherapeutic agents, there is only one dose, rather than a dose range. This dose-escalation design to MTD intrinsically narrows the distance between toxic and effective doses. It is natural to expect that one single dose might not be tolerated by some patients and during the entire duration of the scheduled cycles of treatment. As a result, the onset of serious grade 3 or 4 toxicities (as defined by the National Cancer Institute Common Terminology Criteria for Adverse Events, CTCAE, version 4.03[6]) can largely affect drug dosing and dose intensity.

In all therapeutic areas, ADRs account for an estimated 7% of all hospital admissions, occur in 10-20% of all patients, and result in significantly increased hospital stays.[7] The costs to the health system associated with ADR-related morbidity exceed $3.6 billion annually in the United States. ADRs are consistently among the top 10 reasons for death among patients.[8] Several studies have reported on the incidence of cancer chemotherapy-related ADRs. A meta-analysis of 1,219 colorectal cancer patients, who received 5-fluorouracil as

part of their regimen, revealed that up to 31% experienced grade 3-4 neutropenia.[9] A more recent meta-analysis disclosed that approximately 30% of 2,090 non-small cell lung cancer patients who received docetaxel experienced grade 3-4 neutropenia.[10] It is estimated that greater than 60,000 oncology patients are hospitalized due to severe neutropenia and its complications each year in the United States, and that neutropenia is associated with an inpatient mortality rate of 6.8% annually.[7] Dose-toxicity curves often differ between traditional cytotoxic chemotherapies and targeted molecular agents, and efficacy may occur in targeted agents at doses that do not cause significant toxicities.[5] Although the development of dosing strategies for targeted agents and biologics are still largely dependent on the identification of the MTD, there is significant debate concerning the practicality of abandoning traditional dosing strategies in favor of the identification of a minimal effective dose or the optimal biologic dose (OBD).[11] OBD can be defined as the dose associated with a pre-specified most desirable effect on a biomarker among all doses studied. [5] There are those who argue that the dosing of targeted agents should continue to follow the traditional MTD dose-finding model because they believe there is still a lack of knowledge concerning OBD strategies.[12] But, others favor investigation of a minimal effective dose,[13] or optimum biologic dose.[14,15]

Since dose-toxicity curves and OBD dosing strategies for targeted agents could be starkly different from traditional cytotoxic chemotherapy, the ADRs associated with these agents are rarely life-threatening, but still no less serious. These drugs tend to be associated with more non-hematological ADRs and less neutropenia than the traditional cytotoxics. ADRs, which adversely affect quality of life and activities of daily living, can result in non-compliance to oral targeted agents. For instance, a post-approval survey of epidermal growth

factor receptor (EGFR) inhibitors revealed that 76% of respondents were non-compliant to therapy due to rash, and that up to 32% of clinicians discontinued therapy due to rash.[7] Moreover, at least 11% of patients who receive cetuximab experience grade 3-4 rash, [16] which could lead to dose modifications by the clinician or non-adherence by the patient and ultimately complicate efficacy and survival outcomes. Up to 15% of patients treated with bevacizumab experience grade 3-4 hypertension, [7] and are at a significantly higher risk of grade 3-4 proteinuria and hypertension at both low and high doses.[17]

These common ADRs are still very concerning because medication non-adherence compromises dose-intensity and ultimately survival.[18,19] Maintenance of dose intensity in patients for the entire course of their treatment is extremely important because it has been shown both prospectively and retrospectively, in a variety of tumor types, to be correlated with more positive outcomes.[20,21] Dose intensity is a function of both dose and frequency, and modifications to either chemotherapy dosing or frequency, as a result of ADRs, are detrimental to outcomes.[22] Dosing is often adjusted based on renal or hepatic function, hematologic counts, results of basic metabolic panels, or drug levels measured in plasma. In addition, chemotherapy package inserts also provide recommendations on how to adjust dosing (usually dose reductions or dose interruptions) based upon the onset of toxicity. Although very pragmatic, this approach of personalizing chemotherapy dose remains inefficient, expensive, and adversely affects the quality of life of patients. However, certain ADRs have been researched as surrogate pharmacodynamic markers, and seem to possess a predictive role for chemotherapy response and efficacy. The relationship between drug-induced rash, dose escalation, and survival has been investigated in EGFR inhibitors such as erlotinib and cetuximab. Likewise, the relationships linking drug-induced hypertension, the

dose of the vascular endothelial growth factor (VEGF) inhibitor bevacizumab, and survival have been explored. These studies support both the paradigm that these specific ADRs are likely pharmacodynamic proxies for efficacy, and the approach of dosing patients to OBD rather than MTD.

Pharmacogenetics examines interindividual genetic variability that influences the course of drug action so that medication regimens may be optimized to maximize response, while minimizing drug-induced toxicity. Personalized medicine through genetics and genomics is the overarching goal for the field of pharmacogenetics and vital to the evolution of effective treatment paradigms. The ability to predict and mitigate ADRs in oncology is of paramount importance because a majority of chemotherapeutic agents are dosed to their MTD, and possess a narrow therapeutic index between efficacy and toxicity. There is currently a dearth of clinically relevant prognostic and predictive pharmacogenetic markers derived from germline variants that have been translated for routine incorporation into clinical practice.[23-25] There is an opportunity to better understand the relationship between germline variants, toxicity, and chemotherapy dose so that more of these pharmacogenetic markers may be utilized clinically to individualize therapies for cancer patients.

The purpose of this review is to analyze the available literature to elucidate pertinent germline variant-ADR relationships, and to describe how these associations may impact the delivery of an optimal chemotherapy dose to patients and, consequently, affect survival outcomes. We have chosen to systematically concentrate on a focused number of validated germline pharmacogenetic variants that associate with ADRs and the manner in which those variants might influence downstream effects, such as chemotherapy dosing and survival outcomes. This review should not be construed as a complete evaluation of all

pharmacogenetic variants detailed in oncology-related literature, as this has been addressed extensively in many recent reviews.[26-28] However, the resulting effects of these associations on dosing and survival have never been analyzed in detail. For each specific gene discussed, we will not review a comprehensive list of all alleles in each of the genes. Rather, we have highlighted several specific pharmacogenetic variants that account for the majority of variability in each gene to highlight our approach.

A1.3 Establishing the Relationship Between Germline Variants and Toxicity

Validation of pharmacogenetic markers related to ADRs is of utmost importance to clinicians so they may personalize the most efficacious and safe regimen for each patient, without sacrificing dose intensity. With the advent of hypothesis-driven candidate gene studies and more unbiased genome-wide approaches, oncology researchers have begun to characterize the genetics that underlie serious ADRs. The labels for over twenty chemotherapeutic agents have been rewritten or revised to include information regarding pharmacogenetic markers. Specifically, seven of these agents (6-mercaptopurine, 6-thioguanine, 5-fluorouracil, capecitabine, irinotecan, nilotinib, and imatinib) include language describing a gene-toxicity relationship, which emphasizes the growing importance of validating clinically useful pharmacogenetic markers to prevent the onset or mitigate the severity of ADRs.[29]

Numerous pharmacogenetic studies have been conducted in an attempt to identify novel germline variants that conclusively associate with ADRs. A majority of the validated pharmacogenetic markers, and those primarily discussed in this review, are germline variants that contribute to differences in drug metabolizing enzymes of drug inactivation. A classic

example is the thiopurine antimetabolite, 6-mercaptopurine, an integral component in the treatment of pediatric acute lymphoblastic leukemia (ALL).[30] 6-mercaptopurine is activated into thioguanine nucleotides (TGNs), which are then incorporated into the DNA and interfere with the activity of nucleic acid-processing enzymes.[26] Thiopurine-S-methyltransferase (TPMT) inactivates 6-mercaptopurine through methylation. Genetic variations in the *TPMT* gene (**2*, **3A*, and **3C*) lead to TPMT deficiency and reduced inactivation of 6-mercaptopurine,[31] and usually these patients have increased TGN erythrocyte levels.[32,33]

The germline variant-toxicity relationship between uridine-5'-diphospho-glucuronosyltransferase (UGT) and irinotecan has also been studied extensively. UGT1A1 is responsible for the inactivation of the active metabolite of irinotecan through glucuronidation. Irinotecan is commonly used in advanced colorectal cancer in combination with 5-fluorouracil. An inverse relationship exists between the number of TA repeats in the *UGT1A1* promoter, and the transcriptional efficiency of the gene,[34] its protein[35] and mRNA expression,[36] and the level of glucuronidation of the active metabolite of irinotecan.[37,38] UGT1A1-deficient patients are homozygous carriers of the **28* variant (7 TA repeats, versus 6 repeats in *1*1*). In addition, the germline variant-toxicity relationship has been validated between dihydropyrimidine dehydrogenase (DPD) and fluoropyrimidines. 5-fluorouracil, and its oral equivalent capecitabine, are routinely used in the treatment of gastroesophageal, hepatocellular and colorectal cancers. DPD is responsible for up to 85% of the catabolism of fluoropyrimidine antimetabolites,[23] and the *DPYD*2A* variant has been associated with decreased DPD catabolic activity.

The germline variants described above are associated with increased risk of myelotoxicity (most notably neutropenia), as well as other potentially lethal clinical sequelae.

Myelosuppression is often the DLT of many chemotherapeutic agents due to their mechanisms of action, which target highly replicating cells like those of the bone marrow. Neutropenia can be routinely measured and quantified, which increases the ability to characterize pharmacodynamic effects of these medications with relative precision. Complications associated with neutropenia (i.e. infection and sepsis) are associated with poorer outcomes and increased costs to the health system. It has been estimated that the cost of inpatient treatment of neutropenia for one patient exceeds $13,000 per hospitalization.[39] Overall, severe neutropenia has a detrimental effect on dose intensity, not only in terms of the drug causing neutropenia but also to all other combined chemotherapies administered as part of that regimen.[20,40] *TPMT* deficient patients are at greater risk for 6-mercaptopurine–induced ADRs, mainly severe myelosuppression.[32,33] The positive predictive value of the *TPMT* genotype test has been estimated to be 67-100%.[41] *UGT1A1*-deficient patients are at a 9.3 fold higher risk for the development of irinotecan-induced grade 4 neutropenia, than patients with non-deficient UGT1A1 activity.[42] The positive predictive value of a *UGTA1A1*28/*28* genotype to detect grade 3-4 toxicity has been estimated to be 50%.[43] Approximately 20-50% of patients who experienced a grade 3-4 ADR are carriers of *DPYD*2A*,[44-47] and exhibited decreased DPD activity.[47] They have a 3.4-fold higher risk for grade 4 neutropenia when compared to patients with normal DPD activity.[48] The positive predictive value of *DPYD*2A* for grade 3-4 toxicity has been estimated at 46%.[49] Germline variants have also been associated with chemotherapy-induced ADRs other than myelotoxicities, such as neutropenia. Genetic associations have been investigated linking hypertension with bevacizumab, and skin toxicities with EGFR inhibitors (i.e. cetuximab, panitumumab, erlotinib and gefitinib). Skin rash has been reported in approximately 80% of

patients taking these drugs.[50] Inhibition of EGFR signaling in basal keratinocytes leads to abnormalities in cell growth, differentiation and maturation, as well as an inflammatory response, which ultimately leads to the onset of rash.[51] One proposed germline polymorphism associated with skin rash is a CA dinucleotide repeat in intron 1 of *EGFR*.[52] The number of CA repeats is inversely proportional to expression of *EGFR*, and shorter CA repeats have been linked to greater incidence of skin toxicity.[53-55] Bevacizumab is a monoclonal antibody that binds to VEGF and inhibits angiogenesis. Evidence suggests a link between hypertension and impaired angiogenesis,[56] and treatment with bevacizumab has been associated with increased incidence of grade 3-4 hypertension.[17,57,58] A meta-analysis of 12,656 patients demonstrated that bevacizumab significantly increases the risk of all grade hypertension by 23.6% (95% CI: 20.5-27.1%) and grade 3-4 hypertension by 7.9% (CI: 6.1-10.2%), relative to control patients on concurrent traditional chemotherapies.[57] In another study of patients who took bevacizumab, two *VEGF* germline genotypes (-634CC and -1498TT) were associated with lower rates of grade 3-4 hypertension.[59] This was the first trial to investigate biomarkers associated with bevacizumab-induced hypertension. Emerging data have confirmed a similar relationship and showed that patients with either the -634CCor -1498TT genotypes, who were treated with the VEGF-inhibitor sunitinib, had less incidence of hypertension ($P = 0.03$, in both cases).[60] Nevertheless, at this stage, the variants in *EGFR* and *VEGF* require additional validation before they can be used as predictive markers of toxicity.

The process of pharmacogenetic marker validation still faces several major hurdles before these markers are integrated into routine clinical practice. Conflicting data generated from pharmacogenetic trials have been reported for many of the most well characterized

potential markers. For example, the *UGT1A1*28/*28* genotype was not associated with severe diarrhea (another common side effect of irinotecan) in one study, but it was in others.[42,61,62] In addition, inconsistencies concerning the functional variant associated with a particular ADR complicate the validation of pharmacogenetic markers. For example, data points to -216G/T and -191C/A *EGFR* variants, rather than CA dinucleotide repeats in intron 1, as the causative single nucleotide polymorphisms (SNPs) responsible for EGFR inhibitor-induced skin toxicity.[52] The relationship between 5-fluorouracil toxicity and *DPYD*2A* also illustrates another problem in the process of pharmacogenetic marker validation: rare germline variants, which have limited predictive power. Data suggests that alternate germline variants in *DPYD*,[63] non-coding region genomic variants found through haplotype assessment,[64] *DPYD* promoter hypermethylation,[65-67] or the influences of *TYMS*, *MTHFR* and other genes[23,27] may play a significant role in explaining the remainder of the germline variant-toxicity relationship involving neutropenia and 5-fluorouracil.

A1.4 The Effect of the Germline Variant-Toxicity Relationship on Chemotherapy Dosing

While over twenty chemotherapeutic agents have references to pharmacogenetic markers in their FDA label,[29] a clear disconnect exists between these pharmacogenetic markers and dosing recommendations.[68] Only 6-mercaptopurine, 6-thioguanine, nilotinib and irinotecan include pharmacogenetic-based dose reduction recommendations in their drug labels.[68] But, given the limited scale of clinical investigations on genotype-based dose individualization, they did not recommend mandated pharmacogenetic testing, nor did they provide guidance concerning the optimal timing for testing. This same limited scale of clinical investigations led the FDA to exclude explicit dose reduction guidelines from the

irinotecan or 6-mercaptopurine labels.[68] While the package insert for irinotecan includes language endorsing dose reductions by one level for patients with the *UGT1A1*28/*28* genotype, these recommendations are not sufficiently specific.[69] Genotype-based dose modification information should be far more precise because vague dose reduction language could conceivably compromise the efficacy of irinotecan if clinicians reduce doses too low.

Despite the fact that the 6-mercaptopurine label does not contain specific dose reductions based on genotype, accumulating data regarding its dosing in *TPMT*-deficient heterozygous[70] and homozygous[71] patients led to guidelines supporting 50% and at least 90% reductions, respectively, of the initial dose of 6-mercaptopurine in pediatric patients treated for ALL. This genotype-based dosing strategy has successfully limited ADRs.[72] These conclusions were based on studies showing that *TPMT*-deficient heterozygotes can tolerate 65% of the standard 75 mg/m^2 6-mercaptopurine dose used in pediatric ALL (50 mg/m^2 daily),[71,73,74] while homozygotes can receive 1/10th to 1/15th of the standard dose (i.e. 10 mg/m^2 every third day).[70] In subsequent studies, patients were prospectively screened for *TPMT* status, and doses of 6-mercaptopurine were reduced in patients with *TPMT*-deficient alleles, according to a pre-specified dosing algorithm.[75,76] As a result, these patients were successfully treated with 6-mercaptopurine, with rates of grade 3-4 ADRs comparable to patients with wild-type *TPMT*.[72] In addition, the identification of *TPMT*-deficient patients not only permits clinicians to rationally reduce doses of 6-mercaptopurine to avoid severe ADRs, it also allows them to preserve the dose intensity of the regimen by administering concurrent chemotherapies at unadjusted doses and without interruptions.[77,78]

The irinotecan doses used in current regimens were previously identified through phase I development without stratifying patients for *UGT1A1*28*. It has been postulated that

the dose of irinotecan might be suboptimal in patients with **1/*28* and **1/*1* genotypes, and these patients may benefit from higher, safe doses. This hypothesis has been tested in colorectal cancer patients treated with irinotecan combined with 5-fluorouracil (FOLFIRI regimen).[79] Dose-escalation of irinotecan was assessed in **1/*1* and **1/*28* patients, and results from these studies showed they were safely and effectively treated with irinotecan doses of 370 and 310 mg/m^2, respectively. These doses were considerably higher than the 180 mg/m^2 dose currently used in the clinic. Similar dose levels were safely given in another study of FOLFIRI (390 and 340 mg/m^2, respectively).[80] This study also demonstrated that the MTD for **28/*28* patients was 130 mg/m^2, which is 30% lower than the standard dose. A similar conclusion has been also obtained in Japanese patients with **28* and/or **6*, the common deficient allele in Asians. **28/*28*, **28/*6*, and **6/*6* patients also achieved an MTD of 150 mg/m^2.[81] Meta-analyses demonstrated that **28/*28* patients have a significantly higher risk for grade 3-4 neutropenia, when compared to **1/*28* and **1/*1*, at low doses (RR = 2.43; P = 0.003),[82] at medium doses (OR = 3.22; P = 0.008) and high doses of irinotecan (OR = 27.8; P = 0.005).[83]

There are additional examples where a germline-variant toxicity relationship might alter dosing. In a capecitabine study of 568 patients with advanced colorectal cancer, all of the patients with *DPYD*2A* experienced at least one grade 3-4 ADR (mostly diarrhea).[84] The cumulative percentage dose reduction in these patients was significantly higher (50%) than wild-type patients (10%). For EGFR and VEGF-pathway inhibitors, there are not data on the effect of variants for skin rash and hypertension (respectively) on dosing and dose intensity. This represents a significant gap in current knowledge that should be addressed.

A1.5 The Downstream Effect of Germline Variants of Toxicity on Dosing and Efficacy

The assumption that patients should be treated at the highest, safe doses of chemotherapy is still the foundation of a successful treatment regimen. Although we believe a pharmacogenetic marker of toxicity has high clinical utility, translation of such markers to clinical practice will be increased dramatically if ADRs can be linked to efficacy and survival. Currently, only the example of *TPMT* and 6-mercaptopurine exists where the knowledge of a germline variant-toxicity-dosing relationship is a tool to optimize efficacy. When 6-mercaptopurine doses were determined prospectively, based on *TPMT* genetic status, relapse rates were not significantly different between wild-type patients and those with *TPMT*-deficient genotypes (13.2% $\pm$ 2.3% versus 6.7% $\pm$ 6.7%; P = 0.46). In addition, patients with *TPMT*-deficient genotypes were not at significantly higher rates of grade 3-4 neutropenia when compared to wild-type patients (OR = 1.4; 95% CI = 0.3–6.9; P = 0.71).[72,74] Most importantly, this genotype-based dosing strategy has successfully limited potentially life-threatening ADRs, while not sacrificing efficacy in this patient population.[72,77] The interaction between 6-mercaptopurine and the bone marrow is unique because it is not only activated intracellularly in the bone marrow, but is also the site for its anti-leukemic efficacy and its myelotoxic DLT. This idiosyncrasy of 6-mercaptopurine allows the downstream effects of *TPMT* genetics on survival to be more directly assessable. But aside from the *TPMT* example, the relationship between germline variant-toxicity, chemotherapy dosing and survival outcomes have been extremely difficult to establish. For other relationships, the consequences on survival are tested as a secondary endpoint or have not been tested at all in clinical trials.

Very few genotype-driven dose-optimization studies have prospectively assessed objective response rate, progression-free survival, overall survival, or other measures of efficacy, as their primary endpoints. For example, little evidence exists to conclusively link *UGT1A1* germline variants to ADRs, dose and survival. In a retrospective analysis, dose reductions of irinotecan due to *UGT1A1*28*-related toxicity did not affect progression-free survival (10 versus 11 months) or overall survival (19 versus 18 months).[85] A second study retrospectively evaluated the association between UGT1A1 genotype, prevalence of grade 3-4 toxicity, and survival outcomes in colorectal cancer patients treated with irinotecan. This study demonstrated a significantly higher rate of ADRs, coupled with worse survival, for the *28/*28 patients. Grade 3-4 neutropenia was significantly more common in *28/*28 patients (24% compared with 8.2% and 5.5% in the *1/*28 and *1/*1 genotypes, respectively; P = 0.019), while the median overall survival for the three genotypes was also significantly different (2.4, 2.0 and 1.6 years for the *1/*1, *1/*28 and *28/*28 genotypes, respectively; P = 0.019).[86] It is currently unclear whether the results of this study point to a true downstream effect of the *28/*28 genotype-neutropenia relationship on survival, or if the decrease in median overall survival for patients with the *28/*28 genotype is simply due to neutropenia-induced dose modifications that decrease overall exposure to irinotecan. Two recent studies of irinotecan, dose-escalated by genotype, showed that patients receiving higher doses experienced higher response rates than those treated at lower doses. In the first study, 65% of the patients treated with doses greater than or equal to the MTD of irinotecan achieved a complete or partial response compared with only 25% of patients treated with doses below the MTD of irinotecan (P = 0.14). Irinotecan dosed above MTD was an independent predictor of response in all patients (OR = 4.38; 95% CI 1.13-17.03; P = 0.03). Median time

to progression (TTP) was not significantly different between the patients treated above and below the irinotecan MTD (HR = 0.85; 95% CI 0.40-1.80).[79] In the second study, 67% of patients treated with irinotecan doses ≥260 mg/m^2 achieved a complete or total response, compared to only 24% of patients treated with doses <260 mg/m^2 (P = 0.001). But unlike the previous study, TTP was higher in patients treated with doses ≥260 mg/m^2, when compared to patients treated <260 mg/m^2 (16 versus 7 months; P = 0.003).[80] These data are premature, and intervention prospective studies should test the superiority of genotype-driven dosing of irinotecan versus standard dosing, in relation to survival in metastatic colorectal cancer.

The evidence is still too sparse to provide a complete understanding of the relationship between germline variants associated with ADRs and survival in patients treated with EGFR and VEGF-pathway inhibitors, largely due to a lack of validated SNPs predictive of toxicity. Several phase II-III studies have linked the occurrence and severity of EGFR inhibitor-induced rash to improvements in survival.[87-91] One notable phase III trial of NSCLC patients revealed a substantial survival benefit when patients experienced grade 2 and 3 rash compared to when no rash was observed (7 and 11 months, respectively, versus 3 months).[87] Moreover, several studies have reported an association between higher response and survival with less CA repeats, independently from its effect of skin rash or other toxicities,[52,92] but other studies failed to validate this relationship.[93] A prospective phase II dose-driven erlotinib study in NSCLC patients attempted to discover whether dose escalation beyond MTD would confer a survival advantage. The results from this dose-to-rash erlotinib dosing study revealed that all patients who achieved at least a partial response also experienced rash. Progression free survival was significantly extended in patients who experienced grade 2 rash, when compared to patients without rash (3.5 versus 1.9 months; HR = 0.52; P =

0.051).[94] But, this study did not investigate the role of germline variants on incidence and severity of rash, dose escalation or survival. Furthermore the effects of dose reductions of EGFR inhibitors, as a result of grade 3-4 rash, on efficacy and survival have not been assessed. Clearly in the case of EGFR inhibitors, the germline variant-rash-dose escalation (or reduction) relationship and its effects on efficacy and survival require further and deeper examination.

Studies of patient populations have linked the incidence of bevacizumab-induced grade 3-4 hypertension to improved survival outcomes.[95,96] More recently, associations between *VEGF* variants and survival have been explored.[97,98] A trial of advanced breast cancer patients treated with bevacizumab tested *VEGF* germline variants (-634G/C, -1498T/C, -2578C/A and -1154G/A) for associations with hypertension and survival. This study demonstrated that patients with the -634CC and 1498TT genotypes, who experienced grade 3-4 hypertension, achieved superior median overall survival compared to patients with no hypertension (38.7 versus 25.3 months; P = 0.002), while the -634G/C and -1498C/T variants were not associated with improved survival.[59] But, the results of these associations have not been readily replicated, and data from other studies have associated improved survival outcomes to the -634GG [98] and 1498CC genotypes.[99] However, data from this trial did reveal the potentially important impact of the -2578AA and -1154AA genotypes on improved survival outcomes.[59] The -2578A and the -1498T alleles are in high linkage disequilibrium in the 1000 Genomes reference population of Caucasians (LD, r^2= 0.966).[100] This suggests that an LD block of genetic variation, containing these variants, may impact both phenotypes, and could potentially be useful as predictive markers of both hypertension and survival.

High VEGF expression and MVD have been associated with decreased survival,[101] and the results from the previous study showed that the -2578A allele trended toward lower VEGF expression [59]. A plausible mechanistic explanation is that the -2578AA genotype causes lower VEGF expression and lower levels of VEGF may slow angiogenesis and the vascularization of the tumor, therefore contributing to increased survival. A similar process may also account for the bevacizumab-induced hypertension. It can be postulated that inhibition of the VEGF pathway, as a result of bevacizumab administration and in combination with the -1498TT and -634CC genotypes, may result in impaired vascularization. Lower VEGF levels leads to rarefaction, which is a process that results in reduced MVD, increased peripheral vascular resistance, and ultimately the onset of hypertension.[56] The impact of these germline variants, in terms of their effects on hypertension and survival, has just begun to be analyzed in studies of other VEGF-pathway inhibitors, such as sunitinib.[60] Such studies could be extremely useful to prioritize validation of these variants in patients prescribed other targeted therapies from this class, which all share similar rates of high-grade hypertension. These findings exemplify the challenges facing researchers and clinicians, and illustrate the difficulty in applying such information to guide dosing. Although we start to understand the intricacy of the effects of *VEGF* genetics on both hypertension and survival, more efforts should address whether genotype-driven hypertension of bevacizumab or sunitinib can negatively impact dosing and compromise survival.

A1.6 Conclusions

Incorporation of validated pharmacogenetic markers into routine practice is of paramount importance to clinicians as they strive to reduce severe and life-threatening ADRs. Prevention and mitigation of ADRs improves patient quality of life and compliance to chemotherapy, which ultimately may lead to better survival outcomes. Understanding the putative relationship between germline variants, ADRs, chemotherapy dosing, and survival is essential to the identification and translation of pharmacogenetic markers. We believe that any validation approach of pharmacogenetic markers, with potential clinical utility, must reflect incorporation of these four key components.

The examples of *TPMT* and 6-mercaptopurine, *UGT1A1*28* and irinotecan, *DPYD* and the fluoropyrimidines, *EGFR* and EGFR inhibitors, and *VEGF* and bevacizumab provide illustrations of the varying levels of success that researchers and clinicians have had in translating pharmacogenetic molecular markers to the clinic. Currently, the example of prospective *TPMT* screening and subsequent dose adjustments in pediatric ALL patients has been the model for pharmacogenetic success in oncology. Increased numbers of prospective clinical trials, similar to those used to justify prospective *TPMT* testing, are direly needed to validate many more pharmacogenetic markers. Moreover, specific dosing recommendations need to be included into drug labels to help guide clinician decision-making processes. Dosing recommendations based on *TPMT* and *UGT1A1* genotypes should be clarified and prospective trials should be conducted to solidify the relationship linking *DPYD*, *EGFR*, and *VEGF* variants to ADRs.

While a few smaller trials have explored the effect of genotype on ADRs, dosing and survival, only a handful of studies have conducted well-powered prospective trials to explore

this relationship. This could be due to a number of reasons, ranging from low frequency of a particular germline variant (i.e. *DPYD*2*), to a lack of a convincing body of evidence conclusively linking a germline variant to ADRs and dose modifications (i.e. EGFR and VEGF-pathway inhibitor germline variants). The *DPYD* example illustrates how the current approach of trying to identify highly penetrant, single gene variants to explain complex phenotypes, such as ADRs, to be used as pharmacogenetic markers, may be partially flawed. Research to identify novel markers may rely on more sophisticated methods to characterize polygenic effects. In addition, research to identify translatable pharmacogenetic makers with downstream effects on survival may be complicated by the inability to take into account the effects of somatic mutations, tumor heterogeneity, epigenetic factors or other undiscovered prognostic confounders. But, identification of pharmacogenetic markers, validated through approaches that consider the germline variant-toxicity-survival relationship, will help clinicians craft strategies that rationally dose drugs within an enlarged therapeutic window to optimize efficacy while mitigating ADRs.

A1.7 Future Perspectives

Adoption of validated pharmacogenetic markers into routine clinical practice has been slow, in part due to the scarcity of genotype-driven trials that implement the approach we have outlined. As the field of pharmacogenetics continues to evolve, retrospective meta-analyses of promising gene variants might validate or refute their use in the clinic. Prospective interventional studies involving patients stratified by genotype will be a means of systematically collecting phenotype data, while also collecting data on potential confounding factors that would inhibit translation of germline variants into clinically useful

pharmacogenetic markers. Finally, secondary, pharmacogenetically driven post-hoc analyses of phase III studies (which are sufficiently powered for genomic analysis) will further analyze the relationship between genetics, dosing strategies, ADRs and survival outcomes. Candidate gene and genome-wide association studies are two methods that will continue to drive identification of pharmacogenetic markers.[102] As the number of approved oral chemotherapeutic medications increases, strategies based on germline variant-toxicity-dose paradigms need to be developed to help identify patients at increased risk for the onset of common ADRs that adversely affect patient quality of life and compliance to therapy.

Existing statistical models might aid in the process of elucidating the links that connect germline variants to toxicity, dosing, and survival. These relationships will be better understood by evaluating a cumulative incidence of an event in the context of a regression model of competing risks. Interdependence of variables limits the ability of existing statistical models to extricate one event of interest, for analysis, from all other competing risks. Frequently in cancer studies, several competing events (phenotypes) are present that cause treatment failures. The event of interest might occur later in treatment than the occurrence of severe ADRs, disease progression, or death, which could result in early withdrawal from the study. Use of a standard case-control approach for analyzing the data might be flawed because competing risks and cumulative incidences are ignored, and has the potential to shroud the identification of pharmacogenetic markers. In recent years, cancer researchers have successfully implemented novel methodologies[103,104] into cumulative incidence competing risks model in their analyses.[96,105-107] Future studies, which analyze pharmacogenetic markers, should employ these types of analyses because the intricacies of the model allow for a more accurate depiction of probabilities of events of interest occurring

in the presence of competing risks. By employing this statistical model, researchers will be able to compare cumulative incidence curves, between two or more groups stratified by genotype, for a phenotype of interest.

Another reason for the slow inclusion of pharmacogenetic biomarkers into clinic is the conspicuous absence of explicit pharmacogenetic-based dosing guidelines published by regulatory bodies. The Clinical Pharmacogenetics Implementation Consortium (CPIC) of the National Institutes of Health's Pharmacogenomics Research Network (http://www.pgrn.org) and the Pharmacogenomics Knowledge Base (PharmGKB, http://www.pharmgkb.org) provide peer-reviewed, updated, evidence-based, freely accessible pharmacogenetic guidelines for chemotherapeutic agents.[77] These guidelines will facilitate the translation of knowledge derived from laboratory-based pharmacogenetic and pharmacogenomic research into the clinic, and provide explicit instructions regarding dosing and testing of these agents. Dosing information will be customized for patients based on treatment response. Guidelines have already been developed for *TPMT* and 6-thioguanine and 6-mercaptopurine dosing,[78] but the CPIC and other consortia will continue to develop guidelines for gene-chemotherapy pairings into the future. In addition to CPIC, other consortia, such as the Pharmacogenetics Working Group from the Netherlands, will also continue to publish pharmacogenetic-based dosing guidelines for chemotherapy and targeted agents. To date, the Pharmacogenetics Working Group has published dosing guidelines for tamoxifen, irinotecan, 6-mercaptopurine, and 5-fluorouracil/capecitabine, based on pharmacogenetic markers, but will continue to publish guidelines that will complement the work of the CPIC and PharmGKB.[108] Published pharmacogenetic-based dosing guidelines will undoubtedly aid clinicians in the optimization

of treatment efficacy, mitigation of toxicity, and improvement of both medication adherence as well as quality of life for patients taking chemotherapy.

Finally, interethnic differences in the distribution of variant alleles could also potentially obstruct the adoption of pharmacogenomic standards and impede adoption of pharmacogenetic biomarkers into the clinic. Projects, such as the Pharmacogenomics for Every Nation Initiative (PGENI, http://www.pgeni.org), will serve to address interethnic differences in allele frequency distributions. Currently, PGENI has established regional centers in six countries, and are determining allele frequencies of known polymorphisms in different populations. Ultimately, initiatives like PGENI will help identify interethnic differences that aid in medication formulary decisions, population stratification in genotype-driven clinical trials, and even target identification early on in the global drug development process.

A1.8 Executive Summary

Adverse Drug Reactions (ADRs)

- ADRs account for 7% of all hospital admissions, occur in 10-20% of all patients and increase hospital stays
- ADR-related morbidity exceeds $3.6 billion annually in the United States
- Common toxicities occur frequently in oncology patients and affect medication adherence
- Greater than 60,000 oncology patients are hospitalized annually in the United states due to severe neutropenia and neutropenia has an annual mortality rate of 6.8%
- Prevention and mitigation of ADRs improves quality of life and medication adherence, and may lead to better survival outcomes

Chemotherapy Dose

- Maintaining dose for the entire course of treatment has been shown prospectively and retrospectively, in various tumor types, to be correlated with better outcomes
- Traditional cytotoxic chemotherapies are dosed to MTD, but novel molecularly-targeted agents could be dosed to OBD

TPMT and 6-mercaptopurine

- *TPMT*2*, **3A*, and **3C* account for over 95% of inherited TPMT deficiency.
- Reduced TPMT function places patients at significantly greater risk for 6-mercaptopurine–induced ADRs, including neutropenia
- Patients heterozygous and homozygous for reduced TPMT function alleles receive 50% and ≥90% reductions of the initial dose of 6-mercaptopurine, respectively
- Dose reductions in patients with reduced TPMT function results in less grade 3-4 neutropenia, while preserving positive survival outcomes

UGT1A1 and Irinotecan

- TA repeats in the *UGT1A1* promoter have been associated with *UGT1A1* transcriptional efficiency and glucuronidation levels of the active metabolite of irinotecan
- The risk of grade 3-4 neutropenia is ~9.3-fold higher in patients with the *UGTA1A1*28/*28* genotype, when compared to all other patients, *UGTA1A1*28/*28* patients are at significantly higher risk for grade 3-4 neutropenia at low and medium doses of irinotecan, when compared to *UGTA1A1*1/*28* and *UGTA1A1*1/*1* patients

- *UGT1A1*1* patients were able to tolerate doses that exceed the standard irinotecan MTD, and dose reductions to *UGT1A1*28* patients due to neutropenia did not compromise survival outcomes
- Explicit dosing recommendations for patients with *UGT1A1*1* versus *UGT1A1*28* have not been published

DPYD and 5-fluorouracil/Capecitabine

- Germline variants in *DPYD* have been associated with increased ADRs
- The *DPYD*2* allele is not prevalent in the general population, but is associated with a 3.4-fold higher risk for grade 4 neutropenia, when compared to patients with normal DPD activity
- The relationship between *DPYD*2* genotype, grade 3-4 neutropenia and survival remains unclear and requires further investigation

EGFR and Erlotinib/Cetuximab

- Skin rash is an ADR that is experienced by ~80% of patients taking EGFR inhibitors
- A CA repeat in the intron 1 region of *EGFR* has been associated with increased incidence of rash in patients taking EGFR inhibitors
- Increased incidence of rash has been associated with increased progression free survival
- The role of germline variants in EGFR inhibitor-induced rash, dosing and survival remains unclear

VEGF and Bevacizumab

- The onset of hypertension has been correlated with increased survival in patients treated with bevacizumab

- Bevacizumab-induced hypertension and better survival outcomes may be related to germline variations in the VEGF pathway
- *VEGF* germline variants have been associated with decreased incidence of hypertension (-634CC and -1498TT) and increased survival (-2578AA and -1154AA)
- The -2578C/A and -1498C/T variants are in high LD, and could be contributing to both bevacizumab-induced hypertension and survival outcomes
- The relationship between VEGF-pathway germline variants and bevacizumab dosing is still unclear

Conclusions

- Any approach to clinical validation of pharmacogenetic markers must reflect an understanding of the relationships between germline variants, ADRs, chemotherapy dosing, and survival
- Consortia, such as PharmGKB, CPIC and the Pharmacogenetics Working Group will provide pharmacogenetically-based dosing recommendations for chemotherapeutics and aid in the translation of pharmacogenetic markers into clinical practice
- Initiatives such as PGENI will help identify interethnic differences that aid in medication formulary decisions, population stratification in genotype-driven clinical trials, and even target identification early on in the global drug development process

REFERENCES

1. Hoskins JM, Carey LA, McLeod HL. CYP2D6 and tamoxifen: DNA matters in breast cancer. *Nat Rev Cancer*. 2009;9:576-586.

2. Evans WE, Relling MV. Moving towards individualized medicine with pharmacogenomics. *Nature*. 2004;429:464-468.

3. Evans WE, McLeod HL. Pharmacogenomics--drug disposition, drug targets, and side effects. *N Engl J Med*. 2003;348:538-549.

4. Sassi G, Striano B, Merlo UA. A reporting system for the assessment of chemotherapy toxicity. *J Oncol Pharm Pract*. 2005;11:63-67.

5. Le Tourneau C, Lee JJ, Siu LL. Dose escalation methods in phase I cancer clinical trials. *J Natl Cancer Inst*. 2009;101:708-20.

6. National Cancer Institute. Cancer Therapy Evaluation Program NCI. Common Terminology Criteria for Adverse Events (CTCAE) v.4.03. Available at: http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_8.5x11.pdf. Accessed: September 12, 2011

7. Davies EC, Green CF, Taylor S, Williamson PR, Mottram DR, Pirmohamed M. Adverse drug reactions in hospital in-patients: a prospective analysis of 3695 patient episodes. *PLoS ONE*. 2009;4(2):e4439.

8. Bennett CL, Nebeker JR, Yarnold PR, et al. Evaluation of serious adverse drug reactions: a proactive pharmacovigilance program (RADAR) vs safety activities conducted by the Food and Drug Administration and pharmaceutical manufacturers. *Arch Intern Med*. 2007;167:1041-9.

9. Toxicity of fluorouracil in patients with advanced colorectal cancer: effect of administration schedule and prognostic factors. Meta-Analysis Group In Cancer. *J Clin Oncol*. 1998;16:3537-3541. No authors listed.

10. Qi WX, Shen Z, Yao Y. Meta-analysis of docetaxel-based doublet versus docetaxel alone as second-line treatment for advanced non-small-cell lung cancer. *Cancer Chemother Pharmacol.* 2011;69(1):99-106.

11. Jain RK, Lee JJ, Hong D, et al. Phase I oncology studies: evidence that in the era of targeted therapies patients on lower doses do not fare worse. *Clin Cancer Res*. 2010;16:1289-1297.

12. Sleijfer S, Wiemer E. Dose selection in phase I studies: why we should always go for the top. *J Clin Oncol*. 2008;26:1576-1578.

13. Haines IE. Dose selection in phase I studies: why we should always go for the most effective. *J Clin Oncol*. 2008;26:3650-3652.

14. Shaked Y, Emmenegger U, Man S, et al. Optimal biologic dose of metronomic chemotherapy regimens is associated with maximum antiangiogenic activity. *Blood*. 2005;106:3058-3061.

15. Carducci MA, Musib L, Kies MS, et al. Phase I dose escalation and pharmacokinetic study of enzastaurin, an oral protein kinase C beta inhibitor, in patients with advanced cancer. *J Clin Oncol*. 2006;24:4092-4099.

16. Su X, Lacouture ME, Jia Y, Wu S. Risk of high-grade skin rash in cancer patients treated with cetuximab--an antibody against epidermal growth factor receptor: systemic review and meta-analysis. *Oncology*. 2009;77:124-133.

17. Zhu X, Wu S, Dahut WL, Parikh CR. Risks of proteinuria and hypertension with bevacizumab, an antibody against vascular endothelial growth factor: systematic review and meta-analysis. *Am J Kidney Dis*. 2007;49:186-193.

18. Noens L, van Lierde MA, De Bock R, et al. Prevalence, determinants, and outcomes of nonadherence to imatinib therapy in patients with chronic myeloid leukemia: the ADAGIO study. *Blood*. 2009;113:5401-5411.

19. Hershman DL, Shao T, Kushi LH, et al. Early discontinuation and non-adherence to adjuvant hormonal therapy are associated with increased mortality in women with breast cancer. *Breast Cancer Res Treat*. 2011;126:529-537.

20. Lyman GH. Impact of chemotherapy dose intensity on cancer patient outcomes. *J Natl Compr Canc Netw*. 2009;7:99-108.

21. Kaestner SA, Sewell GJ. Chemotherapy dosing part I: scientific basis for current practice and use of body surface area. *Clin Oncol (R Coll Radiol)*. 2007;19:23-37.

22. Foote M. The Importance of Planned Dose of Chemotherapy on Time: Do We Need to Change Our Clinical Practice? *Oncologist*. 1998;3:365-368.

23. Coate L, Cuffe S, Horgan A, Hung RJ, Christiani D, Liu G. Germline genetic variation, cancer outcome, and pharmacogenetics. *J Clin Oncol*. 2010;28:4029-4237.

24. Innocenti F, Schilsky RL. Translating the cancer genome into clinically useful tools and strategies. *Dis Model Mech*. 2009;2:426-429.

25. Marsh S, McLeod HL. Pharmacogenomics: from bedside to clinical practice. *Hum Mol Genet*. 2006;15 Spec No 1:R89-93.

26. Paugh SW, Stocco G, McCorkle JR, Diouf B, Crews KR, Evans WE. Cancer pharmacogenomics. *Clin Pharmacol Ther*. 2011;90:461-466.

27. Houtsma D, Guchelaar HJ, Gelderblom H. Pharmacogenetics in oncology: a promising field. *Curr Pharm Des*. 2010;16:155-163.

28. Yong WP, Innocenti F, Ratain MJ. The role of pharmacogenetics in cancer therapeutics. *Br J Clin Pharmacol*. 2006;62:35-46.

29. U.S. Food and Drug Administration. Table of Pharmacogenomic Biomarkers in Drug Labels. Available at: http://www.fda.gov/drugs/scienceresearch/researchareas/pharmacogenetics/ucm083378.htm. Accessed: October 12, 2011.

30. Pui CH, Evans WE. Treatment of acute lymphoblastic leukemia. *N Engl J Med*. 2006;354:166-178.

31. Krynetski E, Evans WE. Drug methylation in cancer therapy: lessons from the TPMT polymorphism. *Oncogene*. 2003;22:7403-7413.

32. Kishi S, Cheng C, French D, et al. Ancestry and pharmacogenetics of antileukemic drug toxicity. *Blood*. 2007;109:4151-4157.

33. Relling MV, Hancock ML, Rivera GK, et al. Mercaptopurine therapy intolerance and heterozygosity at the thiopurine S-methyltransferase gene locus. *J Natl Cancer Inst.* 1999;91:2001-2008.

34. Beutler E, Gelbart T, Demina A. Racial variability in the UDP-glucuronosyltransferase 1 (UGT1A1) promoter: a balanced polymorphism for regulation of bilirubin metabolism? *Proc Natl Acad Sci U S A*. 1998;95:8170-8174.

35. Ritter JK, Kessler FK, Thompson MT, Grove AD, Auyeung DJ, Fisher RA. Expression and inducibility of the human bilirubin UDP-glucuronosyltransferase UGT1A1 in liver and cultured primary hepatocytes: evidence for both genetic and environmental influences. *Hepatology*. 1999;30:476-484.

36. Ramirez J, Mirkov S, Zhang W, et al. Hepatocyte nuclear factor-1 alpha is associated with UGT1A1, UGT1A9 and UGT2B7 mRNA expression in human liver. *Pharmacogenomics J*. 2008;8:152-161.

37. Iyer L, Hall D, Das S, et al. Phenotype-genotype correlation of in vitro SN-38 (active metabolite of irinotecan) and bilirubin glucuronidation in human liver tissue with UGT1A1 promoter polymorphism. *Clin Pharmacol Ther*. 1999;65:576-582.

38. Iyer L, King CD, Whitington PF, et al. Genetic predisposition to the metabolism of irinotecan (CPT-11). Role of uridine diphosphate glucuronosyltransferase isoform 1A1 in the glucuronidation of its active metabolite (SN-38) in human liver microsomes. *J Clin Invest.* 1998;101:847-854.

39. Caggiano V, Weiss RV, Rickert TS, Linde-Zwirble WT. Incidence, cost, and mortality of neutropenia hospitalization associated with chemotherapy. *Cancer*. 2005;103:1916-1924.

40. Lyman GH. A comparison of international guidelines for the prevention of chemotherapy-induced neutropenia. *Curr Opin Hematol*. 2011;18(1):1-10.

41. Donnan JR, Ungar WJ, Mathews M, Rahman P. Systematic review of thiopurine methyltransferase genotype and enzymatic testing strategies. *Ther Drug Monit*. 2011;33:192-199.

42. Innocenti F, Undevia SD, Iyer L, et al. Genetic variants in the UDP-glucuronosyltransferase 1A1 gene predict the risk of severe neutropenia of irinotecan. *J Clin Oncol*. 2004;22:1382-1388.

43. Innocenti F, Ratain MJ. Pharmacogenetics of irinotecan: clinical perspectives on the utility of genotyping. *Pharmacogenomics*. 2006;7:1211-1221.

44. Watson RG, McLeod HL. Pharmacogenomic contribution to drug response. *Cancer J*. 2011;17:80-88.

45. Salgueiro N, Veiga I, Fragoso M, et al. Mutations in exon 14 of dihydropyrimidine dehydrogenase and 5-Fluorouracil toxicity in Portuguese colorectal cancer patients. *Genet Med*. 2004;6:102-107.

46. Ahluwalia R, Freimuth R, McLeod HL, Marsh S. Use of pyrosequencing to detect clinically relevant polymorphisms in dihydropyrimidine dehydrogenase. *Clin Chem*. 2003;49:1661-1664.

47. Van Kuilenburg AB, Meinsma R, Zoetekouw L, Van Gennip AH. High prevalence of the IVS14 + 1G>A mutation in the dihydropyrimidine dehydrogenase gene of patients with severe 5-fluorouracil-associated toxicity. *Pharmacogenetics*. 2002;12:555-558.

48. Van Kuilenburg AB, Meinsma R, Zoetekouw L, Van Gennip AH. Increased risk of grade IV neutropenia after administration of 5-fluorouracil due to a dihydropyrimidine dehydrogenase deficiency: high prevalence of the IVS14+1g>a mutation. *Int J Cancer*. 2002;101:253-258.

49. Schwab M, Zanger UM, Marx C, et al. Role of genetic and nongenetic factors for fluorouracil treatment-related severe toxicity: a prospective clinical trial by the German 5-FU Toxicity Study Group. *J Clin Oncol*. 2008;26:2131-2138.

50. Saif MW, Merikas I, Tsimboukis S, Syrigos K. Erlotinib-induced skin rash. Pathogenesis, clinical significance and management in pancreatic cancer patients. *JOP*. 2008;9:267-274.

51. Li T, Perez-Soler R. Skin toxicities associated with epidermal growth factor receptor inhibitors. *Target Oncol*. 2009;4:107-119.

52. Liu G, Gurubhagavatula S, Zhou W, et al. Epidermal growth factor receptor polymorphisms and clinical outcomes in non-small-cell lung cancer patients treated with gefitinib. *Pharmacogenomics J*. 2008;8:129-138.

53. Huang CL, Yang CH, Yeh KH, et al. EGFR intron 1 dinucleotide repeat polymorphism is associated with the occurrence of skin rash with gefitinib treatment. Lung *Cancer*. 2009;64:346-351.

54. Rudin CM, Liu W, Desai A, et al. Pharmacogenomic and pharmacokinetic determinants of erlotinib toxicity. *J Clin Oncol*. 2008;26:1119-1127.

55. Amador ML, Oppenheimer D, Perea S, et al. An epidermal growth factor receptor intron 1 polymorphism mediates response to epidermal growth factor receptor inhibitors. *Cancer Res*. 2004;64:9139-9143.

56. Veronese ML, Mosenkis A, Flaherty KT, et al. Mechanisms of hypertension associated with BAY 43-9006. *J Clin Oncol*. 2006;24:1363-1369.

57. Ranpura V, Pulipati B, Chu D, Zhu X, Wu S. Increased risk of high-grade hypertension with bevacizumab in cancer patients: a meta-analysis. *Am J Hypertens*. 2010;23:460-468.

58. Jubb AM, Hurwitz HI, Bai W, et al. Impact of vascular endothelial growth factor-A expression, thrombospondin-2 expression, and microvessel density on the treatment effect of bevacizumab in metastatic colorectal cancer. *J Clin Oncol.* 2006;24:217-227.

59. Schneider BP, Wang M, Radovich M, et al. Association of vascular endothelial growth factor and vascular endothelial growth factor receptor-2 genetic polymorphisms with outcome in a trial of paclitaxel compared with paclitaxel plus bevacizumab in advanced breast cancer: ECOG 2100. *J Clin Oncol*. 2008;26:4672-4678.

60. Kim JJ, Vaziri SA, Rini BI, et al. Association of VEGF and VEGFR2 single nucleotide polymorphisms with hypertension and clinical outcome in metastatic clear cell renal cell carcinoma patients treated with sunitinib. *Cancer*. 2011;118(7):1946-1954.

61. Massacesi C, Terrazzino S, Marcucci F, et al. Uridine diphosphate glucuronosyl transferase 1A1 promoter polymorphism predicts the risk of gastrointestinal toxicity and fatigue induced by irinotecan-based chemotherapy. *Cancer*. 2006;106:1007-1016.

62. Marcuello E, Altes A, Menoyo A, Del Rio E, Gomez-Pardo M, Baiget M. UGT1A1 gene variations and irinotecan treatment in patients with metastatic colorectal cancer. *Br J Cancer*. 2004;91:678-682.

63. Kristensen MH, Pedersen PL, Melsen GV, Ellehauge J, Mejer J. Variants in the dihydropyrimidine dehydrogenase, methylenetetrahydrofolate reductase and thymidylate synthase genes predict early toxicity of 5-fluorouracil in colorectal cancer patients. *J Int Med Res*. 2010;38:870-883.

64. Amstutz U, Farese S, Aebi S, Largiader CR. Dihydropyrimidine dehydrogenase gene variation and severe 5-fluorouracil toxicity: a haplotype assessment. *Pharmacogenomics*. 2009;10:931-944.

65. Zhang X, Soong R, Wang K, et al. Suppression of DPYD expression in RKO cells via DNA methylation in the regulatory region of the DPYD promoter: a potentially important epigenetic mechanism regulating DPYD expression. *Biochem Cell Biol*. 2007;85:337-346.

66. Yu J, McLeod HL, Ezzeldin HH, Diasio RB. Methylation of the DPYD promoter and dihydropyrimidine dehydrogenase deficiency. *Clin Cancer Res*. 2006;12:3864.

67. Ezzeldin HH, Lee AM, Mattison LK, Diasio RB. Methylation of the DPYD promoter: an alternative mechanism for dihydropyrimidine dehydrogenase deficiency in cancer patients. *Clin Cancer Res*. 2005;11:8699-8705.

68. Maitland ML, Vasisht K, Ratain MJ. TPMT, UGT1A1 and DPYD: genotyping to ensure safer cancer therapy? *Trends Pharmacol Sci*. 2006;27:432-437.

69. Solution for intravenous injection, irinotecan hydrochloride solution for intravenous injection. Pfizer Ltd; New York, NY, USA: 2009. CAMPTOSAR®, package insert.

70. Evans WE, Horner M, Chu YQ, Kalwinsky D, Roberts WM. Altered mercaptopurine metabolism, toxic effects, and dosage requirement in a thiopurine methyltransferase-deficient child with acute lymphocytic leukemia. *J Pediatr.* 1991;119:985-989.

71. Schmiegelow K, Forestier E, Hellebostad M, et al. Long-term results of NOPHO ALL-92 and ALL-2000 studies of childhood acute lymphoblastic leukemia. *Leukemia*. 2010;24:345-354.

72. Relling MV, Pui CH, Cheng C, Evans WE. Thiopurine methyltransferase in acute lymphoblastic leukemia. *Blood*. 2006;107:843-834.

73. Schmiegelow K, Heyman M, Kristinsson J, et al. Oral methotrexate/6-mercaptopurine may be superior to a multidrug LSA2L2 Maintenance therapy for higher risk childhood acute lymphoblastic leukemia: results from the NOPHO ALL-92 study. *J Pediatr Hematol Oncol*. 2009;31:385-392.

74. Stocco G, Cheok MH, Crews KR, et al. Genetic polymorphism of inosine triphosphate pyrophosphatase is a determinant of mercaptopurine metabolism and toxicity during treatment for acute lymphoblastic leukemia. *Clin Pharmacol Ther.* 2009;85:164-172.

75. Stocco G, Crews KR, Evans WE. Genetic polymorphism of inosine-triphosphate-pyrophosphatase influences mercaptopurine metabolism and toxicity during treatment of acute lymphoblastic leukemia individualized for thiopurine-S-methyl-transferase status. *Expert Opin Drug Saf*. 2010;9:23-37.

76. Pui CH, Sandlund JT, Pei D, et al. Improved outcome for children with acute lymphoblastic leukemia: results of Total Therapy Study XIIIB at St Jude Children's Research Hospital. *Blood*. 2004;104:2690-2696.

77. Relling MV, Klein TE. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. *Clin Pharmacol Ther*. 2011;89:464-467.

78. Relling MV, Gardner EE, Sandborn WJ, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for thiopurine methyltransferase genotype and thiopurine dosing. *Clin Pharmacol Ther*. 2011;89:387-391.

79. Toffoli G, Cecchin E, Gasparini G, et al. Genotype-driven phase I study of irinotecan administered in combination with fluorouracil/leucovorin in patients with metastatic colorectal cancer. *J Clin Oncol*. 2010;28:866-871.

80. Marcuello E, Paez D, Pare L, et al. A genotype-directed phase I-IV dose-finding study of irinotecan in combination with fluorouracil/leucovorin as first-line treatment in advanced colorectal cancer. *Br J Cancer*. 2011;105:53-57.

81. Satoh T, Ura T, Yamada Y, et al. Genotype-directed, dose-finding study of irinotecan in cancer patients with UGT1A1*28 and/or UGT1A1*6 polymorphisms. *Cancer Sci*. 2011;102:1868-1873.

82. Hu ZY, Yu Q, Pei Q, Guo C. Dose-dependent association between UGT1A1*28 genotype and irinotecan-induced neutropenia: low doses also increase risk. *Clin Cancer Res*. 2010;16:3832-3842.

83. Hoskins JM, Goldberg RM, Qu P, Ibrahim JG, McLeod HL. UGT1A1*28 genotype and irinotecan-induced neutropenia: dose matters. *J Natl Cancer Inst*. 2007;99:1290-1295.

84. Deenen MJ, Tol J, Burylo AM, et al. Relationship between single nucleotide polymorphisms and haplotypes in DPYD and toxicity and efficacy of capecitabine in advanced colorectal cancer. *Clin Cancer Res*. 2011;17:3455-3468.

85. Liu CY, Chen PM, Chiou TJ, et al. UGT1A1*28 polymorphism predicts irinotecan-induced severe toxicities without affecting treatment outcome and survival in patients with metastatic colorectal carcinoma. *Cancer*. 2008;112:1932-1940.

86. Shulman K, Cohen I, Barnett-Griness O, et al. Clinical implications of UGT1A1*28 genotype testing in colorectal cancer patients. *Cancer*. 2011;117:3156-3162.

87. Wacker B, Nagrani T, Weinberg J, Witt K, Clark G, Cagnoni PJ. Correlation between development of rash and efficacy in patients treated with the epidermal growth factor receptor tyrosine kinase inhibitor erlotinib in two large phase III studies. *Clin Cancer Res*. 2007;13:3913-3921.

88. Gibson TB, Ranganathan A, Grothey A. Randomized phase III trial results of panitumumab, a fully human anti-epidermal growth factor receptor monoclonal antibody, in metastatic colorectal cancer. *Clin Colorectal Cancer*. 2006;6:29-31.

89. Moore MJ, Goldstein D, Hamm J, et al. Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. *J Clin Oncol*. 2007;25:1960-1966.

90. Giaccone G, Gallegos Ruiz M, Le Chevalier T, et al. Erlotinib for frontline treatment of advanced non-small cell lung cancer: a phase II study. *Clin Cancer Res*. 2006;12:6049-6055.

91. Perez-Soler R, Chachoua A, Hammond LA, et al. Determinants of tumor response and survival with erlotinib in patients with non--small-cell lung cancer. *J Clin Oncol*. 2004;22:3238-3247.

92. Nie Q, Wang Z, Zhang GC, et al. The epidermal growth factor receptor intron1 (CA) n microsatellite polymorphism is a potential predictor of treatment outcome in patients with advanced lung cancer treated with Gefitinib. *Eur J Pharmacol*. 2007;570:175-181.

93. Gregorc V, Hidalgo M, Spreafico A, et al. Germline polymorphisms in EGFR and survival in patients with lung cancer receiving gefitinib. *Clin Pharmacol Ther.* 2008;83:477-484.

94. Mita AC, Papadopoulos K, de Jonge MJ, et al. Erlotinib 'dosing-to-rash': a phase II intrapatient dose escalation and pharmacologic study of erlotinib in previously treated advanced non-small cell lung cancer. *Br J Cancer*. 2011;105:938-944.

95. Miller K, Wang M, Gralow J, et al. Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer. *N Engl J Med*. 2007;357:2666-2676.

96. Dahlberg SE, Sandler AB, Brahmer JR, Schiller JH, Johnson DH. Clinical course of advanced non-small-cell lung cancer patients experiencing hypertension during treatment with bevacizumab in combination with carboplatin and paclitaxel on ECOG 4599. *J Clin Oncol.* 2010;28:949-954.

97. Dassoulas K, Gazouli M, Rizos S, et al. Common polymorphisms in the vascular endothelial growth factor gene and colorectal cancer development, prognosis, and survival. *Mol Carcinog*. 2009;48:563-569.

98. Kim JG, Chae YS, Sohn SK, et al. Vascular endothelial growth factor gene polymorphisms associated with prognosis for patients with colorectal cancer. *Clin Cancer Res*. 2008;14:62-66.

99. Loupakis F, Ruzzo A, Salvatore L, et al. Retrospective exploratory analysis of VEGF polymorphisms in the prediction of benefit from first-line FOLFIRI plus bevacizumab in metastatic colorectal cancer. *BMC Cancer*. 2011;11:247.

100. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*. 2008;24:2938-2939.

101. Des Guetz G, Uzzan B, Nicolas P, et al. Microvessel density and VEGF expression are prognostic factors in colorectal cancer. Meta-analysis of the literature. *Br J Cancer*. 2006;94:1823-1832.

102. Innocenti F, Cox NJ, Dolan ME. The use of genomic information to optimize cancer chemotherapy. *Semin Oncol*. 2011;38:186-195.

103. Fine JP, Gray, R.J. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94:496-509.

104. Klein JP, Andersen PK. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*. 2005;61:223-229.

105. Lee JJ, Feng L, Reshef DS, et al. Mortality in the randomized, controlled lung intergroup trial of isotretinoin. *Cancer Prev Res (Phila)*. 2010;3:738-744.

106. Nguyen PL, Taghian AG, Katz MS, et al. Breast cancer subtype approximated by estrogen receptor, progesterone receptor, and HER-2 is associated with local and distant recurrence after breast-conserving therapy. *J Clin Oncol.* 2008;26:2373-2378.

107. Pui CH, Pei D, Sandlund JT, et al. Risk of adverse events after completion of therapy for childhood acute lymphoblastic leukemia. *J Clin Oncol*. 2005;23:7936-7941.

108. Swen JJ, Nijenhuis M, de Boer A, et al. Pharmacogenetics: from bench to byte--an update of guidelines. *Clin Pharmacol Ther*. 2011;89:662-673.

APPENDIX 2[2]: CLINICAL VALIDITY OF NEW GENETIC BIOMARKERS OF IRINOTECAN NEUTROPENIA: AN INDPENDENT REPLICATION STUDY

A2.1 Overview

The overall goal of this study was to provide evidence for the clinical validity of nine genetic variants in five genes previously associated with irinotecan neutropenia and pharmacokinetics. Variants associated with ANC nadir and/or irinotecan pharmacokinetics in a discovery cohort of cancer patients were genotyped in an independent replication cohort of 108 cancer patients. Patients received single-agent irinotecan every three weeks. For ANC nadir, we replicated *UGT1A1*28*, *UGT1A1*93* and *SLCO1B1*1b* in univariate analyses. For irinotecan $AUC_{0\text{-}24}$, we replicated *ABCC2 -24C>T*; however *ABCC2 -24C>T* only predicted a small fraction of the variance. For SN-38 $AUC_{0\text{-}24}$ and the glucuronidation ratio, we replicated *UGT1A1*28* and *UGT1A1*93.* In addition to *UGT1A1*28*, this study independently validated *UGT1A1*93* and *SLCO1B1*1b* as new predictors of irinotecan neutropenia. Further demonstration of their clinical utility will optimize irinotecan therapy in cancer patients.

A2.2 Introduction

Irinotecan is an anticancer agent commonly used for the treatment of metastatic colorectal cancer and other solid tumors. Irinotecan is a potent inhibitor of topoisomerase I,

[2] This chapter was published *The Pharmacogenomics Journal* (Crona DJ, et al. *Pharmacogenomics J*. 2015 Apr 14 [Epub ahead of print].

and is initially hydrolyzed to its active metabolite, SN-38, which is then subsequently inactivated through UGT1A1-mediated glucuronidation. A significant proportion of patients treated with irinotecan develop toxicities, including severe neutropenia. Neutropenia is a common, serious, dose-dependent and dose-limiting toxicity of irinotecan.[1]

A common, germline genetic variation in *UGT1A1* predisposes patients to an increased risk of irinotecan-induced toxicities.[2, 3] The number of TA repeats in the *UGT1A1* promoter is inversely proportional to the transcriptional efficiency of the gene,[4] mRNA expression,[5] and protein levels.[6] Patients with the *UGT1A1*28* variant have seven TA repeats (compared to six repeats in patients with *UGT1A1*1*), have decreased SN-38 glucuronidation,[7] and experience increased systemic exposure to SN-38, which results in a higher risk of severe neutropenia.[1] As a result, an FDA-approved *UGT1A1*28* genotyping test has been made commercially available,[8] and the irinotecan label has been revised to include *UGT1A1*28* as a predisposing factor for severe neutropenia.[9]

Irinotecan-induced neutropenia is a complex, polygenic phenotype. There is significant interindividual variation in systemic exposure to both irinotecan and SN-38 that cannot be explained solely by *UGT1A1*28.* Several additional genetic variants contribute to both variability in irinotecan pharmacokinetics and the risk of severe neutropenia.[10-16] The FDA-approved *UGT1A1*28* genetic test has only moderate predictive power for severe toxicity due to its low positive predictive value,[8] and therefore the genetic test has not been incorporated into routine clinical practice. The discovery of additional variants associated with neutropenia is needed to improve the utilization of irinotecan genetic testing.

Pharmacogenetic studies have identified a vast set of genetic variants as predictors of chemotherapy efficacy and toxicity. The majority of these proposed variants have failed to

produce similar results across different studies, which has limited the clinical utility of pharmacogenetics.[17, 18] Therefore, prospective replication of pharmacogenetic findings in independent and external cohorts of patients is essential to hasten the implementation of pharmacogenetics into routine clinical practice.

In a previous study of cancer patients treated with single-agent irinotecan, novel gene variants that were associated with irinotecan disposition and toxicity were identified.[16] In addition to *UGT1A1*28*, other variants, mostly in drug transporter genes, were associated with neutropenia and irinotecan pharmacokinetics. Therefore, we conducted a replication study to test the clinical validity of these variants in an external cohort of cancer patients treated with single-agent irinotecan.

A2.3 Materials and Methods

A2.3.1 Study design

The overall goal of the study was to replicate genetic associations for irinotecan neutropenia and pharmacokinetics previously identified in a discovery cohort.[16] The primary objective was to validate the associations between four genetic variants and absolute neutrophil count (ANC) nadir by testing them in an external replication cohort. The secondary objective was to validate the effects of eight genetic variants previously associated with pharmacokinetic parameters in the discovery cohort by analyzing them in the replication cohort. Thus, a total of nine common variants in five genes (ANC nadir and the pharmacokinetic phenotypes shared two variants) were genotyped in the replication cohort and tested for associations. Variants for replication testing were selected based on significant genotype-phenotype associations ($p \leq 0.05$) observed in the discovery cohort. All patients in

the replication cohort were White, and therefore only the previously genotyped White patients comprised the discovery cohort (n=67).[16]

A2.3.2 Patient characteristics

In the discovery cohort, advanced solid tumor patients were treated at the University of Chicago (Chicago, USA) with a 90-min infusion of single-agent irinotecan every three weeks at 300 mg/m^2 (n=18) or 350 mg/m^2 (n=49). Eligibility criteria included adequate hematopoietic function (white blood cell count $\geq$3,500/μL, ANC $\geq$1,500/μL, platelets $\geq$100,000/μL), normal renal and hepatic function (creatinine $\leq$1.5 mg/dL, total bilirubin $\leq$1.25 x upper limit of normal (ULN), and AST/ALT <5 x ULN), and adequate performance status (Karnofsky score $\geq$70%). Plasma pharmacokinetic parameters of irinotecan and metabolites were measured during and after the first cycle infusion of irinotecan. Forty-two genetic variants in 12 candidate genes of the irinotecan pathway were previously genotyped and tested for association with irinotecan pharmacokinetics and ANC nadir, measured during cycle 1.

In the replication cohort, 108 White advanced solid tumor patients were treated at the Erasmus University Medical Center, Erasmus MC Cancer Institute (Rotterdam, The Netherlands).[19-21] Patients received a 90-min infusion of single-agent irinotecan every three weeks at 600 mg (flat dose, n=58), 350 mg/m^2 (n=31), or 380-1060 mg (flat dose calculated according to an algorithm,[19] n=19). Eligibility criteria included adequate hematopoietic function (ANC $\geq$2,000/μL, platelets $\geq$100,000/μL), and normal renal and hepatic function (creatinine clearance $\geq$60 mL/min, total bilirubin $\leq$1.25 x ULN, and AST/ALT $\leq$3 x ULN).

Plasma pharmacokinetics of irinotecan and metabolites were measured during and after the first cycle infusion.

All patients in the discovery and replication cohorts provided written informed consent and the local institutional review boards approved the clinical protocols. Patient characteristics from the discovery and replication cohorts are provided in Table A2.1.

A2.3.3 Patient phenotyping: pharmacokinetic parameters and ANC nadir

In both cohorts, pharmacokinetic parameters included: irinotecan area under the concentration-time curve to the last time of sampling (AUC_{0-24}), AUC_{0-24} of the active SN-38 metabolite, AUC_{0-24} of the inactive SN-38 glucuronide (SN-38G), and the ratio of SN-38G AUC_{0-24} to SN-38 AUC_{0-24} (glucuronidation ratio).

For the discovery cohort, samples were collected on day 1 of cycle 1 at baseline prior to irinotecan infusion, during the infusion (30, 60, and 90 min), and after the infusion (10, 20, 30 and 45 min, 1, 1.5, 2, 4, 6, 8, 12 and 24 h). Plasma concentrations of irinotecan and metabolites were measured, as previously reported.[10] Pharmacokinetic parameters were calculated by non-compartmental analysis (WinNonlin®, Pharsight Corp., Cary, NC, USA).

For the replication cohort, samples were collected on day 1 of cycle 1 at baseline prior to infusion, during the infusion (30 and 90 min) and after the infusion (10, 20 and 30 min, and 1, 1.5, 2, 3, 4, 5, 6, 8, 10, 12 and 24 h). Plasma concentrations of irinotecan and metabolites were measured, as previously reported.[20, 22, 23] Pharmacokinetic parameters were calculated by non-compartmental analysis (PK Solutions v2.0, Summit Research Services, Montrose, CO, USA).

In both cohorts, complete blood counts were taken at baseline, weekly throughout cycle 1, and then prior to the start of cycle 2 to obtain the measurements of the ANC nadir.

A2.3.4 Genotype data

Nine common variants, previously associated with irinotecan pharmacokinetics and ANC nadir in the discovery cohort, were genotyped in the replication cohort: *ABCB1 IVS9 44A>G*, *ABCC1 1684T>C*, *ABCC1 IVS11 -48C>T*, *ABCC2 3972C>T*, *ABCC2 -24C>T*, *SLCO1B1*1b*, *SLCO1B1*5*, *UGT1A1*28*, and *UGT1A1*93*. DNA isolated from peripheral blood was used for genotyping. All genotyping assays were performed on an Applied Biosystems TaqMan 7500 (Life Technologies, Grand Island, NY, USA). *UGT1A1*93* was genotyped by restriction fragment length polymorphism PCR, using 5'-ACCTCTAGTTACATAACCTGAA-3' as the forward primer sequence and 5'-ATAAACCCGACCTCACCAC-3' as the reverse primer sequence. *UGT1A1*28* genotyping methods for the replication cohort have been previously described.[20] All other variants were genotyped using TaqMan SNP genotyping assays (Life Technologies, Grand Island, NY, USA) as per the manufacturer's instructions. Positive controls of known genotypes were used in the assays.

A2.3.5 Statistics

Data for all phenotypes for both cohorts were $\log_{10}$ transformed. Hardy-Weinberg Equilibrium was evaluated for all nine variants genotyped in both the discovery and replication cohorts (Table A2.S1). In the discovery cohort, associations between genetic variants and clinical phenotypes were analyzed using linear regression, and were adjusted for

sex, age, and irinotecan dose (300 or 350 mg/m^2). All ANC nadir analyses were also adjusted for baseline ANC.

In the replication cohort, we prospectively tested associations between the nine gene variants described above and phenotypes of ANC nadir and irinotecan pharmacokinetics. The same statistical methodologies employed for the discovery cohort were applied: linear regression adjusted for sex, age, and irinotecan dose (350 mg/m^2, 600 mg flat dose, or dose by an algorithm[19]), with baseline ANC used to adjust the ANC nadir analysis. Flat doses were converted to mg/m^2 according to the body-surface area of each patient. The same mode of inheritance (dominant, recessive, or additive) used in the discovery cohort was also used in the replication cohort.

No general consensus exists to provide standardized criteria for replication cohort analyses. We considered a given variant's association to be replicated based on direct comparison of the observed estimates of effect in the discovery and replication cohorts: an association's estimate of effect in the replication cohort had to be in the same direction as in the discovery cohort (an increased or decreased estimate of phenotype change in both cohorts), and lie within the 95% confidence interval (CI) of the discovery cohort's estimate. Two-sided p-values are reported for reference. Since comparisons between discovery and replication cohort estimates of effect were pre-specified and rely on 95% CIs from the discovery cohort, not on hypothesis testing in the replication cohort, issues related to multiplicity are not present. Therefore, no correction for multiple comparisons was performed.

A2.4 Results

This study sought to replicate, in an independent, external cohort of White cancer patients from the Netherlands, nine variants from five genes that had previously associated with ANC nadir or irinotecan pharmacokinetics[16]. Baseline clinical patient characteristics and pharmacokinetic data (Table A2.1), as well as allele and genotype frequencies (Table A2.S1) were comparable between the two cohorts. Below we report the replication results of each variant for neutropenia and irinotecan pharmacokinetics (Table A2.2).

A2.4.1 Replication of variants previously associated with ANC nadir

For ANC nadir, four variants that previously associated with ANC nadir in the discovery cohort were tested in the replication cohort. In the discovery cohort, *UGT1A1*28* (additive model), *UGT1A1*93* (recessive model), and *ABCC1 IVS11 -48C>T* (recessive model) were associated with decreased ANC nadir; *SLCO1B1*1b* (dominant model) was associated with increased ANC nadir. In the replication cohort, we considered *UGT1A1*28*, *UGT1A1*93* and *SLCO1B1*1b* replicated, since the direction of the estimate of the effect for each variant was consistent between both cohorts (decreased ANC nadir for *UGT1A1*28* and *UGT1A1*93*, as well as increased ANC nadir for *SLCO1B1*1b*) and each was within the 95% CIs for its respective discovery cohort estimate. *ABCC1 IVS11 -48C>T* failed to replicate (Table A2.2).

A2.4.2 Replication of variants associated with the pharmacokinetic parameters of irinotecan

For irinotecan $AUC_{0\text{-}24}$, two variants that were previously associated with irinotecan $AUC_{0\text{-}24}$ in the discovery cohort were tested in the replication cohort. In the discovery cohort, *ABCC2 -24C>T* and *SLCO1B1*5* (both dominant model) were associated with increased

irinotecan $AUC_{0\text{-}24}$. In the replication cohort, we considered *ABCC2 -24C>T* replicated since the direction of the estimate of the effect was consistent between both cohorts (increased $AUC_{0\text{-}24}$ for both variants), and was within the 95% CIs for the discovery cohort estimate. *SLCO1B1*5* failed to replicate (Table A2.2).

For SN-38 $AUC_{0\text{-}24}$, three variants that were previously associated with SN-38 $AUC_{0\text{-}24}$ in the discovery cohort were tested in the replication cohort. In the discovery cohort, *UGT1A1*28* and *UGT1A1*93* (both additive model) were associated with increased SN-38 $AUC_{0\text{-}24}$, while *ABCB1 IVS9 -44A>G* (dominant model) was associated with decreased SN-38 $AUC_{0\text{-}24}$. In the replication cohort, we considered *UGT1A1*28* and *UGT1A1*93* replicated since the direction of the estimate of the effect for each variant was consistent between both cohorts (increased $AUC_{0\text{-}24}$ for both variants), and each was within the 95% CIs for its respective discovery cohort estimate. *ABCB1 IVS9 -44A>G* failed to replicate (Table A2.2).

For SN-38G $AUC_{0\text{-}24}$, although *ABCC2 3972C>T* (recessive model) was associated with increased SN-38G $AUC_{0\text{-}24}$ in the discovery cohort, it failed to replicate when tested in the replication cohort (Table A2.2).

For the glucuronidation ratio, three variants that associated with the glucuronidation ratio in the discovery cohort were tested in the replication cohort. In the discovery cohort, *UGT1A1*28* (additive model), *UGT1A1*93* (additive model), and *ABCC1 1684T>C* (dominant model) were associated with a decreased glucuronidation ratio. In the replication cohort, we considered *UGT1A1*28* and *UGT1A1*93*, replicated, since the direction of the estimate of the effect for each variant was consistent between both cohorts (decreased glucuronidation ratio for all variants), and each was within the 95% CIs for its respective discovery cohort estimate. Although the association between *ABCC1 1684T>C* (dominant

model) and glucuronidation ratio also satisfies our criteria for replication, we are less convinced of the association, given the 84% reduction in the magnitude of the estimate as compared to that of the discovery cohort (Table A2.2).

A2.5 Discussion

In this replication study, we validated the clinical effects of new germline genetic variants for neutropenia and irinotecan pharmacokinetics using an independent, external cohort of White cancer patients treated with single-agent irinotecan.

The most important result of this study was the clinical validation of *SLCO1B1*1b*. To our knowledge, this provides the first replicated data implicating *SLCO1B1*1b* as a protective marker against irinotecan-induced neutropenia. *SLCO1B1* encodes for organic anion transporter family member 1B1 (OATP1B1), and mediates hepatic uptake of both endogenous[24, 25] and xenobiotic compounds.[26] OATP1B1 is a hepatic uptake transporter of SN-38,[27, 28] but not irinotecan.[28] In this study, we have replicated results from the discovery cohort, and have shown that the variant **1b* allele was associated with a higher ANC nadir compared to the reference sequence **1a* allele (Figure A2.1A). Since *SLCO1B1*1b* is a non-synonymous variant (asparagine to aspartate amino acid change), and *SLCO1B1* is primarily expressed in the liver,[29] we postulate this variant might associate with reduced neutropenia by altering systemic SN-38 exposure. The effect of *SLCO1B1*1b* on SN-38 $AUC_{0\text{-}24}$ was -0.083±0.076 (mean±SE) in the White patients of the discovery cohort (n=67; p=0.278), and because the p-value was >0.05, this association was not selected for analysis in the replication cohort. However, an exploratory univariate analysis (adjusted for dose (mg/m^2), age and sex) revealed that *SLCO1B1*1b* was associated with decreased SN-38 $AUC_{0\text{-}24}$ in

the replication cohort (n=84; -0.128±0.055, p=0.023). These results support the hypothesis that the protective effect of *SLCO1B1*1b* against neutropenia could be due to increased hepatic uptake of SN-38, resulting in increased SN-38 elimination from the plasma after irinotecan infusion.

While the pharmacokinetic data are supportive of the protective effect of *SLCO1B1*1b* against neutropenia, the functional effect of this variant is less clear. Using RNA expression data from human livers,[30] *SLCO1B1*1b* (as well as variants in linkage disequilibrium $r^2 \geq 0.8$) did not associate with changes in the mRNA expression of *SLCO1B1* (results not shown). In oocyte studies, the uptake of SN-38 was higher for *SLCO1B1*1b* than *SLCO1B1*1a* (the reference sequence allele), but the observed difference was not statistically significant (see Figure 6a of Nozawa *et al.*[28]). Our results provide evidence that *SLCO1B1*1b* results in a gain of function, which leads to increased hepatic uptake of SN-38 from the plasma. Although this seems the most plausible hypothesis, other mechanisms related to the widespread functions of this transporter on several endogenous constituents cannot be ruled out.

Another important conclusion of this study is that *UGT1A1*93* confers an increased risk of irinotecan-induced neutropenia. We replicated results from the discovery cohort, and have shown that the **93* variant was associated with a lower ANC nadir compared to the reference sequence **1* allele (Figure A2.1B). *UGT1A1*93* is a -3156G>A change discovered during a resequencing study of the region 5′ to the *UGT1A* exon 1.[31] According to an analysis of more than 150 human livers where genome-wide genotyping data were available, *UGT1A1*93* is a major determinant of decreased levels of the UGT1A1 protein (Pearson's $r=-0.46$, $p=3.5 \times 10^{-9}$),[30, 32] and additional preliminary data corroborate these findings.[33]

Because *UGT1A1*93* is in partial linkage disequilibrium with *UGT1A1*28* among White patients (r^2=0.68),[34] our results suggest that *UGT1A1*93,* based on its greater estimate of effect for ANC nadir, may be a more robust marker for neutropenia than *UGT1A1*28* (Table A2.2). While the *UGT1A1*93* variant has not yet been included in the FDA-revised irinotecan label, we envision that recommendations supporting *UGT1A1*93* genotyping could eventually replace *UGT1A1*28* in the irinotecan drug label.

The association between *ABCC2 -24C>T* and increased irinotecan $AUC_{0\text{-}24}$ was also replicated (Figure A2.1C). *ABCC2* encodes for the multidrug resistant protein-2 and contributes to the biliary clearance of irinotecan, SN-38 and SN-38G.[35, 36] The *-24 C>T* variant has been associated with a nearly 20% reduction in promoter activity.[37] This observation is consistent with our results, where the variant T allele was associated with increased irinotecan $AUC_{0\text{-}24}$, likely due to decreased biliary clearance. However the estimate of effect size was relatively small (Table A2.2), and additional studies should be conducted to elucidate the extent of its clinical relevance.

Established criteria for conducting pharmacogenetic replication studies do not currently exist, but we provide a general framework for conducting such studies. Pharmacogenetic replication studies are beset with numerous challenges, including dosing and population heterogeneity between the discovery and replication cohorts. In our study, we attempted to control for population heterogeneity by comparing patients in the replication cohort to only the White patients from the original discovery cohort.[16] Dosing heterogeneity between the two cohorts may have affected our ability to replicate some variants, but it did not confound all associations, as evidenced by the detection of associations serving as "positive controls", such as *UGT1A1*28* versus SN-38 $AUC_{0\text{-}24}$ and *UGT1A1*28* versus

glucuronidation ratio (but not irinotecan $AUC_{0\text{-}24}$). Moreover, we are confident that dosing heterogeneity did not significantly confound our replication results because irinotecan has been shown to demonstrate dose linear pharmacokinetics over a wide range of doses.[38] Regarding our statistical approach, the assessment of replicated associations is not based on hypothesis testing, and therefore using p-values as our main criteria for replication would have been inappropriate. Moreover, given the influence of sample size on p-values, utilization of p-values as the main criteria for replication could have resulted in false negative results. We also cannot exclude the possibility that between-cohort differences limited our ability to detect phenotypic differences and replicate several variants.

This replication study allowed us to demonstrate the clinical validity of associations between *UGT1A1*93* and *SLCO1B1*1b* and neutropenia. The effects of these two variants on neutropenia should be confirmed in studies where irinotecan is given in combination with other anticancer agents that have neutropenic effects (for example with 5-fluorouracil). Additionally, the effects of these replicated variants can currently be applied only to White patients. Efforts should be made to validate these variants in patients from other races who receive irinotecan. Further validation of their clinical utility will aid in the implementation of routine irinotecan pharmacogenetic testing and optimization of personalized treatments for cancer patients.

TABLES

Table 1. Baseline patient characteristics and pharmacokinetic data from the discovery and the replication cohorts. Flat dosing and dosing by algorithm[19] were used only in the replication cohort. The distribution of the algorithm-derived doses includes: 380 mg (n=1), 500 mg (n=1), 520 mg (n=2), 540 mg (n=1), 560 mg (n=1), 620 mg (n=2), 640 mg (n=1), 660 mg (n=1), 680 mg (n=1), 720 mg (n=2), 740 mg (n=3), 780 mg (n=1), 900 mg (n=1), 1060 mg (n=1).

	Discovery Cohort (n=67)		**Replication Cohort (n=108)**	
Dose				
300 mg/m^2	18 (26.9%)		-	
350 mg/m^2	49 (73.1%)		31 (29.7%)	
Flat dose (600 mg)	-		58 (53.7%)	
Dose by algorithm (380-1060 mg)	-		19 (17.1%)	
Sex				
Male	42 (62.7%)		60 (55.6%)	
Female	25 (37.3%)		48 (44.4%)	
	Median	**Range**	**Median**	**Range**
Age (years)	57	34-85	58	26-75
BSA (m^2)	1.87	1.46-2.55	1.88	1.36-2.50
Baseline ANC (cells/μL)	5.27	2.18-14.36	5.12	1.30-13.60
ANC Nadir (cells/μL)	2.21	0.05-7.83	1.79	0.03-7.13
Pharmacokinetic Parameters	**Mean**	**Range**	**Mean**	**Range**
Irinotecan AUC_{0-24} (h*ng/mL)	23251	8857-65305	22776	11422-67560
SN-38 AUC_{0-24} (h*ng/mL)	385	38-1957	364	79-1776
SN-38G AUC_{0-24} (h*ng/mL)	1824	360-8214	2238	396-6912
SN-38G AUC_{0-24} to SN-38 AUC_{0-24} ratio (glucuronidation ratio)	5.85	0.78-37.57	7.32	1-24.07

Table 2. Univariate analyses of the associations between genetic variants and phenotypes. Data were adjusted for age, sex, and dose (mg/m^2). ANC nadir was also adjusted for baseline ANC. The genotype reference groups were the same for all discovery and replication cohort analyses, with the exception of *ABCC1 IVS11 -48C>T* and ANC nadir. For *ABCC1 IVS11 -48C>T* and ANC nadir in the replication cohort, the reference genotype was only CC, as there were no TT genotypes. The estimates of effect of replicated variants are denoted in bold. The number of patients genotyped per variant in the replication cohort varied due to insufficient DNA quantity. *Although the association between *ABCC1 1684T>C* (dominant model) and glucuronidation ratio satisfies our criteria for replication, we are less convinced of the association, given the 84% reduction in the magnitude of the estimate as compared to that of the discovery cohort. Abbreviation: Ref = Reference.

		Discovery Cohort (n=67)			Replication Cohort (n=74-103)		
Variant	**Ref**	**Estimate±SE**	**95% CI**	**P-value**	**n**	**Estimate±SE**	**P-value**
Log_{10} ANC Nadir							
ABCC1 IVS11 -48 (C>T)	CC/CT	-0.489±0.201	(-0.095, -0.883)	0.018	74	0.053±0.117	0.652
*UGT1A1*28 (TA6>TA7)*	*1*1=1 *1*28=2 *28*28=3	-0.257±0.061	(-0.137, -0.377)	<0.001	75	**-0.139±0.081**	0.090
*UGT1A1*93 (G>A)*	GG/AG	-0.607±0.129	(-0.354, -0.860)	<0.001	103	**-0.417±0.170**	0.016
*SLCO1B1*1b (A>G)*	AA	0.240±0.106	(0.032, 0.448)	0.027	84	**0.278±0.107**	0.012
Log_{10} irinotecan AUC_{0-24}							
ABCC2 -24 (C>T)	CC	0.090±0.035	(0.021, 0.159)	0.012	94	**0.067±0.030**	0.040
*SLCO1B1*5 (T>C)*	TT	0.084±0.036	(0.013, 0.155)	0.023	83	-0.010±0.035	0.769
Log_{10} SN-38 AUC_{0-24}							
*UGT1A1*28 (TA6>TA7)*	*1*1=1 *1*28=2 *28*28=3	0.140±0.046	(0.050, 0.190)	0.004	75	**0.189±0.043**	<0.001
*UGT1A1*93 (G>A)*	GG=1 AG=2 AA=3	0.130±0.047	(0.038, 0.222)	0.007	103	**0.171±0.033**	<0.001
ABCB1 IVS9 -44 (A>G)	AA	-0.180±0.070	(-0.043, -0.317)	0.013	77	0.021±0.054	0.703

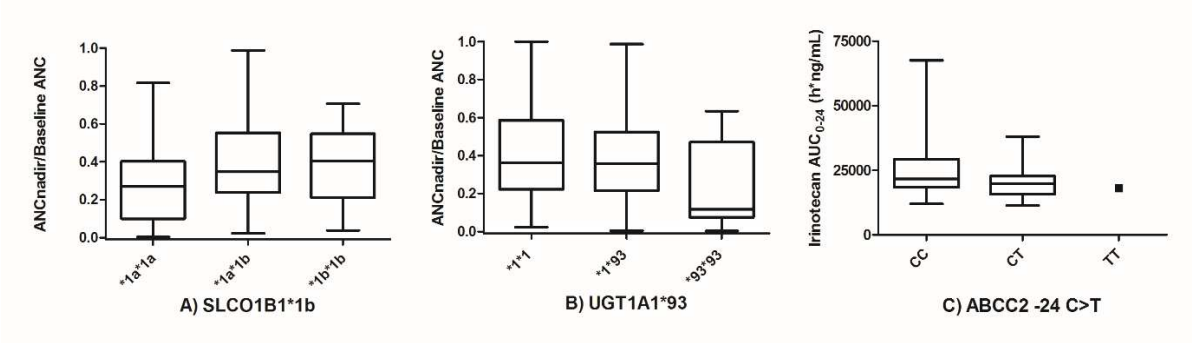
Log_{10} SN-38G AUC_{0-24}							
ABCC2 3972 (C>T)	CC/CT	0.250±0.100	(0.054, 0.446)	0.019	105	-0.104±0.086	0.232
Log_{10} SN-38G AUC_{0-24} /Log_{10} SN-38 AUC_{0-24}							
ABCC1 1684 (T>C)	TT	-0.316±0.153	(-0.016, -0.616)	0.043	95	**-0.052±0.059***	0.382
*UGT1A1*28 (TA6>TA7)*	*1*1=1 *1*28=2 *28*28=3	-0.170±0.048	(-0.076, -0.264)	<0.001	103	**-0.243±0.053**	<0.001
*UGT1A1*93 (G>A)*	GG=1 AG=2 AA=3	-0.150±0.051	(-0.050, -0.250)	0.004	75	**-0.214±0.042**	<0.001

Table A2.S1. Allele and genotype frequencies of the nine variants genotyped in both the discovery and the replication cohorts. In the replication cohort, DNA was insufficient to genotype the nine variants in all patients; therefore, the number of genotyped patients varies by variant. HWE=Hardy Weinberg Equilibrium; MAF=minor allele frequency

			Discovery Cohort (n=67)			Replication Cohort (n=74-104)		
Gene	**Variant (rs#)**	**Function**	**N (%)**	**MAF**	**HWE**	**N (%)**	**MAF**	**HWE**
ABCB1	IVS9 -44 A>G (rs10276036)	Intronic		0.46	0.01	77	0.35	0.82
AA			25 (37)			32 (41)		
AG			23 (34)			36 (47)		
GG			19 (28)			9 (12)		
ABCC1	IVS11 -48 C>T (rs3765129)	Intronic		0.17	0.08	74	0.17	0.08
CC			48 (72)			49 (66)		
CT			15 (22)			25 (34)		
TT			4 (6)			0 (0)		
ABCC1	1684 T>C (rs35605)	Synonymous, Exon 13		0.23	0.78	95	0.22	0.03
TT			40 (60)			54 (57)		
TC			23 (34)			40 (42)		
CC			4 (6)			1 (1)		
ABCC2	-24 C>T (rs717620)	5' UTR		0.18	0.34	94	0.17	0.21
CC			44 (66)			63 (67)		
CT			22 (33)			30 (32)		
TT			1 (1)			1 (1)		
ABCC2	3972C>T (rs3740066)	Synonymous, Exon 28		0.34	0.81	104	0.36	0.84
CC			30 (45)			43 (41)		
CT			29 (43)			47 (45)		
TT			8 (12)			14 (14)		
SLCO1B1*1 b	388 A>G (rs2306283)	Nonsynonymous, Exon 5 (N130D)		0.46	0.58	84	0.40	0.73
AA			21 (31)			29 (35)		
AG			31 (46)			42 (50)		
GG			15 (22)			13 (15)		
SLCO1B1*5	521 T>C (rs4149056)	Nonsynonymous, Exon 6 (V174A)		0.19	0.79	83	0.13	0.74
TT			44 (66)			63 (76)		
TC			21 (31)			19 (23)		
CC			2 (3)			1 (1)		
UGT1A1*93	-3156 G>A (rs10929302)	Intronic		0.32	0.24	103	0.28	0.57
GG			33 (49)			52 (50)		
GA			25 (37)			44 (43)		
AA			9 (13)			7 (7)		
UGT1A1*28	TA6>TA7 (rs8175347)	TA repeat, Promoter		0.34	0.25	75	0.33	0.11
*1/*1			31 (46)			31 (41)		
*1/*28			26 (39)			39 (52)		
*28/*28			10 (15)			5 (7)		

FIGURES

Figure A2.1. Associations between *SLCO1B1*1b* and ANC nadir (A), *UGT1A1*93* and ANC nadir (B), and *ABCC2 -24C>T* and log_{10} irinotecan AUC_{0-24} (C) in the replication cohort. For the purpose of illustrating the replicated genetic associations, the data are not adjusted for the same factors used in the univariate analyses, and the differences among genotypes might not be the same as the ones reported in Table A2.2. ANC nadir is normalized to the baseline pretreatment ANC. Data are expressed as medians, 25th and 75th percentiles, minimums and maximums.



REFERENCES

1. Innocenti F, Undevia SD, Iyer L, Chen PX, Das S, Kocherginsky M, et al. Genetic variants in the UDP-glucuronosyltransferase 1A1 gene predict the risk of severe neutropenia of irinotecan. *J Clin Oncol*. 2004;22(8):1382-1388.

2. Kim TW, Innocenti F. Insights, challenges, and future directions in irinogenetics. *Ther Drug Monit*. 2007;29(3): 265-270.

3. Hoskins JM, Goldberg RM, Qu P, Ibrahim JG, McLeod HL. UGT1A1*28 genotype and irinotecan-induced neutropenia: dose matters. *J Natl Cancer Inst*. 2007;99(17):1290-1295.

4. Beutler E, Gelbart T, Demina A. Racial variability in the UDP-glucuronosyltransferase 1 (UGT1A1) promoter: a balanced polymorphism for regulation of bilirubin metabolism? *Proc Natl Acad Sci U S A*. 1998;95(14):8170-8174.

5. Ramirez J, Mirkov S, Zhang W, Chen P, Das S, Liu W, et al. Hepatocyte nuclear factor-1 alpha is associated with UGT1A1, UGT1A9 and UGT2B7 mRNA expression in human liver. *Pharmacogenomics J.* 2008;8(2):152-161.

6. Ritter JK, Kessler FK, Thompson MT, Grove AD, Auyeung DJ, Fisher RA. Expression and inducibility of the human bilirubin UDP-glucuronosyltransferase UGT1A1 in liver and cultured primary hepatocytes: evidence for both genetic and environmental influences. *Hepatology*. 1999;30(2):476-484..

7. Iyer L, King CD, Whitington PF, Green MD, Roy SK, Tephly TR, et al. Genetic predisposition to the metabolism of irinotecan (CPT-11). Role of uridine diphosphate glucuronosyltransferase isoform 1A1 in the glucuronidation of its active metabolite (SN-38) in human liver microsomes. *J Clin Invest*. 1998;101(4):847-854.

8. Innocenti F, Ratain MJ. Pharmacogenetics of irinotecan: clinical perspectives on the utility of genotyping. *Pharmacogenomics*. 2006;7(8):1211-1221.

9. Product Information. Camptosar (irinotecan). New York NPUC, July 2012. Available at: http://labeling.pfizer.com/ShowLabeling.aspx?id=533.

10. Iyer L, Das S, Janisch L, Wen M, Ramirez J, Karrison T, et al. UGT1A1*28 polymorphism as a determinant of irinotecan disposition and toxicity. *Pharmacogenomics J*. 2002;2(1):43-47.

11. Han JY, Lim HS, Shin ES, Yoo YK, Park YH, Lee JE, et al. Influence of the organic anion-transporting polypeptide 1B1 (OATP1B1) polymorphisms on irinotecan-pharmacokinetics and clinical outcome of patients with advanced non-small cell lung cancer. *Lung Cancer*. 2008;59(1):69-75.

12. Han JY, Lim HS, Yoo YK, Shin ES, Park YH, Lee SY, et al. Associations of ABCB1, ABCC2, and ABCG2 polymorphisms with irinotecan-pharmacokinetics and clinical outcome in patients with advanced non-small cell lung cancer. *Cancer*. 2007;110(1):138-147.

13. Xiang X, Jada SR, Li HH, Fan L, Tham LS, Wong CI, et al. Pharmacogenetics of SLCO1B1 gene and the impact of *1b and *15 haplotypes on irinotecan disposition in Asian cancer patients. *Pharmacogenet Genomics*. 2006;16(9):683-691.

14. Mathijssen RH, Marsh S, Karlsson MO, Xie R, Baker SD, Verweij J, et al. Irinotecan pathway genotype analysis to predict pharmacokinetics. *Clin Cancer Res*. 2003;9(9):3246-3253.

15. Rosner GL, Panetta JC, Innocenti F, Ratain MJ. Pharmacogenetic pathway analysis of irinotecan. *Clin Pharmacol Ther*. 2008; 4(3): 93-402.

16. Innocenti F, Kroetz DL, Schuetz E, Dolan ME, Ramirez J, Relling M, et al. Comprehensive pharmacogenetic analysis of irinotecan neutropenia and pharmacokinetics. *J Clin Oncol.* 2009;27(16):2604-2614.

17. Coate L, Cuffe S, Horgan A, Hung RJ, Christiani D, Liu G. Germline genetic variation, cancer outcome, and pharmacogenetics. *J Clin Oncol*. 2010;28(26):4029-4037.

18. Innocenti F, Schilsky RL. Translating the cancer genome into clinically useful tools and strategies. *Dis Model Mech* . 2009;2(9-10):426-429.

19. van der Bol JM, Mathijssen RH, Creemers GJ, Planting AS, Loos WJ, Wiemer EA, et al. A CYP3A4 phenotype-based dosing algorithm for individualized treatment of irinotecan. *Clin Cancer Res*. 2010;16(2):736-742.

20. Mathijssen RH, de Jong FA, van Schaik RH, Lepper ER, Friberg LE, Rietveld T, et al. Prediction of irinotecan pharmacokinetics by use of cytochrome P450 3A4 phenotyping probes. *J Natl Cancer Inst*. 2004;96(21):1585-1592.

21. de Jong FA, Kehrer DF, Mathijssen RH, Creemers GJ, de Bruijn P, van Schaik RH, et al. Prophylaxis of irinotecan-induced diarrhea with neomycin and potential role for UGT1A1*28 genotype screening: a double-blind, randomized, placebo-controlled study. *Oncologist*. 2006;11(8):944-954.

22. de Bruijn P, Verweij J, Loos WJ, Nooter K, Stoter G, Sparreboom A. Determination of irinotecan (CPT-11) and its active metabolite SN-38 in human plasma by reversed-phase high-performance liquid chromatography with fluorescence detection. *J Chromatogr B Biomed Sci Appl*. 1997;698(1-2):277-285.

23. de Bruijn P, Willems EW, Loos WJ, Verweij J, Sparreboom A. Indirect determination of the irinotecan metabolite 7-ethyl-10-O-glucuronyl-camptothecin in human samples. *Anal Biochem*. 2004;328(1):84-86.

24. Abe T, Kakyo M, Tokui T, Nakagomi R, Nishio T, Nakai D, et al. Identification of a novel gene family encoding human liver-specific organic anion transporter LST-1. *J Biol Chem*. 1999;274(24):17159-17163.

25. Hsiang B, Zhu Y, Wang Z, Wu Y, Sasseville V, Yang WP, et al. A novel human hepatic organic anion transporting polypeptide (OATP2). Identification of a liver-specific human organic anion transporting polypeptide and identification of rat and human hydroxymethylglutaryl-CoA reductase inhibitor transporters. *J Biol Chem*. 1999;274(52):37161-37168.

26. Oshiro C, Mangravite L, Klein T, Altman R. PharmGKB very important pharmacogene: SLCO1B1. *Pharmacogenet Genomics*. 2010;20(3):211-216.

27. Iusuf D, Ludwig M, Elbatsh A, van Esch A, van de Steeg E, Wagenaar E, et al. OATP1A/1B transporters affect irinotecan and SN-38 pharmacokinetics and carboxylesterase expression in knockout and humanized transgenic mice. *Mol Cancer Ther.* 2014;13(2):492-503.

28. Nozawa T, Minami H, Sugiura S, Tsuji A, Tamai I. Role of organic anion transporter OATP1B1 (OATP-C) in hepatic uptake of irinotecan and its active metabolite, 7-ethyl-10-hydroxycamptothecin: in vitro evidence and effect of single nucleotide polymorphisms. *Drug Metab Dispos*. 2005;33(3):434-439.

29. Niemi M, Pasanen MK, Neuvonen PJ. Organic anion transporting polypeptide 1B1: a genetically polymorphic transporter of major importance for hepatic drug uptake. *Pharmacol Rev*. 2011;63(1):157-181.

30. Innocenti F, Cooper GM, Stanaway IB, Gamazon ER, Smith JD, Mirkov S, et al. Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. *PLoS Genet*. 2011;7(5):e1002078.

31. Innocenti F, Grimsley C, Das S, Ramirez J, Cheng C, Kuttab-Boulos H, et al. Haplotype structure of the UDP-glucuronosyltransferase 1A1 promoter in different ethnic groups. *Pharmacogenetics*. 2002;12(9):725-733.

32. Gillis N, Seiser E, Fallon J, Smith P, Innocenti F. Genome-wide analysis of the variation in hepatic protein expression of 22 key drug metabolizing enzymes. American Society of Clinical Pharmacology and Therapeutics (ASCPT) 2014 Annual Meeting (March 14-16, 2014). Abstract number 765 and poster number LBII-017.

33. Peterkin V, Bauman JN, Goosen TC, Paulauskis J, Williams JA, Myrand SP. Genetic variant UGT1A1*93 (-3156 G>A) is predictive of UGT1A1 enzyme activity and protein expression in human liver microsomes. 15th North American Regional International Society for the Study of Xenobiotics (ISSX) Meeting (October 12-16, 2008). Abstract number 331.

34. Cecchin E, Innocenti F, D'Andrea M, Corona G, De Mattia E, Biason P, et al. Predictive role of the UGT1A1, UGT1A7, and UGT1A9 genetic variants and their

haplotypes on the outcome of metastatic colorectal cancer patients treated with fluorouracil, leucovorin, and irinotecan. *J Clin Oncol*. 2009;27(15):2457-2465.

35. Nguyen TD, Markova S, Liu W, Gow JM, Baldwin RM, Habashian M, et al. Functional characterization of ABCC2 promoter polymorphisms and allele-specific expression. *Pharmacogenomics J*. 2013;13(5):396-402.

36. Sugiyama Y, Kato Y, Chu X. Multiplicity of biliary excretion mechanisms for the camptothecin derivative irinotecan (CPT-11), its metabolite SN-38, and its glucuronide: role of canalicular multispecific organic anion transporter and P-glycoprotein. *Cancer Chemother Pharmacol*. 1998;42 Suppl:S44-49.

37. Haenisch S, Zimmermann U, Dazert E, Wruck CJ, Dazert P, Siegmund W, et al. Influence of polymorphisms of ABCB1 and ABCC2 on mRNA and protein expression in normal and cancerous kidney cortex. *Pharmacogenomics J*. 2007;7(1):56-65.

38. Chabot GG, Abigerges D, Catimel G, Culine S, de Forni M, Extra JM, et al. Population pharmacokinetics and pharmacodynamics of irinotecan (CPT-11) and active metabolite SN-38 during phase I trials. *Ann Oncol*. 1995;6(2):141-151.