

UNIVERSITY OF CALIFORNIA SAN DIEGO

Integrative Multi-omics Analysis to Understand Cancer and Anticancer Therapy

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

in

Bioinformatics and Systems Biology
with a Specialization in Biomedical Informatics

by

Michelle Ting Dow

Committee in charge:

Professor Hannah Carter, Chair
Professor Ludmil Alexandrov
Professor Olivier Harismendy
Professor Jill Mesirov
Professor Maurizio Zanetti

2020

Copyright

Michelle Ting Dow, 2020

All rights reserved.

The Dissertation of Michelle Ting Dow is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2020

DEDICATION

This dissertation is dedicated to my parents, Chyi-ren and Yenshan Dow, for devoting their entire lives to give their children the best education and a life filled with promise and opportunity.

I am incredibly blessed to be their daughter.

TABLE OF CONTENTS

SIGNATURE PAGE	iii
DEDICATION	iv
TABLE OF CONTENTS.....	v
LIST OF SUPPLEMENTAL FILES	viii
LIST OF FIGURES	ix
LIST OF TABLES.....	xi
ACKNOWLEDGEMENTS.....	xii
VITA.....	xv
ABSTRACT OF THE DISSERTATION.....	xvii
INTRODUCTION	1
References.....	5
CHAPTER 1: Integrative genomic analysis of mouse and human hepatocellular carcinoma	6
1.1 Abstract.....	6
1.2 Introduction.....	7
1.3 Results.....	11
1.4 Discussion.....	28
1.5 Materials and Methods.....	34
1.6 Figures.....	43
1.7 Tables.....	48
1.8 Supplemental Data, Tables and Figures.....	49
1.9 Author Contributions	59
1.10 Acknowledgements.....	60
1.11 References.....	61

CHAPTER 2: p300/CBP-mediated activation of MHC-I machinery and IFN γ signaling controls anti-tumor immunity	66
2.1 Abstract.....	66
2.2 Introduction.....	67
2.3 Results.....	70
2.4 Discussion.....	81
2.5 Materials and Methods.....	85
2.6 Figures.....	93
2.7 Supplemental Data, Tables and Figures.....	101
2.8 Author Contributions	106
2.9 Acknowledgements.....	107
2.10 References.....	108
CHAPTER 3: Using <i>in vitro</i> evolution to probe the genome-wide basis of drug resistance to chemotherapy	113
3.1 Abstract.....	113
3.2 Introduction.....	114
3.3 Results.....	117
3.4 Discussion.....	132
3.5 Materials and Methods.....	136
3.6 Figures.....	144
3.7 Tables.....	151
3.8 Supplemental Tables and Figures	153
3.9 Author Contributions	168
3.10 Acknowledgements.....	169
3.11 References.....	170

EPILOGUE.....	176
Limitations and Future Directions	176
Conclusions.....	178
References.....	179

LIST OF SUPPLEMENTAL FILES

Table S1.2. Mutation presentation in mouse models and human.

Table S1.3. Covariate enrichment analysis for RNAseq.

Table S3.3. Sequencing sample characteristics and statistics.

Table S3.5. Summary of average number of mutations for each individual clone.

Table S3.7. All SNVs called from the 28 samples.

Table S3.8. All CNVs called from the 28 samples.

LIST OF FIGURES

Figure 1.1. Mutational profiles for the human and mouse cohorts and corresponding mutagenic signatures.	43
Figure 1.2. The driver landscape of four HCC mouse models.	44
Figure 1.3. Comparison of somatic alteration across HCC signaling pathways.....	45
Figure 1.4. Pairwise correlations between mouse tumors and human HCCs based on gene expression profiles.	46
Figure 1.5. Immune activities for human and mouse.....	47
Figure S1.1. Mutational signatures of mouse models and human.	54
Figure S1.2. Top 10 significantly mutated genes in human from ICGC and TCGA cohorts , only patients with mutations in the 10 genes are included.	55
Figure S1.3. Unfiltered mutation profiles of the four HCC mouse models.....	56
Figure S1.5. Details on expression and immune analysis.....	58
Figure 2.1. Expression of MHCI related genes in prostate, liver, and lung cancer.	93
Figure 2.2. Platnoid-induced expression of MHCI component is potentiated by IFNg.	95
Figure 2.3. STAT1 and IFNgR2 mediate the synergistic response to oxali+IFNg.	96
Figure 2.4. Low dose Oxali alters chromatin accessibility of MHCI related genes.	97
Figure 2.5. Oxali stimulated HAT activity and nuclear localization.	98
Figure 2.6. NF-kB mediates IFNgR2 induction and the synergistic response to platinoids + IFNg.	99
Figure 2.7. IFNg2 induction is needed for the Oxali-potentiated response to anti-PD-L1 therapy.	100
Figure S2.1: Differential expression of MHCI molecules and components of their antigen processing and presentation machinery correlates with anti-PD-LA.	101
Figure S2.2: Platinoid induced expression of MHCI components and MHCI peptides in mouse cancer cell lines.....	103
Figure S2.3: Platinoid-induced expression of MHCI antigen processing and presentation components in human cancer cell lines.	105
Figure 3.1. Workflow for experimental design.....	144

Figure 3.2. Sex- and age-specific MHC-based mutation selection.....	145
Figure 3.3. DOX-specific analysis and results.....	146
Figure 3.4. GEM-specific analysis and results.	147
Figure 3.5. PTX-specific analysis and results.....	148
Figure 3.6. TPT-specific analysis and results.	149
Figure S3.1. Hierarchical tree of the parent clones and the different drug-specific replicates...	160
Figure S3.2. Near-normal distribution of the mutations in respect to each chromosome.	161
Figure S3.3. Complex alteration for regions surrounding WWOX gene across all samples with WWOX-CNV events.	162
Figure S3.4. Cross-drug resistance EC ₅₀ curves.	163
Figure S3.5. ETP target genes and validation results.	164
Figure S3.6. Additional mutations likely contribute to drug resistance levels.	165
Figure S3.7. IGV views for the TPT target genes.....	166
Figure S3.8. Multidrug resistance EC ₅₀ curves.....	167

LIST OF TABLES

Table 1.1. Description of mouse models.	48
Table S1.1A. Mouse sequencing characteristics.....	50
Table S1.1A. Mouse sequencing characteristics (continued).	51
Table S1.1B. Unsupervised mutational signature computations from mouse samples.	52
Table S1.4. Summary of findings.	53
Table 3.1. Summary of average number of mutations.....	151
Table 3.2. Validation (knockdown) results for selected genes.	152
Table S3.1. Summary of all drugs tested for IVIEWGA.....	154
Table S3.2. EC ₅₀ values for all drug-specific resistant lines and their isogenic parental cell line (WT).....	155
Table S3.4. Summary of effect classification (as referred in the manuscript) of the SNVs and Indels from snpEff annotations.....	157
Table S3.6. Summary of the number of mutations along the steps of the filtering pipeline for the five candidate drugs.	158
Table S3.9. Potential target genes fitting the filtering criteria.....	159

ACKNOWLEDGEMENTS

First, I would like to thank my advisor Hannah Carter for her extensive mentorship and support over the past five years. She focused on not only training me to conduct rigorous science, but also in skills for my overall career development. Throughout my Ph.D., I had the opportunities to practice presenting my work from on-campus meetings to international conferences. Early in my career, she taught me the value of persistence, and that making mistake is part of the process – as long as we work hard to fix it. It was her guidance and supportiveness that brought me to be the scientist I am today.

Next, I would also like to thank my dissertation committee for their advices and valuable inputs in my research. Olivier Harismendy has been the mentor I seek suggestions from since my pre-Ph.D. summer internship. His encouragement has pushed me to ask more questions, to think more critically, but at the same time be confident in my ideas. Jill Mesirov is a role model who inspires me to work harder and explore further for my projects. Ludmil Alexandrov provided countless valuable feedbacks, patiently guide me through my exploration of mutational signatures, and advised me from his experiences working with large databases. Maurizio Zannetti shared his knowledge to help me understand the complex tumor-immune interactions.

I would like to thank all past and current members of Carter Lab for their unyielding support and friendship. In particular, I would like to thank Rachel Marty, Brian Tsui, Kivil Ozturk, Andrea Castro, Jessica Zhou, Su Xian, Cameron Waller, Meghana Pagadala, Clarence Mah, James Talwar, Shengyu Li, James Han, and Lucia Bazan-Williamson. They endlessly encouraged me through the process, brainstormed strategies to help solve both my scientific and interpersonal problems, and shared all the snacks I stress shopped.

I would like to thank my colleagues and collaborators for all the advises and support. I would like to thank all past and present members of Trey Ideker's lab. In particular, thank you to Yue Qin and Adriana Pitea for reminded me to stay on top of my diets and exercises. Furthermore, I would like to thank Barry Demchak and Willy Markuske for answering all my emails and helping me trouble shoot problems on the cluster. I would like to thank my collaborators Betty Shih, Chi-Ping Day, Shabnam Shalapour, Elizabeth Winzeler, Juan Carlos Jado for the experimental insights and new perspectives on interdisciplinary communications.

I would like to thank my family. My parents, Chyi-ren and Yenshan Dow, for their never-ending love and support. Their wisdom has helped me through tough times and reminded me to relax. I would also like to thank my brother Harrison Dow, and sister Madeline Dow, for encouraging the best out of me. I would like to thank my fiancé, Jenhan Tao, for always being the voice of reason. He celebrated my accomplishments and endured my hysterical moments. Thank you for all the bad science jokes and the strange optimism. Furthermore, I would like to thank my roommate Miin Lin for encouraging me and making sure I am eating nutritiously.

Chapter 1, in full, is a reformatted reprint of the material as it appears as "Integrative genomic analysis of mouse and human hepatocellular carcinoma" in *PNAS*, 2018 by Michelle Dow, Rachel M. Pyke, Brian Y. Tsui, Ludmil B. Alexandrov, Hayato Nakagawa, Koji Taniguchi, Ekihiro Seki, Olivier Harismendy, Shabnam Shalapour, Michael Karin, Joan Font-Burgada, and Hannah Carter. The dissertation author was the primary investigator and author of this paper.

Chapter 2, in full, is a reformatted presentation of the material currently being prepared for submission for publication as "p300/CBP-mediated activation of MHC-I machinery and IFN γ signaling controls" by Michelle Dow, Yixuan Zhou, Ingmar Niels Bastian, Mark D Long, Weihua Li, Tao Liu, Rachael Katie Ngu, Laura Antonucci, Jian Yu Huang, Qui T Phung, Xi-he Zhao,

Daniel Karin, Brian Dang, Sourav Banerjee, Xue-Jia Lin, Hua Su, Sylvia Choi, Hongxia Want, Christina Jamieson, Marcus Bosenberg, Zhang Cheng, Amanda Birmingham, Johannes Haybaeck, Lukas Kenner, Kathleen M. Fisch, Richard Bourgon, Genevive Hernandex, Jennie R. Lill, Ira Mellman, Hannah Carter, Song Liu, Michael Karin, and Shabnam Shalapour. The dissertation author was one of the primary investigators and author of this material.

Chapter 3, in full, is a reformatted presentation of the material currently being prepared for submission for publication as “Discovering biomarkers of chemotherapy resistance using *in vitro* evolution in haploid human cells” by Michelle Dow, Juan Carlos Jado, Krypton Carolino, Gregory J. Fonseca, Trey Ideker, Hannah Carter, and Elizabeth A. Winzeler. The dissertation author was the primary investigator and author of this material.

VITA

- 2015 University of British Columbia
Bachelor of Science, Computer Science: Bioinformatics
- 2016 Boston University
Bachelor of Science, Computer Science: Bioinformatics
- 2020 University of California San Diego
*Doctor of Philosophy, Bioinformatics and Systems Biology
with a Specialization in Biomedical Informatics*

PUBLICATIONS

Y. Zhou*, I. N. Bastian*, M. D Long*, **M. Dow***, W. Li*, T. Liu, R. K. Ngu, L. Antonucci, J. Y. Huang, Q. T. Phung, X. Zhao, D. Karin, B. Dang, S. Banerjee, X. Lin, H. Su, S. Choi, H. Wang, C. Jamieson, M. Bosenberg, Z. Cheng, A. Birmingham, J. Haybaeck, L. Kenner, K. M. Fisch, R. Bourgon, G. Hernandez, J. R. Lill, H. Carter, S. Liu, I. Mellman, M. Karin[#], S. Shalapour[#]. “p300/CBP-mediated activation of MHC-I machinery and IFN γ signaling controls.” *Nature Immunology*, *In Submission 2020*.

J. C. Jado*, **M. Dow***, G. J. Fonseca, K. Carolino, T. Ideker[#], H. Carter[#], E. A. Winzeler[#]. “Discovering biomarkers of chemotherapy resistance using *in vitro* evolution in haploid human cells.” *In Submission 2020*.

S. Stoner, E. Andrews, M. Liu, K. Arimoto, M. Yan, A. Davis, S. Weng, **M. Dow**, S. Xian, R. Dekelver, H. Carter, and D.-E. Zhang. “The RUNX1-ETO target gene RASSF2 suppresses t(8;21) AML development and regulates Rac GTPase signaling.” *Blood Cancer Journal*, *In Press 2020*.

N. Nguyen, **M. Dow**, B. Woodside, J. B. German, O. Quehenberger, and P.-A. B. Shih. “Food-Intake Normalization of Dysregulated Fatty Acids in Women with Anorexia Nervosa.” *Nutrients*, 11(9), sep 2019.

M. Dow, R. M. Pyke, B. Y. Tsui, L. B. Alexandrov, H. Nakagawa, K. Taniguchi, E. Seki, O. Harismendy, S. Shalapour, M. Karin[#], H. Carter[#], and J. Font-Burgada[#]. “Integrative genomic analysis of mouse and human hepatocellular carcinoma.” *Proceedings of the National Academy of Sciences of the United States of America*, 115(42):E9879–E9888, oct 2018.

K. Ozturk, **M. Dow**, D. E. Carlin, B. Rafael, and H. Carter. “The Emerging Potential for Network Analysis to Inform Precision Cancer Medicine”. *Journal of Molecular Biology*, 430(18):2875-2899, sep 2018.

F. Zare, **M. Dow**, N. Monteleone, A. Hosny, and S. Nabavi. “An evaluation of copy number variation detection tools for cancer using whole exome sequencing data.” *BMC Bioinformatics*, 18(1):286, dec 2017.

S. Wang, X. Jiang, S. Singh, R. Marmor, L. Bonomi, D. Fox, **M. Dow**, and L. Ohno-Machado. “Genome privacy: challenges, technical approaches to mitigate risk, and ethical considerations in the United States.” *Annals of the New York Academy of Sciences*, 1387(1):73–83, jan 2017.

E. Levy, R. Marty, V. Ga´rate Caldero´n, B. Woo, **M. Dow**, R. Armisen, H. Carter, and O. Harismendy. “Immune DNA signature of T-cell infiltration in breast tumor exomes.” *Scientific Reports*, 6, Jul. 2016.

F. Chen, **M. Dow**, S. Ding, Y. Lu, X. Jiang, H. Tang, and S. Wang. “PREMIX: PRivacy-preserving EstiMation of Individual admixture. AMIA ... Annual Symposium proceedings.” *AMIA Symposium*, 2016:1747–1755, 2016.

*These authors contributed equally to this work. #Co-corresponding authors.

ABSTRACT OF THE DISSERTATION

Integrative Multi-omics Analysis to Understand Cancer and Anticancer Therapy

by

Michelle Ting Dow

Doctor of Philosophy in Bioinformatics and Systems Biology
with a Specialization in Biomedical Informatics

University of California San Diego, 2020

Professor Hannah Carter, Chair

Precision cancer medicine promises better treatments to a disease as complex and heterogenous as cancer. Many anti-cancer therapies are beneficial to only a subset of patients due to the variability in patient genetic and tumor heterogeneity. Thus, we need better frameworks for understanding underlying genomic and transcriptomic patterns influencing differential patient

outcomes, yet our understanding of how genetic alterations connect to treatment in *in vivo* and *in vitro* models remains understudied. To address this gap, I utilized human patient data from The Cancer Genome Atlas (TCGA), hepatocellular carcinoma (HCC) models, prostate cancer (PCa) models, and chronic myelogenous leukemia (CML) cell lines. Through the integration of multi-omic data, I identified parallel features of human and model organism data that could reveal disease specific characteristics. Additionally, I characterized the landscape of acquired resistance for a panel of chemotherapeutic treatments and revealed potential alleles and genes that mediate the process. The analyses I conducted expose the role of genetic information and suggest future applications for development of precision medicine.

INTRODUCTION

Cancer continues to be one of the major causes of mortality worldwide and there is a need to develop effective therapeutics. Cancer is a complex and heterogeneous genetic disease; the use of next generation sequencing (NGS) has revolutionized our understanding of tumor biology as well as the mechanisms associated with tumor response to treatments. However, the translation of cancer genomic data to therapeutic treatment remains challenging due to patient genetic heterogeneity. Despite the complexity and variability, the majority of cancers are treated with the same generic regimens. “Precision medicine” uses patient information to guide selection of tailored treatments. Molecular understanding of the tumor-immune interaction can be used to fight cancer, and modulate standard treatments based on an individual’s oncogenic driver mutations. For example, a high response rate for Vemurafenib has been observed in metastatic melanoma patients with BRAF V600 mutations (Flaherty et al., 2010). A major research focus has been integrating genomic information to better improve response rate for immunotherapy (Postow et al., 2018). It has become increasingly clear that no two patients’ cancers are exactly the same, and this highlights the need to understand personal molecular features before selecting a cancer treatment (Krzyszczuk et al., 2018). There is an urgent need for better knowledge of tumor biology that can then be translated into clinical applications.

In vivo and *in vitro* models have played a critical role in increasing our understanding of tumor progression and responses to treatments (Cheon and Orsulic, 2011). Mouse models have revolutionized our ability to study gene function, aided in identifying cancer biomarkers, and providing better clinical models for testing novel therapeutic strategies. They have advantages over other mammalian models in that they are inexpensive to maintain, reproduce rapidly, and can be genetically manipulated (Cekanova and Rathore, 2014). Therefore, humanized mouse models have

been developed to mimic the pathophysiology of human cancers. These models are commonly used as a platform to evaluate therapies, however, the extent to which specific models represent human disease has been based largely on histology and the presence of particular genotypes. The comprehensive genetic and molecular similarities between mice and humans have not been thoroughly evaluated. To gain insights into how mouse models can be used to develop effective immunotherapies, I will analyze the characteristics of human and mouse tumors. I hypothesize that multi-omic data can be used to not only provide a better assessment of the similarity of model systems to human diseases, but also to discover new ways to improve patient outcomes with effective precision medicine.

Due to the complexity of risk factors and tumor development, diseases such as hepatocellular carcinoma (HCC) have a lower rate of finding precise treatments. Specifically, HCC research has been hampered by the absence of consensus mouse models with clearly defined molecular features faithfully simulating human HCC. In chapter one, I will implement a comparative analysis between a large cohort of patients and four diverse mouse models that focuses on clinically and therapeutically relevant aspects of their genomic and transcriptomic profiles. I provide a framework for the analysis of the mutation impact to recognize appropriate in vivo models and identify the optimal model for cross-species inference.

Due to its heterogeneity, cancer is rarely treated with a single therapy. Instead, a variety of therapeutic approaches have shown that combining two or more therapeutic agents enhances efficacy compared to the mono-therapy approach (Bayat Mokhtari et al., 2017; Hodi et al., 2010). With precision medicine, information about the genetic changes can help identify multiple key pathways which could be targeted in a synergistic or additive manner (Bayat Mokhtari et al., 2017). Eigentler et al. report higher response rates but no improvement of overall survival in groups

treated with a combinations of immunotherapy and chemotherapy agents (Eigentler et al., 2003). On the other hand, recent reports have shown that advanced staged lung, gastrointestinal, and kidney cancer show better response and prolonged survival period with involvement of immune therapy than conventional chemotherapy alone (Melero et al., 2015, Hodi et al., 2010; Ventola, 2017). However, the underlying genetic changes influencing the results remain unclear.

Recent studies have shown that HCC patients treated with immune checkpoint inhibitors (ICI) are responsive in HCC despite low mutational burden, while similarly treated prostate cancer (PCa) patients are not. However, with low doses of platinoid, especially oxaliplatin, response to ICI improves in PCa patients. In this case, combined therapy could be the key to induce ICI response in PCa patients. Without understanding of the heterogeneous mechanisms underlying tumorigenesis, it is difficult to establish more effective molecular therapeutic targets for individual patients. In chapter one, we found that while similar immune characteristics have been found in between mice and humans, the four mouse models have distinct molecular characteristics and correlations to HCC human subgroups. Here, the mouse models are designer tumors that are made to mimic human diseases. We are interested in studying the transcriptomic profile differences between HCC and PCa. The proposed studies will provide novel insight into how mouse models can characterize the genomics of disease and be used to identify cancer drivers that could be useful for selecting or developing therapies. In chapter two, I evaluate the transcriptomic implication of immune selection for individuals responding to ICI and platinoid combination therapy in custom models of liver and prostate. I will identify the potential genes and mechanisms that contribute to the response rate differences between liver and prostate tumors in mono- and combined- treatment conditions.

Finally, we would like to understand the broader impact of acquired resistance given our newfound knowledge of the differences between mono- and combination therapies response and resistance. In the majority of cases, if the patient progresses after those agents and shows resistance to standard chemotherapy agents (Tsao et al., 2004), there is no second-line treatment available. While there are many known drug resistance genes in chemotherapy, the reason for the disparity in resistance to these agents remains unclear (Luqmani, 2005). In cancer, this variation means that some individuals can present a set of mutations to the immune system while other individuals cannot present those same mutations. In chapter three, I will identify variants and genes potentially responsible for chemotherapy resistance.

In this dissertation, I aim to provide a framework to better characterize model systems used to study cancer. Furthermore, I aim to expose the underlying biological mechanisms elements associated with the response and resistance to antitumor therapies by systematically analyzing and studying the relationship between individual genomics and transcriptomic profile. My studies will help generate mechanistic insight into antitumor therapeutic thus facilitating the development of improved therapies, and advancing precision medicine in treatment of cancer.

References

- Bayat Mokhtari, R., Homayouni, T.S., Baluch, N., Morgatskaya, E., Kumar, S., Das, B., and Yeager, H. (2017). Combination therapy in combating cancer. *Oncotarget* 8, 38022–38043.
- Cekanova, M., and Rathore, K. (2014). Animal models and therapeutic molecular targets of cancer: utility and limitations. *Drug Des. Devel. Ther.* 8, 1911–1921.
- Cheon, D.-J., and Orsulic, S. (2011). Mouse models of cancer. *Annu. Rev. Pathol.* 6, 95–119.
- Eigentler, T.K., Caroli, U.M., Radny, P., and Garbe, C. (2003). Palliative therapy of disseminated malignant melanoma: a systematic review of 41 randomised clinical trials. *Lancet Oncol.* 4, 748–759.
- Flaherty, K.T., Puzanov, I., Kim, K.B., Ribas, A., McArthur, G.A., Sosman, J.A., O'Dwyer, P.J., Lee, R.J., Grippo, J.F., Nolop, K., Chapman, P.B. (2010). Inhibition of mutated, activated BRAF in metastatic melanoma. *N. Engl. J. Med.* 363, 809–819.
- Hodi, F.S., O'Day, S.J., McDermott, D.F., Weber, R.W., Sosman, J.A., Haanen, J.B., Gonzalez, R., Robert, C., Schadendorf, D., Hassel, J.C., Akerley, W., van den Eertwegh A.J.M., Lutzky, J., Lorigan, P., Vaubel, J.M., Linette, G.P., Hogg, D., Ottensmeier, C.H., Lebbé, C., Peschel, C., Quirt, I., Clark, J.I., Wolchok, J.D., Weber, J.S., Tian, J., Yellin, M.J., Nichol, G.M., Hoos, A., Urba, W.J. (2010). Improved survival with ipilimumab in patients with metastatic melanoma. *N. Engl. J. Med.* 363, 711–723.
- Krzyszczczyk, P., Acevedo, A., Davidoff, E.J., Timmins, L.M., Marrero-Berrios, I., Patel, M., White, C., Lowe, C., Sherba, J.J., Hartmanshenn, C., O'Neill, K.M., Balter, M.L., Fritz, Z.R., Androulakis, I.P., Schloss, R.S., Yarmush, M.L. (2018). The growing role of precision and personalized medicine for cancer treatment. *Technology* 6, 79–100.
- Luqmani, Y.A. (2005). Mechanisms of drug resistance in cancer chemotherapy. *Med. Princ. Pract.* 14 *Suppl 1*, 35–48.
- Melero, I., Berman, D.M., Aznar, M.A., Korman, A.J., Pérez Gracia, J.L., and Haanen, J. (2015). Evolving synergistic combinations of targeted immunotherapies to combat cancer. *Nat. Rev. Cancer* 15, 457–472.
- Postow, M.A., Sidlow, R., and Hellmann, M.D. (2018). Immune-Related Adverse Events Associated with Immune Checkpoint Blockade. *N. Engl. J. Med.* 378, 158–168.
- Tsao, H., Atkins, M.B., and Sober, A.J. (2004). Management of cutaneous melanoma. *N. Engl. J. Med.* 351, 998–1012.
- Ventola, C.L. (2017). Cancer Immunotherapy, Part 2: Efficacy, Safety, and Other Clinical Considerations. *P T* 42, 452–463

CHAPTER 1: Integrative genomic analysis of mouse and human hepatocellular carcinoma

1.1 Abstract

Cancer genomics has enabled the exhaustive molecular characterization of tumors and exposed hepatocellular carcinoma (HCC) as among the most complex cancers. This complexity is paralleled by dozens of mouse models that generate histologically similar tumors but have not been systematically validated at the molecular level. Accurate models of the molecular pathogenesis of HCC are essential for biomedical progress; therefore, we compared genomic and transcriptomic profiles of four separate mouse models [MUP transgenic, TAK1-knockout, carcinogen-driven diethylnitrosamine (DEN), and Stelic Animal Model (STAM)] with those of 987 HCC patients with distinct etiologies. These four models differed substantially in their mutational load, mutational signatures, affected genes and pathways, and transcriptomes. STAM tumors were most molecularly similar to human HCC, with frequent mutations in *Ctnnb1*, similar pathway alterations, and high transcriptomic similarity to high-grade, proliferative human tumors with poor prognosis. In contrast, TAK1 tumors better reflected the mutational signature of human HCC and were transcriptionally similar to low-grade human tumors. DEN tumors were least similar to human disease and almost universally carried the *Braf* V637E mutation, which is rarely found in human HCC. Immune analysis revealed that strain-specific MHC-I genotype can influence the molecular makeup of murine tumors. Thus, different mouse models of HCC recapitulate distinct aspects of HCC biology, and their use should be adapted to specific questions based on the molecular features provided here.

1.2 Introduction

Primary liver cancer is the 5th most common cancer worldwide but is the second leading cause of cancer deaths, being associated with very poor prognosis (Ferlay et al. 2015a). Hepatocellular carcinoma (HCC) accounts for 85-90% of all primary liver cancers (Ferlay et al. 2015b; El-Serag and Rudolph 2007). In most cases, HCC presents as a double disease since it is usually accompanied by cirrhosis caused by a variety of risk factors. Globally, HCC epidemiology is driven by chronic HBV and HCV infections, which account for 80% of cases (Bosch et al. 2005; Holmberg et al. 2013). Other major risk factors include alcohol, aflatoxin, parasites, obesity and genetic disorders. Incidence trends have been changing across the globe depending on regional variation in prevalence of underlying etiologies. For instance, incidence rates in the United States have tripled from 1977-79 to 2005-07 (El-Serag and Davila 2011). Furthermore, 30-35% of HCC patients in the US do not have viral infection. Non-alcoholic fatty liver disease (NAFLD) and its advanced presentation: non-alcoholic steatohepatitis (NASH), is emerging as the driving force behind this growth. Current estimates project 1.5-2% of the US population as having cirrhosis due to NASH, and this is expected to become the leading indication for liver transplantation in the not-so-distant future (Ascha et al. 2010a). Considering the annual cumulative incidence of HCC in NASH-cirrhosis (Ascha et al. 2010b) an upsurge of new HCC cases is expected in the US in the coming decades. Analogous trends are expected in regions that have similar socioeconomic characteristics with coinciding increases in obesity rates.

Thirty to forty percent of HCC patients are indicated for potential curative treatment (surgical resection and transplantation) while the remaining 60-70% are only eligible for palliative and symptomatic treatment (Llovet et al. 2016). The sole targeted therapy for HCC is the pan-kinase inhibitor Sorafenib, affecting Raf kinases and VEGF and PDGF receptors, indicated for

patients with advanced diseases (Wilhelm et al. 2008). However, the response to sorafenib is dismal and the median overall survival is only extended by 2.8 months (Llovet et al. 2008). Similar small increase in overall survival has been observed with another FDA approved pan-kinase inhibitor, Regorafenib (Bruix et al. 2017). However, Nivolumab, a programmed cell death protein-1 (PD-1) immune checkpoint inhibitor, has shown promising effects for patients who failed to respond to Sorafenib (El-Khoueiry et al. 2017).

Due to the complex links between different risk factors and HCC development, the molecular drivers and combinations thereof involved in hepatocarcinogenesis are still poorly understood. Previous studies have characterized the transcriptomic landscape of HCC tumors, identifying expression signatures associated with HBV, TP53 signaling, and WNT or AKT pathway activation (Hoshida et al. 2009). Genomic analyses have identified major pathways altered in HCC, such as the WNT, PIK3/Ras, and cell cycle pathways (Ahn et al. 2014; Guichard et al. 2012), and more recent studies have characterized mutational signatures associated with diverse risk factors such as alcohol consumption, smoking, and exposure to aflatoxin B1, as well as putative driver genes including TERT promoter mutations, TP53, CTNNB1, and AXIN1 (Schulze et al. 2015; Totoki et al. 2014; Letouzé et al. 2017; Fujimoto et al. 2016). Among the largest HCC cohort studies are The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC) which together have undertaken four independent projects to help catalogue the molecular alterations found in HCC patients (Zhang et al. 2011; Cancer Genome Atlas Research Network. 2017). HCC genetic aberrations show substantial heterogeneity, likely reflecting effects of etiology, ethnicity and environmental exposures. Remarkably, the genetics of NASH induced HCC remains poorly characterized.

The development of well-defined mouse models of HCC that accurately reproduce human disease has been a priority and is essential for studying basic tumor biology and therapeutic response. However, dozens of HCC mouse models have been developed by different interventions. Several carcinogens have been historically used such as diethylnitrosamine (DEN), carbon tetrachloride (CCl₄), 2-Acetylaminofluorene (2-AAF), thioacetamide (TAA), and aflatoxins among others with large variations in protocols for dosage and dosage schedule, use of tumor promoters and different genetic backgrounds all producing phenotypic variations. The alternative, genetically modified mice encompass two main categories, one in which a genetic alteration produces chronic tissue damage that spontaneously generates HCC and a second group in which different oncogenic variants or tumor suppressors are introduced specifically into all or a subset of hepatocytes. Other models include tumor cell implantation or complex systems to recapitulate chronic infections by HBV and HCV (Santos, Colaço, and Oliveira 2017). An early attempt to characterize the similarity of 68 tumors from 7 HCC mouse models and 91 human HCC samples based on gene expression microarrays found that some mouse models shared gene expression patterns similar to HCC patients with good prognosis and other with poor prognosis (Lee et al. 2004), but the analysis was limited to expression data and evaluated mouse models rarely used in HCC studies. Unfortunately, the genetic makeup of full-blown HCC in mouse models is generally unknown and therefore it is unclear which animal models are molecularly more similar to human disease. Thus, the large multiplicity of models and the inconsistency in experimental parameters along with their unknown genetic component have obscured a better understanding of HCC biology. In addition, our current knowledge of human NASH-induced HCC is limited and although some mouse models have been proposed, most do not replicate the complete spectrum of clinical features and have been poorly characterized molecularly (Nakagawa 2015).

Here, we provide the first comprehensive comparative genomics characterization of mouse HCC in four independent mouse models with four human HCC cohorts. We catalogue mutational signatures, mutated genes and pathways together with the associated transcriptional perturbations to determine mouse-human similarities. This work uncovers the benefits and limitations of the use of mouse models of HCC and provides a reference to better tailor future mouse research on HCC and preclinical studies to reflect the molecular characteristics of the human disease.

1.3 Results

Somatic mutational signatures. Genomic and transcriptomic profiles were obtained for tumors from four different mouse models of HCC: DEN (Heindryckx, Colle, and Van Vlierberghe 2009a; He et al. 2013a), TAK1^{HEP} (TAK1) (Inokuchi et al. 2010; He et al. 2013b), MUP-uPA + HFD (MUP) (Nakagawa et al. 2014), and STAM (Fujii et al. 2013) which rely on different approaches to induce liver tumors within 9 months, except STAM which develops tumors within 5 months (**Table 1.1**). The DEN model has been widely used for its ease and consistency in generating tumors. However, this model does not present with common risk factors for human HCC, such as chronic liver injury or fibrosis. Conversely, the TAK1 model induces HCC in the context of a highly similar microenvironment to HCC patients with chronic liver disease. The MUP and STAM models are designed to mimic NASH-induced HCC. The MUP model reproduces all of the clinical features of human NASH, but most of the tumors appear “adenoma-like”, while the STAM model rapidly develops NASH-like cirrhosis in the absence of obesity with rapid development of HCCs histologically similar to human disease.

We performed whole exome sequencing (WES) on a total of 56 mouse tumors and 6 normal spleens reaching an average coverage depth of 100x (minimum 70x, *SI Appendix*, Table S1A). Initially, we sequenced 9 samples from each model with 10 additional samples sequenced at a later date for STAM and DEN models, based on the greater potential of these models relative to the others to generate insights into HCC mutagenesis. To systematically compare the properties of somatic alterations observed in mouse models with those found in human patients, human HCC exome-based (total N=987; from LICA-CN LICA-FR, LIHC-US, and LINC-JP, N=163, 234, 377, and 213, respectively) and genome-based somatic mutation data (total N=314; from LICA-FR, LIHC-US, LINC-JP, and LIRI-JP, N=5, 53, 28, and 228, respectively) was retrieved from the

International Cancer Genome Consortium (ICGC) data portal (Zhang et al. 2011) and The Cancer Genome Atlas (TCGA) (Cancer Genome Atlas Research Network 2017).

We identified somatic mutations in the four mouse models by comparing tumor DNA against splenic DNA for each model. Median somatic mutation rates in MUP and TAK1 mice were similar to those observed in human HCCs (2-3 mutations/Mb) while DEN had a much higher median burden of somatic mutations (~27 mutations/Mb) as expected of a well characterized carcinogen (**Fig. 1.1A**; left). Strikingly, the STAM model had on average a 5-fold higher mutation rate compared that observed in DEN (~122.56 mutations/Mb). We also observed large differences in the somatic mutation profiles amongst the different mouse models (**Fig. 1.1B**; left). The DEN model had a very low rate of C>G transversions (mutations referred to by the pyrimidine nucleotide of the Watson-Crick base-pair), and fewer C>T transitions but an excess of T>A transversions. The TAK1 model had a notable excess of T>G transversions and the STAM model mutational profile was dominated by C>T transitions relative to all other categories. The differences are likely due to the diverse mechanisms by which HCCs are induced in these models. For example, the DEN model is treated with diethylnitrosamine, a carcinogen found in tobacco smoke and the STAM model is treated with Streptozotocin to impair pancreatic beta cell function (Fujii et al. 2013) and fed with a high fat diet to recapitulate NASH-associated cirrhosis. In contrast, TAK1^{HEP} and MUP-uPA models are transgenic mouse models relying on the induction of chronic hepatocyte injury with spontaneous hepatocarcinogenesis. Thus, mutations in these models are expected to accumulate more slowly through chronic injury and inflammatory processes as compared with acute exposure to DNA damaging agents.

Amongst human cohorts, we observed a fairly consistent median somatic mutation rate in the range of 2-3 mutations/Mb except for the LINC-JP cohort that had a higher median rate of 6.26

mutations/Mb (**Fig. 1.1A**; right). This cohort includes samples from HCC patients treated with cisplatin, a chemotherapy drug that crosslinks DNA to inhibit replication and cause apoptosis, and is associated with a higher burden of mutations (Zhang et al. 2011; Szikriszt et al. 2016). Further, signature 12 was highly elevated in LINC-JP compare to other cohorts (3-fold elevation; Wilcoxon signed-rank test p-value: 1.65e-14). The etiology of signature 12 remains unknown and this elevation might reflect environmental mutagen that is more prevalent in Japan. The mutation profiles of the four human exome sequencing cohorts were also similar, however, we observed an excess of T>A transversions in the LICA-CN cohort, which was less prevalent in the LIHC-US cohort. Furthermore, a higher rate of C>T transitions was observed in the LIHC-FR cohort and a slightly elevated rate of T>G transversions was found in the LIHC-JP and LIHC-US cohorts (**Fig. 1.1B**; right). Overall, the observed variation in mutation rates and spectra suggest that different factors drive mutagenesis between mouse models. While human tumors look less heterogeneous, there were still differences in the abundance of certain base substitutions.

The somatic mutations in a cancer genome are the results of the activity of multiple mutational processes. Previously, more than 30 characteristic patterns, termed mutational signatures, of single nucleotide substitutions have been found across human cancer and many of them have been attributed to the activity of endogenous and/or exogenous carcinogens. To determine the signatures of the mutational processes operative in the cancer genomes of the four mouse models, we decomposed the single nucleotide substitutions in their immediate context (i.e., 96 possible classes of mutations) into mutational signatures. We applied the method described by Alexandrov *et al.* across all 56 mouse samples, and identified 5 distinct signatures (A, B, C, D, and E) (**Table S1.1B and Fig. S1.1A**), of which four corresponded to signatures previously reported in humans (Forbes et al. 2017) (cosine similarity greater than 0.92). Signature A was

different from any previously observed mutational signatures in human cancer (maximum cosine similarity of 0.76). Signature A was specific to the DEN samples; thus, this signature is likely associated with exposure to diethylnitrosamine. Signature B corresponded to human signature 44 (cosine similarity =1.0), which is attributable to a known sequencing artifact. Two signatures (Signature C and D), were dominated by an alkylating agent specific signal, where signature C was most similar to human signature 11 (cosine similarity =1.0) and was predominantly found in STAM mice but also in a few other samples. The remarkably high mutation rate of STAM tumors and the strong signature of an alkylating agent suggests that Streptozotocin is a potent carcinogen of the liver. Signature D was mostly observed in MUP and some TAK1 samples, and shared similarity with human Signature 12, previously associated with HCC (cosine similarity = 0.93). Signature E, present in the four mouse models, was similar to two as yet unexplained human signatures (signatures 19 and 32). Given that the mutation profiles of the DEN and STAM models are dominated by mutations attributable to mutagens, the more subtle contributions of other mutagenic processes may be masked in these samples (**Fig. S1.1B**).

We next sought to determine how similar mutation profiles observed in mouse models are to known mutational signatures previously reported in human HCC. We used a supervised approach that attempts to find a weighted combination of existing signatures that best explains a mutation profile (Methods). After mapping mutational profiles for mouse models and human cohorts to known human mutational signatures (Alexandrov et al. 2015), we clustered the groups based on overall similarity of the signature weights (Fig. 1C and *SI Appendix*, Fig. S1C). Mutational signatures have previously been analyzed in several human HCC cohorts (Schulze et al. 2015; Letouzé et al. 2017; Ng et al. 2017), implicating several signatures in human HCC, including 4, 6, 16, 22, 23 and 24. Our analysis revealed similar patterns of mutation signature

across human cohorts with clear cohort-specific risk factors (Fig. 1C). While all human cohorts were dominated by signature 5, known to correlate with age in liver (Alexandrov et al. 2015; Blokzijl et al. 2016), the LICA-FR WGS cohort was most strongly associated with a smoking signature, with the LIHC-US WGS showing weaker evidence of similar exposures. In contrast, the LICA-CN cohort showed stronger influence of a signature attributed to reactive oxygen species compared to other human cohorts.

Interestingly, signatures 1, 5 and 19 were seen almost universally across both human and mouse models. While signature 5 accounted for a high proportion of mutations in most samples, only small proportion of mutations was attributed to signatures 1 and 19. Nevertheless, all three signatures were found in most human HCC WXS, WGS, MUP, and TAK1 samples. Signatures 1 and 5 are associated with clock-like mutational processes accumulating with age, while the etiology of signature 19 remains unknown. Most signatures were either observed only in human HCCs (signatures 4, 8, 18, 22, 26, 30, 35, and 40) or in mouse HCCs (signatures 11, 32, and mouse signature A). As expected, STAM was dominated by signature 11 associated with exposure to alkylating agents and DEN mutations were predominantly attributable to the DEN-specific mutagenic signature. Although DEN is a carcinogenic compound found in tobacco smoke, its mutational signature has not been observed in human cancers associated with tobacco smoking. Rather, the patterns of somatic mutations in directly-exposed tobacco associated cancers is predominately signature 4, which has been experimentally matched with exposure to benzo[a]pyrene (Alexandrov et al. 2016). Interestingly, MUP and TAK1 were associated with signature 12 which was also observed in all of the human whole genome sequencing (WGS) cohorts.

Mutations in mouse models thus strongly reflect the etiology of HCC in the model, with some etiologies creating mutation profiles that differ widely from those in human disease. Human HCCs have the most diversity in mutational signatures among human cancers (Letouzé et al. 2017), potentially reflecting the liver's central role in metabolism and detoxification and its unique physiology being the first organ to receive blood influx from the intestinal tract. Thus, the liver is potentially the target of a wide variety of tissue damaging agents as well as mutagens which may account for the molecular heterogeneity of human HCC.

Somatically Altered Genes. Despite considerable differences in mutation signatures between human HCC and mouse models, the resulting mutations could nonetheless impact common genes in both species, causing disease through similar mechanisms. To evaluate this hypothesis, we analyzed the occurrence of mutations in 17,046 genes that are 1:1 orthologs in human and mouse, which included 92% of solid-tumor relevant genes (Methods) from the cancer gene census (CGC) (Forbes et al. 2017). Consistent with previous reports (Cancer Genome Atlas Research Network 2017; Vogelstein et al. 2013), we observed considerable heterogeneity in somatic alterations across human HCC tumors. *TP53* and *CTNNB1*, the most frequently mutated genes, were only mutated in ~20-25% of tumors without significant co-occurrence (**Fig. S1.2A**), while other known cancer genes were mutated in fewer than 10% of tumors. Among the top 10 significantly mutated genes (MutSigCV, $q < 0.1$), only *CTNNB1* showed significantly different patterns of mutation between human HCC cohorts (**Fig. S1.2B**), with > 20% of tumors harboring *CTNNB1* mutations in three of four cohorts but only 2% of LICA-CN tumors having *CTNNB1* mutations. Across human cohorts, MutSigCV (Lawrence et al. 2013) identified 13 genes with higher than expected mutation rates (*TP53*, *CTNNB1*, *AXIN1*, *ALB*, *BAP1*, *EEF1A1*, *NFE2L2*, *RPS6KA3*, *CDKN2A*, *RBI*, *ACVR2A*, *KEAP1*, and *ARID1A*).

We next mapped all somatic SNVs and indels obtained for mouse models to annotated genes. Overall, MUP and TAK1 mouse models had very few non-synonymous mutations (N= 306, 876) whereas STAM and DEN showed much higher numbers (N=83,910, 17,859), consistent with the observed somatic mutation rates in these models. We therefore sought to identify genes enriched for positive signatures of selection for each mouse model. Since MutSigCV requires large sample sizes (Lawrence et al. 2013), we used an alternative method to identify significantly mutated genes (SMGs; Methods) that seeks to detect a bias toward mutations expected to perturb protein activity by using the ratio of nonsynonymous to synonymous mutations (Greenman et al. 2006). In total, we evaluated 12,995 genes and identified 377 SMGs across mouse models, of which 1 was found in MUP (*Hist1h4j*, ortholog of human *HIST1H4K*), 25 were found in DEN and 351 were found in STAM (**Table S1.2A**). In total, 11 of the 377 were previously implicated cancer genes in the CGC. These included *Abll*, *Bcl9l*, *Cbl*, *Foxa1*, *Mycn*, *Nutm1*, *Olig2*, *Tert*, *Tshr*, *Braf*, and *Sh2b3*.

Since STAM and DEN models far exceeded the average number of mutated genes in human HCC, we concluded that most detected SMGs could be the result of the high mutation rate caused by mutagenic agents used in these models. As a first step to simplify the analysis, we focused the mouse-human comparison on the subset of tier 1 and 2 cancer driving genes in the Cancer Gene Consensus (CGC). Supplementary Fig. S3 summarizes the total alterations observed in each mouse model, contrasting mutation frequencies across mouse and human HCC tumors. Again, the majority of mutations in CGC genes were contributed by the DEN and STAM models with extremely high mutation rates. Importantly, although only CGC genes were considered, a large proportion of mutations had very low CHASM scores (Carter et al. 2010) indicating that many of their mutations are likely to be passenger mutations. We therefore sought to eliminate

likely passenger mutations in CGC genes using a liver cancer-specific CHASM classifier. Mutations were ranked according to CHASM score, and a cut-off was selected that retained only highly probable driver mutations (FDR \leq 0.1) unless previously documented in human tumors in COSMIC. This was repeated for human cohorts, with the additional condition that samples from the four studies were removed from COSMIC. After this procedure, we compared mutation frequencies human-mouse CGC orthologs to evaluate potential similarities. In general, most of the top 31 most frequently mutated genes in human tumors were also found frequently mutated across mouse models (**Fig. 1.2A**). However, although *TP53* was the most frequently mutated gene across human HCCs, it was only mutated in 2 of the 56 mouse tumors. *CTNNB1*, the second most common mutated gene in human HCC, was mutated at a comparable frequency in the STAM model (26%), but no mutations were observed in the other 3 models. *Ctnnb1* mutations mostly corresponded to recognized driver mutations in human *CTNNB1* including S33F, G34E, T41I, and D32N (**Table S1.2B**). Interestingly, we also observed a higher overall mutation rate in mouse tumors of several known cancer genes (Forbes et al. 2017) that were rarely mutated in human HCC (**Fig. 1.2B** and **Fig. S1.3B**). Most notable among these was *BRAF*, which was mutated in 17 out of 19 DEN tumors, all carrying the V637E substitution that is equivalent to the human V600E mutation.

Somatic DNA copy number alterations (SCNAs) have also been implicated in human HCC pathogenesis. Previous HCC studies have reported recurrent homozygous deletions in *MAP2K3*, *MAP2K4*, *PTEN*, *CDKN2A/B*, *RBI*, and *ARID1A*, and high-copy amplifications in *MDM4*, *MYC*, *CCND1*, *TERT*, and *FGF19* (Totoki et al. 2014; Ahn et al. 2014; Schulze et al. 2015). We evaluated mouse SCNAs from whole exome sequencing data using Control-FreeC (45). Overall patterns of broad and focal alterations across the four mouse cohorts varied, with copy number

losses observed more often than gains. The only large recurrent chromosomal alteration in mouse (>25% of chromosome arm) was copy loss of 14q in MUP (**Fig. S1.4A**). We also identified deletions affecting *Rb1* (14qD3; 9% mice) and *Cdkn2a* (4qC4; 5% mice) across the four mouse models, and amplifications affecting *Myc*, *Tert*, and *Vegfa* (Fig. 2C). Only SCNAs where the expression of the gene relative to the median value for the model was consistent with the direction of the SCNAs were considered drivers (Methods). Other common human SCNAs including deletions in *ERRF1* (13% human), *NCOR1* (22% human) and amplification of *CCND1* (6% human) were not observed in any of the mouse tumors.

Alteration of pathways. Pathway analysis in human HCC has previously implicated WNT signaling, Receptor Tyrosine Kinase (RTK)/PI-3 kinase signaling and Cell Cycle in HCC oncogenesis (Totoki et al. 2014; Guichard et al. 2012). To determine whether partial overlap in mutated HCC genes between human and mouse models could imply perturbation of similar pathways in mouse liver tumorigenesis we aggregated mutations from each model on these human HCC pathways. In addition to key pathways described in previous human HCC studies, we included elements of the MAPK pathway implicated by mutation frequency in mouse tumors (e.g. *Braf* in DEN mice) (**Fig. 1.3**). The mutation rate of a pathway was calculated as the fraction of tumors that harbored a likely driver mutation in one or more genes of the pathway. Among human tumors, pathway mutation rates were fairly uniform across the four cohorts with the exception of WNT signaling, which was altered in only 8% of LINC-CN tumors but in ~30% to 40% of samples in other human cohorts (**Fig. S1.4B**). Overall, the pathways with the highest mutation rate in human HCC were Cell Cycle, WNT signaling and the SWI/SNF complex, with other pathways covering less than 20% of tumors.

STAM mice had similar pathway mutation rates to humans for most pathways, but had more alterations targeting RTK signaling and chromatin modification genes (**Fig. S1.4B**). Despite the low mutation rate of *Trp53*, the Cell Cycle pathway was perturbed at a similar rate between human tumors and STAM mice, suggesting that mouse tumors may target the Cell Cycle more often through alternative mechanisms than *TP53*. In contrast to STAM, DEN mice did not frequently harbor mutations in Cell Cycle genes. Instead, DEN mice tended to have mutations in MAPK signaling and RTK signaling, with less frequent alterations in WNT signaling, chromatin remodeling and SWI/SNF. We observed few mutations in the MUP mouse model, but the few observed mutations affected the RTK and SWI/SNF pathways. TAK1 mice had no somatic alterations affecting the most frequently altered genes in human and mouse HCC. Overall these findings suggest a more prominent role for RTK signaling in mouse tumorigenesis, and suggest that in the DEN model, dual activation of MAPK signaling and RTK signaling could promote tumor formation in the absence of point mutations to Cell Cycle and Wnt genes.

In general, a single driver mutation is sufficient to perturb the activity of a pathway, thus it is uncommon to observe more than a single driver per pathway in a tumor. By aggregating mutations from multiple tumors at the pathway level, we therefore expect that driver mutations should be mutually exclusive (ME), with only one driver mutation per pathway per tumor (Canisius, Martens, and Wessels 2016; Ciriello et al. 2012; Vandin, Upfal, and Raphael 2011). Mutual exclusivity is a signature of positive selection at the pathway level and implicates the pathway in the disease process. We therefore evaluated ME of mutations within 6 pathways in Figure 3 (Cell Cycle, WNT pathway, MAPK pathway, RTK/RAS/PI3K pathway, chromatin modifiers and the SWI/SNF complex) using a groupwise DISCOVER test (Canisius, Martens, and Wessels 2016) (**Table S1.2C**). Human patients showed significant ME of mutations affecting all

6 pathways, consistent with the established role of these pathways in HCC pathogenesis and confirming that mutations retained after removing likely passengers are enriched for driver events. Applying this analysis to mouse HCC samples, we found that among STAM tumors, five of the six pathways (all but MAPK signaling) showed significant patterns of ME. In contrast, DEN tumors showed no significant ME of mutations in the 6 pathways. The DISCOVER test is intended to detect ME across pathway genes, however MAPK mutations in DEN tumors were nearly all in a single gene (*Braf*). TAK1 and MUP models had too few mutations in genes belonging to the pathways considered to produce robust results.

In order to aid in the future development of mouse models recapitulating similar molecular alterations to human HCC, we evaluated patterns of ME mutation amongst the major HCC driver genes, focusing on the 13 genes identified as mutated above expected background levels. At an FDR < 0.05, we observed ME between 12 gene pairs involving 10 of the 13 significantly mutated genes (*SI Appendix*, Table S2D). *TP53* mutations were ME with mutations to *CTNNB1*, *ALB*, *BAP1* and *RPS6KA3* while *CTNNB1* mutations were ME with *TP53*, *AXIN1*, *RBI* and *BAP1*. *AXIN1* mutations were ME with *ALB*, *ACVR2A*, *KEAP1* and *BAP1* mutations, and *BAP1* mutations also tended not to coincide with *ARIDIA* mutations. These patterns are consistent with human HCC being predominantly driven by deficiency of a factor associated with cell cycle progression (*TP53*, *RBI* and *RPS6KA3*) or by major regulators of the Wnt signaling pathway (*CTNNB1*, *AXIN1*) and suggest that secondary mutations such as those observed in *KEAP1*, *ARIDIA* or *ALB* may be biased by the primary driver pathway. Extending this analysis to include SCNAs (Methods), we observed additional ME between deletion of *RBI* and amplification of *MYC*, and between deletion of *PTEN* and mutation of *CTNNB1*.

Transcriptomic characterization. Transcriptomic changes reflect the cumulative effects of the genomic and epigenomic alterations that accumulate in cancer cells. Thus, we compared RNAseq-derived transcriptomic similarity of human HCCs (N = 371) from TCGA and the 56 mouse tumor samples. Mouse tumors clustered predominantly according to mouse model; STAM and DEN were mostly isolated in single clusters while TAK1 and MUP samples formed distinct but overlapping clusters. This configuration suggests that the largest differences in transcription across mouse HCC samples result from differences in etiology, especially for DEN and STAM and their characteristic genetic aberrations (**Fig. S1.5A**).

To determine whether tumor transcriptional profiles were similar among human tumors and mouse models, we calculated pairwise correlations between the gene expression profiles of all mouse tumors with the core TCGA 196 human HCCs and 50 matched normal liver samples restricting our analysis to the top 800 genes that are characteristically expressed in human HCC (Methods). We then clustered human samples according to their similarity to mouse tumors (**Fig. 1.4**). The majority of human normal liver samples clustered as a distinct group which included a small subset of human HCC samples. These samples fell within a larger cluster of tumors that was designated H1. Two additional larger clusters of tumors were apparent and were designated H2 and H3, with H3 being the largest cluster. Each of these clusters could be further partitioned into subclusters (H1a, H1b, H1c, H2a, H2b, H3a, H3b, H3c). We further annotated human and mouse tumors with proliferative status, as determined from expression of a gene set associated with cell proliferation (Andrisani, Studach, and Merle 2011) (**Fig. S1.5B**).

Since this clustering was the result of human-mouse transcriptome similarities, we sought to determine if these clusters contained any biologic or clinical correlates that could define them and therefore explain the basis of the inter-species similarities (**Table 1.3A,B**). H1, and particularly

the subgroup interspersed with normal samples (H1a), was enriched for lower grade tumors, tumors with low expression of proliferative genes and had a very low prevalence of *TP53* mutations. H1 was also enriched in previously characterized Hoshida clusters C1 and C3, as well as TCGA iCluster 2. H2 was comprised predominantly male tumor samples and H2 tumors were characterized by low expression of proliferative genes, silencing of *CDKN2A* and mutations in *CTNNB1*. Hoshida 3 was the predominant expression-based subtype associated with H2. Finally, H3 was enriched for high grade tumors, tumors with high expression of proliferative genes and *TP53* mutations. Hoshida 1 and 2 as well as iCluster 1 and 2 defined H3. We observed no association between clusters and HCC risk-related exposures, including alcohol consumption, HBV/HCV infection, cirrhosis or smoking. However, we did observe a bias for male tumors to be enriched for alcohol consumption as well as *CTNNB1* and *TP53* mutations (Fisher's Exact test $p < 0.01$).

The characterization of these clinically and molecularly defined human HCC clusters allowed for their meaningful mapping to mouse to interpret similarities between human clusters and mouse models. The STAM model displayed the highest general correlation with cluster H3 and the lowest with H1. Two thirds of samples were highly correlated with H3A and had low correlation with all other clusters. The other third of the samples divided into two groups, one enriched in *Ctnnb1* mutations with high correlation with H2B and another with additional similarity to H2A and H1. DEN samples did not show strong similarity with any human cluster, with most samples displaying medium correlation to normal samples and H1/H2. The TAK1 model had the highest correlation with normal human liver samples while showing medium correlation with H1 and H2. Most MUP samples showed high correlation with normal and H1

groups with low correlation with H2 or H3. Only two MUP samples showed high correlation with H3 and low correlation with normal samples.

The clustering of STAM tumors with high-grade proliferative and *TP53* mutation-enriched H3 HCCs versus *CTNNB1*-mutation enriched H2 HCCs suggested the possibility of shared patterns of gene expression downstream of inactivated *TP53* or activated *CTNNB1*. Indeed, 10 *TP53* response genes, 13 *CTNNB1* response genes and 7 WNT pathway genes numbered among the 800 HCC-associated genes used for evaluating sample correlation. Ranking mouse tumors according to expression of *TP53* response genes, STAM models were associated with higher activity of 6 genes inactivated by *TP53* including mouse orthologs of *CCNB1*, *CCNB2*, *CDK1*, *EZH2*, *FOXM1* and *PLK1* ($p < 0.01$), suggesting that *TP53* activity is frequently impaired in this model despite the paucity of *TP53* mutations (**Fig. S1.5C**). Interestingly, STAM tumors with *Ctnnb1* mutations had similar patterns of downstream expression, but tumors with mutations to *Apc* appeared to have the highest overall expression of *CTNNB1*/WNT response genes. STAM tumors with neither *Ctnnb1* nor *Apc* mutations were more highly correlated with H3 tumors and had higher overall expression of a subset of genes including *Loxl2*, *Cd14*, *Spp1*, and *Mmp14* that were expressed at low levels in the *CTNNB1* mutant STAM samples that had high correlation with H2 but low correlation with H3 (**Fig. S1.5D**).

In summary, TAK1 and MUP tumors were most similar to low grade human tumors, with a small number of MUP tumors beginning to show characteristics of higher grade tumors to which the STAM model was the most similar. Notably, the DEN model does not appear to recapitulate the expression signature of any human tumors well. Overall, these results demonstrate that integration of human molecular and clinical data with mutational and transcriptomics data from

mouse models allows for a useful mapping of mouse models to the most suitable subgroups of patients.

s The role of the immune system in cancer initiation, development and clinical manifestation has recently taken center stage because of the demonstrated potential for immune manipulation to generate strong positive therapeutic outcomes (Mittal et al. 2014). Human HCCs were reported to differ ostensibly in the immune cell content of their tumor (Cancer Genome Atlas Research Network 2017); some tumors were characterized by high levels of immune cell infiltrate, while others were almost devoid of immune cells, and cellular composition of the immune infiltrate differed substantially across tumors. Indeed, the important effects that inflammation induced B-cells exert on HCC development were recently revealed in the Mup-uPA HFD and STAM models by inhibiting T-cell immune surveillance (Shalapour et al. 2017), demonstrating that mouse models can be effective for studying mechanisms of anti-tumor immunity and point to new therapeutic strategies to unleash robust immune responses. We therefore assessed the characteristics of the immune infiltration in the four mouse models of HCC from expression levels of immune marker genes to evaluate similarities among models and with human HCC.

We estimated the relative abundance of 9 immune cell types in human and mouse tumors based on expression of established cell type specific markers previously used for this purpose (Davoli et al. 2017) (**Fig. 1.5A**). Levels of mature CD4 T cells, T regulatory cells and dendritic cells were comparable in human and mouse tumors. In contrast, mouse tumors had lower levels of CD8 T cells, natural killer cells and B cells, but higher levels of all types of macrophage. Consistent with lower CD8 T cell levels, we observed lower levels of cytotoxicity, as approximated from granzyme A and perforin expression (Rooney et al. 2015), in mouse tumors versus human HCCs (**Fig. 1.5B**). Cytotoxicity did not correlate with mutation burden across mouse or human tumors

(**Fig. S1.5E**). We did not observe major differences in immune cell infiltrates when comparing between mouse models (**Fig. S1.5F**), suggesting that the molecular differences among the four models do not result in systematic differences in immune infiltration.

Innate and adaptive immunity can strongly promote or suppress cancer initiation, progression and dissemination. Importantly, T-cell mediated elimination of cancer cells depends strictly on their T cell receptor (TCR) specifically recognizing an MHC-I bound neoantigen on cancer cells. We recently reported that individual genotype of the Human Leukocyte Antigen (HLA) locus that encodes the Major Histocompatibility Complex Class I (MHC-I), imposes a restriction on which cancer driver mutations are more likely to arise during carcinogenesis. This restriction results in a personal blind spot to specific driver mutations generated as a consequence of the different binding patterns displayed by the more than 3000 human MHC-I alleles. Using the Best Rank (BR) mutation presentation scoring scheme we previously developed and the same sets of driver and passenger mutations we defined based on human tumors (Marty et al. 2017), we compared BR score distributions for driver- and passenger mutations across different mammalian species' MHC-I alleles ($p < 1.0e-5$ for all except pig (*Sus scrofa*)), confirming that the general trend of higher presentation for cancer driving mutations is evolutionary conserved (**Fig. 1.5C**).

We next focused the analysis on HCC patients' MHC-I alleles and mutations. We used the Patient Harmonic-mean Best Rank (PHBR) score (Marty et al. 2017) to estimate the genotype-specific immune presentation of mutations. Again, when we compared the score distribution of drivers and passengers found in HCC patients, there was a significantly higher score for HCC driver mutations ($p < 0.05$; **Fig. 1.5D**). Interestingly, when we considered mouse HCC mutations and alleles, we observed the same trend suggesting that mouse MHC-I genotype exerts similar constraints in HCC development as found in human. These observations indicate that mouse

MHC-I genotype can have a strong influence in selecting the driving mutations a given mouse model will develop. Since different mouse strains commonly used in preclinical cancer models have completely different sets of MHC-I alleles, it could be that the same cancer model can present different mutations depending on the genetic background. To analyze this possibility, we calculated PHBR scores for all mutations found in the four mouse models considering C57BL/6 or BALB/C for which binding predictions can be obtained. C57BL/6 has the MHC-I b allele for H-2K and H-2D and is null for H-2L. BALB/C has the d allele for all three loci. PHBR distributions show a significantly higher PHBR bias for C57BL/6 ($p < 0.01$) indicating that if tumor samples had been obtained from BALB/C instead of C57BL/6, the mutational content could have been different (**Fig. 1.5E,F**), perhaps because of the additional H-2L alleles. Taken together these results suggest that the C57BL/6 background with only 2 homozygous MHC-I alleles could be driving the lower CD8 infiltration and cytotoxic activity observed in mice compared to human.

1.4 Discussion

HCC mouse modeling has historically relied heavily on the use of a variety of hepatotoxins which efficiently induce liver tumors in rats and mice. However, the genetic make-up of these tumors remains unknown and therefore validity to faithfully model human HCC is uncertain. An outstanding difficulty in mouse modeling of human HCC is the universal link with cirrhosis and the diverse conditions associated with the development of liver disease. Recent genomic analyses have uncovered an unexpected complexity in the genetics of HCC that hampers a direct translation into mice using conventional transgenesis (Bakiri and Wagner 2013). All these antecedents have contributed to the accumulation of countless mouse models of HCC generated by a large variety of chemicals, genetic engineering, dietary factors, and administration regimes (Heindryckx, Colle, and Van Vlierberghe 2009b; Chen and Calvisi 2014). Nowadays, it is quite unusual to find two independent studies in which the same model or regime for inducing mouse HCC is employed. In order to mitigate this extraordinary diversity in HCC experimental models that hampers advancement in the field, we performed a systematic investigation of the similarities and differences between human HCCs and four mouse models of HCC at the genomic and transcriptomic scales. This study provides a molecular reference for tailoring the use of HCC mouse models to specific experimental hypothesis or clinical testing.

Mutation signature analysis revealed subtle differences across human cohorts. However, among the four mouse models there was large variation in signatures suggesting major differences in DNA damage mechanisms and mutation rates. DEN-induced tumors had an expected high mutation rate with a unique mutation signature which has never been observed in human cancer. STAM-induced tumors surprisingly contained an even higher mutational rate expected of a very potent hepatotoxin. The mechanisms of carcinogenesis in STAM model have been attributed to

streptozotocin-induced diabetes combined with a high fat diet, thought to induce severe chronic liver injury leading to rapid HCC development. Our findings suggest that streptozotocin is also a potent liver carcinogen with a distinct alkylating mutational signature. Notably, STAM mice do not develop tumors unless liver damage is subsequently induced by a high fat diet. Since streptozotocin is an approved anti-neoplastic agent, we speculate that patients treated with this drug may benefit from preventive screening for HCC, particularly upon subsequent liver injury. Mutation signature clustering showed that TAK1 and MUP induced tumors present similar signatures to human HCC. This observation indicates that long term chronic liver damage without any carcinogen typical of these two models activates similar mechanisms of mutagenesis to that observed in patients that suffer from chronic liver disease. Taken together, these analyses show that mouse models of HCC based on administration of mutagens are not adequate models to study tumor initiation and mutational processes observed in human HCC. However, mouse models based on the spontaneous development of tumors after long lasting liver injuries tend to more faithfully recapitulate the mutational processes observed in human HCC patients as was previously postulated for MUP-uPA + HFD (Shalapour et al. 2017) .

Contrary to the mutation signature analysis, gene and pathway-centric analysis determined that TAK1 and MUP tumors lacked oncogenic mutations and pathway alterations typically observed in human HCC. DEN molecular characteristics included mutations to the MAPK pathway, almost universally the equivalent of *BRAF* V600E, and frequent alterations to PI3K signaling. In a recent mouse study of DEN induced HCC, Conner *et al.* similarly reported recurrent *Braf* mutation was the predominant driver in C57BL/6 mice (Connor et al. 2018). Of note, *TP53* and *CTNNB1*, the most frequently altered genes in human were never found mutated in DEN-induced tumors. The DEN model thus results in tumors that are clearly distinct and different from

human HCC. STAM tumors, with their high mutation rate, represented the only model containing oncogenic mutations in the prototypical genes found mutated in human HCC, concomitantly altering Wnt, Cell Cycle and Chromatin Modification pathways. While STAM tumors carried *Cttnb1* mutations at a comparable rate to human tumors, *Trp53* mutations were less frequent. Nonetheless, STAM was the only mouse model that closely recapitulated the molecular characteristics of human HCC.

Mutation data is by nature sparse and therefore, it is not appropriate to quantitatively measure the degree of similarity between mouse and human HCC, thus we compared tumors evaluating similarity at the transcriptomic level. Clustering human and mouse tumors according to pair-wise transcriptomic correlations and consideration of molecular and clinical features allowed for the new definition of groups of patients based on their similarity to mouse HCC tumors. The H1 human cluster, which includes normal and low-grade samples, is mainly defined by its similarity to TAK1 and, to a lesser extent, MUP tumors along with a few samples from the DEN and STAM models. Indeed, around 70% of tumors >2mm in HFD fed MUP-uPA were histologically reported to be adenomas which agrees with the lowest mutational load, lack of oncogenic mutations and low general similarity to human HCC (Nakagawa et al. 2014). The absence of oncogenic mutations in TAK1 and MUP and high correlation with “normal like” human tumors and normal human liver invalidates these models for clinical testing or analysis of tumor biology. The H2 human cluster which is characterized by enrichment of *CTNNB1* mutations and silencing of *CDKN2A* was defined by its similarity to half of the STAM tumors, including all *Cttnb1* mutated samples, as well as to TAK1 tumors although at lower correlation values. Finally, cluster H3, characterized by high grade tumors and enrichment in *TP53* mutations, was uniquely defined by its high correlation with all STAM samples. We observed a sex-specific bias toward

males in the *CTNNB1* mutation associated H2 tumor group (*SI Appendix*, Table S3) and found that *CTNNB1* was more frequently mutated in male tumors. This raises the possibility of sex-differences that could influence the molecular characteristics of tumors. Previous studies have suggested that the more frequent occurrence of HCC in men could be attributable to differences in environmental exposure (greater alcohol consumption, smoking, and higher body mass index) or hormone levels (Totoki et al. 2014). Schulze *et al.* (2015) also suggested that the mutation of *CTNNB1* in HCC is higher in cohorts with high levels of alcohol consumption (Schulze et al. 2015). Interestingly, we found that *TP53* mutations were biased towards male tumors as well. We note that the mouse tumors studied here were almost universally derived from male mice, but we did not observe any bias for these samples to correlate more strongly with male or female tumors on the basis of the expression of the 800 HCC genes ($p \gg 0.05$; **Fig. S1.5G**). This confirms that the transcriptional similarities observed among mouse and human tumors were not confounded by gender.

Anti-tumor immunity is emerging as an important research area, and mouse models are likely to play a central role in determining the optimal application of immunotherapy in HCC. Although we see evidence that mouse MHC I genotype restricts mutations in tumors in a manner similar to humans, we nonetheless see differences in immune cell infiltration and corresponding differences in cytotoxicity. It is possible that these differences are rooted in the rate at which mouse tumors develop (Shalpour et al. 2017), or could reflect differences in presentation associated with the diversity of HLA alleles carried by a particular strain. Specifically, the C57BL/6 genetic background of the tumor samples analyzed here includes only two homozygous HLA alleles, potentially reducing the breadth of peptides presented and therefore minimizing the substrate available to drive T-cell infiltration and recognition. Interestingly, Connor *et al.* found

that DEN induced mutations differed between mouse strains, with Braf mutations being more prevalent on C57BL/6 while Hras and Egfr were far more frequent in CH3 mice (Connor et al. 2018). We note that these two mouse strains carry different MHC genotypes, further suggesting that genetic background could be an important consideration for using mice to study tumor development and preclinical studies. Nevertheless, further work is needed to determine whether conclusions about immune response drawn from specific mouse models will generalize well to human disease.

In summary, we performed a comprehensive human-murine HCC molecular comparison to determine the similarity of a given mouse model to different subgroups of patients. We analyzed four different models diverging from carcinogen treatment to genetic induction of chronic liver injury. Our results indicate that mouse models spontaneously inducing HCC through chronic liver damage tend to be most similar to human HCC in terms of mutational processes but these tumors tend to be low grade with very low frequency of oncogenic mutations typically found in standard human HCC. Thus, these models appear to be only suitable to study HCC initiation or microenvironmental determinants of mutational processes. In contrast, our analysis implicated the STAM model as the most suitable model for research on tumor biology and preclinical studies. Despite its high mutational burden and biased mutation signatures, it generates tumors most molecularly similar to high grade human HCC as well as with human tumors with *CTNNB1* mutation. Our results also indicate that in general, the DEN model should be avoided as a model of human HCC because it is dominated by mutational mechanisms that are never found in human cancer and DEN-induced tumors have clearly a different molecular profile at the gene, pathway and transcriptomic level compared to the other models analyzed. Of note, we also found that the genetic background of the mice can have an impact on the molecular characteristics of the tumors

that develop based on mouse strain-specific MHC-I genotype. In conclusion, we expect that these analyses will help the HCC field to better tailor experimental analysis to the most adequate mouse models and at the same time serve as an initial workflow by which additional mouse models can be reliably characterized.

1.5 Materials and Methods

Samples. Mouse studies were performed in accordance with NIH guidelines for the use and care of laboratory animals and approved by the UCSD Institutional Animal Care and Use Committee, S00218. All mice used were of C57BL/6 background. Tumors >4mm were excised from all mice without any other consideration to avoid biases. TAK1 samples consisted of 9 tumors and 3 spleens from five 9 months old Alb-Cre; Tak1F/F mice (Inokuchi et al. 2010; He et al. 2013a) . MUP samples, 1 spleen and 9 tumors, were obtained from five MUP-uPA male mice fed with HFD as previously reported (Nakagawa et al. 2014). DEN samples, 19 tumors and 1 spleen, were obtained from 8 male mice injected with DEN at P15 as previously described (Heindryckx, Colle, and Van Vlierberghe 2009b; He et al. 2013a). STAM samples, 1 spleen and 19 tumors, were obtained from six male mice treated with STZ at P2 and fed with HFD as previously described (Fujii et al. 2013). DNA and RNA were extracted and purified with Qiagen AllPrep DNA/RNA Mini Kit. Exome capture and sequencing were performed at UCSD genomics core for TAK1, DEN and STAM samples while MUP samples were processed at BGI. RNAseq for all samples was performed at UCSD genomics core.

Exome Sequencing and Data Analysis. For mouse data, the quality of the raw FASTQ files was checked with FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). WES reads were mapped to GRCm38 using BWA (v.0.7.5). Duplicates were removed using Picard (v.1.94); normal (spleen)-tumor BAM files were used as input for The Genome Analysis Toolkit (GATK, v3.2-2). Local realignment and base quality recalibration were performed using default parameters. Somatic single nucleotide variants (SSNVs) and small insertion and deletion (indels) were called using GATK MuTect2 (<https://ccbr.github.io/Pipeliner/Tools/MuTect2.html>). Only the variants with PASS status were considered for further analysis.

Small-Variant Annotation for WES. Somatic variants were annotated into proteins and amino acid changes using Variant Effect Predictor v92 (1) to determine the synonymous or nonsynonymous nature of each alteration for both mouse and human to have consistent definitions of mutational effects (human = hg38, mouse = mm10). The list of 17,046 mouse-human 1:1 orthologous genes was obtained from MGI (<http://www.informatics.jax.org>) and Ensembl (2). Mouse variants were mapped to orthologous positions in human proteins using liftOver (3) from GRCm38/mm10 to GRCh38/hg38. 82% of mutations were successfully mapped to the orthologous position, 2% of mutations were mapped to the corresponding position but the reference AA differed between species, and 16% could not be mapped to an orthologous position. Human mutations and human orthologs of mouse mutations were further annotated as likely drivers if they received a CHASM driver prediction (4) with FDR ≤ 0.1 or had previously been observed at least once in an independent set of tumors in the COSMIC database.

COSMIC Cancer Census Genes. We used COSMIC (5) to identify known HCC genes as well as known cancer genes. Cancer-related genes were obtained from the COSMIC Cancer Gene Census v84. Genes which were included on the basis of translocations were omitted from the list, resulting with a total of 474 unique genes.

Mutational Signature Analysis. Mutational profiles were generated by considering the immediate 5' and 3' sequence context for each somatic single point mutation resulting in a mutational classification with 96 categories. *De novo* extraction based on somatic substitutions and their immediate sequence context was performed to derive the set of mutational signatures in mouse models and human cohorts using our previously developed methodology (6). The set of identified mutational signatures was compared to the global set of consensus mutational signatures (7) and these signatures were reintroduced in the samples following our prior approach (8).

Documented mutational signatures were obtained from COSMIC (5). The sources of the somatic mutations were as follows: mouse HCC exome-based (total N=56; from TAK1, MUP, STAM, and DEN, N = 9, 9, 19, and 19, respectively), human HCC exome-based (total N=987; from LICA-CN, LICA-FR, LIHC-US, and LINC-JP, N=163, 234, 377, and 213, respectively) and genome-based (total N=314; from LICA-FR, LIHC-US, LINC-JP, and LIRI-JP, N=5, 53, 28, and 228, respectively). Mutation % for each of the signatures within each mouse and human cohorts are calculated based on their total mutation burden.

Identification of Significantly Mutated Genes. MutSigCV v1.0 (9) was applied to the simple somatic mutation sets of human exome-sequenced cohorts from both ICGC and TCGA. Genes were considered significantly mutated at an FDR < 0.1. Pan-HCC significantly mutated genes were determined by meta-analysis using Fisher's Method (10). Mouse sample sizes were too small to apply MutSigCV. Instead, significantly mutated genes in mice were determined based on dN/ds ratio (11) and empirical significance was determined against a random model. Specifically, we generated random mutation profiles 1000 times by sampling base substitution and context from the underlying 96 category mutation profile and creating the same number of mutations per gene as observed in the tumor. This generated a distribution of 1000 random dN/dS scores for each gene that was then used to assess an empirical p-value. We applied multiple testing corrections (Benjamini-Hochberg) to adjust for the p-values. Genes with adjusted-p < 0.05 were considered significantly mutated genes for each of the mouse models.

A second analysis was performed considering only driver mutations. Human and mouse genes were selected for comparison based on driver alteration frequency, approximated as the fraction of tumors harboring a driver mutation in that gene. Driver alteration frequency was estimated

separately for human and mouse tumors, and the top genes for each group were included in Figure 2 and added to Figure 3.

Somatic Copy Number Alteration (SCNA) Analysis. Copy number regions for mouse exomes were called by ControlFreeC (12) using the default setting for exome sequencing data. Background spleen samples for each model served as the control. Recurrent CNV regions were defined as regions which have been observed in more than 1 tumor, but not in all of the tumors from a single mouse model (as these are more likely to indicate potential sequencing artifacts). For the DEN and STAM models that were sequenced in two batches, we removed any regions that were present only in the first or second batch (*SI Appendix*, Fig. S3A). For selected genes, SCNA status was compared with the median expression of the gene in the associated mouse model. An amplification or deletion was deemed likely functional if the expression of the particular gene was above or below the median expression value respectively for the relevant mouse model. For the majority of these SCNAs, we observed that the expression of the gene relative to the median value for the model was consistent with the direction of the SCNAs (Fig. 2C).

Pathway Analysis of Somatic Mutations. We used the HCC pathways previously published by TCGA to define groups of functionally related genes. Recurrent somatically altered genes in mice that participate in these pathways were added according to annotated interactions in KEGG (<http://www.genome.jp/kegg/>). Pathway mutation rates were estimated using only predicted driver mutations. We calculated pathway alteration frequency as the percentage of samples with one or more genes altered in the specific pathway.

Mutual Exclusive and Co-occurrence Analysis. We compared the alteration frequency of the four human cohorts (considered as ‘Human’ group) to the pathway alteration frequencies of each of the mouse models. We used the DISCOVER algorithm(13) to test for mutual exclusivity

of mutations within pathways. Each HCC pathway was assessed for mutual exclusivity of driver mutations using a single impurity-based group test. We also performed pairwise tests for mutual exclusivity for the top 13 frequently mutated and 5 most recurrently CNV-altered genes in human HCCs. The DISCOVER algorithm provides adjusted p-values based on an adaptation of the Benjamini-Hochberg procedure.

RNAseq Expression Analysis. Transcript levels were quantified by Sailfish (14) in Transcripts per Million (TPM) and only 1:1 mouse human gene expression levels were retained. We focused the human RNAseq analysis on 196 patients analyzed in the TCGA publication for which clinical and molecular annotation was available and 50 tissue-matched normal samples from TCGA (15). To identify genes with expression changes relevant to HCC, we applied DESeq2(16) to RNAseq derived raw read count data for human HCCs and tissue-matched normal samples. We first identified genes associated with HCC by comparing tumor samples against tissue matched normals. Next we extended the differential expression analysis to include genes that might be relevant for identifying distinct HCC subtypes. This was accomplished by labeling the 196 tumors according to the expression based clustering analysis performed by the Broad Firehose Pipeline (doi:10.7908/C1H994MZ), and then identifying the genes that were differentially expressed in each cluster versus all others. The top 5% of genes with an expression fold change greater than 2 and FDR < 0.05 were selected as significantly differentially expressed genes (DEGs). In total this resulted in 800 genes from these comparisons as follows: (a) all HCC tumors versus all tissue-matched normal (N genes=303), (b) TCGA group 1 (N genes=295), (c) TCGA group 2 (N genes=117), and (d) TCGA group 3 (N genes=152) versus all other patients.

Unsupervised Clustering of Human and mouse tumors based on expression correlations. TPMs for the 800 HCC associated genes were subsequently log₂ transformed after

adding a pseudocount of 1. The log₂ transformed expression values were then z-score transformed and pairwise similarity between human and mouse tumors was determined using the Pearson correlation between profiles. Clustering was performed by agglomerative hierarchical clustering with euclidean distance and average linkage method using seaborn clustermap in python. Human tumors were divided into 3 major subgroups (H1, H2, H3) and 8 minor subgroups (H1a, H1b, H1c, H2a, H2b, H3a, H3b, H3c) by cutting the dendrogram.

Proliferative Status. We determined proliferative status of each tumor based on unsupervised hierarchical clustering on a set of 104 proliferation associated genes (17). Based on the unsupervised clustering result, we separate the patients into two groups: high proliferative (high overall expression) and low proliferative (low overall expression). To define the mouse proliferative status, we performed unsupervised clustering on the same set of genes simultaneously for both the mouse samples and the human patients. The mouse samples were then assigned to proliferative or nonproliferative status based on the labels of the nearby human tumors.

Covariate Enrichment Analysis of Human Subgroups. We evaluated each of the human subgroups for enrichment of key HCC-associated molecular events and clinical covariates and liver-damaging exposures. The covariates tested for the major and minor subgroups included pre-existing clustering assignments (iClusters and Hoshida clusters), relevant environmental exposures (hepatitis B virus (HBV) and hepatitis C virus (HCV) infection status, smoking history, alcohol), clinical (gender, ethnicity, tumor grade, tumor stage, tumor cell purity), and molecular (*CDKN2A* silencing and mutation status of *TP53*, *CTNNB1*, *TERT*, *APC*, and *AXIN1*, and proliferative status) details for the 196 HCC patients. For discrete variables (exposure status, gender, mutations), we performed Fisher's Exact Tests on to evaluate enrichment for each group/subgroup. Tumor stage was defined as low (stage I and II) or high (stage III and IV). For continuous variables (purity), we

used the Wilcoxon rank-sum test to evaluate enrichment. We used the Benjamini-Hochberg method to adjust for testing multiple subgroups and covariates and only reported findings at FDR < 0.05. We analyzed association with covariates at three scales: (1) between major groups, (2) between minor subgroups within a major group, and (3) between all minor subgroups, using a one versus rest approach.

TP53 and CTNNB1 activity analysis. We evaluated the 800 HCC associated genes used for correlation-based clustering for overlap with curated sets indicative of *TP53* activity (20 induced and 10 repressed *TP53* target genes). A *TP53* activity score for each mouse tumor was assessed based on the summed Z-score of expression of 10 genes in the overlap. Since this score was driven predominantly by higher expression of *TP53* suppressed genes (*SI Appendix*, Fig. S5C), a higher score indicates reduced *TP53* activity. Mouse tumors were ranked according to *TP53* activity score and enrichment of the STAM model for reduced *TP53* activity was evaluated using the Wilcoxon rank-sum test. We similarly assessed overlap between the 800 HCC associated genes and *CTNNB1*/WNT response genes based on several curated gene sets from MSigDB (<http://software.broadinstitute.org/gsea/msigdb>) (SANSOM_WNT_PATHWAY_REQUIRE_MYC, KENNY_CTNNB1_TARGETS_UP, and KENNY_CTNNB1_TARGETS_DN). In total, these pathways included 160 genes (58 WNT pathway, 50 *CTNNB1* up- and 52 down- regulated genes), of which 20 were included in the overlap. We clustered mouse samples based on the 20 overlapping genes using agglomerative hierarchical clustering with euclidean distance and average linkage method using seaborn clustermap in python (*SI Appendix*, Fig. S5D). For both *TP53* and *CTNNB1* response heatmaps (*SI Appendix*, Fig. S5C,D) we annotated mouse samples according to median correlation with human HCCs in major subgroups H1, H2 and H3.

Allele-based Immune Presentation Analysis. To evaluate the conservation of immune presentation across species, we compared immune presentation scores for driver and random mutations as described in Marty *et al.* (Marty et al. 2017) across all alleles present in IEDB for several species: mouse (*Mus musculus*), chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*), and pig (*Sus scrofa*). The 1,018 driver mutations were mapped to their orthologous position in mouse (*Mus musculus*, *GRCm38*), monkey (*Rhesus macaque* or *Macaca mulatta*, *Mmul_8.0.1*), chimpanzee (*Pan troglodytes*, *Pan_tro3.0*), and pig (*Sus scrofa*, *Sscrofa11.1*) using liftOver. Random mutations (N=10,000) were generated for each species by sampling random positions from the protein fasta downloaded from the Ensembl Release 91 Databases (<ftp://ftp.ensembl.org/pub/release-91/>) and creating random amino acid changes. Each HLA allele-mutation pair was assigned an immune presentation score using the Best Rank formulation computed from affinities predicted by NetMHCpan3.0 (Nielsen and Andreatta 2016). Only classic alleles from each species are considered. Best Rank scores (Marty et al. 2017) across 2,924 human alleles, 409 monkey alleles, 8 mouse alleles, 105 chimpanzee alleles, and 55 pig alleles were calculated for the 1,018 driver mutations and 10,000 random mutations. Across categories, this resulted in 38.7 million residue scores (human: 2,924,000 driver, 29,210,000 random; mouse: 7,416 driver, 80,000 random; chimpanzee: 101,220 driver, 1,050,000 random; gorilla: 397,155 driver, 4,150,000 random; pig: 69,750 driver, 750,000 random). Best Rank scores were then pooled across HLAs.

Patient Harmonic-mean Best Rank (PHBR) score. PHBR scores (Marty et al. 2017) were calculated for mutations observed in human and mouse models and distributions were compared between driver and passenger mutations within each species. Observed mutations were considered drivers if they overlapped the 1,018 driver mutations from the human pan-cancer study

(Marty et al. 2017) and the remainder of the mutations were considered passengers (human: 182 driver, 28216 passenger; mouse: 27 driver, 44601 passenger). Within each species, the difference in distribution of PHBR scores for drivers versus passenger mutations was evaluated using the Wilcoxon rank-sum test. We also evaluated the effect of mouse strain-specific MHC-I genotypes on PHBR score distributions for driver and passenger mutations. The mouse models in the current study are from the C57BL/6 strain which carries the k allele for H-2K and H-2D and is null for H-2L. We compared presentation by the C57BL/6 genotype to presentation by the BALB/C genotype which consists of the d allele for all three HLA loci using the Wilcoxon rank-sum test. Parametric and non-parametric statistics tests were used to analyze significance of results, including Wilcoxon rank-sum test to evaluate enrichment, TP53 and CTNNB1 activities, and driver passenger mutations distributions. We applied multiple testing corrections (Benjamini-Hochberg) to adjust for the p-values. * signifies $P < 0.05$ unless otherwise noted.

Immune Signatures. The log₂-transformed TPM gene expression data was z-scored within human (196 patients) and mouse cohorts (56 samples) respectively and used to calculate cell type specific scores (summed z-scores for cell type-specific marker genes) approximating immune cell infiltration in each tumor as described by Davoli *et al.* (18). Immune cell type-specific markers described by Davoli *et al.* were filtered to retain only the 1:1 orthologous genes shared between human and mice prior to estimating infiltration in order to allow cross-species comparison. Cytotoxicity was measured as described in Rooney *et al.* (19), by summing the z-scored expression of the key cytolytic effectors, granzyme A (*GZMA*) and perforin (*PRF1*).

1.6 Figures

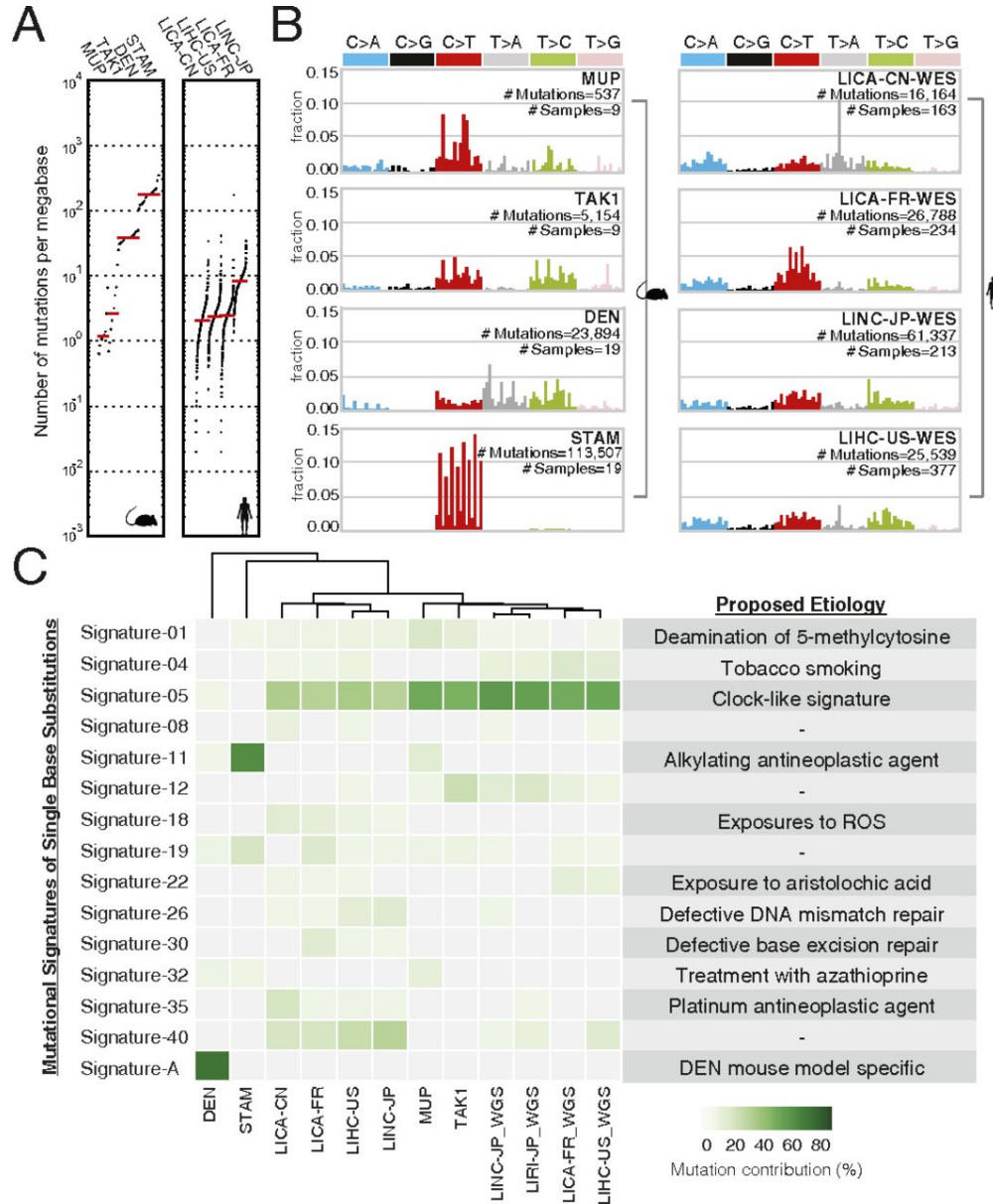


Figure 1.1. Mutational profiles for the human and mouse cohorts and corresponding mutagenic signatures.

(A) Mutational burden of mouse models and human HCC patient cohorts (WES). Design inspired by Lawrence *et al.* (Lawrence *et al.* 2013). (B) Nucleotide substitutions for each cohort were binned into 96 categories by combining the 6 possible base substitutions (C>A, C>G, C>T, T>A, T<C, T<G) with 5' and 3' flanking bases, following the standard/alphabetic order of trinucleotides. Human cohorts are separated by geographical origins of the ICGC data. (C) Mutation contribution (%) for the identified signatures are shown for each human and mouse cohort. Signatures are shown if they account for >5% of mutations in one or more of the 12 cohorts.

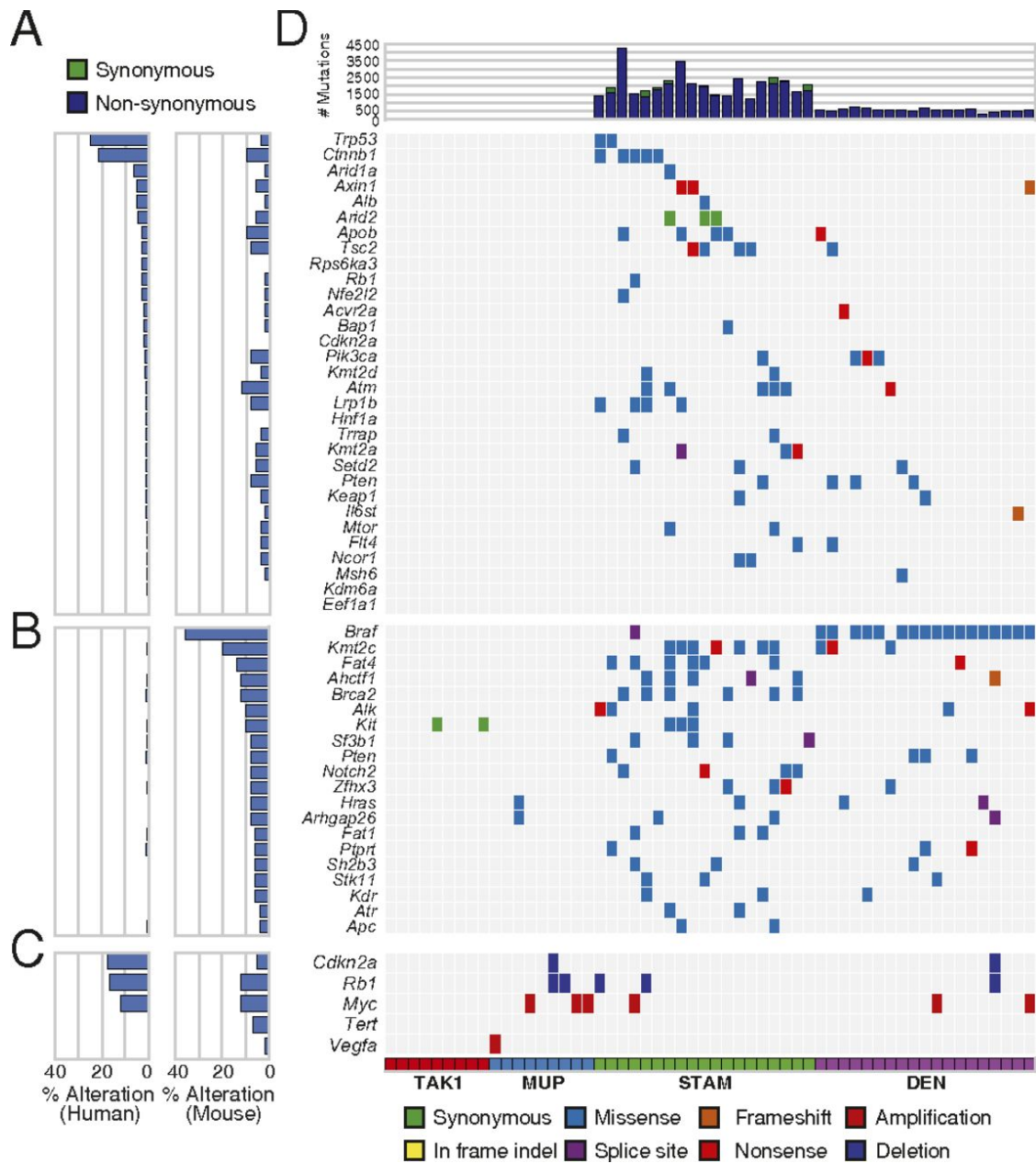


Figure 1.2. The driver landscape of four HCC mouse models.

Mutational effects (synonymous, missense, etc.) are indicated for each gene and sample. Silent mutations were excluded unless the identical mutation was observed in multiple independent mouse tumors. The mouse models are identified by a color bar at the bottom of the figure. The top bar plot shows unfiltered individual tumor mutation rates across tumors. Bar plots on left side show human and mouse alteration frequencies across: **(A)** frequently mutated genes from human HCC cohorts, **(B)** frequently mutated cancer genes in the four mouse models, and **(C)** somatic copy-number alterations (SCNAs) of genes in mouse which overlapped with recurrent SCNAs previously reported in human HCC.

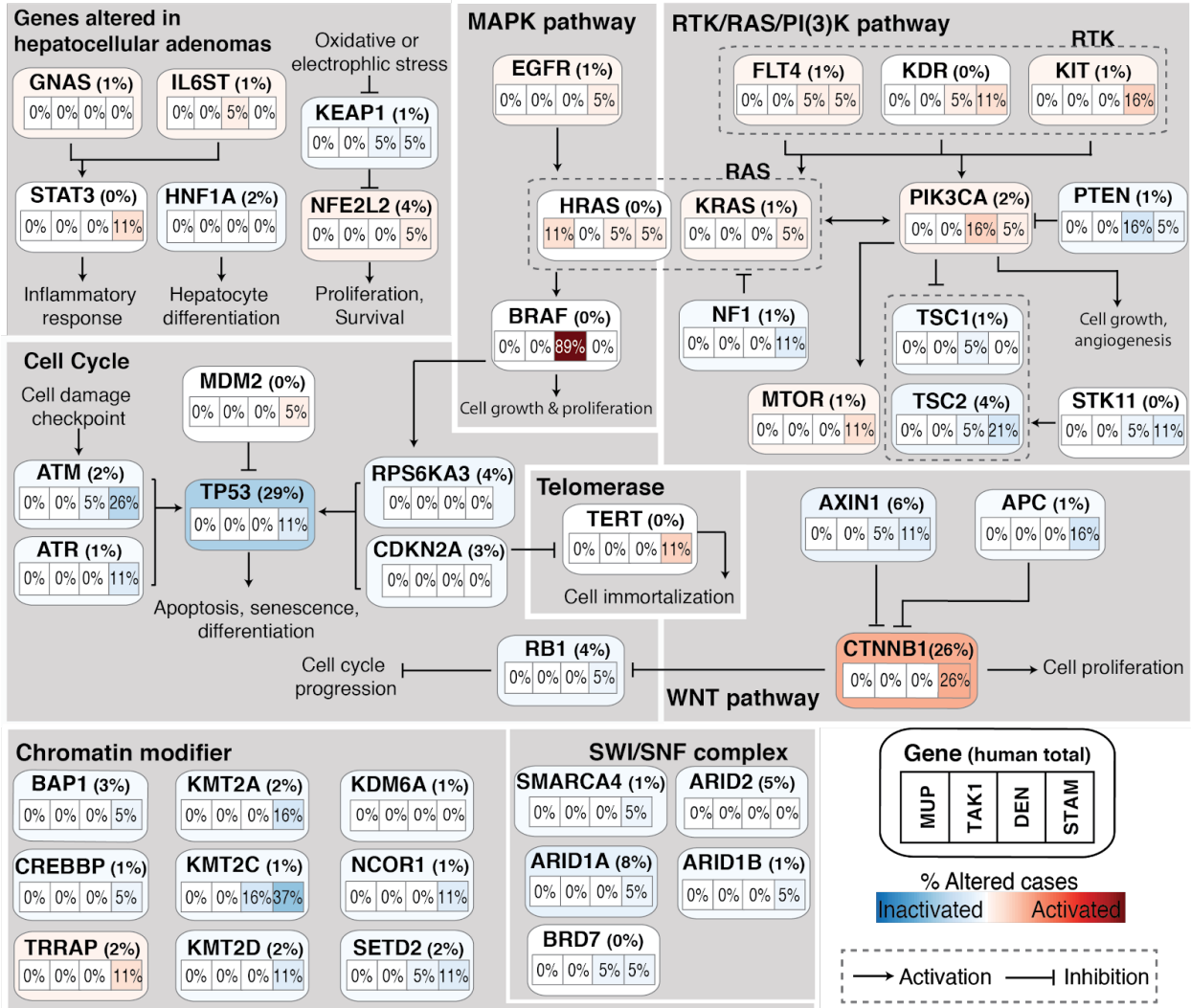


Figure 1.3. Comparison of somatic alteration across HCC signaling pathways.

Somatic mutations are indicated as the percentages of altered cases. Frequencies across 4 human cohorts are indicated as a single percentage to the right of the gene name (human gene nomenclature is used), while frequencies in the four mouse models are indicated separately by cells inside each gene. Missense mutations were only included if they were likely drivers (Methods). Genes are grouped by signaling pathways, with edges indicating direction and effect of gene activity within the pathway.

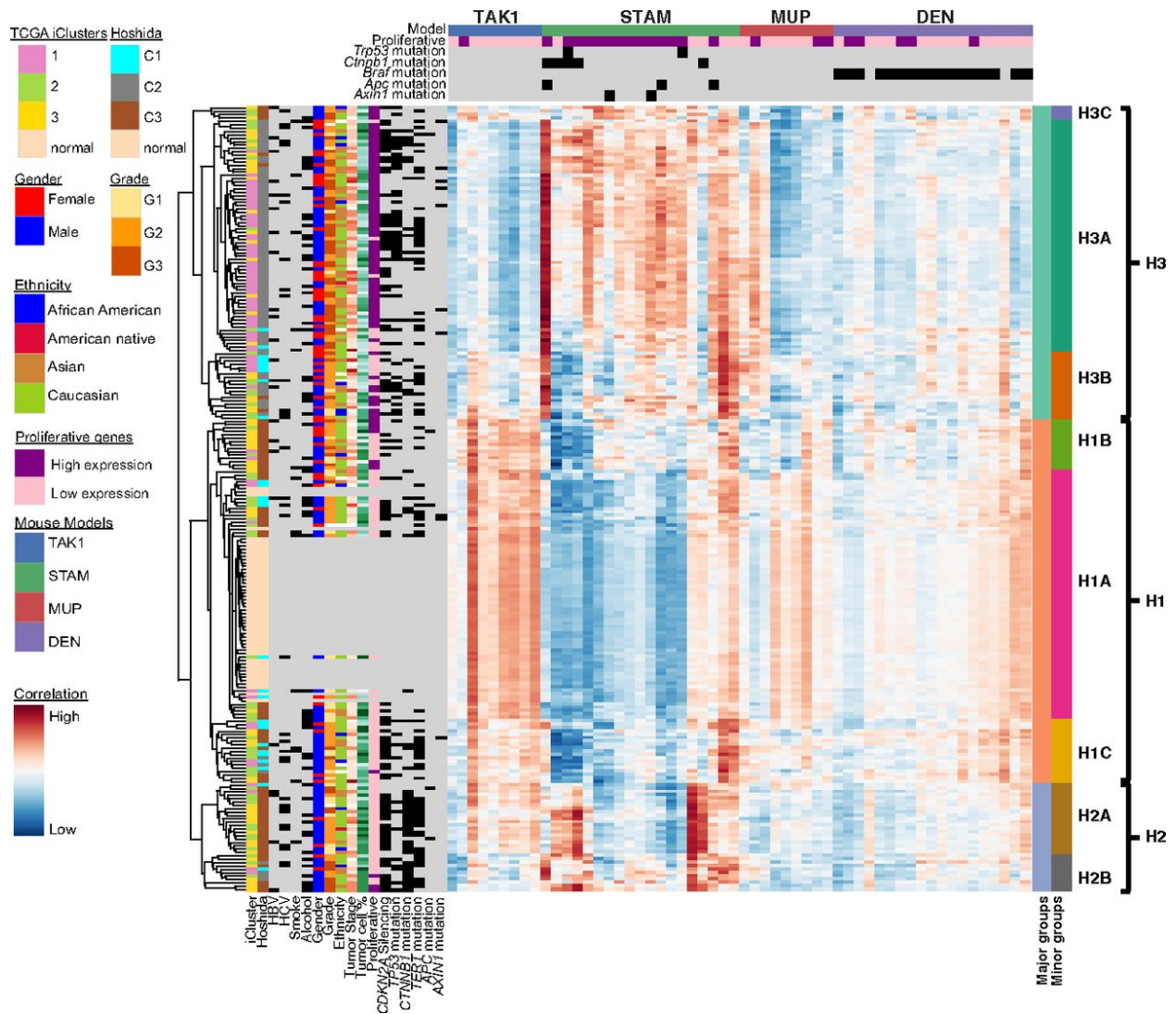


Figure 1.4. Pairwise correlations between mouse tumors and human HCCs based on gene expression profiles.

Human clusters are indicated by a color bar on the right side of the heatmap and the mouse models are indicated by a color bar at the top. Mutations for selected HCC genes in mouse tumors are shown in the top panel. The left side includes a dendrogram of human HCCs and matched normal liver samples clustered according to correlation with mouse tumors based on the expression of 800 HCC related genes. Clusters are labeled on the right side of the heatmap. Also left of the heatmap is a panel with clustering assignments (iCluster and Hoshida cluster), relevant environmental exposures (smoking history, alcohol), clinical (gender, ethnicity, tumor grade, tumor stage, tumor cell purity) and molecular (*CDKN2A* silencing and mutation status of *TP53*, *CTNNB1*, *TERT*, and proliferative status) details for the 194 HCC patients. In the heatmap, red represents higher correlation and blue shows lower correlation between a human sample and a mouse tumor

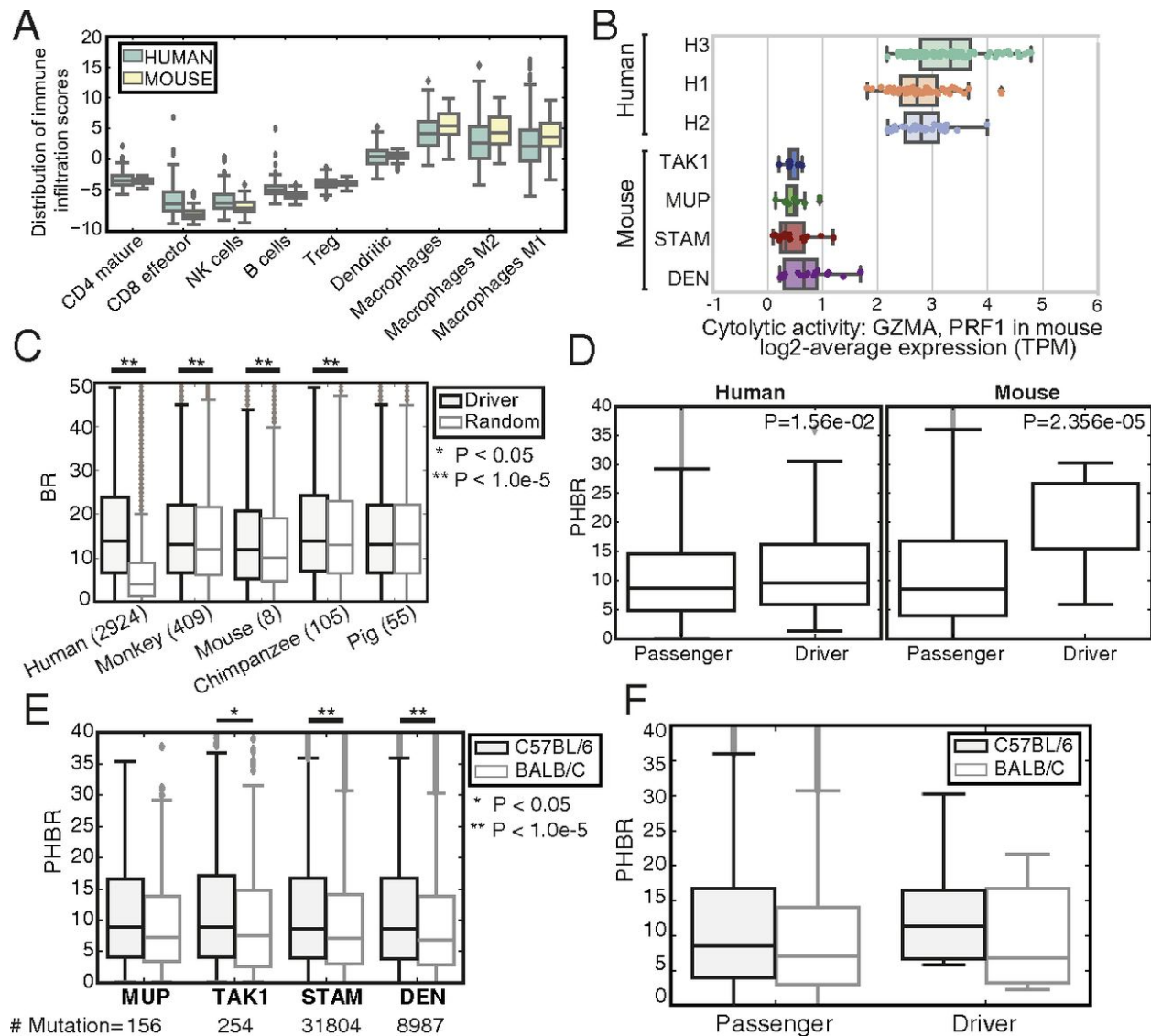


Figure 1.5. Immune activities for human and mouse.

(A) Immune infiltration scores for nine immune cell types in mouse and human tumors. (B) Cytolytic activities [log₂-average expression (transcripts per million)] of GZMA and PRF1 of the mouse models and human clusters. (C) BR MHC-I presentation score distribution of driver mutations (inferred from frequency in human tumors) compared with random mutations in different species with number of alleles shown in parentheses. (D) PHBR score distributions for HCC driver mutations versus passenger mutations observed for human ($n = 182$ for driver and $n = 28,216$ for passenger) (Left) and mouse ($n = 27$ driver and $n = 44,601$ passenger) (Right). (E) PHBR score distributions for observed mutations for MHC-I alleles of C57BL/6 and BALB mouse models. (F) PHBR distribution of observed HCC driver and random mutations for MHC-I alleles of C57BL/6 and BALB mouse models.

1.7 Tables

Table 1.1. Description of mouse models.

Mouse Model	Treatments/traits	HFD	Chronic damage	Cirrhosis	Time to Cancer (month)	Histology	#Samples
DEN	Carcinogen N-nitrosodiethylamine (DEN)	No	No	No	9	Histologically hard to classify	9 + 10
MUP	uPA under MUP promoter, expressed only in hepatocytes. Hepatocyte death	Yes	Yes	Yes (mild)	9	Adenoma like, low malignancy	9
STAM	Streptozocin treatment to WT mice, generate insulin deficiency	Yes	Yes	Yes (severe)	5	Malignant phenotype, with human like features	9 + 10
TAK1	Genetic deletion of TAK1 gene in liver. Hepatocyte death	No	Yes	Yes	9	Adenoma like, low malignancy	9

1.8 Supplemental Data, Tables and Figures

Table S1.2. Mutation presentation in mouse models and human.

(A) Significantly mutated genes. **(B)** Descriptive table of mouse mutations mentioned in this study, including known human HCC genes and significantly mutated cancer consensus genes. Information in the table includes mouse mutations, amino acid changes, orthologous human mutations, orthologous amino acid changes, and COSMIC mutation ID if reported. **(C)** DISCOVER results for pathway. **(D)** DISCOVER results genes with frequent mutations and recurrent CNVs.

Table S1.3. Covariate enrichment analysis for RNAseq.

(A) Summary of enrichment tests performed, **(B)** All enrichment results aggregated, and **(C-H)** enrichment results of defined clusters, exposures, clinical traits, molecular characteristics, purity level, and gender specific results.

Table S1.1A. Mouse sequencing characteristics.

Model	Animal	Sample	Total Reads	Mapped Reads	%Mapped	avg_Depth	%Bases >20x	Note
TAK1	TAK1-C	TAK1-C-spleen	176289533	175171398	0.974	116.135	0.989	TAK1-spleen
TAK1	TAK1-C	TAK1-C1	93436769	93212853	0.992	80.617	0.942	Batch1
TAK1	TAK1-C	TAK1-C2	54850952	54653660	0.990	86.554	0.956	Batch1
TAK1	TAK1-C	TAK1-C3	59121742	58977852	0.992	95.532	0.967	Batch1
TAK1	TAK1-D	TAK1-D1	190659285	189457446	0.992	70.848	0.909	Batch1
TAK1	TAK1-A	TAK1-A-spleen	174614850	173626049	0.994	82.819	0.952	TAK1-spleen
TAK1	TAK1-A	TAK1-A1	60925747	60749004	0.992	77.116	0.931	Batch1
TAK1	TAK1-A	TAK1-A2	77568257	77296960	0.990	100.636	0.973	Batch1
TAK1	TAK1-A	TAK1-A3	73050314	72854794	0.992	79.728	0.941	Batch1
TAK1	TAK1-B	TAK1-B1	195354151	194231726	0.991	95.077	0.967	Batch1
TAK1	TAK1-B	TAK1-B-spleen	170637577	169605142	0.971	129.918	0.992	TAK1-spleen
TAK1	TAK1-E	TAK1-E1	168038771	166780002	0.991	84.557	0.948	Batch1
MUP	MUP-D	MUP-D1	140669291	139801733	0.996	119.692	0.978	Batch1
MUP	MUP-E	MUP-E1	146818099	146587878	0.995	134.056	0.985	Batch1
MUP	MUP-C	MUP-C1	164024552	159967713	0.996	124.693	0.981	Batch1
MUP	MUP-C	MUP-C2	132635810	131960711	0.996	111.421	0.971	Batch1
MUP	MUP-A	MUP-A1	123865602	123087715	0.995	111.528	0.972	Batch1
MUP	MUP-A	MUP-A2	187380276	183373388	0.996	110.402	0.971	Batch1
MUP	MUP-B	MUP-B-spleen	164867335	163272584	0.998	124.827	0.982	MUP-spleen
MUP	MUP-B	MUP-B1	172507725	167182948	0.994	111.131	0.972	Batch1
MUP	MUP-B	MUP-B2	196938434	193742002	0.995	114.854	0.976	Batch1
MUP	MUP-B	MUP-B3	170143744	168370084	0.994	119.547	0.978	Batch1
STAM	STAM-A	STAM-A1	137131728	136092008	0.995	113.345	0.986	Batch1
STAM	STAM-A	STAM-A2	148991057	148350021	0.975	122.622	0.990	Validation
STAM	STAM-B	STAM-B1	125557591	124775525	0.994	110.647	0.981	Batch1
STAM	STAM-C	STAM-C1	119092979	118536771	0.979	133.132	0.993	Validation
STAM	STAM-C	STAM-C2	202797130	201605990	0.969	117.799	0.989	Validation
STAM	STAM-C	STAM-C3	167316740	154264739	0.984	137.952	0.993	Validation
STAM	STAM-C	STAM-C4	131355899	130693789	0.990	126.522	0.992	Validation
STAM	STAM-D	STAM-D-spleen	203087370	176973775	0.996	124.716	0.990	STAM-spleen
STAM	STAM-D	STAM-D1	187436708	164416671	0.992	117.040	0.986	Batch1
STAM	STAM-F	STAM-F2	182093924	150942584	0.994	113.363	0.983	Batch1
STAM	STAM-D	STAM-D3	144564309	143697996	0.995	107.746	0.983	Batch1
STAM	STAM-D	STAM-D4	64298630	64194955	0.994	150.615	0.993	Batch1
STAM	STAM-D	STAM-D2	87467227	87008171	0.990	125.208	0.990	Validation
STAM	STAM-E	STAM-E1	165614603	164683834	0.995	118.876	0.984	Batch1
STAM	STAM-E	STAM-E3	154231721	152976658	0.995	87.893	0.962	Batch1
STAM	STAM-E	STAM-E2	153491615	152388405	0.922	116.999	0.988	Validation
STAM	J759	STAM-F1	83041666	82849906	0.994	120.742	0.987	Batch1
STAM	J759	STAM-F5	183960506	183115434	0.871	119.568	0.989	Validation
STAM	J759	STAM-F3	83582926	83383361	0.877	113.898	0.990	Validation
STAM	J759	STAM-F4	73646323	73478896	0.829	102.184	0.984	Validation
DEN	DEN-F	DEN-F-spleen	152608004	152051871	0.998	90.027	0.952	DEN-spleen
DEN	DEN-A	DEN-A2	84149211	83279439	0.993	115.590	0.988	Validation
DEN	DEN-B	DEN-B1	175607623	171091540	0.994	124.165	0.989	Validation
DEN	DEN-B	DEN-B2	62800491	62290738	0.992	113.909	0.986	Validation
DEN	DEN-B	DEN-B3	71332005	70925107	0.995	135.244	0.992	Validation
DEN	DEN-C	DEN-C1	72046124	71501664	0.998	106.316	0.975	Batch1
DEN	DEN-C	DEN-C2	101739118	100769163	0.998	97.569	0.966	Batch1

Table S1.1A. Mouse sequencing characteristics (continued).

Model	Animal	Sample	Total Reads	Mapped Reads	%Mapped	avg_Depth	%Bases >20x	Note
DEN	DEN-C	DEN-C3	72024634	71484193	0.998	106.537	0.975	Batch1
DEN	DEN-D	DEN-D1	96457659	95595462	0.998	115.002	0.981	Batch1
DEN	DEN-D	DEN-D2	79779954	79022218	0.996	78.578	0.930	Batch1
DEN	DEN-D	DEN-D3	186409345	181032816	0.998	85.360	0.941	Batch1
DEN	DEN-E	DEN-E1	143556779	142964022	0.994	126.838	0.992	Validation
DEN	DEN-A	DEN-A1	95157717	94373195	0.994	134.426	0.992	Validation
DEN	DEN-E	DEN-E2	181151843	180317289	0.994	129.675	0.992	Validation
DEN	DEN-F	DEN-F1	128902915	128389897	0.997	85.307	0.945	Batch1
DEN	DEN-F	DEN-F2	128317933	127732252	0.997	100.069	0.969	Batch1
DEN	DEN-F	DEN-F3	126247483	125803110	0.997	97.641	0.964	Batch1
DEN	DEN-G	DEN-G1	126948447	126157530	0.994	137.424	0.992	Validation
DEN	DEN-G	DEN-G2	136569489	135905999	0.993	123.628	0.991	Validation
DEN	DEN-G	DEN-G3	74426657	73796388	0.994	126.243	0.992	Validation

Table S1.1B. Unsupervised mutational signature computations from mouse samples.

Extracted Mouse Signatures	Most Similar Human Mutational Signatures	Cosine similarity	Note
Signature A	68.3% Signature 25; 31.7% Signature 46	0.76	New signature; No match with human signatures
Signature B	100.0% Signature 44	0.93	Known sequencing artifact
Signature C	100.0% Signature 11	0.98	Dominated by alkylating agent (C>T)
Signature D	17.6% Signature 1; 53.6% Signature 5; 28.8% Signature 12	0.93	Signature 12 is specific to human HCC. Dominated by alkylating agent (C>T)
Signature E	33.7% Signature 19; 66.3% Signature 32	0.94	Signature appeared in human liver. Unknown aetiology

Table S1.4. Summary of findings.

Model	Mutational Signature	Significantly mutated genes	Pathway	Expression	Suitable studies
DEN	Low similarity with mutational signatures found in human HCC	Some similarity with significantly mutated genes as human HCC	High alteration in MAPK signaling and RTK signaling pathways	Low correlation with all HCC human groups	Should be avoided as a human HCC model
MUP	High similarity with mutational signatures found in human HCC	Similar SCNA genes as human HCC	Low alteration in RTK and SWI/SNF pathways	Some correlation with HCC groups H1	HCC initiation; microenvironmental determinants of mutational processes
STAM	Some similarity with mutational signatures found in human HCC	High similarity with significantly mutated genes as human HCC	High alteration in most human HCC pathways	High correlation with HCC groups H3 and low with H1	Most suitable for tumor biology and preclinical studies
TAK1	High similarity with mutational signatures found in human HCC	No similarity of somatic alterations affecting human HCC	No somatic alterations affecting human HCC	High correlation with HCC groups H1 and some with H2	HCC initiation; microenvironmental determinants of mutational processes

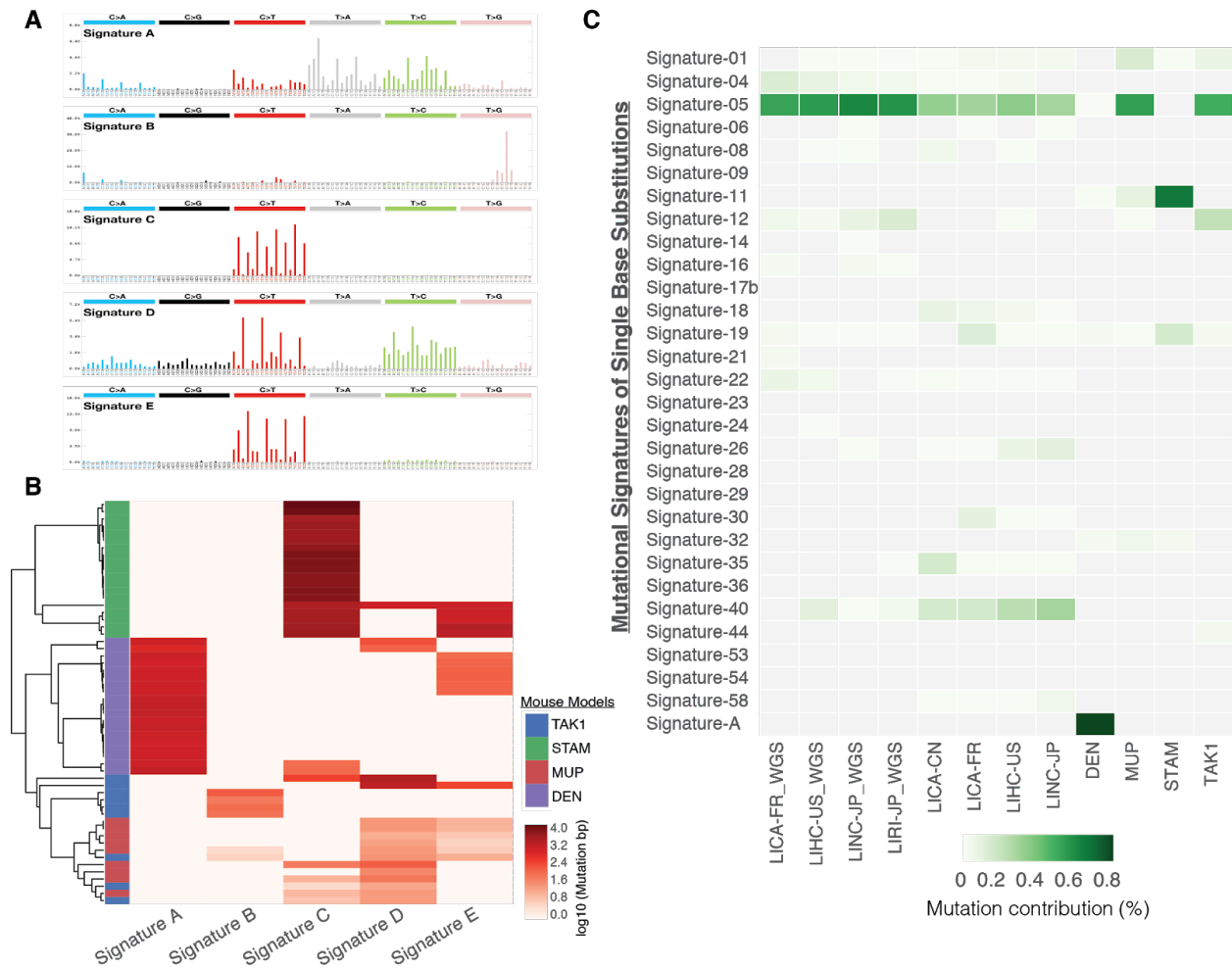


Figure S1.1. Mutational signatures of mouse models and human.

(A) Unsupervised mutational signatures decomposition from the overall mutation profiles of the four mouse models. **(B)** Unsupervised signatures for mouse samples. Colors represent the weight (\log_{10} number of somatic mutations attributed to signatures) of the signatures in each samples. **(C)** Mutation contribution (%) for all identified signatures are shown for each human and mouse cohort.

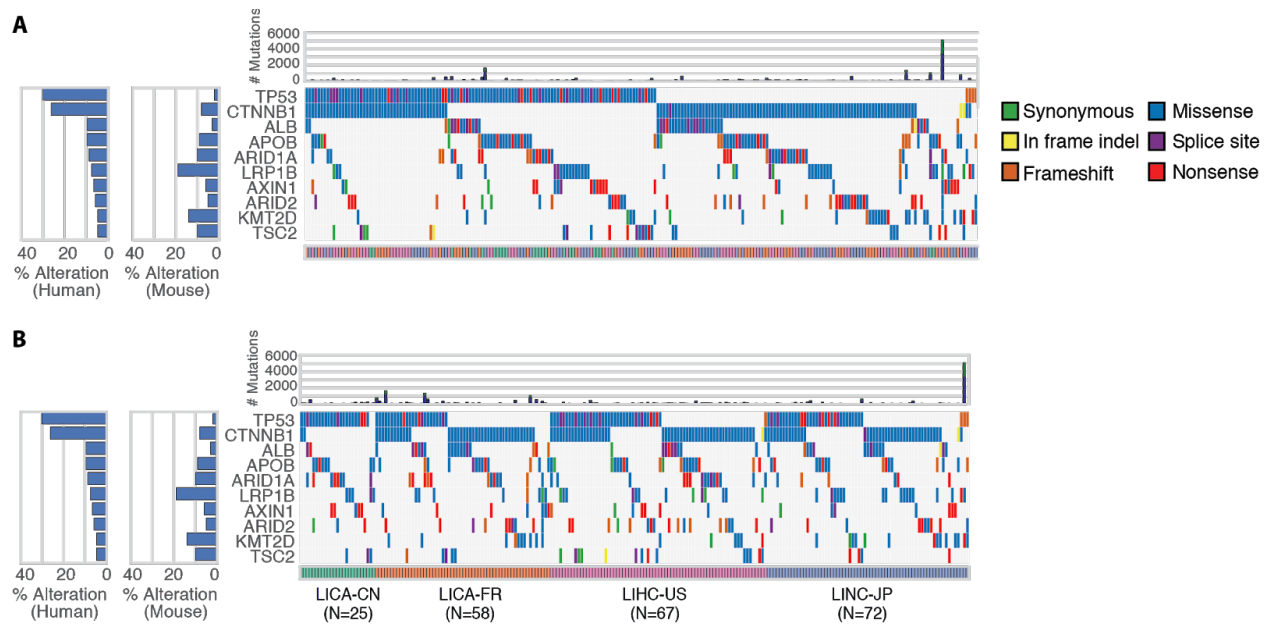


Figure S1.2. Top 10 significantly mutated genes in human from ICGC and TCGA cohorts , only patients with mutations in the 10 genes are included.

(A) Mutation profiles of the patients in the ICGC and TCGA cohorts, ordered by the mutation frequencies in the genes. **(B)** Mutation profiles of the patients in the ICGC and TCGA cohorts, ordered by first the specific cohorts then the mutation frequencies in the genes.

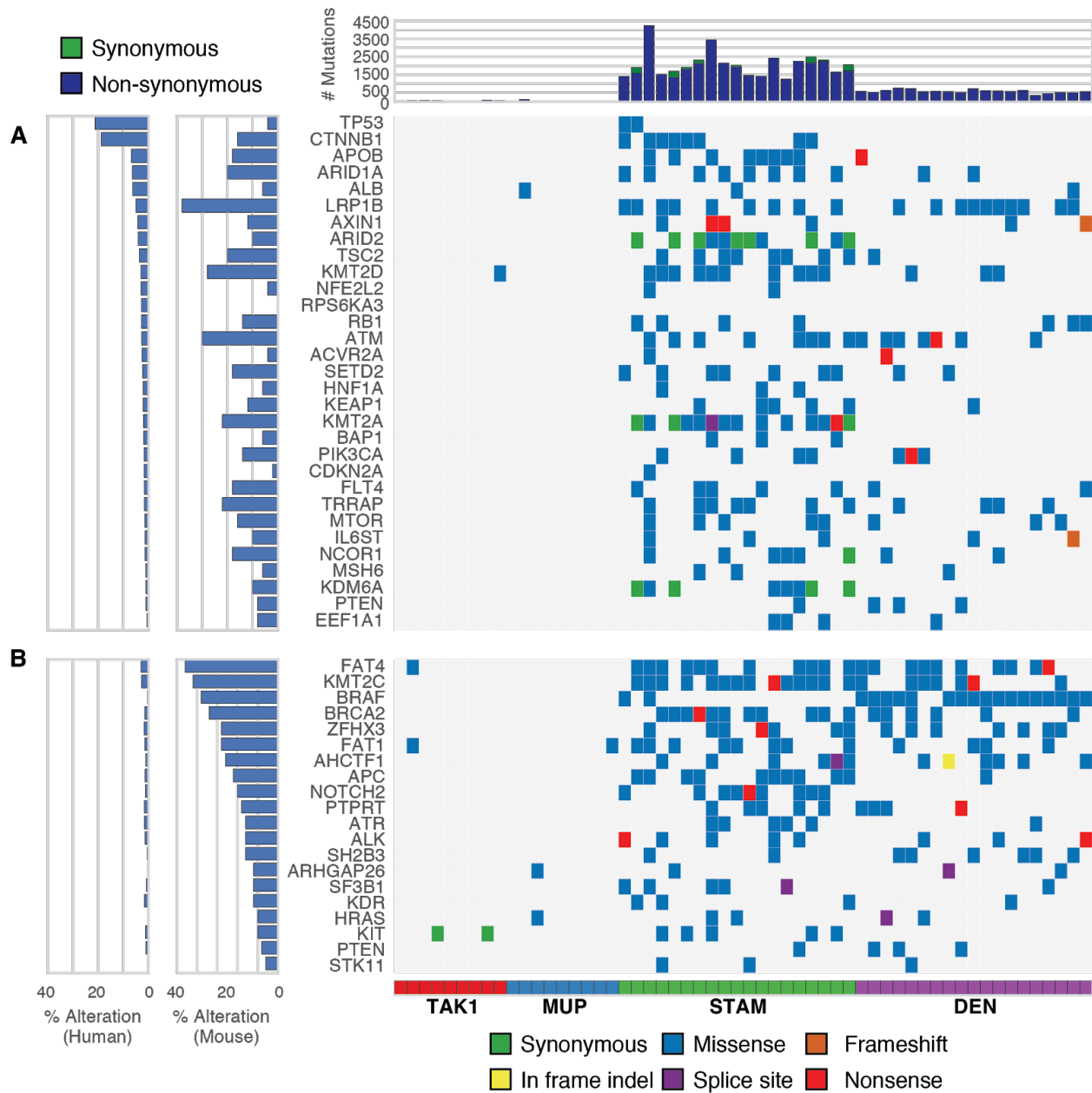


Figure S1.3. Unfiltered mutation profiles of the four HCC mouse models.

(A,B) The genomic landscape of the four HCC mouse models before driver mutations filtering in Figure 2. Mutational effects (synonymous, missense, etc.) are labelled accordingly for each gene and sample.

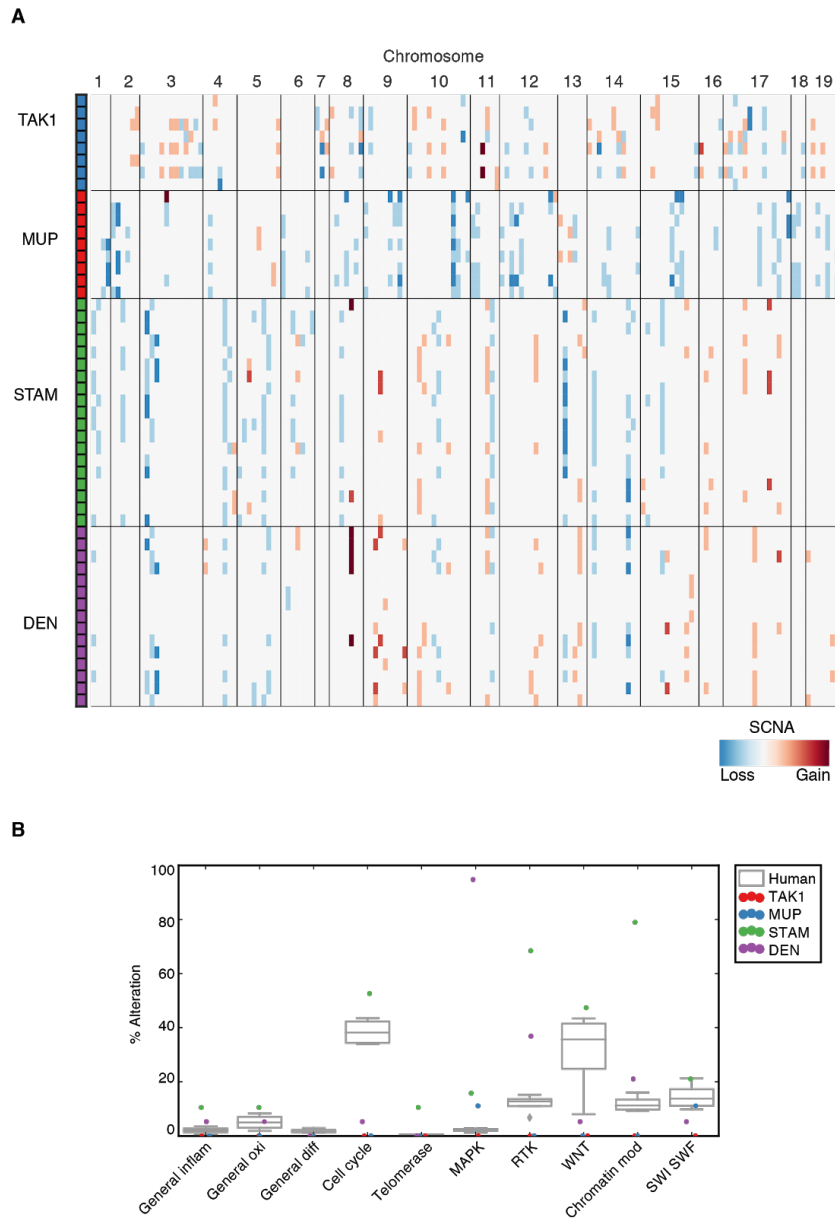


Figure S1.4. Evaluating the association between CNV and pathway in mouse samples. **(A)** CNV deletions across samples in each chromosomes and **(B)** Alteration frequencies for the complete pathway genes (boxplot represents the four human cohorts and dots represent each mouse cohort).

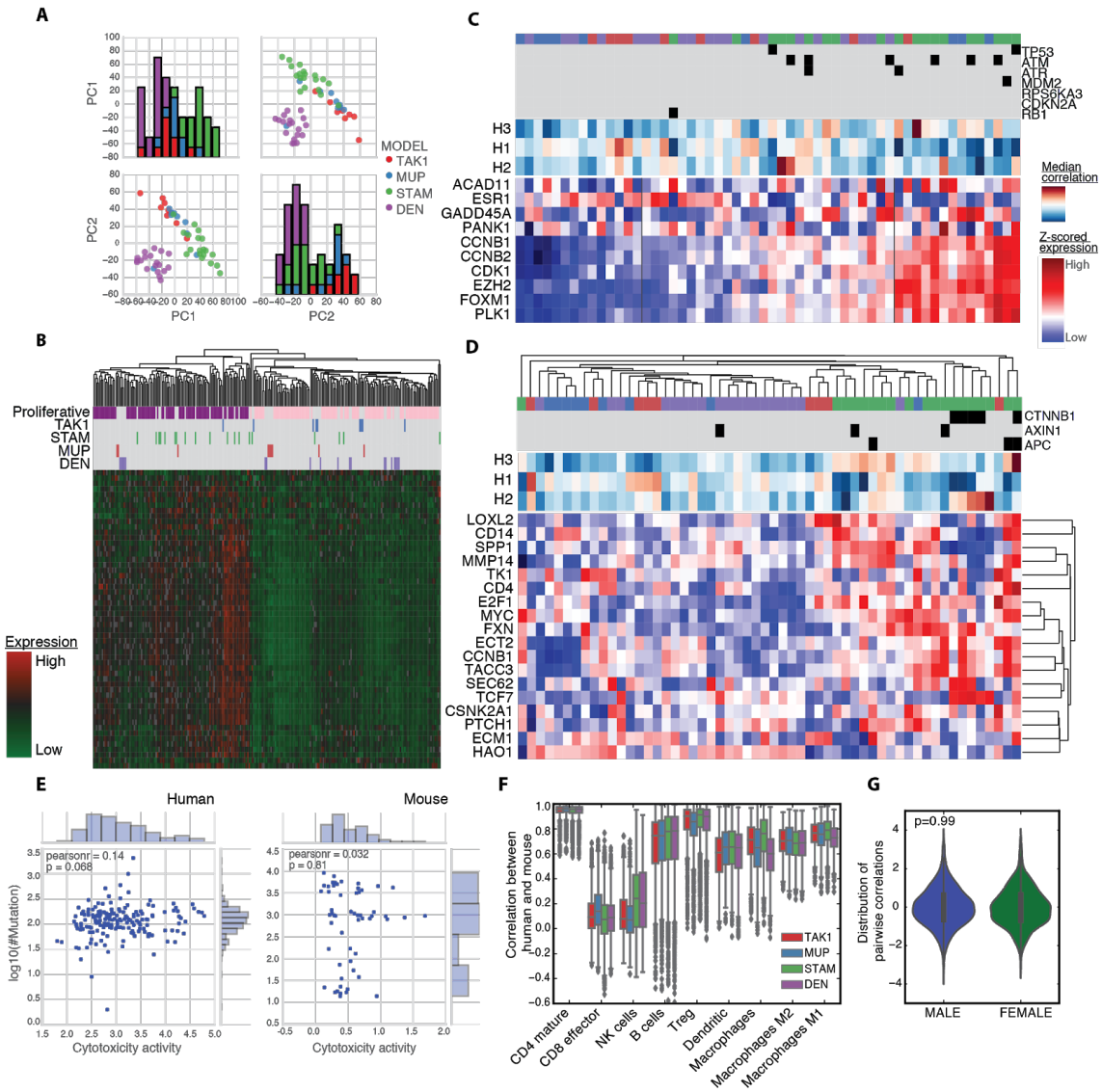


Figure S1.5. Details on expression and immune analysis.

(A) PCA clustering found using RNAseq gene expression from mouse. (B) Proliferative gene expression-based clustering across HCC patients and mouse tumors. Top panel shows the dendrogram of tumors grouped by their expression while the middle panel details human subtypes of high/low proliferation subgroups for the 194 HCC patients, and the model of the 56 mice. Red represents higher expression and green shows lower expression. (C) Overlapping genes from p53 response genes and top 800 DEGs for Figure 4. Top panel describes the mutation profile of cell cycle pathway genes. Middle panel shows the median correlations of the three human clusters. Bottom panel shows the z-scored gene expression of the p53 response genes. (D) Overlapping genes from CTNNB1 response genes and top 800 DEGs for Figure 4. Top panel describes the mutation profile of WNT signaling pathway genes. Middle panel shows the median correlations of the three human clusters. Bottom panel shows the z-scored gene expression of the CTNNB1 response genes. (E) Correlations between cytotoxicity scores and log10 (mutation burden) of the human patients and mouse samples. (F) Immune infiltration level between the four mouse models. (G) Distribution of pairwise correlations between male and female.

1.9 Author Contributions

M.D., M.K., H.C., and J.F.-B. designed research; M.D., H.C., and J.F.-B. performed research; M.D., R.M.P., B.Y.T., L.B.A., H.N., K.T., E.S., O.H., and S.S. contributed new reagents/analytic tools; M.D., R.M.P., B.Y.T., L.B.A., H.C., and J.F.-B. analyzed data; and M.D., L.B.A., M.K., H.C., and J.F.-B. wrote the paper.

1.10 Acknowledgements

The results published here are in part based upon data generated by The Cancer Genome Atlas managed by the NCI and NHGRI. Information about TCGA can be found at <http://cancergenome.nih.gov>.

The research was supported by the National Institutes of Health (NIH), National Library of Medicine Training Grant T15LM011271 (M.D.), NSF graduate fellowship #2015205295 (R.M.), CIFAR fellowship (H.C.), JSPS KAKENHI Grant Number JP15K21775 and the “Kibou Projects” Startup Support for Young Researchers in Immunology to K.T., Irvington Cancer Research Institute (S.S.), Prostate Cancer Foundation Young Investigator Award (S.S.), Southern California Research Center for ALPD and Cirrhosis grant (S.S.), M.K. holds the Ben and Wanda Hildyard Chair. NIH funding was as follows: CA127923, CA118165, AI043477, and U01AA022614 to M.K.; P42ES010337 to M.K.; DP5-OD017937 to H.C.; K99/R00CA191152 to J.F-B.

Chapter 1, in full, is a reformatted reprint of the material as it appears as "Integrative genomic analysis of mouse and human hepatocellular carcinoma" in *PNAS*, 2018 by Michelle Dow, Rachel M. Pyke, Brian Y Tsui, Ludmil B. Alexandrov, Hayato Nakagawa, Koji Taniguchi, Ekihiro Seki, Olivier Harismendy, Shabmam Shalpour, Michael Karin, Joan Font-Burgada, and Hannah Carter. The dissertation author was the primary investigator and author of this paper.

1.11 References

- Ahn, S.-M., Jang, S.J., Shim, J.H., Kim, D., Hong, S.-M., Sung, C.O., Baek, D., Haq, F., Ansari, A.A., Lee, S.Y., Chun, S.M., Choi, S., Choi, H.-J., Kim, J., Kim, S., Hwang, S., Lee, Y.-J., Lee, J., Jung, W., Jang, H.Y., Yang, E., Sung, W.-K., Lee, N.P., Mao, M., Lee, C., Zucman-Rossi, J., Yu, E., Lee, H.C., Kong, G. (2014). Genomic portrait of resectable hepatocellular carcinomas: implications of RB1 and FGF19 aberrations for patient stratification. *Hepatology* 60, 1972–1982.
- Alexandrov, L.B., Jones, P.H., Wedge, D.C., Sale, J.E., Campbell, P.J., Nik-Zainal, S., and Stratton, M.R. (2015). Clock-like mutational processes in human somatic cells. *Nat. Genet.* 47, 1402–1407.
- Alexandrov, L.B., Ju, Y.S., Haase, K., Van Loo, P., Martincorena, I., Nik-Zainal, S., Totoki, Y., Fujimoto, A., Nakagawa, H., Shibata, T., Campbell, P.J., Vineis, P., Phillips, D.H., Stratton, M.R. (2016). Mutational signatures associated with tobacco smoking in human cancer. *Science* 354, 618–622.
- Andrisani, O.M., Studach, L., and Merle, P. (2011). Gene signatures in hepatocellular carcinoma (HCC). *Semin. Cancer Biol.* 21, 4–9.
- Ascha, M.S., Hanouneh, I.A., Lopez, R., Tamimi, T.A.-R., Feldstein, A.F., and Zein, N.N. (2010a). The incidence and risk factors of hepatocellular carcinoma in patients with nonalcoholic steatohepatitis. *Hepatology* 51, 1972–1978.
- Ascha, M.S., Hanouneh, I.A., Lopez, R., Tamimi, T.A.-R., Feldstein, A.F., and Zein, N.N. (2010b). The incidence and risk factors of hepatocellular carcinoma in patients with nonalcoholic steatohepatitis. *Hepatology* 51, 1972–1978.
- Bakiri, L., and Wagner, E.F. (2013). Mouse models for liver cancer. *Mol. Oncol.* 7, 206–223.
- Blokzijl, F., de Ligt, J., Jager, M., Sasselli, V., Roerink, S., Sasaki, N., Huch, M., Boymans, S., Kuijk, E., Prins, P., Nijman, I.J., Martincorena, I., Mokry, M., Wiegerinck, C.L., Middendorp, S., Sato, T., Schwank, G., Nieuwenhuis, E.S., Verstegen, M.M.A., van der Laan, L.J.W., de Jonge, J., Ijzermans, J.N.M., Vries, R.G., van de Wetering, M., Stratton, M.R., Cleves, H., Cuppen, E., van Boxtel, R. (2016). Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* 538, 260–264.
- Bosch, F.X., Ribes, J., Cléries, R., and Díaz, M. (2005). Epidemiology of hepatocellular carcinoma. *Clin. Liver Dis.* 9, 191–211, v.
- Bruix, J., Qin, S., Merle, P., Granito, A., Huang, Y.-H., Bodoky, G., Pracht, M., Yokosuka, O., Rosmorduc, O., Breder, V., Gerolami, R., Masi, G., Ross, P.J., Song, T., Bronowicki, J.-P., Ollivier-Hourmand, I., Kudo, M., Cheng, A.-L., Llovet, J.M., Finn, R.S., LeBerre, M.-A, Baumhauer, A., Meinhardt, G., Han, G., RESORCE Investigators. (2017). Regorafenib for patients with hepatocellular carcinoma who progressed on sorafenib treatment (RESORCE): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet* 389, 56–66.

Cancer Genome Atlas Research Network. Electronic address: wheeler@bcm.edu, and Cancer Genome Atlas Research Network (2017). Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* 169, 1327–1341.e23.

Canisius, S., Martens, J.W.M., and Wessels, L.F.A. (2016). A novel independence test for somatic alterations in cancer shows that biology drives mutual exclusivity but chance explains most co-occurrence. *Genome Biol.* 17, 261.

Carter, H., Samayoa, J., Hruban, R.H., and Karchin, R. (2010). Prioritization of driver mutations in pancreatic cancer using cancer-specific high-throughput annotation of somatic mutations (CHASM). *Cancer Biol. Ther.* 10, 582–587.

Chen, X., and Calvisi, D.F. (2014). Hydrodynamic transfection for generation of novel mouse models for liver cancer research. *Am. J. Pathol.* 184, 912–923.

Ciriello, G., Cerami, E., Sander, C., and Schultz, N. (2012). Mutual exclusivity analysis identifies oncogenic network modules. *Genome Res.* 22, 398–406.

Connor, F., Rayner, T.F., Aitken, S.J., Feig, C., Lukk, M., Santoyo-Lopez, J., and Odom, D.T. (2018). Mutational landscape of a chemically-induced mouse model of liver cancer. *J. Hepatol.*

Davoli, T., Uno, H., Wooten, E.C., and Elledge, S.J. (2017). Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science* 355, eaaf8399.

El-Khoueiry, A.B., Sangro, B., Yau, T., Crocenzi, T.S., Kudo, M., Hsu, C., Kim, T.-Y., Choo, S.-P., Trojan, J., Welling, T.H., Rd, Meyer, T., Kang, Y.-K., Yeo, W., Chopra, A., Anderson, J., Cruz, C.D., Lang, L., Nelly, J., Tang, H., Dastani, H.B., Melero, I. (2017). Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): an open-label, non-comparative, phase 1/2 dose escalation and expansion trial. *Lancet* 389, 2492–2502.

El-Serag, H.B., and Davila, J.A. (2011). Surveillance for hepatocellular carcinoma: in whom and how? *Therap. Adv. Gastroenterol.* 4, 5–10.

El-Serag, H.B., and Rudolph, K.L. (2007). Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology* 132, 2557–2576.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F. (2015a). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* 136, E359–E386.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F. (2015b). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* 136, E359–E386.

Forbes, S.A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., Cole, C.G., Ward, S., Dawson, E., Ponting, L., Stefancsik, R., Harsha, B., Kok, C.Y., Jia, M., Jubb, H., Sondka, Z., Thompson, S., De, T., Campbell, P.J. (2017). COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* 45, D777–D783.

Fujii, M., Shibazaki, Y., Wakamatsu, K., Honda, Y., Kawauchi, Y., Suzuki, K., Arumugam, S., Watanabe, K., Ichida, T., Asakura, H., Yoneyama, H. (2013). A murine model for non-alcoholic steatohepatitis showing evidence of association between diabetes and hepatocellular carcinoma. *Med. Mol. Morphol.* *46*, 141–152.

Fujimoto, A., Furuta, M., Totoki, Y., Tsunoda, T., Kato, M., Shiraishi, Y., Tanaka, H., Taniguchi, H., Kawakami, Y., Ueno, M., Gotoh, K., Ariizumi, S.-I., Wardell, C.P., Hayami, S., Nakamura, T., Aikata, H., Arihiro, K., Borojevich, K.A., Abe, T., Nakano, K., Maejima, K., Sasaki-Oku, A., Ohsawa, A., Shibuya, T., Nakamura, H., Hama, N., Hosoda, F., Arai, Y., Ohashi, S., Urushidate, T., Nagae, G., Yamamoto, S., Ueda, H., Tatsuno, K., OOjima, H., Hiraoka, N., Okusaka, T., Kubo, M., Marubashi, S., Yamada, T., Hirano, S., Yamamoto, M., Ohdan, H., Shimada, K., Ishikawa, O., Yamaue, H., Chayama, K., Miyano, S., Aburatani, H., Shibata, T., Nakagawa, H. (2016). Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat. Genet.* *48*, 500–509.

Greenman, C., Wooster, R., Futreal, P.A., Stratton, M.R., and Easton, D.F. (2006). Statistical analysis of pathogenicity of somatic mutations in cancer. *Genetics* *173*, 2187–2198.

Guichard, C., Amaddeo, G., Imbeaud, S., Ladeiro, Y., Pelletier, L., Maad, I.B., Calderaro, J., Bioulac-Sage, P., Letexier, M., Degos, F., Clément, B., Balabaud, C., Chevet, E., Laurent, A., Couchy, G., Letouzé, E., Calvo, F., Zucman-Ross, J. (2012). Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat. Genet.* *44*, 694–698.

He, G., Dhar, D., Nakagawa, H., Font-Burgada, J., Ogata, H., Jiang, Y., Shalpour, S., Seki, E., Yost, S.E., Jepsen, K., Frazer, K.A., Harismendy, O., Hatziapostolou, M., Iliopoulos, D., Suetsugu, A., Hoffman, R.M., Tateishi, R., Koike, K., Karin, M. (2013a). Identification of liver cancer progenitors whose malignant progression depends on autocrine IL-6 signaling. *Cell* *155*, 384–396.

Heindryckx, F., Colle, I., and Van Vlierberghe, H. (2009b). Experimental mouse models for hepatocellular carcinoma research. *Int. J. Exp. Pathol.* *90*, 367–386.

Holmberg, S.D., Spradling, P.R., Moorman, A.C., and Denniston, M.M. (2013). Hepatitis C in the United States. *N. Engl. J. Med.* *368*, 1859–1861.

Hoshida, Y., Nijman, S.M.B., Kobayashi, M., Chan, J.A., Brunet, J.-P., Chiang, D.Y., Villanueva, A., Newell, P., Ikeda, K., Hashimoto, M., Watanabe, G., Gabriel, S., Friedman, S. L., Kumada, H., Llovet, J.M., & Golub, T.R. (2009). Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. *Cancer Res.* *69*, 7385–7392.

Inokuchi, S., Aoyama, T., Miura, K., Osterreicher, C.H., Kodama, Y., Miyai, K., Akira, S., Brenner, D.A., and Seki, E. (2010). Disruption of TAK1 in hepatocytes causes hepatic injury, inflammation, fibrosis, and carcinogenesis. *Proc. Natl. Acad. Sci. U. S. A.* *107*, 844–849.

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., Kiezun, A., Hammerman, P.S., McKenna, A., Drier, Y., Zou, L., Ramos, A. H., Pugh, T. J., Stransky, N., Helman, E., Getz, G.(2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* *499*, 214–218.

- Lee, J.-S., Chu, I.-S., Mikaelyan, A., Calvisi, D.F., Heo, J., Reddy, J.K., and Thorgeirsson, S.S. (2004). Application of comparative functional genomics to identify best-fit mouse models to study human cancer. *Nat. Genet.* 36, 1306–1311.
- Letouzé, E., Shinde, J., Renault, V., Couchy, G., Blanc, J.-F., Tubacher, E., Bayard, Q., Bacq, D., Meyer, V., Semhoun, J., Bioulac-Sage, P., Prévôt, S., Azoulay, D., Paradis, V., Imbeaud, S., Deleuze, J.-F., and Zucman-Rossi, J. (2017). Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat. Commun.* 8, 1315.
- Llovet, J.M., Ricci, S., Mazzaferro, V., Hilgard, P., Gane, E., Blanc, J.-F., de Oliveira, A.C., Santoro, A., Raoul, J.-L., Forner, A., Schwartz, M., Porta, C., Zeuzem, S., Bolondi, L., Greten, T. F., Galle, P. R., Seitz, J.-F., Borbath, I., Häussinger, D., SHARP Investigators Study Group. (2008). Sorafenib in advanced hepatocellular carcinoma. *N. Engl. J. Med.* 359, 378–390.
- Llovet, J.M., Zucman-Rossi, J., Pikarsky, E., Sangro, B., Schwartz, M., Sherman, M., and Gores, G. (2016). Hepatocellular carcinoma. *Nat Rev Dis Primers* 2, 16018.
- Marty, R., Kaabinejadian, S., Rossell, D., Slifker, M.J., van de Haar, J., Engin, H.B., de Prisco, N., Ideker, T., Hildebrand, W.H., Font-Burgada, J., Carter, H. (2017). MHC-I Genotype Restricts the Oncogenic Mutational Landscape. *Cell* 171, 1272–1283.e15.
- Mittal, D., Gubin, M.M., Schreiber, R.D., and Smyth, M.J. (2014). New insights into cancer immunoediting and its three component phases--elimination, equilibrium and escape. *Curr. Opin. Immunol.* 27, 16–25.
- Nakagawa, H. (2015). Recent advances in mouse models of obesity- and nonalcoholic steatohepatitis-associated hepatocarcinogenesis. *World J. Hepatol.* 7, 2110–2118.
- Nakagawa, H., Umemura, A., Taniguchi, K., Font-Burgada, J., Dhar, D., Ogata, H., Zhong, Z., Valasek, M.A., Seki, E., Hidalgo, J., Koike, K., Kaufman, R. J., and Karin, M. (2014). ER stress cooperates with hypernutrition to trigger TNF-dependent spontaneous HCC development. *Cancer Cell* 26, 331–343.
- Ng, A.W.T., Poon, S.L., Huang, M.N., Lim, J.Q., Boot, A., Yu, W., Suzuki, Y., Thangaraju, S., Ng, C.C.Y., Tan, P., Pang, S.-T., Huang, H.-Y., Yu, M.-C., Lee, P.-H., Hsieh, S.-Y., Chang, A. Y., Teh, B. T., and Rozen, S. G. (2017). Aristolochic acids and their derivatives are widely implicated in liver cancers in Taiwan and throughout Asia. *Sci. Transl. Med.* 9.
- Nielsen, M., and Andreatta, M. (2016). NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med.* 8, 33.
- Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 160, 48–61.
- Santos, N.P., Colaço, A.A., and Oliveira, P.A. (2017). Animal models as a tool in hepatocellular carcinoma research: A Review. *Tumour Biol.* 39, 1010428317695923.

- Schulze, K., Imbeaud, S., Letouzé, E., Alexandrov, L.B., Calderaro, J., Rebouissou, S., Couchy, G., Meiller, C., Shinde, J., Soysouvanh, F., Calatayud, A.-L., Pinyol, R., Pelletier, L., Balabaud, C., Laurent, A., Blanc, J.-F., Mazzaferro, V., Calvo, F., Villanueva, A., and Zucman-Rossi, J. (2015). Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat. Genet.* *47*, 505–511.
- Shalapour, S., Lin, X.-J., Bastian, I.N., Brain, J., Burt, A.D., Aksenov, A.A., Vrbanac, A.F., Li, W., Perkins, A., Matsutani, T., Zhong, Z., Dhar, D., Navas-Molina, J. A., Xu, J., Loomba, R., Downes, M., Yu, R. T., Evans, R. M., Dorrestein, P. C., and Karin, M. (2017). Inflammation-induced IgA⁺ cells dismantle anti-liver cancer immunity. *Nature* *551*, 340–345.
- Szikriszt, B., Póti, Á., Pipek, O., Krzystanek, M., Kanu, N., Molnár, J., Ribli, D., Szeltner, Z., Tusnády, G.E., Csabai, I., Szallasi, Z., Swanton, C., and Szüts, D. (2016). A comprehensive survey of the mutagenic impact of common cancer cytotoxics. *Genome Biol.* *17*, 99.
- Totoki, Y., Tatsuno, K., Covington, K.R., Ueda, H., Creighton, C.J., Kato, M., Tsuji, S., Donehower, L.A., Slagle, B.L., Nakamura, H., Yamamoto, S., Shinbrot, E., Hama, N., Lehmkuhl, M., Hosoda, F., Arai, Y., Walker, K., Dahdouli, M., Gotoh, K., and Shibata, T. (2014). Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. *Nat. Genet.* *46*, 1267–1273.
- Vandin, F., Upfal, E., and Raphael, B.J. (2011). De novo discovery of mutated driver pathways in cancer. *Genome Res.* *22*, 375–385.
- Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Jr, and Kinzler, K.W. (2013). Cancer genome landscapes. *Science* *339*, 1546–1558.
- Wilhelm, S.M., Adnane, L., Newell, P., Villanueva, A., Llovet, J.M., and Lynch, M. (2008). Preclinical overview of sorafenib, a multikinase inhibitor that targets both Raf and VEGF and PDGF receptor tyrosine kinase signaling. *Mol. Cancer Ther.* *7*, 3129–3140.
- Zhang, J., Baran, J., Cros, A., Guberman, J.M., Haider, S., Hsu, J., Liang, Y., Rivkin, E., Wang, J., Whitty, B., Wong-Erasmus, M., Yao, L., & Kasprzyk, A. (2011). International Cancer Genome Consortium Data Portal--a one-stop shop for cancer genomics data. *Database* *2011*, bar026.

CHAPTER 2: p300/CBP-mediated activation of MHC-I machinery and IFN γ signaling controls anti-tumor immunity

2.1 Abstract

Certain chemotherapeutic drugs potentiate immune checkpoint inhibition with PD-(L)1 antibodies, but how this synergism is achieved is poorly understood. Many cancers evade immune rejection by suppressing MHC-I antigen processing and presentation (AgPP). We show that two distinct classes of DNA damaging drugs, platinoids (e.g. oxaliplatin) and topoisomerase inhibitors (e.g. mitoxantrone), induce NF- κ B activation, nuclear translocation of the histone and lysine acetyl transferases p300/CBP, interferon (IFN) regulatory factors and IFN γ receptor 2 (IFN γ R2). This results in transcriptional activation of loci coding for the MHC-I transactivator NLRC5 and components of the MHC-I AgPP machinery. Accordingly, p300 inactivation prevents drug-induced stimulation of MHC-I antigen presentation leading to impaired recognition of cancer cells by effector CD8⁺ T cells, whose tumor rejecting activity is IFN γ R2-dependent. Interestingly, *EP300* loss especially in HLA-expressing human cancers correlates with reduced expression of MHC-I/HLA related genes and/or changes in neoantigen amounts and expression, suggesting that p300 is an immune-directed oncosuppressor.

2.2 Introduction

Immune checkpoint inhibitors (ICI), such as antibodies that block negative regulators of T-cell activation, have radically transformed cancer treatment (Eggermont et al., 2018; Gandhi et al., 2018; Schachter et al., 2017). However, even in metastatic melanoma and non-small cell lung cancer (NSCLC), malignancies that respond well to ICI, response rates rarely exceed 40% (Conforti et al., 2018). Furthermore, many common malignancies, including prostate cancer (PCa) and pancreatic ductal adenocarcinoma (PDAC), are ICI refractory (Guo et al., 2017; Hossain et al., 2018; Isaacsson Velho and Antonarakis, 2018) but causes of treatment failure are largely unknown. Early work correlated ICI responsiveness with mutational burden, which dictates production of neoantigens that are recognized by CD8⁺ cytotoxic T lymphocytes (CTL) (Chabanon et al., 2016; Snyder et al., 2014). Although this correlation may hold for a single tumor type, several malignancies initially predicted to be non-responsive due to low mutational burdens, e.g. renal cell carcinoma (RCC) and hepatocellular carcinoma (HCC), were found to be nearly as responsive to PD-1 inhibitors as highly mutated NSCLC (El-Khoueiry et al., 2017; Motzer et al., 2018). Recent clinical trials show that ICI responsiveness is significantly augmented by combining PD-1 signaling inhibitors with platinoid chemotherapeutics (Gandhi et al., 2018; Langer et al., 2016; Paz-Ares et al., 2018). Such results have led to approval of ICI + platinoid combination in NSCLC, but the basis for this synergism remains unknown.

Having found that PD-L1 blockade is highly effective in a mouse model of non-alcoholic steatohepatitis (NASH)-driven HCC (Shalapour et al., 2017), we searched for factors that influence the efficacy of this response. Anti-PD-1/PD-L1 drugs function by inducing reinvigoration of exhausted or dysfunctional CD8⁺ T cells (Keir et al., 2008). Effector CD8⁺ T cells can only recognize and kill tumors that present antigens via major histocompatibility complex (MHC) class

I molecules (Tschärke et al., 2015; Wang et al., 2009). MHC I antigens originate from endogenously synthesized proteins (self or viral) through a process shared by all nucleated mammalian cells, or exogenous proteins that are engulfed by antigen-presenting cells and delivered via cross presentation (van Montfort et al., 2014; Cresswell et al., 2005). Antigen processing and loading of the resulting peptides onto MHC I: b₂-microglobulin (b2m) heterodimers requires a complex and intricate molecular machinery that includes immunoproteasomes, which differ from conventional proteasomes by three alternative subunits (Rock et al., 2004), peptide transporters, peptide loaders, peptide trimmers, and vesicles that transport peptide-loaded MHC I molecules to the cell surface (Jongsma et al., 2017). Expression of most of these molecules is induced by interferon (IFN) γ through a poorly understood pathway (Zhou, 2009) that depends on NLRC5 or CITA, a transcriptional regulator that belongs to the Nod-like receptor family (Kobayashi and Elsen, 2012). NLRC5 loss-of-function (LOF) mutations or epigenetic modifications that reduce its expression, such as promoter methylation, are common immune evasion mechanisms (Yoshihama et al., 2016). Correspondingly, many cancers under express NLRC5 and MHC I (Kobayashi and Elsen, 2012). LOF mutations in the IFN γ signaling pathway also confer ICI resistance (Sharma et al., 2017).

We found that mouse models of PCa that are ICI refractory become responsive to PD-L1 blockade or ablation after co-treatment with low doses of the platinum drug oxaliplatin (Oxali) (Shalpour et al., 2015). The Oxali dose used in our experiments elicited minimal cell death *in vitro* and did not cause tumor regression *in vivo*, unless tumor-bearing mice were depleted of PD-L1- and IL-10-expressing IgA⁺ immunosuppressive plasmacytes (ISP). Without low-dose Oxali, the effect of ISP depletion on PCa growth was negligible and did not differ from the effect of PD-L1/PD-L1 inhibitors. In the past, Oxali was studied as a prototype for anticancer drugs that can

induce immunogenic cell death (ICD) and T-cell priming (Galluzzi et al., 2015). However, the exact mechanism of ICD induction is poorly defined, and it is not clear whether Oxali and similar drugs exert their immunogenic activity solely via ICD. Other studies have shown that Oxali and other platinoids induce the integrated stress response (ISR) (Bruno et al., 2017; Kepp et al., 2015). However, the mechanism of ISR activation by platinoids and its relevance for their immunostimulatory activity are unknown. Here we show that low dose Oxali, and to a lesser extent other platinoids, possess a unique ability to activate the transcriptional program that controls MHCI antigen processing and presentation. Activation of the MHCI program correlates with induction of histone acetylation and enhanced chromatin accessibility. Oxali treatment also results in induction of IFN γ R2, which potentiates the response of MHCI-expressing cancer cells to IFN γ produced by reinvigorated CD8⁺ T cells. These results provide a potential explanation for ICI-platinoid synergy in human NSCLC and other solid malignancies.

2.3 Results

Platinoids upregulate MHCI antigen processing and presentation. Like human PCa (Bilusic et al., 2017), mouse models of PCa are refractory to anti-PD-L1 monotherapy or ISP ablation (Shalpour et al., 2015). By contrast, mouse NASH-driven HCC is highly responsive to either treatment, undergoing near-complete regression (Shalpour et al., 2017). To understand the basis for these marked differences in ICI responsiveness, which also apply to human HCC and PCa (El-Khoueiry et al., 2017; Goswami et al., 2016), we examined both cancer types, mouse and human, for expression of MHCI (HLA-ABC) molecules. HLA-ABC expression was considerably lower in PCa than HCC (**Figures 2.1A, S2.1A and S2.1B**) and as previously described (Ylitalo et al., 2017). Analysis of TCGA data from HCC and PCa patients confirmed higher expression of genes related to MHC machinery in HCC tumor samples compared to their normal tissue, however, no clear relationship between these genes was found in samples from PCa patients (**Figures S2.1C**). PCa malignant progression was associated with a further decline in HLA-ABC expression (**Figures 2.1B and S2.1D**). Of note, PCa HLA-ABC expression positively correlated with expression of TAP1 and ERAP1, molecules needed for MHCI antigen presentation, which also declined during malignant progression (**Figures S2.1D and S2.1E**). Of note, increased expression of HLA-C, TAP1, and the immunoproteasome component PSMB9 correlated with a significantly higher response of NSCLC patients to anti-PD-L1 (Atezolizumab), carboplatin (Carbo), and Pemetrexed (Pem) combination therapy (**Figures 2.1C and 2.1D**). Patients that responded to the above combination also showed high expression of genes involved in CTL effector function, as well as high CD8⁺ T cell infiltration and PD-L1 expression (**Figures 2.1D and S2.1F**).

In mice, the poor response of PCa to anti-PD-L1 therapy is strongly potentiated by co-treatment with low-dose Oxali (Shalpour et al., 2015). To investigate how low-dose Oxali affects

the PCa transcriptome, we subjected mouse PCa cells (Myc-CaP and TRAMP-C2) incubated with 2 mM of either Oxali or Cisplatin (Cis), a platinoid that, unlike Oxali, minimally enhances the response to PD-L1 blockade or ISP ablation, (Pfirschke et al., 2016; Shalapour et al., 2015) to RNA sequencing (RNA-seq). Although both drugs strongly induced a gene signature characteristic of NF- κ B-dependent TNF signaling, Oxali was a considerably stronger inducer of IFN γ responsive gene set (**Figure 2.1E**). Similar observations were made in TRAMP-C2 (TRC2) cells, in which both drugs induced a gene signature associated with Kras signaling, while the IFN γ response was preferentially induced by Oxali (**Figure S2.1G**). The distinction between the transcriptional response to Oxali vs. Cis became clearer using “heat map” comparison of Myc-CaP cells treated with the two drugs (**Figures 2.1F and S2.1H**). Remarkably, low-dose Oxali induced numerous genes whose products encompass the MHCI antigen processing and presentation machinery. The very same genes were constitutively upregulated in NASH-driven mouse HCC (**Figure 2.1G**), which is highly responsive to anti-PD-L1 monotherapy (Shalapour et al., 2017). In addition to *Ifnar2*, *Ifngr2*, *Myd88*, *Nfkb1*, *Nfkb2*, *Ikkb*, *Stat1*, *Socs1*, *Irf1*, *Irf2*, and *Ripk2*, whose products are involved in innate immunity, NF- κ B and cytokine signaling, Oxali-induced genes code for peptide transporters (*Tap1* and *Tap2*), immunoproteasome subunits [*Psmb10*, *Psmb9* (*Lmp2*), and *Psmb8* (*Lmp7*)], TAP binding protein (*Tapasin*, *Tapbp*), b2m, and other MHCI antigen processing and presentation components. All of these genes were upregulated by high fat diet (HFD) in the *MUP-uPA* model of NASH-driven HCC (**Figures 2.1G and S2.1I**).

Platinoid-induced expression of MHCI components is potentiated by IFN γ . Platinoid drugs exhibit some cancer type specificity (Dilruba and Kalayda, 2016; McWhinney et al., 2009; Puisset et al., 2014). We therefore compared the induction of MHCI components by the three most commonly used platinoids, Oxali, Cis, and Carbo, in the PCa cell lines Myc-CaP and TRC2

(**Figures 2.2 and S2.2**) and in other cancer types that differ in basal MHCI expression. The latter included mouse melanoma YUMM and B16 cell lines, mouse colon cancer MC-38 cells and human cancer cell lines derived from PCa, PDAC, NSCLC, and melanoma (**Figures S2.2 and S2.3**). Melanoma and NSCLC were chosen because of high ICI responsiveness, whereas PCa, PDAC, and colon cancer were chosen based on ICI resistance.

Among the three platinoids that were tested at 2 mM (Oxali and Cis) or 4 mM (Carbo) in Myc-CaP cells, Oxali most efficiently induced *Nlrc5*, *Psmb9*, *Tap1*, *Ifngr2*, *Tapasin*, and *Erap1* mRNAs (**Figures 2.2A-F**). Low-dose Oxali induced PSMB8 and PSMB9 protein expression in both Myc-CaP and TRC2 cells (**Figure S2.2A**) and stimulated immunoproteasome activity measured by hydrolysis of an LMP7 (PSMB8)-specific substrate (**Figure 2.2G**). Induction of MHCI antigen presentation and processing genes by Oxali was strongly potentiated by exogenous IFNg, not only in Myc-CaP but also in TRC2 cells (**Figures 2.2A-F and S2.2B**). IFNg also enhanced the response to Cis and Carbo, but the effect was not as strong as that of Oxali + IFNg (**Figures 2.2A-F**). One exception, however, was the *Ifngr2* gene, whose expression was induced by Oxali and to a lesser extent Cis, but not by IFNg (**Figure 2.2D**). For most genes, the synergy between Oxali and IFNg became more obvious when the response was examined by RNA-seq analysis (**Figure 2.2H**). Despite the common notion that IFNg is a potent inducer of MHCI genes (Zhou, 2009), its effect at 200 pg/mL was weaker than the effect of low-dose (2 mM) Oxali in Myc-CaP and other cell lines.

Some cell lines, e.g. MC-38, barely responded to platinoids alone. Such cell lines, however, did exhibit potent induction of MHCI molecules and their cognate antigen processing and presentation machinery when the platinoids were combined with low concentrations of IFNg

(**Figures S2.2B-E**). In other cases, for instance the YUMM mouse melanoma lines, we observed considerable variation in the response (**Figure S2.2C**).

To confirm the ability of Oxali to potentiate antigen presentation, we conducted mass spectrometry-based peptidomic profiling of H-2Kb and H-2Db molecules isolated from MC-38 cells. In the H-2Kb experiment, treatment with IFN γ + Oxali resulted in higher amounts (based on area under the curve) of MHCI-bound peptides relative to cells treated with Oxali or IFN γ alone (**Figure S2.2F**). In the H-2Db experiment, IFN γ -treated cells exhibited higher amounts of MHCI-bound peptides than Oxali treated cells, however, some peptides were considerably more abundant only after incubation with IFN γ + Oxali.

Importantly, in Myc-CaP cells, Oxali induced surface expression of H-2Kq, the predominant MHCI molecule expressed by these cells (**Figure 2.2I**). The response to Oxali alone was stronger than the response to low-dose IFN γ , but the combination of Oxali + IFN γ resulted in synergistic H-2Kq induction. Although Cis, and to a lesser extent Carbo, barely induced surface H-2Kq on their own, they potentiated the response to IFN γ (**Figure 2.2I**). Functionally, Oxali-induced antigen presentation was confirmed using TRC2 cells made to express high-, medium-, and low-affinity variants of ovalbumin (Ova). In these cells, 4 mM Oxali stimulated presentation of the Ova-derived SIINFEKL epitope by H-2Kb, especially in TRC2-N4 cells that express the high-affinity (wild-type) variant (**Figure 2.2J**). When incubated with OT-I CD8⁺ T cells, whose T-cell receptor (TCR) is SIINFEKL-specific, Oxali-treated TRC2-N4 cells led to T-cell activation, resulting in tumor cell killing (**Figures 2.2K, L**). In the absence of Oxali, OT-I T cells enhanced presentation of the high-affinity SIINFEKL epitope but had no effect on the medium (TRC2-G4)- or low (TRC2-E1)-affinity variants and did not lead to their killing. These results are consistent with previously published data showing that only the high-affinity SIINFEKL epitope induces

IFN γ secretion by OT-I cells (Denton et al., 2011), and indicate that the effect of Oxali is mechanistically distinct from the effect of IFN γ .

The effect of Oxali on MHCI surface expression was also seen in human PCa metastatic PC3 cells (**Figure S2.3A**), primary melanoma WM793 cells (**Figure S2.3B**), and NSCLC cell lines (**Figures S2.3E,F**). Even in MIA PaCa-2 cells, derived from ICI refractory PDAC, Oxali induced HLA-ABC surface expression (**Figure S2.3C**). Examination of a panel of human melanoma cell lines confirmed the observations made with mouse melanoma cells lines, showing a considerable variation in the response (**Figures S2.3B and S2.4D**).

STAT1 and IFN γ R2 needed for Oxali + IFN γ synergism. We investigated the basis for the synergy between low-dose Oxali and IFN γ . Paralleling induction of PSMB9, Oxali at 2 mM, and to a lesser extent Cis, induced interferon response factor 1 (IRF1) expression in Myc-CaP cells (**Figures S2.4A, B**). Although IFN γ itself led to modest IRF1 induction, that induction was strongly potentiated by Oxali and Cis (**Figure S2.4B**). Likewise, the addition of Cis and especially Oxali to IFN γ -treated cells dramatically increased STAT1 phosphorylation, which was not stimulated by IFN γ alone (**Figure S2.4B**). Given the ability of Oxali to induce *Ifngr2* mRNA expression, not only in Myc-CaP cells (**Figure 2.2D**) but also in mouse and human melanoma cell lines (**Figures S2.4C, D**), we postulated that synergistic IRF1 and STAT1 activation by Oxali plus IFN γ may be due to IFN γ R2 induction. On its own, IFN γ did not induce *Ifngr2* mRNA in any of the cell lines we analyzed (**Figures 2.2D, 2.3D, and S2.4D**).

To examine the role of IRF1, STAT1, and IFN γ R2 in the synergistic induction of MHCI components by Oxali + IFN γ , we used CRISPR-Cas9 genome editing to ablate IRF1, STAT1, and IFN γ R2 in Myc-CaP cells and IRF1 in TRC2 cells. As predicted, IFN γ R2-deficient clones no longer responded to IFN γ (**Figure S2.4E**) and IRF1- and STAT1-ablated clones did not express

IRF1 or STAT1, respectively (**Figure S2.4F**). Remarkably, the synergistic induction of *Nlrc5*, *Psmb9*, and *Tap1* mRNAs by Oxali + IFN γ was minimally dependent on IRF1 but completely abolished by ablation of either STAT1 or IFN γ R2 (**Figures 2.3A-C**). Induction of *Ifngr2* mRNA, however, was unaffected by either IRF1 or STAT1 ablation, whereas *Tapasin* induction by Oxali + IFN γ was modestly reduced only in IFN γ R2 ablated cells and *Erap1* induction was decreased by either IRF1 or IFN γ R2 ablation (**Figures 2.3D, E**). Importantly, ablation of STAT1 or IFN γ R2 prevented synergistic surface H-2Kq induction (**Figure 2.3G**). Ablation of IRF1 also reduced synergistic H-2Kq induction, but this was mainly due to loss of responsiveness to Oxali alone (**Figures 2.3G, H**). As expected, ablation of TAP1 abrogated induction of surface H-2Kq but had no effect on PD-L1, whose expression was induced by Oxali + IFN γ (**Figure 2.3H**). In TRC2-N4 cells, ablation of IRF1 decreased Oxali-induced *Tap1*, *Psmb9*, or *Nlrc5* mRNA expression, but had a modest effect on Oxali + IFN γ -induced *Tap1* mRNA and no effect on Oxali + IFN γ -induced *Psmb9* or *Nlrc5* mRNAs (**Figure S2.2B**). Collectively, these results suggest that prior induction of IFN γ R2 by Oxali (or Cis) via an IRF1- (and STAT1-) independent pathway strongly potentiates the ability of Myc-CaP (and TRC2) PCa cells to respond to exogenous IFN γ .

Consistent with its ability to evoke the ISR (Bruno et al., 2017), Oxali induced eIF2 α phosphorylation and expression of the ER-stress-responsive bZIP transcription factor CHOP (**Figure S2.4G**). IFN γ had no effect on either response and neither IRF1 nor STAT1 ablation prevented their induction (**Figure S2.4H**). Conversely, ablation of CHOP (*Ddit3*) had no effect on H-2Kq induction by Oxali or Oxali + IFN γ (**Figure S2.4I**), suggesting that the response to Oxali depends on other effectors. Not surprisingly, Oxali and especially Cis led to gH2AX induction (**Figure S2.4G**), a marker of DNA damage. Oxali and to a lesser extent Cis also induced I κ B α , suggestive of NF- κ B activation (**Figure S2.4G**). DNA damage can activate cGAS-STING

signaling and induce immune stimulatory type I IFN (Chen et al., 2016; Corrales et al., 2015). Indeed, Oxali treatment increased *Ifna*, *Ifnb*, and *Il1b* expression (**Figure S2.4J**). We used CRISPR-Cas9 to ablate cyclic GMP-AMP synthase (cGAS, encoded by *Mb21d1*) and *Ifnar2* (**Figure S2.4K**), but neither ablation reduced Oxali-induced IFN γ R2 expression (**Figure S2.4L**). Nonetheless, cGAS and *Ifnar2* knockout cells exhibited reduced *Psmb9* mRNA induction and surface H-2Kq expression after Oxali treatment, but the Oxali + IFN γ combination was still synergistic (**Figures S2.4M, N**). Thus, cGAS activation, probably triggered by DNA damage, may have an auxiliary role in the immunogenic response to Oxali.

Low-dose Oxali enhances MHCI related chromatin accessibility. The above results suggest that low-dose Oxali and IFN γ induce MHCI components, NLRC5 and IFN γ R2 through different mechanisms. To better understand the transcriptional mechanisms underlying the response to Oxali, we subjected Myc-CaP cells that were incubated with either 2 mM Oxali, 1 ng/mL IFN γ , or a combination of the two to RNA-seq and ATAC-seq analyses. The coupling of ATAC-seq, a method for assessing transcription factor binding site accessibility (Buenrostro et al., 2015), with RNA-seq allows transcription factor loading to be correlated with actual transcriptional changes. Although the two methods revealed a considerable overlap between the Oxali- and IFN γ -elicited responses, each agent also had a unique effect on the transcriptome and chromatin accessibility (**Figures 2.4A, B, and S2.5A**). By-and-large, the response to Oxali was broader than the response to IFN γ and the combination of IFN γ + Oxali predominantly enhanced the magnitude of gene induction rather than its breadth. In addition to induction of the antigen presenting and processing machinery, RNA-seq analysis confirmed induction of IFN γ , IFN α , ATM, NF- κ B, p53, and oxidative phosphorylation signaling by Oxali + IFN γ . Using aggregate analysis of peaks of accessible chromatin, which provides estimates of frequencies and footprints

of transcription factor binding (Buenrostro et al., 2013), we found a large number of transcription factors whose chromatin accessibility was enhanced by Oxali. These transcription factors included members of the bZIP superfamily, such as AP-1, ATF/CREB and NRF2, forkhead (FOX), RUNX, and NF-kB proteins (**Figures 2.4A, B**).

Of note, the NF-kB pathway, represented by the *Nfkb1* gene, was stimulated by Oxali but not IFN γ . Congruently, Oxali increased access to multiple transcription factor binding sites at the *Nfkb1* locus, while IFN γ alone or together with Oxali barely had any effect (**Figure 2.4C**). To examine activation of the antigen processing and presentation pathway, we analyzed a gene cluster on mouse chromosome 17 harboring *Psmb9*, *Tap1*, *Psmb8*, and *Tap2*. Again, Oxali alone, and to a lesser extent Cis, increased transcription factor accessibility to certain sites within this locus, whose chromatin structure was barely affected by IFN γ alone (**Figure 2.4D**). However, IFN γ further enhanced transcription factor accessibility in Oxali-treated cells (**Figure 2.4D**), an effect that was consistent with the transcriptomic analysis (**Figure 2.4A**). The *Nlrc5* gene, whose expression was induced by both agents, was also made more accessible after Oxali treatment but not IFN γ (**Figure S2.5B**). Similarly, the *Erap1* locus, which responds to both Oxali and IFN γ , was made more accessible by Oxali relative to IFN γ (**Figure S2.5C**). By-and-large, Oxali treatment enhanced accessibility of MYB, IRF8, IRF2, FOXO1, MAFF, GATA, p65/RelA, GFY, DMRT1, RUNX2, OCT4, NR5a2, BORIS, CTCF, AP-1, E2F3, IRF1, IRF3, ATF3, STAT1, STAT3-5, c-Myc, and EBF1 binding sites. Addition of IFN γ expanded this response to include E2F6, JUNB, HIF-1b, KLF3, and DMRT6 binding sites. Of note, the chromosome 17 MHC I region opened by Oxali contained recognition sites for BORIS and CTCF, two general transcription factors responsible for chromatin opening (Li et al., 2012).

Oxali activates histone acetyltransferases. The above results suggest that Oxali may induce chromatin de-compaction, a response most commonly mediated by histone acetylation (Shahbazian and Grunstein, 2007). We therefore examined the effect of low dose Oxali on histone acetyltransferase (HAT) and deacetylase (HDAC) activity. Interestingly, Oxali treatment of Myc-CaP cells increased HAT enzymatic activity within 3 hr and its effect was comparable to that of an HDAC inhibitor (HDACi) (**Figures 2.5A, B**). Remarkably, Oxali led to greater stimulation of HAT activity at 2 mM than at 4 mM. Consistent with these results, Oxali treatment increased the nuclear amount of total p300 and acetylated p300/CBP (**Figure 2.5C**). Cis and IFN γ treatments also increased the nuclear amounts of total p300 and acetylated p300/CBP, but their effect was weaker compared to Oxali (**Figure 2.5D**). Fluorescent microscopy indicated that Oxali treatment induced formation of nuclear and nucleolar foci containing p300 (**Figure 2.5E**). IFN γ treatment alone had a weaker effect. HDACi treatment led to formation of smaller nuclear p300 foci and unlike Oxali or IFN γ - no nucleolar foci were observed. Oxali treatment also enhanced mRNA expression of many chromatin modifiers while inhibiting expression of others, and a similar pattern was observed in NASH-induced HCC (**Figures 2.5F**). Chromatin immunoprecipitation (ChIP) confirmed Oxali-enhanced p65/NF- κ B recruitment to the *Ifngr2* promoter (**Figure 2.5G**), as well as p300 recruitment to the *Ifngr2*, *Tap1*, *Psmb8* and *Psmb9* promoters (**Figures 2.5H-J**). The same promoter regions exhibited increased H3 at lysine 14 and 27 acetylation after Oxali treatment (**Figure 2.5K**). A general increase of H3 lysine 14 acetylation in nuclear foci similar to those revealed by p300 antibody staining was confirmed using IF analysis (**Figure S2.6E**). To determine whether histone acetylation was needed for H-2Kq induction by Oxali, we used the p300/CBP inhibitor C646 and found it to elicit a dose- dependent decrease in H-2Kq induction by Oxali or Oxali + IFN γ treatments (**Figure S2.6F**). Only a minor effect on the response to IFN γ

was observed. Moreover, CTB, a p300/CBP activator, induced H-2Kq surface expression in a dose-dependent manner (**Figure S2.6G**). Further confirming that enhanced histone acetylation potentiates expression of genes encoding MHC I components, treatment of Myc-CaP cells with the HDACi panabinstat at doses that did not induce cell death, led to induction of *Ifngr2*, *Psmb9*, *Nlrc5*, *Tap1* and *Tapasin* mRNAs and surface H-2Kq (**Figures S2.6H-J**). However, combination of HDACi with low dose Oxali did not synergistically increase the effect of each agent alone, indicating that they work through similar pathways (**Figures S2.6H-J**). By contrast, inhibition of the DNA damage response mediators ATM and ATR, whose activity was stimulated by low dose Oxali, did not interfere with Oxali-induced H-2Kq expression (**Figures S2.6K, L**). Thus, Oxali-induced DNA damage, as opposed to histone acetylation, only has a minimal role in MHC I induction.

Role of NF- κ B in Oxali-induced MHC I gene expression. Consistent with the above results, treatment of Myc-CaP cells with 2 mM Oxali led to sustained increases in nuclear p65/NF- κ B, CREB1, ATF3, P-ATM, and JunB (**Figures 2.6A-D**). Consistent with its weaker effect on MHC I gene expression, Cis exerted a more transient and weaker effect on NF- κ B and CREB. Consistent with the ChIP results, NF- κ B activation was needed for full *Ifngr2* mRNA induction, as treatment of Myc-CaP cells with two different IKKb inhibitors (IV and ML120B) led to a 4-fold decrease in *Ifngr2* mRNA in cells incubated with Oxali or Oxali + IFNg (**Figure 2.6E**). IKKb inhibitors also reduced H-2Kq surface expression (**Figure 2.6F**) and attenuated synergistic *Psmb9* and *Nlrc5* mRNA induction (**Figures 2.6G, H**).

IFNgR2 expression is needed for Oxali-enhanced tumor regression. Low-dose Oxali had little effect on subcutaneous (s.c.) growth of Myc-CaP, B16, or YUMM1.7 cells, but it strongly potentiated the inhibitory effect of anti-PD-L1, which was ineffective by itself in Myc-CaP and

B16 tumors (**Figures 2.7A**). The synergistic inhibition of Myc-CaP tumor growth by Oxali + anti-PD-L1 was abrogated by IFN γ R2 ablation (**Figures 2.7A**). Treatment with low-dose Oxali, but not Cis, enhanced expression of *Ifngr2*, *Psmb9*, *Tap1*, and *Nlrc5* mRNAs in s.c. Myc-CaP tumors (**Figures 2.7B, C**). Anti-PD-L1 ICI had no effect on *Ifngr2* mRNA expression, although it potentiated induction of *Psmb9*, *Tap1*, and *Nlrc5* mRNAs by Oxali (**Figures 2.7B, C**). Importantly, *Psmb9*, *Tap1*, and *Nlrc5* induction by Oxali + anti-PD-L1 was abrogated after IFN γ R2 ablation (**Figures 2.7C**). Oxali + anti-PD-L1 induced MHCI (H-2Kq and H-2Dd) surface expression by tumor cells, which was also abolished by IFN γ R2 ablation (**Figure 2.7D**). Although treatment of bone marrow (BM)-derived macrophages with Oxali or Oxali + IFN γ upregulated MHCI machinery expression, MHCII surface expression did not respond to Oxali alone. Nonetheless, Oxali treatment potentiated MHCII expression in macrophages that were co-stimulated with IFN γ . Of note, IFN γ R2 ablation in tumor cells had little effect, if any, on infiltration by effector CD8⁺ T cells, whose numbers were equally increased after Oxali + anti-PD-L1 treatment in IFN γ R2-expressing and non-expressing tumors (**Figures 2.7E-G**). Thus, Oxali-induced upregulation of MHCI genes in malignant cells is important for the final recognition and killing stage of the cancer-immunity cycle (Chen and Mellman, 2013) but has no role in ICI-induced CTL reinvigoration.

2.4 Discussion

Certain chemotherapeutic drugs, including Oxali, are immunostimulatory when used in low, non-lymphoablative doses (Bracci et al., 2014; Galluzzi et al., 2015). The molecular basis for this phenomenon has been enigmatic and was attributed to ICD, a unique form of apoptosis that is immunostimulatory rather than immunosuppressive (Kroemer et al., 2013). Although its mechanistic basis remains obscure, ICD is defined by antigen release and T-cell priming (Kroemer et al., 2013), the first step in the cancer-immunity cycle (Chen and Mellman, 2013). Our results, however, show that Oxali mainly acts by inducing the processing and presentation of class I antigens by viable malignant cells, thereby enhancing their recognition and eventual killing by reinvigorated CTLs. This activity is also exhibited by other platinoids, albeit to a lower extent, and may explain the success of the anti-PD-L1 + Carbo combination in human NSCLC and its correlation with enhanced MHCI component expression. Induction of MHCI associated genes by low dose Oxali correlates with relaxation of their regulatory regions and increased transcription factor accessibility, a response known to depend on histone acetylation. Indeed, we found that Oxali is a potent HAT activator and inducer of H3 K14 acetylation, independent of IFN γ signaling. Importantly, the response to IFN γ depends on STAT1 and IRF1 activation, which, unlike Oxali, causes minimal chromatin remodeling. Nonetheless, by inducing IFN γ R2 in an NF- κ B-dependent manner, Oxali treatment greatly enhances the response to exogenous IFN γ derived from reinvigorated CTLs. These results may explain why PD-L1/PD-1 inhibitors function better in NSCLC patients that were treated with platinoid drugs, such as Carbo. Indeed, the patients who benefited most from anti-PD-L1 + Carbo + Pemetrexed combination therapy showed higher expression of MHCI components. Furthermore, other clinical studies have suggested a role for impaired HLA Class I antigen processing and presentation in acquired ICI resistance (Gettinger et

al., 2017), supporting the concept of MHCI upregulation as a means to potentiate ICI responsiveness.

Mechanistic basis of Oxali-induced immune stimulation. All platinoids form inter- and intra-strand DNA adducts, but Cis- and Oxali-generated adducts are differentially recognized by DNA repair and damage-recognition proteins (Chaney et al., 2005). For instance, certain damage recognition proteins bind with higher affinity to Cis-GG adducts than Oxali-GG adducts, correlating with stronger γ -H2AX induction. Oxali, however, was suggested to have higher affinity to nucleolar DNA than Cis, a property that may be related to its ability to activate the integrated or ribosomal stress responses (Bruno et al., 2017; Kepp et al., 2015). Although the precise mechanism of stress response activation remains unknown, we confirmed that Oxali exposure led to ER expansion and induction of the ER and oxidative stress responsive transcription factor, CHOP. Nonetheless, CHOP ablation had little effect, if any, on induction of MHCI genes. It was also suggested that Oxali may preferentially activate the p53-mediated stress and DNA damage response (Chiu et al., 2009), but in most cancer cells we have examined, there were no significant differences in p53 activation by the two platinoids. Furthermore, inhibition of the DNA damage response mediators ATM and ATR did not affect induction of MHCI components, suggesting that DNA damage per se has little role in Oxali-induced immunogenicity.

Given these negative results, we examined whether Oxali treatment alters chromatin structure of responsive genes using ATAC-seq. The results revealed that Oxali enhanced the accessibility of MHCI related chromatin regions to a diverse collection of transcription factors, many of which, such as NF- κ B, AP-I, and CREB, are known to interact with the co-activators p300 and CBP (Chan and La Thangue, 2001; Mukherjee et al., 2013; Wojciak et al., 2009). Since chromatin accessibility is controlled by H3 lysine acetylation (Shahbazian and Grunstein, 2007),

we examined whether Oxali affects the activity of enzymes that mediate this modification. Strikingly, we found a marked (4-fold) increase in nuclear HAT activity within 3 hr of Oxali addition. Even more surprisingly, Oxali enhanced the nuclear expression and acetylation of p300 and CBP and induced the formation of nuclear foci containing p300 and K14 acetylated H3. Of note, autoacetylation stimulates p300 and CBP activity and may reflect their dimerization (Thompson et al., 2004). Based on these findings, we speculate that Oxali and other platinoids may covalently interact with p300 and/or CBP to enhance their dimerization. Indeed, after its non-enzymatic activation, Oxali can bind different proteins, including histones and ubiquitin (Hartinger et al., 2008; Soori et al., 2015). Furthermore, p300 and CBP share two well conserved metal binding fingers (Park et al., 2013) that may serve as targets for different platinoids. Supporting the role of enhanced histone acetylation in MHCI gene induction, we found that the HDACi panabinstat elicited nearly the same transcriptional response as low dose Oxali, and other studies have shown HDACi to potentiate the response to PD-1 blockade and induce MHCI expression (Terranova-Barberio et al., 2017). Moreover, along with NLRC5, p300/CBP are important components of the transcriptional activation complex responsible for MHCI induction (Meissner et al., 2012), and decreased nuclear p300 is associated with disease progression and worse prognosis of melanoma patients (Rotte et al., 2013).

MHCI induction vs. immunogenic cell death. Importantly, Oxali-induced MHCI antigen presentation takes place in viable cancer cells well before they succumb to CTL-mediated killing. By contrast, platinoid-induced ICD entails the release of damage associated molecular patterns (DAMP) and antigens by dead cancer cells, killed by platinoid-elicited DNA damage. Unlike HCC cells, which efficiently express MHCI molecules and components of their cognate antigen processing and presentation machinery and are readily killed by cancer-directed CTLs (Shalapour

et al. 2017), MHCI expression and antigen presentation are much lower in PCa. Correlating with high expression of MHCI components, HCC responds well to PD-1/PD-L1-inhibitors despite having relatively low mutational burden (El-Khoueiry et al., 2017; Shalapour et al., 2017). By contrast, PCa is ICI refractory (Bilusic et al., 2017) despite having a mutational burden that is not much lower than that of HCC (Schachter et al., 2017). Of note, the Myc-CaP and TRC2 PCa cell lines became highly responsive to PD-L1 blockade after Oxali co-treatment, an effect that required IFN γ R2 induction. Although NF- κ B-dependent IFN γ R2 induction renders Oxali-treated cancer cells much more responsive to IFN γ -expressing effector CTLs, it has no effect on CTL activation and recruitment of tumor-eradicating CD8⁺ T cells. Collectively, these findings demonstrate that the immunogenic activity of low-dose Oxali in established tumors is predominantly ICD-independent, which is needed for T-cell priming and initiation of the cancer-immunity cycle. Importantly, low-dose Oxali promotes in combination with checkpoint inhibitors termination rather than initiation of the cancer immunity cycle, which induces cancer cell killing. It remains to be seen whether more specific HAT activators or HDACi will exhibit the same immunogenic activity as Oxali and other platinoids. Until then, our work suggests that Oxali and other platinoids may be the ideal drugs to combine with PD-1/PD-L1 inhibitors, especially in cancers with insufficient MHCI expression and antigen presentation. Curiously, both human and mouse HCC show higher expression of MHCI components than normal liver tissue. This may be due to the highly inflamed and stressed nature of such tumors and their development in the context of localized immunosuppression (Shalapour et al., 2017), which alleviate the necessity for MHCI downregulation.

2.5 Materials and Methods

Animal models. *C57BL/6* and *FVB/N* control mice were purchased from Charles River Laboratories and *Cd8a^{-/-}* were purchased from the Jackson Laboratory.

Cell culture experiments. Mammalian cells were all grown in a humidified incubator with 5% CO₂ at 37°C. Myc-CaP and HEK293T cells were grown in Dulbecco's Modified Eagle Media (DMEM, Gibco) supplemented with 10% FBS, 1% L-glutamine, and 1% penicillin and streptomycin. TRAMP-C2 cells were grown in DMEM supplemented with 5% FBS, 0.2% L-glutamine, 1% penicillin and streptomycin (Gibco), 5% NuSerum IV (Corning), 0.005 mg/ml bovine insulin (Sigma-Aldrich, Darmstadt, Germany), and 10nM dehydroisoandrosterone (Acros Organics). MC38, PC3, and all YUMM cells were grown in DMEM/F12 (Gibco) supplemented with 10% FBS, 1% L-glutamine, and 1% penicillin and streptomycin. B16, WM793, H2030, and PC9 cells were grown in RPMI 1640 (Gibco) supplemented with 5% FBS, 0.2% L-glutamine, and 1% penicillin and streptomycin. Cells were counted by trypan blue exclusion and seeded into 12 or 24 well plates (Flow Cytometry and RNA processing) or into 6 well plates for Immunoblot analysis. Cells were treated with agents in fresh media as indicated for 3-48 hr after seeding and collected from plates using Acutase (STEMCELL Technologies) for flow cytometry analysis, TRIzol Reagent (Ambion) for RNA processing, or Immunoblot (IB) Lysis Buffer for IB analysis at indicated time points.

Nuclear, Cytoplasmic Extraction and Immunoblot Analysis. NE-Per Nuclear and cytoplasmic extraction kit and EpiQuik Nuclear Extraction Kit I (Epigentek, Farmingdale, NY, USA) were used for acquiring nuclear and cytosolic fractions of cells and tumor tissues and the buffers were prepared accordingly to the manufacturer's protocol. Whole cell or whole tissue lysates were made in RIPA buffer (20mM pH7.5 Tris-HCl, 10mM EDTA, 150mM NaCl, 1%

Triton, 1% sodium deoxycholate) with protease and phosphatase inhibitors. IB analysis was performed on cell or tissue lysates that were separated by SDS-PAGE and transferred to PVDF membranes. Antibody details are provided in the Key Resources Table.

Flow cytometry. Cell suspensions were prepared from prostate, tumor, and spleens as described¹⁴. Lymphocytes were isolated from human blood using Ficoll-Paque PLUS (GE Healthcare Life Science), according to manufacturer's recommendations. For lymphocyte isolation, tissues were cut into small pieces and incubated in dissociation solution [DMEM medium supplemented with 5% FBS, collagenase type I (200 U/mL), collagenase type IV (100 U/mL), and DNase I (100 mg/mL)] for 30 min at 37° C. After incubation, cell suspensions were passed through a 40 mm cell strainer and washed twice. To block Fc-mediated interactions, mouse cells were pre-incubated with 0.5-1 mg of purified anti-mouse CD16/CD32 per 100 mL, and human cells were incubated with FcR blocking reagent (Miltenyi Biotec). Isolated cells were stained with labeled antibodies in PBS with 2% FBS and 2 mM EDTA or cell staining buffer (Biolegend). Dead cells were excluded based on staining with Live/Dead fixable dye (FVD-eFluor780, eBioscience). For intracellular cytokine staining, cells were restimulated with cell stimulation cocktail [eBioscience; containing phorbol 12-myristate 13-acetate (PMA) and ionomycin (Ion)], in the presence of a protein transport inhibitor cocktail containing Brefeldin A and Monensin (eBioscience). For CD107a measurement, a staining antibody was added to the culture during stimulation. After 4-5 hr, cells were fixed and permeabilized either with BD™ Cytofix/Cytoperm reagent for cytokine staining or eBioscience™ Foxp3/Transcription Factor staining buffer for combined staining of cytokines with transcription factors. After fixation/permeabilization, cells were stained with labeled antibodies of interest. Cells were analyzed on a Beckman Coulter Cyan ADP flow cytometer. Data was analyzed using FlowJo 10.2

software (Treestar). Absolute numbers of particular immune cells in spleens were calculated by multiplying the total cell number of the spleen by the percentage of the particular cell type amongst total cells. Absolute numbers of particular immune cells in prostates were calculated by multiplying the cell number of the prostate portion by the percentages of the corresponding cell type in total prostate cells divided by the weight of the analyzed prostate (tumor) fragment (cell number per gr).

Immunostaining and Histology. Tissues were embedded in Tissue-Tek OCT compound (Sakura Finetek, Torrance, CA, USA) and snap-frozen. Tissue sections were fixed in cold acetone/methanol or 3-4% PFA for 3-10 min and washed with PBS. Slides were blocked with 1x PBS/1% normal donkey or goat serum for surface staining or 0.2% gelatin 0.2% BSA (from cold water fish skin; Sigma-Aldrich)/PBS/1% normal donkey or goat serum for intracellular staining for 15-30 min. Sections were incubated with primary antibodies for 1 or 12 hr at RT or 4°C, respectively. After washing with PBS, secondary antibodies were added for 1 hr at RT. As negative controls, samples were incubated with isotype-matched control antibodies or secondary antibodies only. After staining with DAPI, sections were covered with Vectashield Mounting Medium (Vector Laboratories, Burlingame, CA, USA). Paraffin-embedded tissue sections were subjected to de-paraffinization and rehydration, and then were immersed in a pre-heated antigen retrieval water bath with a pH 6.1 citrate buffer or Dako Target Retrieval Solution for 20 min at 95-96°C. All staining was done according to manufacturer's protocols (Vector Laboratories). Nuclei were lightly counterstained with a freshly made haematoxylin solution then further washed in water. Sections were examined using an Axioplan 200 microscope with AxioVision Release 4.5 software (Zeiss, Jena, Germany) or TCS SPE Leica confocal microscope (Leica, Germany).

MIA-Paca-2 cells were cultured on coverslips and fixed in 4% paraformaldehyde for 10 min at room temperature. After washing twice in PBS, cells were incubated in PBS containing 10% FBS and 1% donkey serum to block nonspecific sites of antibody adsorption. The cells were then incubated with the appropriate primary antibodies (diluted 1:100) and secondary antibodies (diluted 1:500) in PBS containing 0.1% saponin, 10% FBS and 1% donkey serum as indicated in the Figure legends. Confocal images were captured in multitracking mode on a SP5 confocal microscope (Leica) with a 63 × Plan Achromat 1.3 NA objective.

Paraffin-embedded specimens from a total of 118 Prostate cancer patients were integrated into a tissue microarray system (TMA) constructed at the Clinical Institute of Pathology at the Medical University of Vienna (MUV). All of the human specimens used for TMA construction were approved by the MUV Research Ethics Committee (1753/2014).

Human liver biopsies were obtained from the Biobank of the Medical University of Graz. Biopsies were registered in the biobank and kept anonymous. The research project was authorized by the ethical committees of the Medical University of Graz (ref. no. 1.0 24/11/2008). The study protocol was in accordance with the ethical guidelines of the Helsinki Declaration.

Slides were de-paraffinized, rehydrated, and incubated in 0.3% H₂O₂. After antigen epitope repairing, sections were blocked with horse serum from ImmPRESS kit (Vector Laboratories). Sections were then incubated with antibodies against human HLA (1:150, rabbit anti-human, Proteintech, catalog number 15240-1-AP), human aSMA (1:300, mouse anti-human, Clone 1A4, DAKO, catalog number M0851), human PSA (1:1, rabbit anti-human, Clone EP15884, NeoMarkers, catalog number RM-2104-R7), human ERAP1 (1:200, rabbit anti-human, Proteintech, catalog number 13821-1-AP), human TAP1 (1:200, rabbit anti-human, Proteintech, catalog number 11114-1-AP), and CD45 (1:200, mouse-anti-human, Proteintech, catalog number

60287-1-Ig). For secondary antibodies, slides were incubated with ImmPRESS-AP REAGENT Anti-Rabbit IgG (Vector Laboratories, MP-5401), ImmPRESS-HRP REAGENT Anti-Rabbit IgG (Vector Laboratories, MP-7451), and ImmPRESS-AP REAGENT Anti-Mouse IgG (Vector Laboratories, MP-5402). DAB (Vector Laboratories, SK-4100) and ImmPACT Vector Red (Vector Laboratories, SK-5105) were used for coloration.

Sections were imaged under a Hamamatsu 2.0-HT Digital slide scanner (Hamamatsu Photonics, EU, Japan, and USA). For measurement, Image J 1.49v was used. Serial cut slides were used to show adjacent tissues. Prostate cancer cells were shown by PSA staining and their HLA abundance was measured according to the PSA positive location. Prostate cancer cells were specifically selected using image J, and its mean optical density was obtained and statistically compared among groups. For hepatic cancer slides, cancer cells were the most abundant cell type and could be identified through their histopathological features.

Subcutaneous tumor models. $1-3 \times 10^6$ Myc-CaP, TRAMP-C2 cells, B16 or YUMM1.7 (purchased from ATCC) were s.c. injected into the right flank. Tumors were measured every 2-3 days using a caliper. Tumor volumes were calculated as $V=(width^2 \times length)/2$. For prostate cancer models, only male mice have been used, for melanoma models, both gender have been used.

RNA-seq processing and analysis. RNA-seq reads were aligned to the mouse genome (GRCm38/mm10) using STAR (v2.5.2b) (Dobin et al., 2013). Only reads that mapped to a single, unique location in the mouse genome were used for downstream analysis. Gene expression counts were counted for reads overlapping exons for all RefSeq transcripts using HOMER (Heinz et al., 2010). Differentially expressed genes were determined using DESeq2 (Love et al., 2014) using thresholds of 5% FDR and a fold change of 1.5 to determine the final lists. Gene expression profile clustering was performed by first using the DESeq2 rlog variance stabilization transform to

normalize the gene expression values. Hierarchical clustering was performed using Cluster3.0 (de Hoon et al., 2004) and visualized using Java TreeView (Saldanha, 2004). Pathway/Functional enrichment calculations were performed using www.gsea-mskcc.org/gsea/index.jsp. Pathway/Functional enrichment calculations were performed using (metascape.org) (Tripathi et al., 2015).genesets (Mootha et al., 2003; Subramanian et al., 2005).

The RNA-seq processing and analysis for liver samples were processed as described previously (Shalapour et al., 2017). Hierarchical clustering of important genes involved in the IFN γ response and allograft rejection hallmark gene sets was further performed using Cluster3.0 (de Hoon et al., 2004) and visualized using Java TreeView (Saldanha, 2004). Gene set enrichment analysis (GSEA) was also performed to identify hallmark gene sets (Mootha et al., 2003; Subramanian et al., 2005).

Quality control of the raw fastq files was performed using the software tool FastQC (Andrews et al., 2010). Sequencing reads were aligned to the mouse genome (mm10) using the STAR v2.5.1a aligner (Dobin et al., 2013). Read quantification was performed with RSEM (Li and Dewey, 2011) v1.3.0 and Ensembl annotation (Mus_musculus.GRCm38.68.gtf). The R BioConductor packages edgeR (Robinson et al., 2010) and limma (Ritchie et al., 2015) were used to implement the limma-voom (Law et al., 2014) method for differential expression analysis. Lowly expressed genes were filtered out (cpm > 1 in at least one sample). Trimmed mean of M-values (TMM) (Robinson and Oshlack, 2010) normalization was applied. The experimental design was modeled upon time point and treatment (~ 0 +treatment). The lmFit function in limma with consensus correlation to account for repeated measures of patient followed by the eBayes function was used to fit the design on voom normalized counts per gene.

Significance was defined by using an adjusted p-value cut-off of 0.05 after multiple testing correction using a moderated t-statistic in Limma.

Immunoproteasome activity assay. Immunoproteasome subunit LMP7 activity assays were analyzed directly on cell lysates from Myc-CaP cells using LMP7 specific fluorogenic peptide substrate Ac-ANW-AMC (Boston Biochem) (Cascio et al., 2001). Cells were treated with or without indicated concentration of oxaliplatin (2 mM), cisplatin (2 mM), or carboplatin (4 mM) alone or in combination with IFN γ (2 ng/mL) for 48 hr. One set of cells treated with IFN γ alone or in combination with each of the platinoids was treated with 200 nM ONX-0914 (Selleckchem) (Muchamuel et al., 2009) for 2 hr prior to lysis to determine background non-specific AMC fluorescence. Cells were lysed directly in proteasome assay buffer (50 mM Tris-HCl, pH 7.5, 40 mM KCl, 5 mM MgCl $_2$, 1mM ATP, 0.1% NP-40, phosphatase inhibitors, 1 mM dithiothreitol). 100 mM substrate peptide Ac-ANW-AMC was added to the lysate and reaction was carried out for 30 min - 1hr at 30°C in an orbital shaker. Fluorogenic AMC release into solution was assessed by a Tecan M200 plate reader (excitation: 380 nm, emission: 440–460 nm) which provides a direct measurement of LMP7 activity in the cell lysates. The measured LMP7 activity was normalized against total protein concentration of cell lysates and represented as fold LMP7 activity to control untreated Myc-CaP cell lysates. Data was represented as bar graphs using Graphpad Prism.

Nuclear and cytosolic extraction. Further steps are performed using EpiQuik Nuclear Extraction Kit I (Epigentek, Farmingdale, NY, USA) and the buffers are prepared accordingly to the manufacturer's protocol.

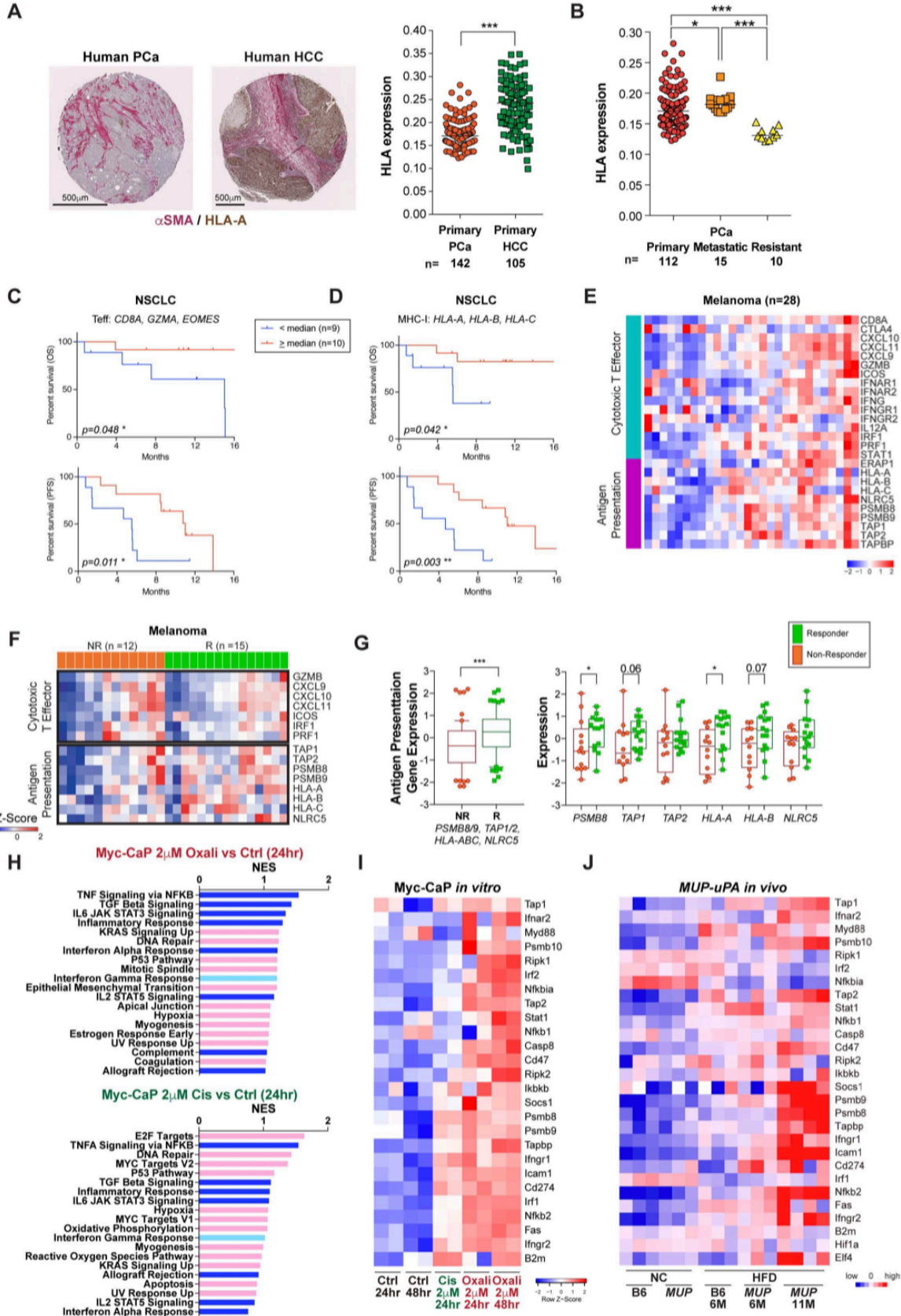
HAT activity assay. HAT activity assays were performed with EpiQuik HAT Activity/Inhibition Assay Kit (Epigentek, Farmingdale, NY, USA). Solutions were prepared according to manufacturer's protocol.

Quantification and statistical analysis. Data is presented as either means \pm SEM or medians of continuous values and was analyzed by two-sided Students' t-test or Mann-Whitney test for comparison of two groups, respectively. Two-way ANOVA was used to compare three or more groups in mouse data analyses. One-way ANOVA or Kruskal-Wallis test was used to compare three or more groups in human data analyses. Bonferroni's or Dunn's multiple comparison test was applied to compare all pairs of groups. D'Agostino & Pearson test and/or Shapiro-Wilk test were used to test the normality of sample distribution. Fisher's exact Chi-square t test was used to calculate statistical significance of categorical values between groups. Two tailed *P* values of ≤ 0.05 were considered significant. Linear regression was used to determine the correlation between two different variables. GraphPad Prism software was used for statistical analyses. Unpaired t test was used to determine the power using GraphPad StatMate software.

2.6 Figures

Figure 2.1. Expression of MHC I related genes in prostate, liver, and lung cancer.

(A) Tumor microarrays encompassing 142 primary PCa and 105 HCC patients (5–6 spots per patient = 3–4 tumor and 2 non-tumor) were stained for HLA-ABC (brown) and aSMA (red). Nuclei were counterstained with haematoxylin. Representative samples are shown on the left. Quantification performed by Image J is shown on the right. **(B)** HLA-ABC expression in primary ($n = 112$), drug resistant ($n = 15$), and metastatic ($n = 10$) PCa. Each dot = one patient; line = median. Mann–Whitney test was used to calculate statistical significance. **(C)** RNA-seq data from human samples from human samples. **(D)** RNA-seq data from human samples. **(E–G)** Myc-CaP cells incubated with indicated agents for 24 hr were subjected to RNA-seq. Top 20 hallmark gene sets were sorted by normalized enrichment score (NES). Immune-related gene sets are in blue (IFN γ signaling in light blue). Results repeated with a 48 hr incubation are shown in Figure S1G. **(E)** Expression of Oxali-induced genes involved in inflammation, antigen presentation, and IFN γ signaling **(F)** was compared to genes that are upregulated in NASH-driven HCC of *MUP-uPA* mice **(G)** ($n=3-4$ mice per group). See also Figure S2.1.



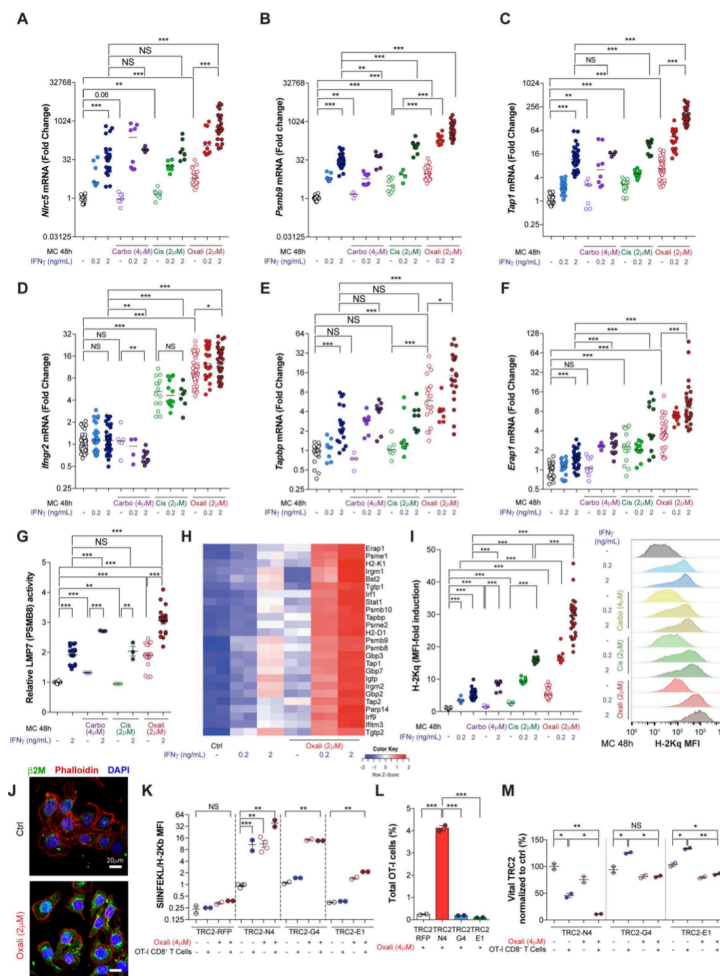


Figure 2.2. Platnoid-induced expression of MHC I component is potentiated by IFN γ .

(A-F) RNA from Myc-CaP cells incubated as indicated with IFN γ , Oxali, Carbo, or Cis for 48 hr was analyzed by qRT-PCR using primers for *Nlrc5*, *Psmb9*, *Tap1*, *Ifngr2*, *Tapasin*, and *Erap1*. The data are representative for at least 5-6 biologically independent experiments. (G) Myc-CaP cells were incubated with IFN γ , Oxali, Carbo, or Cis as above and lysed. LMP7 (PSMB8) immunoproteasome activity was measured using a specific fluorogenic peptide substrate. (H) RNA from Myc-CaP cells treated as above was subjected to RNA-seq analysis. Genes involved in antigen presentation are depicted by heat map representation. (I) Myc-CaP cells treated as above were analyzed for surface MHC I (H-2Kq) expression by flow cytometry. Each dot is a single experiment and horizontal lines are the medians. (J-L) TRC2 cells stably transfected with vectors expressing high, medium, and low affinity variants of ovalbumin were incubated with 4 mM Oxali and/or CFSE-labeled OT-I cells for 72 hr and analyzed by flow cytometry using an antibody that recognizes SIINFEKL bound to H-2Kb (J). The number of OT-I cells in each culture (K) and percentage of vital TRC2 cells (L) were determined by flow cytometry.

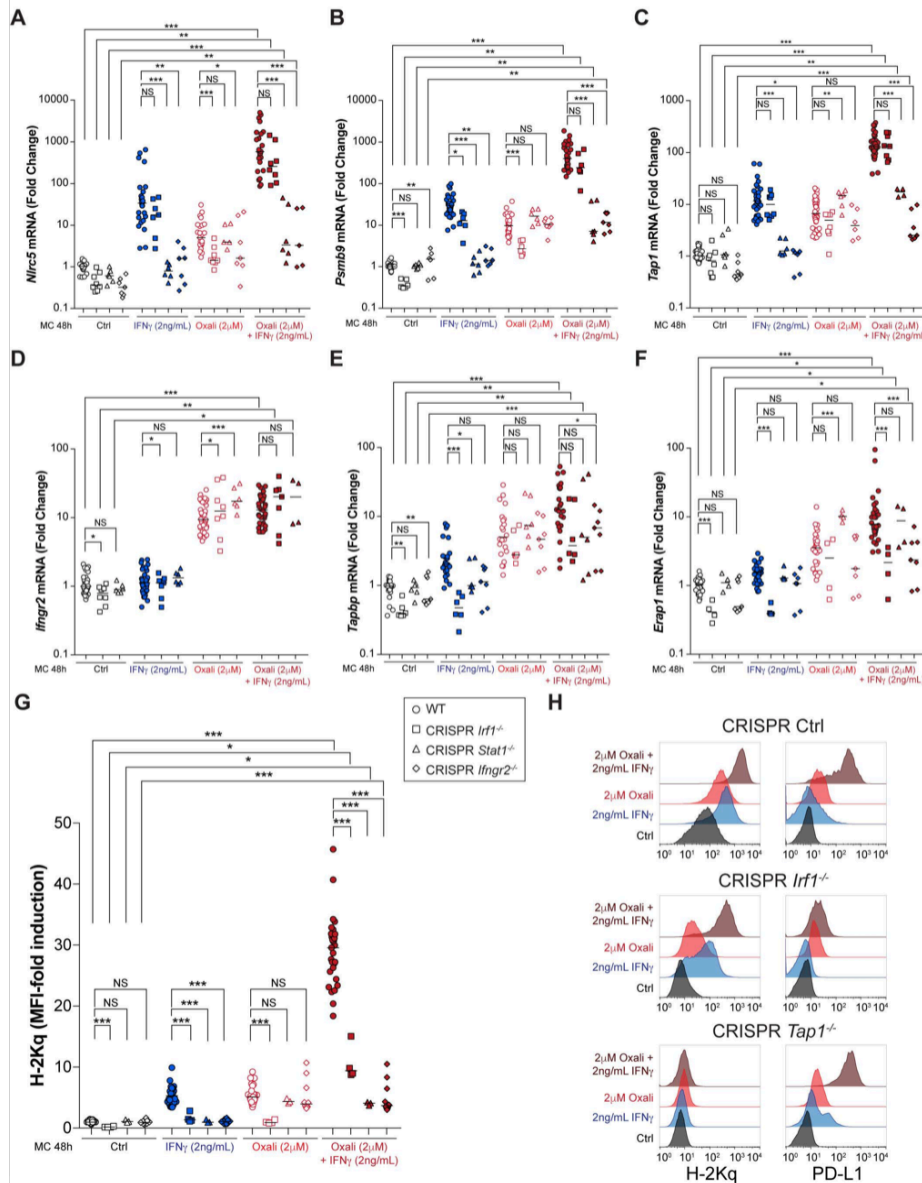


Figure 2.3. STAT1 and IFNGR2 mediate the synergistic response to oxali+IFNg.

(A-F) Myc-CaP cells transfected with lentiviruses containing Cas9 and gRNAs that target *Irf1*, *Stat1*, or *Ifngr2* were expanded under puromycin selection. RNAs extracted from cells that were treated as indicated with IFN γ and/or Oxali for 48 hr were analyzed by qRT-PCR with primers for the indicated MHC I genes. The data are representative for at least 5-6 biologically independent experiments. (G) Parental and gene-edited Myc-CaP cells were treated as above and analyzed for surface MHC I (H-2Kq) expression. Each dot represents an experiment and horizontal lines are the median. (H) Parental and gene edited Myc-CaP cells were treated as above and analyzed by flow cytometry for surface H-2Kq and PD-L1 expression.

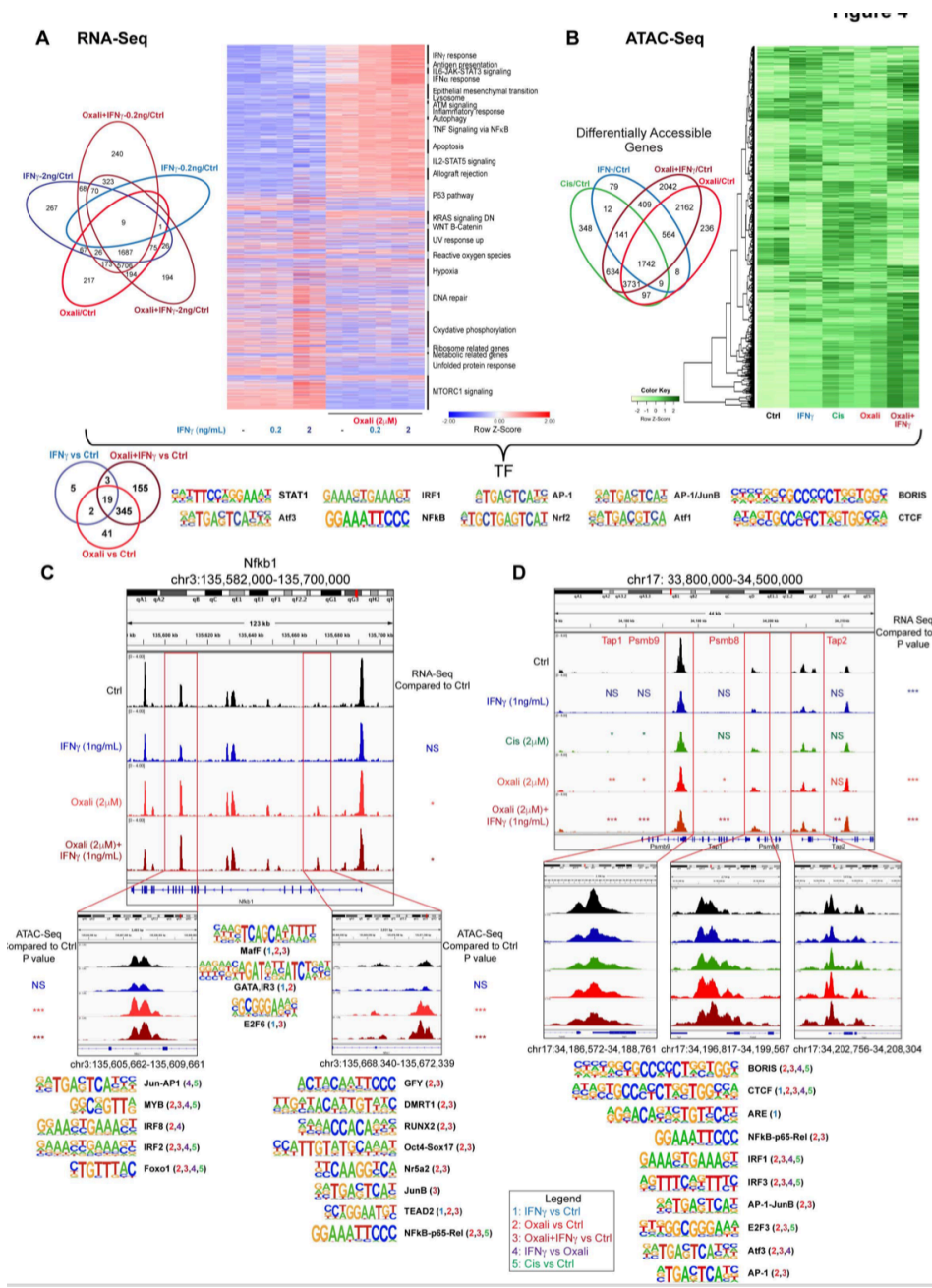


Figure 2.4. Low dose Oxali alters chromatin accessibility of MHCII related genes.

(A) RNAs extracted from Myc-CaP cells treated as indicated were subjected to RNA-seq. The Venn diagram compares gene expression between untreated and treated cells (left). The heatmap depicts differentially expressed genes involved in the indicated pathways (right). (B) Myc-CaP cells treated as above were subjected to ATAC-seq analysis. The Venn diagram presents the number of binding site changes and overlaps after each treatment (left). (C-D) Detailed ATAC seq analyses of the *Nfkb1* gene (C) and a mouse Chr17 gene cluster containing *Psmb8*, *Psmb9*, *Tap1* and *Tap2* (D). Changes in transcription factor (TF) binding site accessibility are compared to RNA seq results. The affected TF binding sites are depicted below each panel.

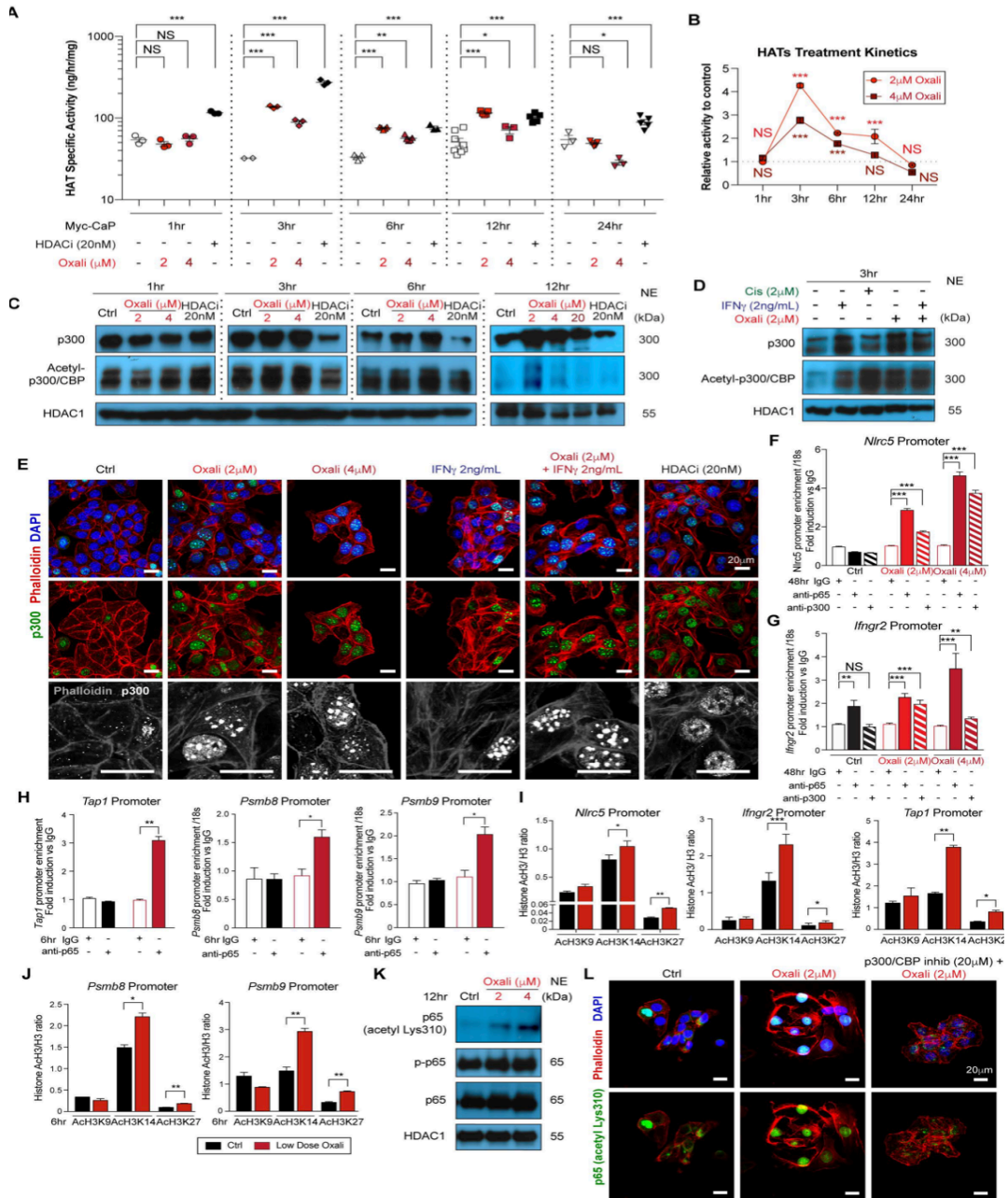


Figure 2.5. Oxali stimulated HAT activity and nuclear localization.

(A-B) Myc-CaP cells incubated with Oxali (2 or 4 mM) or the HDACi panabinostat (20 nM) were lysed and analyzed for HAT activity using H3 as a substrate. (C-D) Myc-CaP cells incubated with Oxali or HDACi were lysed and IB analyzed with antibodies to p300, acetylated CBP/p300 and HDAC1, which served as a loading control. (E) Myc-CaP cells treated as above were stained for p300 (green) and phalloidin to detect actin filaments (red). Nuclei were counter stained with DAPI (blue). (F) Effects of Oxali and IFN γ on expression of genes encoding chromatin modifiers in Myc-CaP cells. (G-J) Untreated and Oxali treated Myc-CaP cells were subjected to ChIP analysis with control IgG, p65/RelA, or p300 antibodies. (K) Untreated and Oxali treated Myc-CaP cells were subjected to ChIP analysis to detect H3 acetylation (lysines 9, 14, and 27) at the indicated promoter regions.

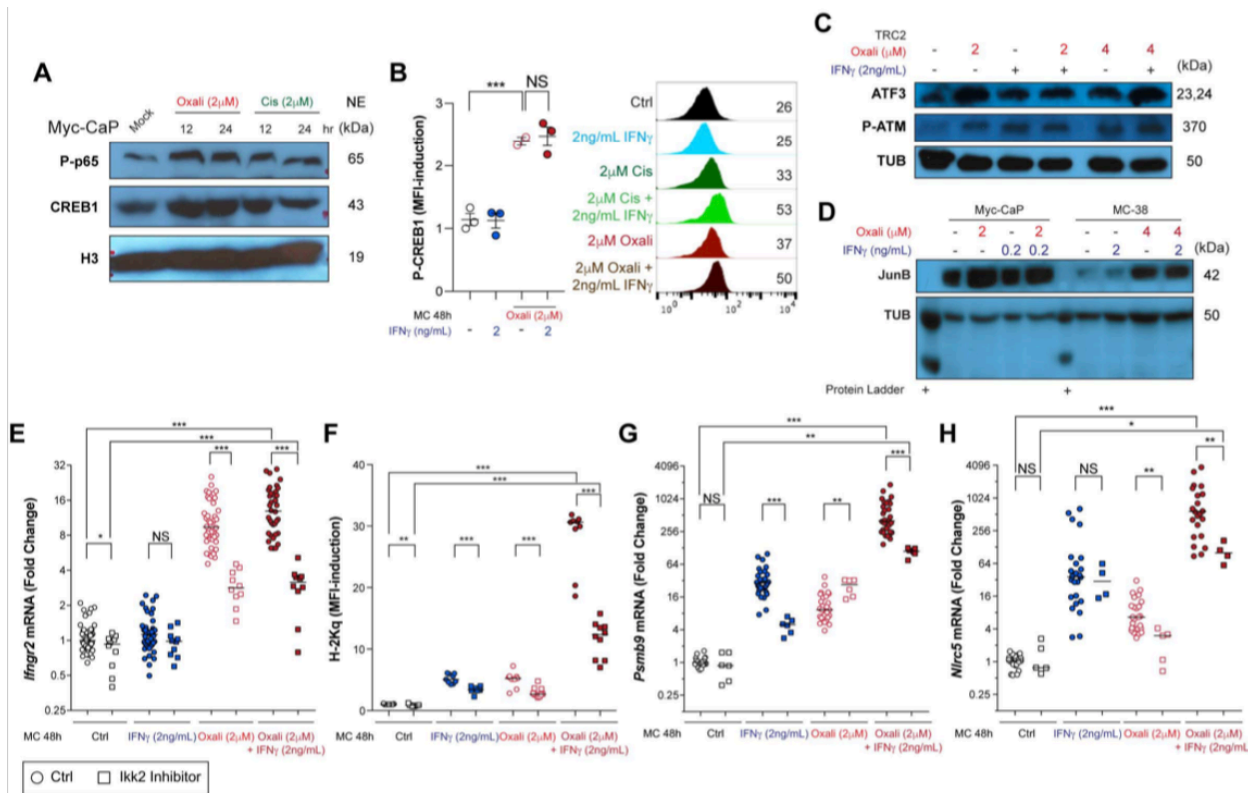


Figure 2.6. NF- κ B mediates IFN γ R2 induction and the synergistic response to platinumoids + IFN γ .

(A) Myc-CaP cells incubated with Oxali or Cis were lysed and IB analyzed with antibodies to phosphorylated p65/RelA, CREB1, and H3. (B) Myc-CaP cells treated as indicated were analyzed for CREB1 expression and phosphorylation by flow cytometry. (C) Myc-CaP cells treated as indicated were IB analyzed for ATF3 and phospho-ATM. (D) Myc-CaP and MC-38 cells treated with Oxali or IFN γ were analyzed for JunB expression. (E-H) Myc-CaP cells treated as indicated with or without IKKb inhibitors, ML120B or IV, were analyzed by qRT-PCR (E, G, and H) or flow cytometry for H-2Kq surface expression (F). Each dot represents an experiment and horizontal lines denote the median.

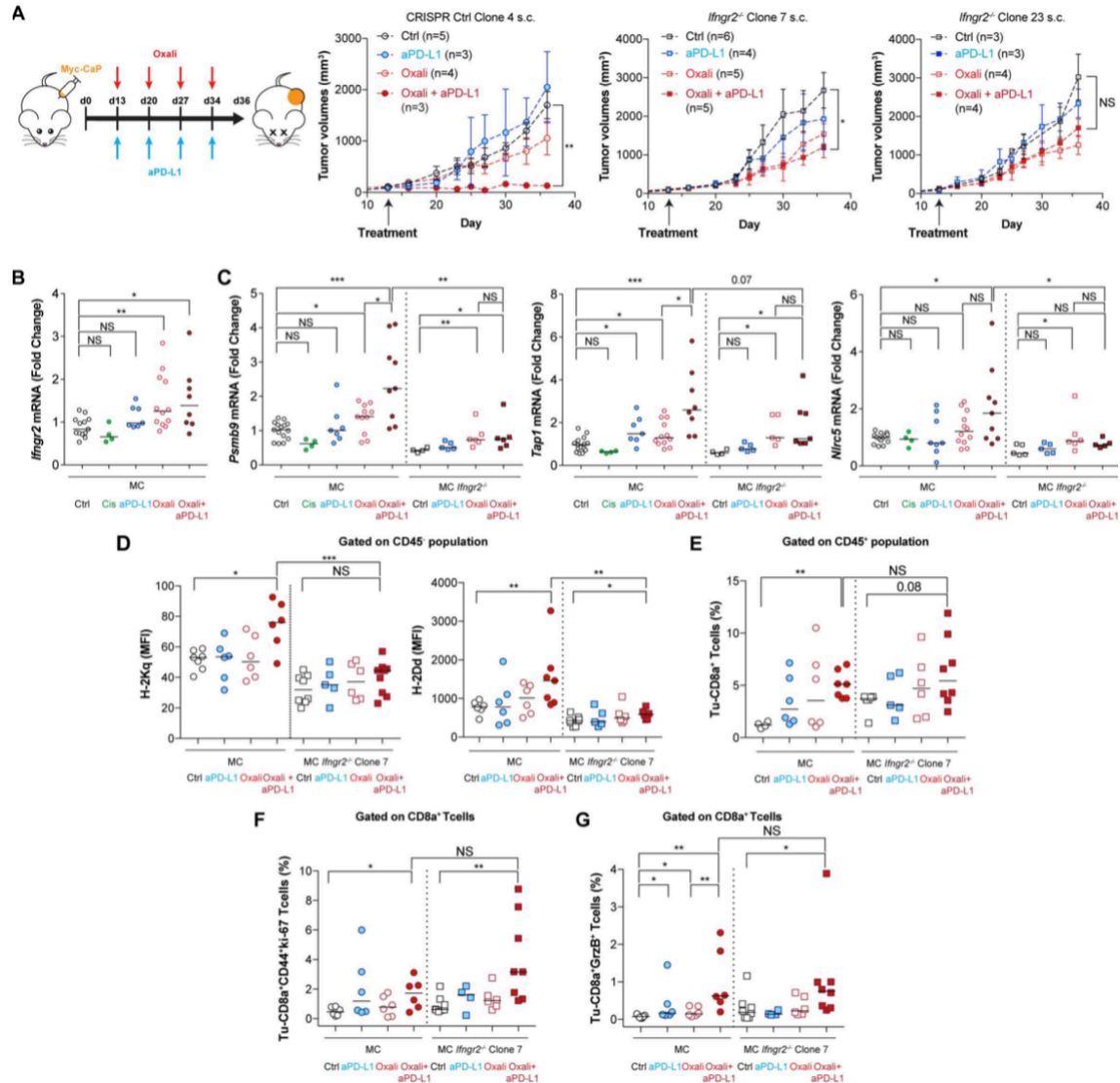


Figure 2.7. IFN γ 2 induction is needed for the Oxali-potentiated response to anti-PD-L1 therapy.

(A) Mice bearing s.c. tumors generated by control or *Ifngr2*-ablated Myc-CaP cells were allocated into 4 treatment groups: (1) control (5% dextrose), (2) Oxali (weekly), (3) a-PD-L1 (weekly), and (4) Oxali plus a-PD-L1 (weekly). After four treatment cycles, during which tumor size was measured, the mice were euthanized and analyzed. Significance was determined by Mann-Whitney and t-tests. Transient Cas9 expression was used to avoid any immune response to Cas9. (B-C) Total tumor RNA was analyzed by qRT-PCR for expression of the indicated genes. (D-G) Single tumor cell suspensions were analyzed by flow cytometry for H-2Kq expression on CD45⁺ cells (D) and presence of effector CD8⁺ T cell subsets (E, F, G). See also Figure S7.

2.7 Supplemental Data, Tables and Figures

Figure S2.1: Differential expression of MHCI molecules and components of their antigen processing and presentation machinery correlates with anti-PD-LA.

(A-B) PCa microarrays were stained for HLA-ABC, aSMA, PSA, and CD45 to determine HLA expression by cancer cells, CD45⁺ cells, and stromal (aSMA⁺) cells. Nuclei were counterstained with haematoxylin. Magnification bar: 100 mm. (C) Expression of genes involved in immune response, antigen presentation, and histone modification in non-tumor and tumor samples of patients with HCC or PCa was analyzed through unsupervised mining of TCGA data base.

(D-E) Low, intermediate, and high risk, and recurrent human PCa specimens were stained with TAP1, ERAP1, and HLA-ABC antibodies (n=20). Nuclei were counterstained with haematoxylin (D). Expression levels were determined using ImageJ software and the correlation between TAP1, ERAP1, and HLA expression was plotted (E). Each dot represents one patient. Magnification bar: 100 mm. (F) NSCLC tissue was stained with H&E or antibodies to CD8 and PD-L1. Magnification bar: 100 mm. (G) Total RNA extracted from TRC2 cells incubated with 2 mM of Oxali or Cis for 24 hr was subjected to RNA-seq analysis. The top 20 hallmark gene sets sorted by normalized enrichment score (NES) are shown to depict the Oxali- and Cis-induced responses determined by GSEA analysis. Immune-related gene sets are colored blue (IFN γ signaling in light blue). (H) Total RNA extracted from Myc-CaP cells treated as indicated for 48 hr was subjected to RNA-seq analysis. Expression of the indicated genes is depicted using heat maps. (I) Total RNAs extracted from s.c. Myc-CaP and spontaneous TRC2 tumors and from NASH-driven HCC were analyzed for expression of the indicated genes by qRT-PCR. Each dot represents a mouse and each horizontal line denotes the median.

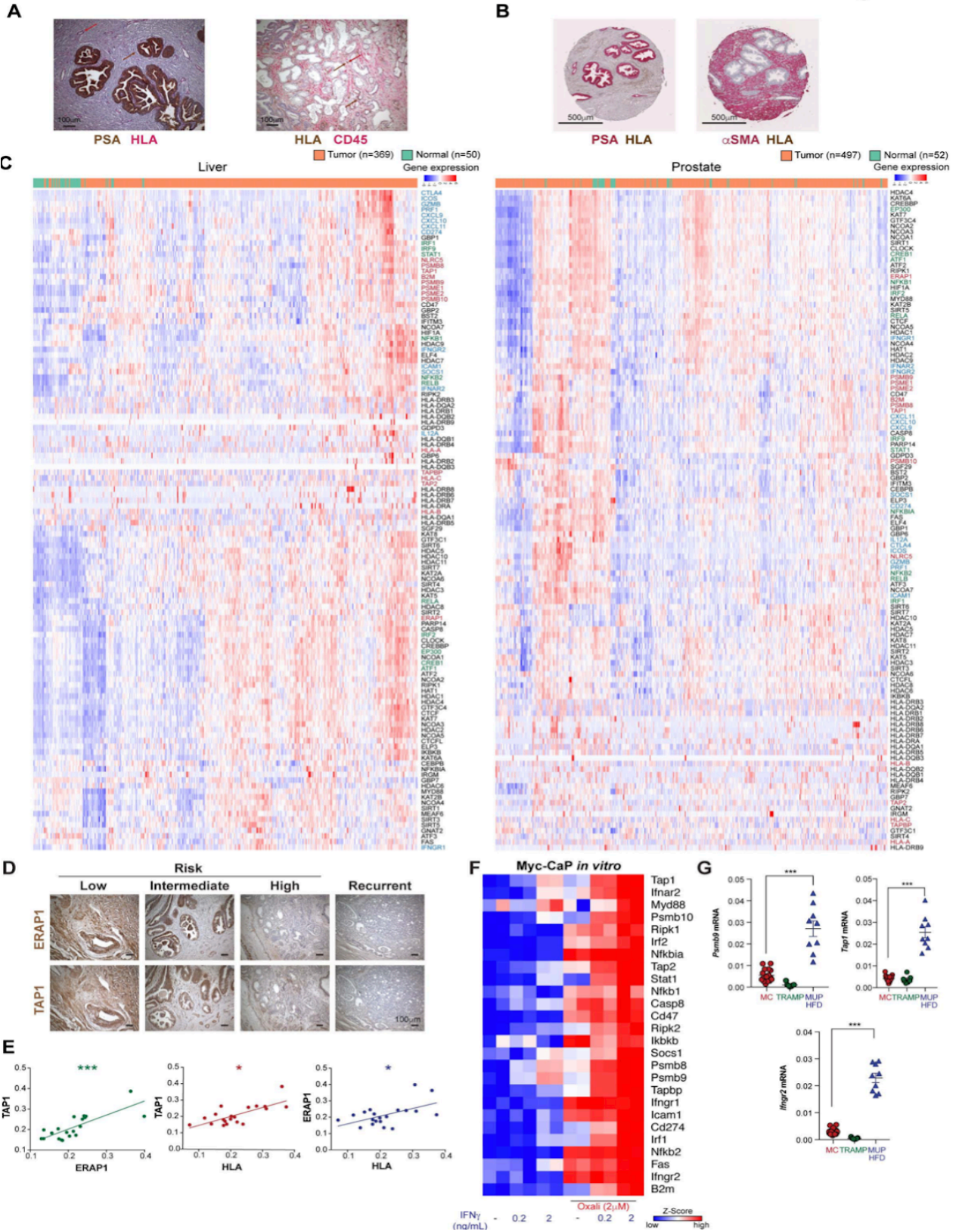
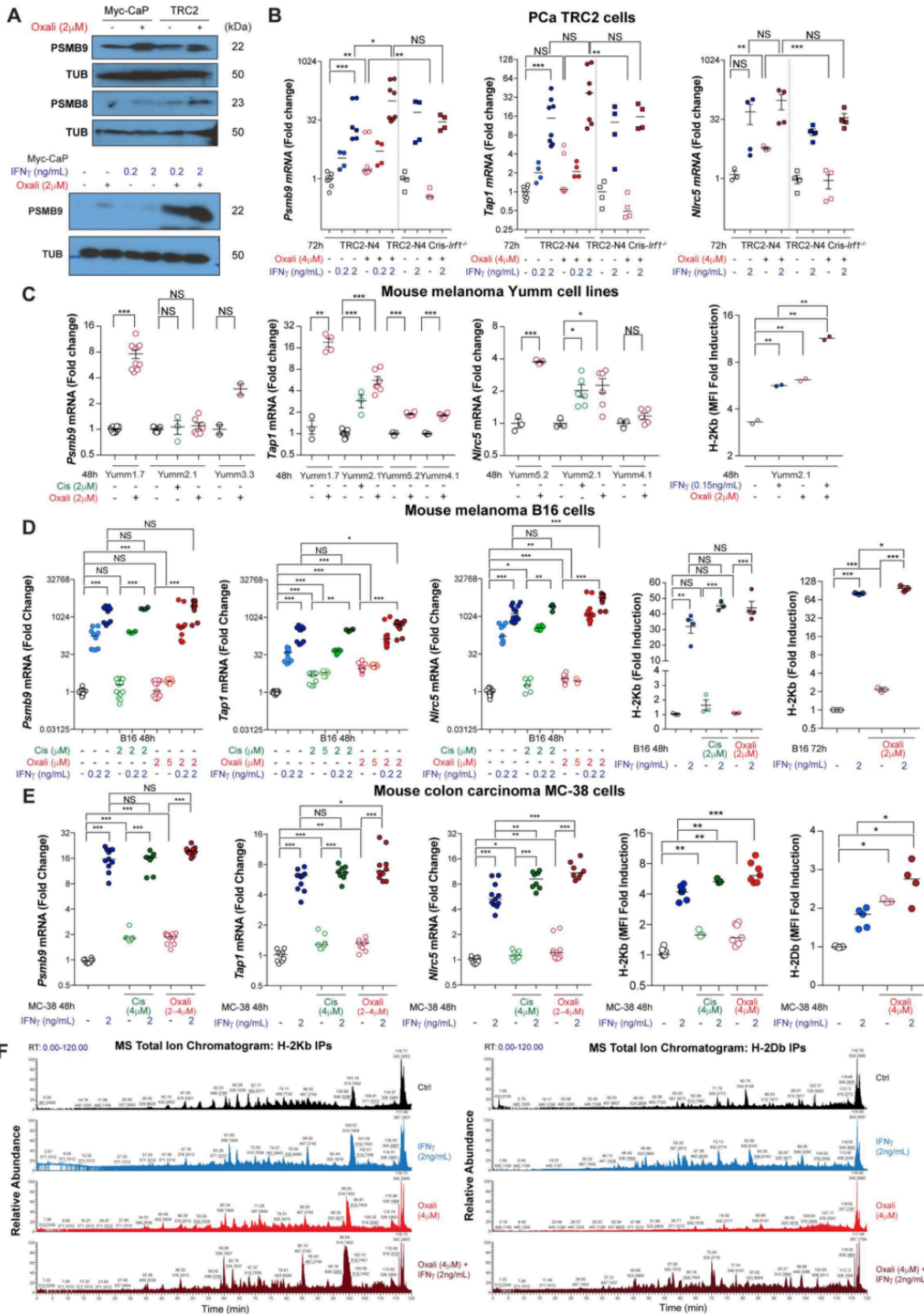


Figure S2.2: Platinoid induced expression of MHCI components and MHCI peptides in mouse cancer cell lines.

(A) Myc-CaP cells were incubated with the indicated Oxali, Cis, or IFN γ concentrations and IB analyzed for immunoproteasome (PSMB8 and PSMB9) subunit expression. Tubulin was used as a loading control. (B) Total RNAs extracted from WT and *Irf1*-ablated TRC2-N4 cells that were incubated with the indicated concentrations of IFN γ and Oxali for 72 hr were analyzed by qRT-PCR for expression of the indicated genes. (C) Mouse melanoma cell lines, Yumm1.7, Yumm2.1, Yumm3.3, Yumm4.1, and Yumm5.2, were incubated with Oxali or IFN γ as indicated and analyzed by qRT-PCR for expression of the indicated genes (left), while surface H-2Kb expression by Yumm2.1 cells was analyzed by flow cytometry (right). (D) B16 melanoma cells were incubated with Cis, Oxali, or IFN γ as indicated and analyzed by qRT-PCR for expression of the indicated genes (left), while surface H-2Kb expression was analyzed after 48 and 72 hr by flow cytometry (right). (E) Colon carcinoma MC-38 cells were incubated with Cis, Oxali, or IFN γ as indicated and analyzed by qRT-PCR (left) and flow cytometry (right) as above. (F) Colon carcinoma MC-38 cells were incubated with Oxali and IFN γ for 48 hr as indicated. Thereafter, cells were lysed and subjected to IP with H-2Kb or H-2Db antibodies. MHCI bound peptides were extracted and analyzed by mass spectrometry.



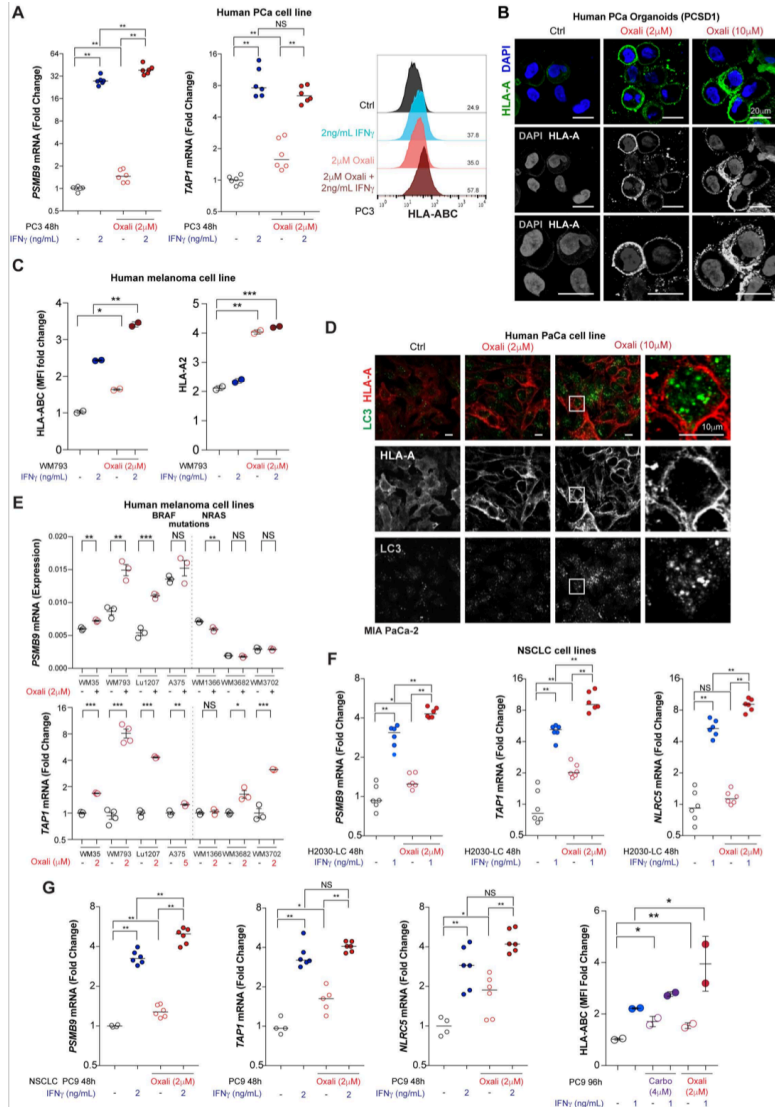


Figure S2.3: Platinoid-induced expression of MHC I antigen processing and presentation components in human cancer cell lines.

(A) Human PCa PC3 cells were incubated with IFN γ and Oxali for 48 hr and analyzed by flow cytometry for surface MHC expression (HLA-ABC) or by qRT-PCR for *PSMB9* and *TAP1* mRNA expression. (B) Human WM793 melanoma cells were incubated with Oxali and IFN γ for 48 hr as indicated and analyzed for surface MHC expression (HLA-ABC and HLA-A2) by flow cytometry. (C) Human MIA PaCa-2 cells were incubated with Oxali for 24 hr and stained with HLA-ABC (red) and LC3 (green) antibodies and counterstained with DAPI. Stained cells were examined by indirect immunofluorescence. Magnification bar: 10 μ m. (D) Human melanoma cell lines bearing *BRAF* (V600E) or *NRAS* mutations were incubated with Oxali and analyzed by qRT-PCR using primers for *PSMB9* and *TAP1*. (E) Human NSCLC H2030 cells were incubated with Oxali or IFN γ as indicated and analyzed by qRT-PCR using *PSMB9* and *TAP1* primers. (F) Human NSCLC PC9 cells were incubated with IFN γ , Oxali, or Carbo for 96 hr and analyzed by qRT-PCR for expression of indicated genes. Surface HLA-ABC expression was determined by flow cytometry.

2.8 Author Contributions

S.S. and M.K. conceived and designed the project. S.S. and I.N.B. designed and performed experiments and analyzed data with M.K. W.L., M.D., J.R.L., Q.T.P., L.A., X.Z., C.R.L., S.B., X.J.L., S.C., A.P., B.D., H.S., M.R.C., H.W., K.W., M.B., Z.C., J.H., L.K., A.B., L.D., K.F., and I.M. assisted with data collection and analysis. M.D. performed the human (TCGA) RNAseq and mouse scRNAseq analysis. J.R.L., Q.T.P., and L.D. performed the peptidomic analysis. J.H., L.K. and I.M collected and analyzed human specimens. A.B., K.F., Z.C., and M.R.C. generated and analyzed sequencing data. L.A. designed and performed ChIP experiments. M.B. provided melanoma cell lines. M.K. and S.S. wrote the manuscript, with all authors contributing and providing feedback and advice.

2.9 Acknowledgements

We thank A. Perkins for support and contribution. We thank C.R. Lichtenstern, D.T. Tam, S. Lee, E. Sanchez-Lopez, S. Pandit, K. Wong, and M. Muldong for technical support and help, and E. Fokas and C.M. Rödel for discussion and advise. S.S. was supported by PCF-Young Investigator Award, Merck MISP 57917 and the NIAAA funded SCRC for ALPD & Cirrhosis (P50 AA011999). Work in M.K. laboratory was supported by grants from the NIH (AI043477, CA128814, CA211794), Tower Cancer Research Foundation and San Diego NCI Cancer Centers Council (C3) and Padres Pedal the Cause #PTC2018. Additional support came from U01AA027681 to S.S. and M.K., integrated training grant for Physician-Scientists T32 GM008692 to I.N.B., P01 CA128814 to M.K./Ze'ev Ronai and M.B., U24CA232979 to S.L., DP5-OD017937 to H.C., 111 project (B16021) to X.L.. Funding for K.F., A.B. and ATAC-seq data analysis was from NIH grant UL1TR001442 of CTSA. S.B., in J. E. Dixon's laboratory, was supported by the Mary Kay Ash Breast Cancer grant 047.16 and the Moores Cancer Center Translational and Clinical Pilot Project.

Chapter 2, in full, is a reformatted presentation of the material currently being prepared for submission for publication as “p300/CBP-mediated activation of MHC-I machinery and IFN γ signaling controls” by Michelle Dow, Yixuan Zhou, Ingmar Niels Bastian, Mark D Long, Weihua Li, Tao Liu, Rachael Katie Ngu, Laura Antonucci, Jian Yu Huang, Qui T Phung, Xi-he Zhao, Daniel Karin, Brian Dang, Sourav Banerjee, Xue-Jia Lin, Hua Su, Sylvia Choi, Hongxia Want, Christina Jamieson, Marcus Bosenberg, Zhang Cheng, Amanda Birmingham, Johannes Haybaeck, Lukas Kenner, Kathleen M. Fisch, Richard Bourgon, Genevive Hernandex, Jennie R. Lill, Ira Mellman, Hannah Carter, Song Liu, Michael Karin, and Shabnam Shalapour. The dissertation author was one of the primary investigators and author of this material.

2.10 References

- Bilusic, M., Madan, R.A., and Gulley, J.L. (2017). Immunotherapy of Prostate Cancer: Facts and Hopes. *Clin. Cancer Res.* *23*, 6764–6770.
- Bracci, L., Schiavoni, G., Sistigu, A., and Belardelli, F. (2014). Immune-based mechanisms of cytotoxic chemotherapy: implications for the design of novel and rationale-based combined treatments against cancer. *Cell Death Differ.* *21*, 15–25.
- Bruno, P.M., Liu, Y., Park, G.Y., Murai, J., Koch, C.E., Eisen, T.J., Pritchard, J.R., Pommier, Y., Lippard, S.J., and Hemann, M.T. (2017). A subset of platinum-containing chemotherapeutic agents kills cells by inducing ribosome biogenesis stress. *Nat. Med.* *23*, 461–471.
- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* *10*, 1213–1218.
- Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* *109*, 21.29.1-9.
- Chabanon, R.M., Pedrero, M., Lefebvre, C., Marabelle, A., Soria, J.-C., and Postel-Vinay, S. (2016). Mutational Landscape and Sensitivity to Immune Checkpoint Blockers. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* *22*, 4309–4321.
- Chan, H.M., and La Thangue, N.B. (2001). p300/CBP proteins: HATs for transcriptional bridges and scaffolds. *J. Cell Sci.* *114*, 2363–2373.
- Chaney, S.G., Campbell, S.L., Bassett, E., and Wu, Y. (2005). Recognition and processing of cisplatin- and oxaliplatin-DNA adducts. *Crit. Rev. Oncol. Hematol.* *53*, 3–11.
- Chen, D.S., and Mellman, I. (2013). Oncology Meets Immunology: The Cancer-Immunity Cycle. *Immunity* *39*, 1–10.
- Chen, Q., Sun, L., and Chen, Z.J. (2016). Regulation and function of the cGAS–STING pathway of cytosolic DNA sensing. *Nat. Immunol.* *17*, 1142–1149.
- Chiu, S.-J., Lee, Y.-J., Hsu, T.-S., and Chen, W.-S. (2009). Oxaliplatin-induced gamma-H2AX activation via both p53-dependent and -independent pathways but is not associated with cell cycle arrest in human colorectal cancer cells. *Chem. Biol. Interact.* *182*, 173–182.
- Conforti, F., Pala, L., Bagnardi, V., De Pas, T., Martinetti, M., Viale, G., Gelber, R.D., and Goldhirsch, A. (2018). Cancer immunotherapy efficacy and patients' sex: a systematic review and meta-analysis. *Lancet Oncol.* *19*, 737–746.
- Corrales, L., Glickman, L.H., McWhirter, S.M., Kanne, D.B., Sivick, K.E., Katibah, G.E., Woo, S.-R., Lemmens, E., Banda, T., Leong, J.J., and Gajewski, T. (2015). Direct Activation of STING in the Tumor Microenvironment Leads to Potent and Systemic Tumor Regression and Immunity. *Cell Rep.* *11*, 1018–1030.

- Cresswell, P., Ackerman, A.L., Giodini, A., Peaper, D.R., and Wearsch, P.A. (2005). Mechanisms of MHC class I-restricted antigen processing and cross-presentation. *Immunol. Rev.* 207, 145–157.
- Denton, A.E., Wesselingh, R., Gras, S., Guillonneau, C., Olson, M.R., Mintern, J.D., Zeng, W., Jackson, D.C., Rossjohn, J., Hodgkin, P.D., Doherty, P.C., and Turner S. J. (2011). Affinity Thresholds for Naive CD8⁺ CTL Activation by Peptides and Engineered Influenza A Viruses. *J. Immunol.* 187, 5733–5744.
- Dilruba, S., and Kalayda, G.V. (2016). Platinum-based drugs: past, present and future. *Cancer Chemother. Pharmacol.* 77, 1103–1124.
- Eggermont, A.M.M., Blank, C.U., Mandala, M., Long, G.V., Atkinson, V., Dalle, S., Haydon, A., Lichinitser, M., Khattak, A., Carlino, M.S., and Robert, C. (2018). Adjuvant Pembrolizumab versus Placebo in Resected Stage III Melanoma. *N. Engl. J. Med.* 378, 1789–1801.
- El-Khoueiry, A.B., Sangro, B., Yau, T., Crocenzi, T.S., Kudo, M., Hsu, C., Kim, T.-Y., Choo, S.-P., Trojan, J., Welling, T.H., and Melero, I. (2017). Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): an open-label, non-comparative, phase 1/2 dose escalation and expansion trial. *Lancet Lond. Engl.* 389, 2492–2502.
- Galluzzi, L., Buqué, A., Kepp, O., Zitvogel, L., and Kroemer, G. (2015). Immunological Effects of Conventional Chemotherapy and Targeted Anticancer Agents. *Cancer Cell* 28, 690–714.
- Gandhi, L., Rodríguez-Abreu, D., Gadgeel, S., Esteban, E., Felip, E., De Angelis, F., Domine, M., Clingan, P., Hochmair, M.J., Powell, S.F., and Garassino, M.C. (2018). Pembrolizumab plus Chemotherapy in Metastatic Non-Small-Cell Lung Cancer. *N. Engl. J. Med.*
- Gettinger, S., Choi, J., Hastings, K., Truini, A., Datar, I., Sowell, R., Wurtz, A., Dong, W., Cai, G., Melnick, M.A., Herbst, R.S. and Politi, K. (2017). Impaired HLA Class I Antigen Processing and Presentation as a Mechanism of Acquired Resistance to Immune Checkpoint Inhibitors in Lung Cancer. *Cancer Discov.* 7, 1420–1435.
- Goswami, S., Aparicio, A., and Subudhi, S.K. (2016). Immune Checkpoint Therapies in Prostate Cancer. *Cancer J. Sudbury Mass* 22, 117–120.
- Guo, S., Contratto, M., Miller, G., Leichman, L., and Wu, J. (2017). Immunotherapy in pancreatic cancer: Unleash its potential through novel combinations. *World J. Clin. Oncol.* 8, 230–240.
- Hartinger, C.G., Tsybin, Y.O., Fuchser, J., and Dyson, P.J. (2008). Characterization of Platinum Anticancer Drug Protein-Binding Sites Using a Top-Down Mass Spectrometric Approach. *Inorg. Chem.* 47, 17–19.
- Hossain, M.K., Nahar, K., Donkor, O., and Apostolopoulos, V. (2018). Immune-based therapies for metastatic prostate cancer: an update. *Immunotherapy* 10, 283–298.
- Isaacsson Velho, P., and Antonarakis, E.S. (2018). PD-1/PD-L1 pathway inhibitors in advanced prostate cancer. *Expert Rev. Clin. Pharmacol.* 11, 475–486.

- Jongsma, M.L.M., Guarda, G., and Spaapen, R.M. (2017). The regulatory network behind MHC class I expression. *Mol. Immunol.*
- Keir, M.E., Butte, M.J., Freeman, G.J., and Sharpe, A.H. (2008). PD-1 and Its Ligands in Tolerance and Immunity. *Annu. Rev. Immunol.* *26*, 677–704.
- Kepp, O., Semeraro, M., Bravo-San Pedro, J.M., Bloy, N., Buqué, A., Huang, X., Zhou, H., Senovilla, L., Kroemer, G., and Galluzzi, L. (2015). eIF2 α phosphorylation as a biomarker of immunogenic cell death. *Semin. Cancer Biol.* *33*, 86–92.
- Kobayashi, K.S., and Elsen, P.J. van den (2012). NLRC5: a key regulator of MHC class I-dependent immune responses. *Nat. Rev. Immunol.* *12*, 813–820.
- Kroemer, G., Galluzzi, L., Kepp, O., and Zitvogel, L. (2013). Immunogenic Cell Death in Cancer Therapy. *Annu. Rev. Immunol.* *31*, 51–72.
- Langer, C.J., Gadgeel, S.M., Borghaei, H., Papadimitrakopoulou, V.A., Patnaik, A., Powell, S.F., Gentzler, R.D., Martins, R.G., Stevenson, J.P., Jalal, S.I., Raftopoulos, H. and Gandhi, L. (2016). Carboplatin and pemetrexed with or without pembrolizumab for advanced, non-squamous non-small-cell lung cancer: a randomised, phase 2 cohort of the open-label KEYNOTE-021 study. *Lancet Oncol.* *17*, 1497–1508.
- Li, P., Spolski, R., Liao, W., Wang, L., Murphy, T.L., Murphy, K.M., and Leonard, W.J. (2012). BATF–JUN is critical for IRF4-mediated transcription in T cells. *Nature* *490*, 543–546.
- McWhinney, S.R., Goldberg, R.M., and McLeod, H.L. (2009). Platinum neurotoxicity pharmacogenetics. *Mol. Cancer Ther.* *8*, 10–16.
- Meissner, T.B., Liu, Y.-J., Lee, K.-H., Li, A., Biswas, A., van Eggermond, M.C.J.A., van den Elsen, P.J., and Kobayashi, K.S. (2012). NLRC5 cooperates with the RFX transcription factor complex to induce MHC class I gene expression. *J. Immunol. Baltim. Md 1950* *188*, 4951–4958.
- van Montfoort, N., van der Aa, E., and Woltman, A.M. (2014). Understanding MHC Class I Presentation of Viral Antigens by Human Dendritic Cells as a Basis for Rational Design of Therapeutic Vaccines. *Front. Immunol.* *5*.
- Motzer, R.J., Tannir, N.M., McDermott, D.F., Arén Frontera, O., Melichar, B., Choueiri, T.K., Plimack, E.R., Barthélémy, P., Porta, C., George, S., George, S., Powles, T., and Donskov, F. (2018). Nivolumab plus Ipilimumab versus Sunitinib in Advanced Renal-Cell Carcinoma. *N. Engl. J. Med.* *378*, 1277–1290.
- Mukherjee, S.P., Behar, M., Birnbaum, H.A., Hoffmann, A., Wright, P.E., and Ghosh, G. (2013). Analysis of the RelA:CBP/p300 Interaction Reveals Its Involvement in NF- κ B-Driven Transcription. *PLoS Biol.* *11*, e1001647.
- Park, S., Martinez-Yamout, M.A., Dyson, H.J., and Wright, P.E. (2013). The CH2 domain of CBP/p300 is a novel zinc finger. *FEBS Lett.* *587*, 2506–2511.

- Paz-Ares, L., Luft, A., Vicente, D., Tafreshi, A., Gümüş, M., Mazières, J., Hermes, B., Çay Şenler, F., Csöszi, T., Fülöp, A., Rodríguez-Cid, J., and Wilson, J. (2018). Pembrolizumab plus Chemotherapy for Squamous Non-Small-Cell Lung Cancer. *N. Engl. J. Med.*
- Pfirschke, C., Engblom, C., Rickelt, S., Cortez-Retamozo, V., Garris, C., Pucci, F., Yamazaki, T., Poirier-Colame, V., Newton, A., Redouane, Y., Zitvogel, L., Weissleder, R., and Pittet, M.J. (2016). Immunogenic Chemotherapy Sensitizes Tumors to Checkpoint Blockade Therapy. *Immunity* *44*, 343–354.
- Puisset, F., Schmitt, A., and Chatelut, E. (2014). Standardization of Chemotherapy and Individual Dosing of Platinum Compounds. *Anticancer Res.* *34*, 465–470.
- Rock, K.L., York, I.A., and Goldberg, A.L. (2004). Post-proteasomal antigen processing for major histocompatibility complex class I presentation. *Nat. Immunol.* *5*, 670–677.
- Rotte, A., Bhandaru, M., Cheng, Y., Sjoestroem, C., Martinka, M., and Li, G. (2013). Decreased expression of nuclear p300 is associated with disease progression and worse prognosis of melanoma patients. *PloS One* *8*, e75405.
- Schachter, J., Ribas, A., Long, G.V., Arance, A., Grob, J.-J., Mortier, L., Daud, A., Carlino, M.S., McNeil, C., Lotem, M., Ibrahim, N., and Robert, C. (2017). Pembrolizumab versus ipilimumab for advanced melanoma: final overall survival results of a multicentre, randomised, open-label phase 3 study (KEYNOTE-006). *Lancet Lond. Engl.* *390*, 1853–1862.
- Shahbazian, M.D., and Grunstein, M. (2007). Functions of Site-Specific Histone Acetylation and Deacetylation. *Annu. Rev. Biochem.* *76*, 75–100.
- Shalapour, S., Font-Burgada, J., Caro, G.D., Zhong, Z., Sanchez-Lopez, E., Dhar, D., Willimsky, G., Ammirante, M., Strasner, A., Hansel, D.E., Birner, P., Kenner, L., and Karin, M. (2015). Immunosuppressive plasma cells impede T-cell-dependent immunogenic chemotherapy. *Nature* *521*, 94–98.
- Shalapour, S., Lin, X.-J., Bastian, I.N., Brain, J., Burt, A.D., Aksenov, A.A., Vrbanac, A.F., Li, W., Perkins, A., Matsutani, T., and Karin, M. (2017). Inflammation-induced IgA⁺ cells dismantle anti-liver cancer immunity. *Nature* *551*, 340–345.
- Sharma, P., Hu-Lieskovan, S., Wargo, J.A., and Ribas, A. (2017). Primary, Adaptive and Acquired Resistance to Cancer Immunotherapy. *Cell* *168*, 707–723.
- Snyder, A., Makarov, V., Merghoub, T., Yuan, J., Zaretsky, J.M., Desrichard, A., Walsh, L.A., Postow, M.A., Wong, P., Ho, T.S., and Chan, T. (2014). Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* *371*, 2189–2199.
- Soori, H., Rabbani-Chadegani, A., and Davoodi, J. (2015). Exploring binding affinity of oxaliplatin and carboplatin, to nucleoprotein structure of chromatin: Spectroscopic study and histone proteins as a target. *Eur. J. Med. Chem.* *89*, 844–850.

Terranova-Barberio, M., Thomas, S., Ali, N., Pawlowska, N., Park, J., Krings, G., Rosenblum, M.D., Budillon, A., and Munster, P.N. (2017). HDAC inhibition potentiates immunotherapy in triple negative breast cancer. *Oncotarget* *8*, 114156–114172.

Thompson, P.R., Wang, D., Wang, L., Fulco, M., Pediconi, N., Zhang, D., An, W., Ge, Q., Roeder, R.G., Wong, J., Cotter, R.J., and Cole, P.A. (2004). Regulation of the p300 HAT domain via a novel activation loop. *Nat. Struct. Mol. Biol.* *11*, 308–315.

Tscharke, D.C., Croft, N.P., Doherty, P.C., and Gruta, N.L.L. (2015). Sizing up the key determinants of the CD8⁺ T cell response. *Nat. Rev. Immunol.* *15*, 705–716.

Wang, R., Natarajan, K., and Margulies, D.H. (2009). Structural basis of the CD8 alpha beta/MHC class I interaction: focused recognition orients CD8 beta to a T cell proximal position. *J. Immunol. Baltim. Md 1950* *183*, 2554–2564.

Wojciak, J.M., Martinez-Yamout, M.A., Dyson, H.J., and Wright, P.E. (2009). Structural basis for recruitment of CBP/p300 coactivators by STAT1 and STAT2 transactivation domains. *EMBO J.* *28*, 948–958.

Ylitalo, E.B., Thysell, E., Jernberg, E., Lundholm, M., Crnalic, S., Egevad, L., Stattin, P., Widmark, A., Bergh, A., and Wikström, P. (2017). Subgroups of Castration-resistant Prostate Cancer Bone Metastases Defined Through an Inverse Relationship Between Androgen Receptor Activity and Immune Response. *Eur. Urol.* *71*, 776–787.

Yoshihama, S., Roszik, J., Downs, I., Meissner, T.B., Vijayan, S., Chapuy, B., Sidiq, T., Shipp, M.A., Lizee, G.A., and Kobayashi, K.S. (2016). NLRC5/MHC class I transactivator is a target for immune evasion in cancer. *Proc. Natl. Acad. Sci.* *113*, 5999–6004.

Zhou, F. (2009). Molecular Mechanisms of IFN- γ to Up-Regulate MHC Class I Antigen Processing and Presentation. *Int. Rev. Immunol.* *28*, 239–260.

CHAPTER 3: Using *in vitro* evolution to probe the genome-wide basis of drug resistance to chemotherapy

3.1 Abstract

Understanding general mechanisms of chemotherapy resistance can assist with the design of new drugs and guide cancer treatment decisions. Here we applied *in vitro* evolution and whole genome analysis (IVIEWGA) to the human, near-haploid cell line (HAP-1) to directly identify resistance alleles. Clones (N=28) were treated with five different anticancer drugs (doxorubicin, gemcitabine, etoposide, topotecan, and paclitaxel) until they developed resistance, and compared to their isogenic drug-sensitive parents via whole genome and whole exome sequencing (WES). High frequency alleles predicted to change protein sequence, or alleles which appeared in the same gene for multiple independent selections with the same compound were identified in only 21 genes: The set included clinically-relevant resistance genes or drug targets (*TOP1*, *TOP2A*, *DCK*, *WDR33*, *SLCO3A1*), as well as new genes (*SLC13A4*). In addition, some lines carried structural variants that encompassed additional known resistance genes (*ABCB1*, *WWOX* and *RRM1*). Gene expression knockdown and knockout experiments (via shRNA and CRISPR-Cas9 respectively) of 10 validation targets showed a high degree of specificity and accuracy from our candidate resistance genes. Our work demonstrates that the drug resistance mechanisms found in diverse clinical samples can be identified and studied in an isogenic background in the laboratory.

3.2 Introduction

Recent advances in the development of anticancer agents have contributed significantly to the improvement of prognosis and remission of cancer patients. However, acquisition of drug-resistance, even in cases of favorable prognosis, leading to worse patient survival and recurrence (Brown et al. 2014; Agarwal and Kaye 2003). Numerous studies have used cancer cell-line panels to screen for modifications to the genome with pharmacological implication across major cancer types (Barretina et al. 2012; Weinstein et al. 1997; Greshock et al. 2010). These studies have typically focused on identifying molecular biomarkers of drug sensitivity but have lacked characterization of the routes of evasion that lead to drug resistance and a loss of efficacy (Ko et al. 2019). The ability to predict the routes of drug resistance would allow design of better therapies, leading to lower incidence of tumor recurrence. Thus, it is important to develop new strategies to uncover molecular characteristics that contribute to drug resistance.

Current advancement of DNA microarrays, proteomics technology and RNA-sequencing have contributed to the discovery of drug resistance genes and provided new avenues and potential clues to develop new targeted therapies to overcome the drug resistance (Mansoori et al. 2017; Januchowski et al. 2017; Ju et al. 2017). Gene expression has been more commonly used to identify resistance associated genes (Tong et al. 2015), however, expression-based analysis does not provide information about the effects of mutations on protein activity or capture genetic heterogeneity between patients.

Work in other organisms has shown that *in vitro* evolution and whole genome analysis (IVIEWGA) is a powerful method to discover both a comprehensive set of drug resistance alleles, as well as the targets of compounds with unknown mechanisms of action (Antonova-Koch et al. 2018; Cowell et al. 2018). In this method, clonal or near clonal organisms are isolated and then

clones are subjected to increasing levels of a drug that inhibits growth. After selection, the organism is cloned again. The genomes of resistant clones are then compared to the sensitive parent clone using next generation sequencing (NGS) methods. In organisms such as *Saccharomyces cerevisiae* (Ottillie et al. 2017), *Plasmodium falciparum* (Antonova-Koch et al. 2018; Cowell et al. 2018), Mycobacteria (Andries et al. 2005), and trypanosomes (Ottillie et al. 2017), this method has been used to comprehensively discover resistance conferring variants. Surprisingly, the data shows that typically only a small number of *de novo* variants are detected after evolution. If multiple selections are performed on independent clones, the same resistance gene will appear repeatedly, although often appearing with different alleles, providing a high level of statistical confidence that the allele has not arisen by chance.

Many of the organisms on which IVIEWGA has been used with success have both haploid and diploid phases of their lifecycle, which means that selections can be performed in a haploid stage. Selecting for resistant clones in a haploid organism greatly simplifies analysis as a driver resistance allele will approach 100% frequency. In addition, for loss of function alleles, only one mutation is needed (versus both copies). Finally, work with clonal lines allows direct hypothesis-testing and the ability to measure the small, yet reproducible changes in resistance that result from reintroducing single mutations. Although metazoans are all diploid, haploid human cells lines are nevertheless available: HAP-1, is a human chronic myelogenous leukemia (CML)-derived cell line that is completely haploid except for a 30 megabase fragment of chromosome 15 (Bürckstümmer et al. 2013). HAP-1 has been used for genetic studies because mutated phenotypes are immediately exposed (Carette et al. 2011).

Using five different anticancer drugs (Doxorubicin, Gemcitabine, Etoposide, Topotecan, and Paclitaxel) as examples, we show that *in vitro* evolution in HAP-1 cells can be used to study

cancer drug resistance. Through our analysis of evolved clones, we detect a limited number of genes that acquire SNVs or CNVs after prolonged, sublethal exposure to our selected xenobiotics. We further demonstrate the power of the approach by using shRNAs and CRISPR-Cas9 to downregulate or reintroduce selected alleles and demonstrate that this confers resistance or sensitivity to the drug which elicited the evolved genomic change. Our work has implications for clinical intervention strategies to prevent the emergence of drug resistance and tumor recurrence through gene mutations acquired through DNA damage from chemotherapeutics or natural variants which exist and persist from the heterogenous tumor cell environment.

3.3 Results

Selection of compounds for resistance studies. To identify xenobiotics with the best efficacy against HAP-1 cells we first measured ATP levels (CellTiterGlo) treating HAP-1 cells with serial dilutions of 16 different drug for 48 hours. Five drugs showed EC_{50} values between 5 to 340 nM (Fig. 3.1A-B, Table S3.1). These included doxorubicin (DOX, $EC_{50} = 46.05 \pm 4.6$ nM), also known as adriamycin, an anthracycline antibiotic that works by inhibiting topoisomerase II alpha (*TOP2A*) (Mandell et al. 2019; Jiang et al. 2019); gemcitabine (GEM, $EC_{50} = 8.7 \pm 0.7$ nM), a synthetic pyrimidine nucleoside prodrug that is used against a variety of hematopoietic malignancies (Drenberg et al. 2019; Alvarellos et al. 2014; Mini et al. 2006); etoposide (ETP, $EC_{50} = 338.6 \pm 39.7$ nM), a semisynthetic derivative of podophyllotoxin that also targets *TOP2A* and prevents re-ligation of the double-stranded DNA (Pommier et al. 2010); paclitaxel (PTX, $EC_{50} = 17.5 \pm 4.0$ nM) also known as taxol, an effective anticancer agent that targets tubulin, perturbing the cytoskeleton and causing M phase cell-cycle arrest (Hayashi and Karlseder 2013), and topotecan (TPT, $EC_{50} = 5.6 \pm 0.1$ nM), a semisynthetic water-soluble derivative of camptothecin that inhibits topoisomerase I (*TOP1*). Our HAP-1 EC_{50} values were similar to those previously reported for other CML cell lines (www.cancerrxgene.org) with the exception of etoposide, which appeared more effective in HAP-1 cells ($EC_{50} = 338.6 \pm 39.7$ nM) relative to other CML cell lines ($> 1 \mu\text{M}$ in BV-173, KU812, EM-2, MEG-01, JURL-MK1, KCL-22, RPMI-8866, LAMA-84, K-562).

Evolution of resistance is readily achieved for all compounds. Our next objective was to create drug resistant lines. Although we have had difficulty creating resistant lines for some drugs in some species (“irresistibles” (Corey et al. 2016)), there is precedent for resistance to the drugs included here (Christowitz et al. 2019; Amrutkar and Gladhaug 2017; Pan et al. 2013). To

reduce the possibility of selecting mutations that were present in a nonhomogenous population of HAP-1 cells and to facilitate later genomic analysis, we first cloned the cells. This was accomplished by diluting cells to an average density of ~0.5 cells per well in a poly-L-lysine treated 96-well plate (**Fig. 3.1A**) and then picking clones from wells that contained single colonies. Selections were initiated with different parent clones for the different drug-specific replicates (**Fig. 3.1C, Fig. S3.1**). To create drug resistant clones, cells were grown in tissue culture dishes (reaching 60-80% semi-confluence) in the presence of concentrations of each drug selected to kill a predefined percentage of cells (**Methods**). Because mutations will arise randomly during long term cell culture, we attempted at least three independent selections for each drug, in each case starting with an identical parental isolate (**Fig. 3.1C**). In a few cases, independent selections could not be achieved and dependent clones with a shared lineage (DOX-R4a and DOX-R4b; PTX-R2a and PTX-R2b; TPT-R4a, TPT-R4b and TPT-R4c) were collected. Resistance emerged after several months depending on the drug and the method used (7-30 weeks approximately, 49-210 generations) (**Fig. S3.1, Methods**).

Once resistance was observed in the batch culture, we isolated clones from batch drug-selected cultures and the drug sensitivity of the clone was measured and compared to the isogenic parent clones (**Fig. 3.1D**). We observed an EC_{50} fold shift between 3.3 and 411.7 (**Table S3.2**) in paired comparisons. To demonstrate that the drug resistance phenotype was stable, drug pressure was removed for 8 weeks (approximately 56 generations) and clones were retested for sensitivity. We observed no changes in the EC_{50} values, indicating that resistance phenotypes were not due to transient adaptation.

Identification of putative resistance variants using next-generation sequencing. We next performed whole genome and exome paired-end read sequencing on the 35 cell lines (both drug-resistant clones and their matched drug-sensitive parent clones). Our IVIEWGA studies in Plasmodium (Cowell et al. 2018), have shown that stable drug resistance is typically conferred by SNVs in coding regions and thus exome sequencing was an efficient mechanism to find causal variants. However, gene amplifications, which contribute to 1/3 of drug resistance events in Plasmodium (Cowell et al. 2018), are more accurately detected with WGS because exact chromosomal recombination sites, which may fall in intergenic regions, can be reconstructed from WGS data. Because of falling costs over the course of the project, more samples (N=21) were ultimately whole genome sequenced than whole exome sequenced (N=14).

Sequencing quality was high for all samples: alignment showed that, on average, 99.9% of 700 million WGS (40 million WES) reads mapped to the hg19 reference genome with 86% of the bases covered by 20 or more reads (**Table S3.3**). By comparing sequences of evolved clones to their respective parental clones, we discovered a total of 26,625 unique SNVs (**Table 3.1, Table S3.5, Methods**). The majority of variants in all cell lines were non-coding (**Table 3.1, Table S3.4, S3.5**) and were evenly distributed with respect to chromosome length (**Fig. S3.2**). We consider mutations from each resistance clone relative to their parental clone with high allele frequencies (AF) (AF > 0.85) as likely driver mutations, as they have higher percentage in the resistance cells, given that the sensitive parents were cloned (to the best of our ability) before resistance selections were initiated. Of the 26,625 mutations most (26,468) were present at AF < 0.85 and are more likely passenger and not relevant to the drug resistance mechanisms.

We next developed a pipeline (**Fig. 3.2A, Methods**) to filter the 26,625 mutations (Table S6) to a final list of potential variants conferring drug resistance (**Fig. 3.2A, Table 3.1**). Previous

analysis suggested that variants present in coding regions are more likely to contribute to drug resistance even though this could exclude the variants associated with certain transcription factor (TF) binding sites. Therefore, our strategy focused on mutations that were in exonic regions, and were drug-specific (**Fig. 3.2A**). We further considered only mutations likely to have a functional impact at the protein level (missense, frameshift, start or stop gain or loss) which reduced the number to an average of 35 and 23 nonsynonymous mutations for WGS and WES, respectively (**Fig. 3.2A**: detailed mutation summary in **Table 3.1**, **Table S3.4**). Reasoning that resistance driver mutations would be present in all haploid cells in the sequenced culture, we selected only the variants with high allele frequency ($AF > 0.85$). Although some with $AF < 0.85$ could confer a beneficial advantage to the cell, most are likely to be random mutations that arose during long term cultures. Finally, based on our experience with microbes whereupon genes with multiple, predicted independent amino acid changes (not expected by chance) are often found for the same drug, we added these genes to our final list of candidates (**Table S3.8**). The five drugs varied in the number of mutations with TPT having the highest overall numbers (**Table 3.1**).

Somatic Copy number variations (CNVs). We next searched for CNVs (both amplifications and deletions) in our WGS and WES data using Control-FreeC (Boeva et al. 2012). Overall patterns of broad and focal alterations across the drugs and conditions varied, with copy number gains observed more often than losses (**Fig. 3.2B**, **Table S3.7**). Using a corrected p-value of less than 0.05, we identified 93 total amplification and 108 deletion events, with most appearing in the TPT resistant samples (123) (**Table S3.7**). The CNVs had an average size of 8.5 Mbp (stdev 19 Mbp), ranged from 15,000 bp to 152 Mbp (**Fig. 3.2B**) and covered ~3% of the genome, on average. More CNVs were called in WES samples because of sequencing gaps—even for WGS samples, some CNVs were separated by short distances and were nearly contiguous (**Fig. 3.2B**).

The number of events was proportional to chromosome size with the exception of the Y chromosome, for which there were ~4X more events (47) per unit length.

The most frequently modified gene in all lines was *WWOX*: we identified 9 recurrent deletions affecting *WWOX* (16q23.1-q23.2). *WWOX* (WW domain-containing oxidoreductase) bears a well-known fragile site (FRA16D) (Krummel et al. 2000) at position chr16:78,569,021 that was adjacent to most CNVs and is discussed below. For WGS, CNV boundaries for rearrangements were manually confirmed by searching for paired end reads with abnormal insert sizes (shown as red bars) allowing the exact location of the recombination event to be derived from the short-read sequences, as shown for the *WWOX* deletion event (**Fig. 3.2C**, lower panel). These confirmatory data highlight the advantage of working on haploid cell lines because phasing does not need to be considered.

Although it can be challenging to determine the causative genes in an amplification or deletion event, some encompassed known candidates. For example, we found three amplifications affecting *ABCBI* (7q21.12, PTX) (**Fig. 3.2B**), and three events associated with *RRMI*. *ABCBI* is well known drug resistance gene and *WWOX* and *RRMI* were validated as described below. Some CNVs could contain enhancer regions. A manually-confirmed amplification was observed beginning within *GMDS* in two lines. *GMDS*, encoding GDP-mannose 4,6 dehydratase is adjacent to the well-known drug resistance gene, *FOXCI*. Some smaller CNVs encompassed only a few coding genes. For example, we observed a small amplification in two DOX-resistant lines that involved only four predicted protein-coding genes *RPS24* and *KCNMA1*, *DLG5* and *POLR3A*, which encodes RNA polymerase III and which has been recently linked to DOX resistance (X. Wang et al. 2019). A small amplification on chromosome IV in PTX-R2 involved only *ARHGAP10* and *PRMT9* (**Table S3.7**). *PRMT9* encodes an arginine methyltransferase, and

arginine methyltransferases are able to add methyl groups to proteins, playing a role in cancer progression (Jarrold and Davies 2019), but also potentially disrupting the interaction between PTX and tubulin (Raposo and Piller 2018). Methyltransferases are key determinants of resistance to antibiotics in bacteria. Further studies (genome-wide knockdowns or additional evolutions) will be needed to comprehensively determine which genes are important for resistance within the intervals.

Doxorubicin resistance is associated with mutations in TOP2A and a solute carrier transporter. We next considered the set of SNVs and CNVs for each drug. For DOX, we had six available selections from two different starting clones (WT-1 (WGS) and WT-5 (WES)) that were analyzed by WGS (DOX-R1, DOX-R2, DOX-R3) and by WES (DOX-R4a, DOX-R4b and DOX-R5) (**Fig. 3.3A**). High allele frequency missense mutations were found in only 11 genes (**Table S3.8**). Of note DOX-R1 and DOX-R2 carried mutations in TOP2A at frequencies of 0.89 and 0.87, respectively. This is the known target of DOX (Pommier et al. 2010; Jeon, Yu, and Kwon 2018) and is known to play a role in drug resistance (Jeon, Yu, and Kwon 2018; Ghisoni et al. 2019; Turner et al. 2016). The amino acid mutation, Pro803Thr (**Fig. 3.3E**), sits within the principal DNA-binding locus, the DNA-gate, a region conserved in type II topoisomerases (TOP2A and TOP2B). It is also adjacent to the catalytic tyrosine (Tyr805), responsible for nucleophilic attack on DNA (Wendorff et al. 2012). While one explanation is that Pro803Thr creates steric hindrance and blocks DOX access to the site, a more likely explanation is that the mutation is a loss-of-function mutation, especially as knockdown of TOP2A activity has previously been shown to confer DOX resistance in a E μ -Myc mouse lymphoma system (Burgess et al. 2008). To reproduce these results in our HAP-1 human cells, TOP2A was downregulated using a shRNA pool containing constructs encoding target-specific shRNA hairpins for TOP2A. Western blots further

showed the expected down regulation of protein levels (**Fig. 3.3F**) and an EC_{50} analysis of the wildtype and the knockdown line revealed a 4.25-fold increase in DOX resistance compared to the isogenic parent (**Fig. 3.3G, H**).

We also found missense mutations present in 100% of the reads for several other attractive but less well characterized genes; *SLC13A4* (Gln165His, DOX-R4b), and *SPG7* (Lys593Asn, DOX-R5), as well as one uncharacterized gene (AC091801.1, His13Asn, DOX-R4a) in the three different clones that were subjected to WES and were derived from WT-5. *SLC13A4* is a solute carrier transport family member and members of this general solute carrier family have appeared in selections conducted in microbes (e.g. the UDP-galactose transporter and the AcetylCoA transporter (Lim et al. 2016)) and are also associated with cancer drug resistance (Mansoori et al. 2017). The Gln165His mutation is located in the disordered region of the protein. To validate *SLC13A4* we performed a gene knockdown using a shRNA pool that targeted *SLC13A4*, containing three expression constructs each encoding target-specific 19-25 nucleotide shRNA hairpins. Protein expression levels of the knockdown line were measured by western blot followed by a dose-response assay to compare its EC_{50} value with the wildtype line (**Fig. 3.3B**). The 4X increase in resistance demonstrates that *SLC13A4* contributes to resistance, although it may not account completely for the level of resistance of the sequenced clones, which ranged from 4X to 11X (**Fig. 3.3C, D**). It should be noted that 28 CNVs were found in DOX resistant lines, including three loss alleles around *WWOX* for DOX-R3, R4a and R5 (**Table S3.7, Fig. S3.3**), and these may further contribute to levels of resistance.

We also investigated *SPG7*, which encodes a matrix AAA peptidase. Mutations in this gene in humans are associated with spastic paraplegia (De la Casa-Fages et al. 2019) but biochemical characterization has demonstrated that the protein is a member of the mitochondrial permeability

transition pore (PTP) (Shanmughapriya et al. 2015) and knockdown of *SPG7* can significantly prevent ionomycin or H₂O₂-induced cell death by blocking Ca²⁺ and ROS-induced mitochondrial membrane depolarization in HEK293T and HeLa cells (Shanmughapriya et al. 2015). Furthermore, mutations in the homologous gene in *C. elegans* confer resistance to some compounds via mitochondrial dysfunction (Zubovych, Straud, and Roth 2010). Although *SPG7* could be a candidate resistance gene for compounds such as DOX that kill cells by inducing apoptosis (S. Wang et al. 2004), we were unable to detect protein expression for this gene in WT HAP-1 cells using two independent antibodies, suggesting that this allele may be a passenger mutation. It is important to consider that a few high allele-frequency passenger mutations may be expected given the genome size.

Gemcitabine resistance is conferred by changes in DCK and RRM1 activity. Six selections were performed with GEM (starting from two different isogenic parents; WT-2 (WGS) and WT-3 (WES)). Among those, three GEM-resistant clones subjected to WGS (GEM-R1, GEM-R2 and GEM-R3) showed an average EC₅₀ shift of 300 to 400-fold (**Fig. 3.4A, Table S3.2**), and the clones showed no change in HAP-1 sensitivity to other drugs (**Fig. 3.4B**). As there were no candidate alleles with AF > 0.85, we looked for genes that acquired mutations in multiple selections without the AF threshold, identifying deoxycytidine kinase (*DCK*) as likely important for resistance. Interestingly, across cell lines several distinct mutations were found in *DCK*, with varying effects (missense and frameshift) across several different positions (**Table 3.2**). In particular, the missense substitution Ser129Tyr, present in GEM-R1 and GEM-R3, not only alters the amino-acid size and charge also falls at the end of exon 3, within the ATP-binding pocket of a phosphorylation site, making it a strong causal candidate for GEM drug resistance (**Fig. 3.4C**). GEM only becomes pharmacologically active if it is phosphorylated and the first phosphorylation

is catalyzed by DCK. shRNA knockdown of DCK was confirmed by western blot analysis (**Fig. 3.4D**). Downregulation of the gene showed a 36.5-fold increase in the EC_{50} value compared to both the isogenic parent line and the shRNA negative control (**Fig. 3.4E, F; Table 3.2**).

The three WT-3 derived GEM-resistant clones (GEM-R4, GEM-R5 and GEM-R6) subjected to WES were not as resistant as those used in WGS (~6X versus ~400X, **Table S3.2**). No high AF SNVs were evident in these lines and *DCK* exons were not mutated. However, multiple mutations were found in the gene encoding the centrosomal protein *STARD9* (**Table S3.8**), which promotes microtubule stability and regulates spindle microtubule dynamics (Srivastava and Panda 2018). Given the size of the human genome, the probability of finding two independent missense mutations (His1021Tyr and Ser1330Ile) in different clones exposed to the same drug by random chance is low, especially for genes that are known not to be easily mutated. On the other hand, the three WES clones contained 20 CNVs that could play a role in drug resistance. Most CNVs were not shared between lines but GEM-R4, GEM-R5 and GEM-R6 all bore overlapping CNVs of varying sizes on chromosome 11, with all three lines bearing 3-4 copies (p value = 1.38e-37 to 2.05e-142) (**Fig. 3.3**). The chromosome 11 CNV was only found in GEM resistant lines and not in any of the other evolved lines (in contrast to CNVs on chromosome 1 or 16, for example). While it is difficult to determine which of the 140 genes in the smallest interval contribute to resistance, a known resistance mediator or target of GEM, ribonucleotide reductase (*RRM1*), is found within the amplified region. *RRM1* is the largest catalytic subunit of ribonucleotide reductase, a key enzyme catalyzing the transformation of ribonucleotide diphosphates to deoxyribonucleoside diphosphates that are required for DNA synthesis and repair, and GEM is known to inhibit DNA polymerase by inhibiting *RRM1* (Bepler et al. 2006). Furthermore,

overexpression of *RRM1* is associated with poorer prognosis after gemcitabine treatment in non-small cell lung cancer (Zeng, Fan, and Zhang 2015) and in bladder cancer (Z. Yang et al. 2019).

Western blot analysis of the evolved lines showed that the amplification was indeed associated with increased protein levels (**Fig. 3.4G**). As an additional validation, we performed a single shRNA knockdown of *RRM1* to reduce protein expression (**Fig. 3.4H**), followed by a dose-response assay comparing EC_{50} values of both wildtype HAP-1 and *RRM1* knockdown lines, which showed that downregulation of *RRM1* made HAP-1 cells 31-fold more sensitive to GEM than their isogenic parent (**Fig. 3.4I, J**). As expected *RRM1* downregulation had no effect on HAP-1 sensitivity to other drugs (**Fig. S3.4**).

Etoposide resistance is modulated by levels of WDR33. We created three independent ETP resistant clones, all of which were subjected to WES, and compared them to one isogenic parent clone (WT-3) (23, 13 and 9-fold increased resistance respectively (**Fig. S3.5A, Table S3.2**). A single gene, *WDR33* (ETP-R3), carried a SNV with a 100% allele frequency. This gene encodes for a member of the *WD* repeat protein family and is one of the six subunits of a multiprotein complex called *CPSF* (cleavage and polyadenylation specific factor) (Chan et al. 2014) involved in cell cycle progression, signal transduction, apoptosis and gene regulation. Disruption of *WDR33* can lead to slowed DNA replication forks (Teloni et al. 2019), which could potentially explain why its disruption protects against topoisomerase inhibitors that block DNA unwinding. Lines in which *WDR33* (Pro622Thr) was knocked down via shRNA acquired an EC_{50} value 3.4 times greater than its parental line or the scrambled control (**Fig. S3.5B-D; Table 3.2**).

No clear candidate SNVs were evident for ETP-R1 and ETP-R2, which did not carry the *WDR33* mutation (**Table S3.6, Table S3.8**). All ETP lines carried multiple CNVs, however, including a large shared amplification on chromosome 15 (ETP-R1 and R3). Approximately 120

protein coding genes are found in this region, including *BUB1B*, the BUB1 mitotic checkpoint serine/threonine kinase B, *BMF*, a BCL-modifying factor, as well as the *RAD51* recombinase, whose overexpression has been previously shown to confer ETP resistance (Hansen et al. 2003). Inhibition of *RAD51* activity sensitizes cells to a variety of cancer drugs, including etoposide, doxorubicin and topotecan (F. Huang and Mazin 2014). Notably, all ETP resistant lines were also cross-resistant to PTX, TPT and DOX, providing support for this general mechanism of resistance. ETP-R2 shows a copy number gain that encompassed a region of chromosome 1 with ABC transporters (*ABCB10*) and cancer associated genes such as *AKT3* and *WNT3A*. This chromosome 1 region was amplified in several other drug resistant lines (GEM-R3, TPT-R1, R3 and R4b), in some cases with smaller boundaries (the GEM-R3 CNV only has 18 genes) and inactivation of the Wnt/ β -catenin pathway has been shown to resensitize some tumor lines to chemotherapy (Z. Zhang et al. 2018). Efforts to disrupt *WNT3A* were unsuccessful. Additional selections may be needed to further winnow down the list of potential ETP resistance genes.

Topotecan resistance is associated with complex alterations in TOP1, deletion of WWOX and SNVs in cytochrome p450s (CYP1B1). The six TPT samples were derived from four independent selection events (TPT-R4a-c are clones from the same selection with levels of resistance ranging from 10-20X; **Fig. 3.6A; Table S3.2**) and all six clones were subjected to WGS together with their parent clones (WT-6 and WT-7). No clear coding SNVs with a high allele frequency were obvious in TPT-R1, R2 and R3 but we noted multiple SNVs (Asp217Glu from TPT-R4a,b,c and Val432Leu from TPT-R1) in *CYP1B1*, which encodes a cytochrome p450 isoform. Gene knockdown experiments showed a small (2X) increase in TPT resistance with shRNA knockdown (**Fig. 3.6B-D**). Overexpression of *CYP1B1* has previously been associated with TPT resistance (Cruz-Muñoz et al. 2014).

For TPT-R4a-c lines, 268 alleles were present with AF > 0.85, but of these, only six were coding mutations and the rest were intergenic. Three of the six coding mutations were frameshift mutations (His81) with AF = 1 in *TOP1* (**Fig. S3.7A**), the known target of topotecan (Pfister et al. 2009). Interestingly, although annotated as a frameshift, we observed no loss of protein with our antibodies (**Fig. S3.7B**), nor size differences, suggesting that the mutation, located in the important but poorly characterized N terminal region, may be a gain-of-function mutation (Wright et al. 2015). Sequencing of the cDNA near the His81 “frameshift” for all three lines and as well as the parent line confirmed the two base deletion in the mutant as well as homozygosity in TPT-R4a-c evolved lines but showed no clear alternative splicing pattern. Two of the three showed a

In addition, TPT resistant lines had a greater number of CNVs than other evolved lines. Control-FreeC analysis showed most of chromosome 20 amplified in TPT-R4c but chromosome 20 deletions in the vicinity of *TOP1* for TPT-R4a and TPT-R4b (**Fig. 3.2B, Table S3.7**). Inspection of the raw reads confirmed a complex structural variant involving *TOP1* in TPT-R4a-c lines (**Fig. 3.6E**) near 20q12 (chr20:36,738,012 to 42,330,882) (**Fig. 3.6F-H, Fig. S3.7C**). By investigating the raw copy number profiles in TPT-R4a-c and examining the region in IGV, we found the frameshift event to be within a “deleted” region occurring within an “amplified” region (**Fig. 3.6F**) for all three lines. These changes provide clues about the process by which TPT resistance was acquired, with amplification of chromosome 20 developing first, followed by a loss of heterozygosity in *TOP1* after the acquisition of the frameshift allele. Control-FreeC and IGV analysis additionally showed two TPT (R4a and c) samples had amplifications on chromosome 13 (from position 89311517 to the end of the chromosome) that included *ABCC4* (*MRP4*), an ABC transporter previously shown to be clinically associated with TPT resistance (Leggas et al. 2004).

As mentioned above, TPT resistant lines (TPT-R1, R2 and R3) also showed large chromosomal abnormalities at *WWOX* (**Fig. 3.2C**) with a clear deletion of the *WWOX* gene region (chr16:78,569,166-78,792,736, exon7 and 8). The absence of protein was confirmed by western blot analysis (**Fig. 3.6I, J**). Knockdown of *WWOX* by shRNA resulted in marked resistance to TPT (**Fig. 3.6K-M**). Several lines also had amplifications, including a region of chromosome 1 with an ABC transporter (*ABCB10*) and cancer associated genes such as *AKT3* and *WNT3A*, also found in ETP lines, as described above.

Additional mutations likely contribute to drug resistance levels. In creating our drug resistant clones, we likely selected for clones that had higher levels of resistance than could be explained by a single mutation. Such contributing mutations would be expected to be present at allele frequencies of less than 0.85. Altogether there were 96 genes which presented with code-changing allele frequencies of between 0.5 and 0.85. Some of these have known roles in resistance. For example, *SGPPI* (Val273Phe), encoding a Sphingosine-1-Phosphate Phosphatase, implicated in drug resistance (X. Huang et al. 2013), was mutated in ETP-R2 and showed an allele frequency of 0.68. The transcription factor, Krüppel-Like Factor 1 *KLF1* (AF = 0.6325) was mutated in ETP-R1 and a solute carrier transporter, *SLC22A25* (AF = 0.58) in ETP-R2. The shRNA knockdown studies of the hematopoietic-specific transcription factor *KLF1* (Gly198fs, ETP-R1) showed a resistance-fold increase of 3.2 (**Fig. S3.6A-C**).

Testing for absolute levels of resistance with shRNA knockdowns, which deliver an incomplete level of inhibition, is difficult. We were able to completely disrupt one contributing allele via CRISPR/Cas9: Although TPT resistance is best explained by mutations in *TOP1*, we nevertheless investigated *USP47*, a ubiquitin-specific peptidase, with stop codon (Arg408*, TPT) in TPT-R4a-c (AF = 0.38, 0.57, and 0.58) (**Table 3.3**). Previous studies have also shown that

USP47 contributes to cell survival and chemoresistance for different cancer types and that downregulation or depletion of *USP47* sensitizes cancer cells to chemotherapeutic agents, suggesting that *USP47* may represent a potential marker to overcome drug resistance (Naghavi et al. 2018; Weinstock et al. 2012; Yu et al. 2019). In addition, because we observed a stop codon, which is easier to recreate, validation was more feasible than for *TOP1*. The dose-response assay for *USP47* KO line showed that it was 2.8 times more sensitive than the WT HAP-1 line (**Fig. S3.6D-F**).

Multidrug resistance genes. Multidrug resistance (MDR) mechanisms provide resistance to functionally and structurally unrelated drugs (Saraswathy and Gong 2013) via activation of drug efflux, or DNA repair, disruption of signaling pathways, and reduction of drug uptake (Y. Yang et al. 2018). To investigate whether our resistance mechanisms were drug dependent, we interrogated all knockdown or knockout lines for cross-resistance. Our data suggests the majority of genes implicated through IVIEWGA were drug-specific (Fig. S3.4) but that disruption of several genes resulted in multidrug resistance (Fig. S3.8). Examples include the hematopoietic-specific transcription factor, KLF-1, which was identified in ETP selections but whose disruption resulted in GEM resistance (EC_{50} fold shift of 6.5X Fig. S3.8B). Likewise, TOP2A shRNA knockdown cells were cross-resistant with ETP, even though we did not observe acquired alterations to TOP2A in ETP-resistant lines. This is not unexpected as ETP is a TOP2A inhibitor. As expected, the DOX lines carrying the TOP2A Pro803Thr mutation were also resistant to ETP (15-fold shift in the EC_{50} relative to the parental line) (Fig. S3.8C). WWOX was identified as frequently mutated and disruption of WWOX using shRNAs resulted in cells that were ~7.5-fold (Fig. S8D-F) more resistant to PTX in addition to TPT (Fig. S3.3). Interestingly, *USP47* stop alleles were found in

TPT resistant lines (TPT-R4a-c) but generated cell susceptibility for PTX (showing 5.3-fold shift compared to the parental line) as well as TPT (Fig. S3.8A).

3.4 Discussion

Here, we show for the first time that *in vitro* evolution and whole genome analysis (IVIEWGA) can readily lead to the identification of drug resistance mechanisms in human cells. Our results show *in vitro* resistance acquisition and provide a framework for the determination of chemotherapy resistance alleles that may arise in patients.

Using IVIEWGA to study drug resistance can be time-consuming but very rewarding if performed correctly and with exact attention to lineage. Unlike bacteria, where mobile genetic elements often confer resistance, eukaryotic pathogens often acquire drug resistance by point mutations and CNVs. *In vitro* evolution has been used in malaria parasites to either discover or rediscover most, if not all (to our knowledge), clinically relevant drug resistance genes including the chloroquine resistance transporter (Cowell et al. 2018), the artemisinin resistance gene, *Pfkelch13* (Ariey et al. 2014), and well-known ABC transporters (Cowell et al. 2018). As with the HAP-1 cells used here, the process can be time-consuming with selection for *Pfkelch13* taking several years and requiring sequencing of intermediate clones to pinpoint genomic changes that correlated with jumps in resistance levels. Despite the challenges of working with malaria parasites (which grow slowly in liquid culture within human red cells, seldom reaching parasitemias beyond 10%), hundreds of *in vitro* evolutions have been performed. Based on our results here, resistance emerges in HAP-1 cells readily but more work will need to be done to determine if this is because of the compounds that were used. The rate of at which resistance arises based on the number of population doublings and the number of starting cells is now used to prioritize drug candidates for development in malaria parasites.

Our work using IVIEWGA in pathogens (see (Luth et al. 2018) for a review) guided our pipeline development: We focused on protein coding alterations that arose in association with a

single treatment condition. Overall, our results are similar to what we have observed in eukaryotic pathogens with a mix of CNVs and single nucleotide variants giving rise to resistance. In contrast to what we have observed in microbes, we find much more heterogeneity and a higher number of CNVs per cell line, potentially because of the much larger genome size. The number of observed CNVs may be an underestimation as well as some CNVs may not have achieved statistical significance, especially in our WES data. This could be addressed in the future by using different algorithms, as well sequencing to somewhat higher coverage.

Another difference is that in eukaryotic microbes three independent selection will often result in the repeated identification of the same gene, often with different alleles. Here, in our first set of GEM selections we created mutations in *DCK*, while in the second set of selections, we created amplifications of *RRMI*. This may be because there are more routes to resistance in HAP-1 cells or with the drugs used. In addition, different protocols (single dose versus step-wise increase) that were used here may have led to different outcomes: Alternatively, it may be the result of using a different starting clone. More work and repetition will be needed to distinguish between these possibilities.

Our initial study presented here ignores noncoding genomic alterations. However, non-coding RNAs such as *EGRF-ASI* and activating cis elements such as enhancers have previously been implicated in evasion of drug response (Si et al. 2019; Calle et al. 2018; Sur and Taipale 2016). The intergenic mutations with high allele frequency are present in our provided datasets and provide opportunity for reanalysis or for querying by those interested in a specific noncoding RNA or enhancer. It is feasible that even synonymous mutations could confer resistance if they altered the rate of protein folding.

As expected from our work in eukaryotic pathogens, variants found at high allele frequencies in resistant cell lines validated with a high rate of success, suggesting that high allele frequency events often contribute to resistance. Given that we frequently observed more than one attractive genetic change per cell line, the final phenotype of most drug resistant lines can be attributed to a complex combination of multiple resistance alleles. Without the painstaking task of introducing the exact selected allele into the genome via CRISPR-Cas9 it is difficult to determine the exact contribution of a given allele: shRNA knockdowns are often incomplete and thus it is difficult to reciprocate the effect of resistance alleles or compare the resulting EC_{50} s in a quantitative way. For clearer results, fewer generations of selection should be performed in the future. In addition, collecting intermediate clones throughout the selection process for sequencing might allow better detection of driver mutations.

Our repeated unbiased (at least in the case of SNVs) identification of genes that are known to confer resistance in both patient isolates as well as in diploid cancer cells lines suggests that our results will be clinically relevant and that the use of a haploid cell line has little bearing on the results. Nevertheless, there may be differences between haploid and diploid cells as well as differences in the set of transcribed gene in HAP-1 cells, relative to other cancer cell lines. While modeling with diploid cells may provide more clinically relevant information, the haploid cells allow us to understand how chemoresistance arises from each mutation without the effect being masked by the other alleles.

Although the mutations that were studied here only appeared with drug pressure, for some genes there could be natural variants within the human population that could affect the response to chemotherapy. Knowing the complete set of genes associated with resistance could allow better therapies to be designed that match a given patients genetic makeup. For example, we identified

the cytochrome p540, CYP1B1, and there is substantial natural variation in this gene. The studies described here could allow more personalized therapies and better therapeutic responses.

Finally, it is important to keep in mind that the compounds examined here are not modern cancer therapies and while still used clinically, they were developed without consideration of resistance rates. In microbes, it is commonplace to evaluate compounds which will be developed into anti-infective therapies for their potential to give resistance using IVIEWGA. We anticipate that this method will be useful for predicting resistance mechanisms for new drugs in clinical development, as well as for new drug combinations and may lead to better classes of drugs for chemotherapy.

3.5 Materials and Methods

Compounds. All chemotherapeutic agents used in this study were obtained from Sigma-Aldrich, dissolved in DMSO at 10mM concentration and stored at -20°C.

Cell cultures. The human chronic myelogenous leukemia cell line, HAP-1, was purchased as authenticated at passage 7 from Horizon Discovery and cultured in tissue culture dishes (Genesee Scientific, Cat# 25-202) as a monolayer at 37°C in a humidified atmosphere with 5% CO₂ using Iscove's Modified Dulbecco's Medium (IMDM) (Life Technologies, CA) supplemented with 10% fetal bovine serum (FBS), 0.29mg/mL L-glutamine, 25mM HEPES, 100U/mL Penicillin and 100µg/mL Streptomycin (1% pencillin/streptomycin). Monoclonal and polyclonal stocks were made and stored in IMDM + 10% DMSO in liquid nitrogen.

In vitro evolution of resistant HAP-1 clones. Prior to selection, an aliquot of the parental line was stocked as a reference for subsequent whole genome sequencing analysis. Three independent clones of HAP-1 cells were cultured in tissue culture dishes exposed to increasing sublethal concentrations of each chemotherapeutic agent at a starting concentration previously determined by the EC₅₀ value for around 7-30 weeks depending on the drug, its speed of action and the method used as two methods were considered: high-pressure intermittent selection method and a step-wise selection method. For high pressure selection, cells were treated at a concentration 3-10 X EC₅₀ value until more than 95% of the cells died. Then treatment was removed and cells were allowed to recover. After reaching around 60% semi-confluence, treatment was reinstalled and EC₅₀ value monitored. For step-wise selection method, drug concentration used was at the EC₅₀ which implied reduced growth rate of approximately 50% and drug pressure was increased in intervals of around 5-10% keeping growth inhibition around 50%. Once the EC₅₀ values of the resistant lines were at least 5 times greater than the one used as control, cells were again cloned by

limiting dilution and further reconfirmed for drug resistance and subsequent DNA extraction for whole genome sequencing analysis. Most cell lines (DOX, GEM, TPT and PTX resistant clones) were subjected to a lethal concentration ($\sim 3\text{-}5\times EC_{50}$ value), killing more than 95% of the cells. Then, treatment was removed until cells reached semi-confluence again (doubling every 22 hours (Andersson et al. 1995)) whereupon drug at \sim the EC_{50} value was reapplied. Alternatively, for ETP-resistant clones a stepwise selection method was used whereby cells were repeatedly exposed to a concentration that killed around 50% of the cell population. Drug concentration was increased by 5-10% every 5 days while keeping the growth rate at 50% of untreated culture. Because mutations will arise randomly during long term cell culture, we attempted at least three independent selections for each drug, in each case starting with an identical parental isolate (**Fig. 3.1C**).

Dose-response assay by EC_{50} determination and bioluminescence quantification. Drug sensitivity and cell viability were assessed by a bioluminescence measurement as follows: twenty-four hours prior to addition of the drugs, 2×10^4 cells/well for every replicate were seeded in a 96-well plate. Once attached, media was removed and 10 different concentrations of drug were added in serial dilutions 1:3 with a starting concentration of $10\mu\text{M}$ or one of which the EC_{50} value of the replicates fell within an intermediate drug concentration. When drug-resistant lines were co-treated in combination with verapamil, a fixed concentration of verapamil ($10\mu\text{M}$) was added to every concentration of the drug. After a 48-hour incubation period at 37°C and 5% CO_2 with the drug, cells were treated with CellTiterGlo (Promega) reagent (diluted 1:2 with deionized water) for quantification of HAP-1 cell viability. Immediately after addition of the luminescence reagent, luminescence was measured using the Synergy HT Microplate Reader Siasfrtd (BioTek). The data was normalized to 100% cell survival and 100% cell death and EC_{50} values were obtained using

the average normalized luminescence intensity of 8 wells per concentration and a non-linear variable slope four-parameter regression curve fitting model in Prism 8 (GraphPad Software Inc.)

Isolation of total DNA from drug resistant lines. Genomic DNA (gDNA) was extracted from drug-specific resistant cell lines together with their isogenic parental lines using the DNeasy[®] Blood & Tissue Kit (Qiagen) following the manufacturer's instructions. Samples were assessed for quantity with the Qubit[™] dsDNA BR Assay Kit (Life Technologies, Carlsbad, CA, USA). All samples (>2.0µg, >50ng/µL, >20µL) were prepared for quality control by testing gDNA degradation or potential contamination using agarose gel electrophoresis (1% Agarose, TAE, 100 Voltage). Then gDNA concentration was again measured using the Qubit[®] DNA Assay Kit with the Qubit[®] 2.0 Fluorometer (Life Technologies). Finally, fragment distribution of the gDNA library was measured using the DNA 1000 Assay Kit with the Agilent Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, USA). DNA libraries were sequenced with 150 base pair (bp) paired single end reads on an Illumina HiSeq 4000 (PE150).

Genome Sequencing and Data Analysis. The quality of the raw FASTQ files was checked with FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Whole genome sequencing (WGS) reads were mapped to GRCh37 (hg19) using BWA (v.0.7.17), specifically with the hs37d5 reference genome from 1000 Genomes project (Phase II). Whole exome sequencing (WES) samples were captured using Agilent SureSelect Human All Exon V6 (58 M), and the reads were also mapped to GRCh37 using BWA (v.0.7.17) with the same reference genome as WGS. Duplicate reads were removed using Picard (v.1.94); paired resistant and parent (WT) BAM files were used as input for The Genome Analysis Toolkit (GATK, v3.8-1). Local realignment and base quality recalibration were performed using default parameters. Somatic single nucleotide variants (SNVs) and small insertion and deletion (indels) were called using

GATK MuTect2 following the state-of-the-art GATK Best Practices pipeline (<https://ccbr.github.io/Pipeline/Tools/MuTect2.html>). In this project, the input to MuTect2 consisted of alignments for the parent and resistant clone in order to call mutations with statistically significant differences in read support in the setting of resistance. Only the variants with PASS status, suggesting confident somatic mutations, were considered for further analysis. Variant allelic fraction was determined as the fraction of reads supporting the variant allele relative to total read depth at the variant position. Minimum callable depth was set to 10 and base quality score threshold was set to 18, following the default from MuTect2. All sequences have been deposited in SRA BioProject PRJNA603390.

Small-Variant Annotation for SNVs and Indels. Somatic variants were annotated using snpEff (v 4.3q) (Cingolani et al. 2012). The annotation was performed with parameters including (1) canonical transcripts and (2) protein coding to enable identification of different functional classes of variant and their impact on protein coding genes (Table 1 showing finalized and consolidated annotations; Table S4 shows the raw annotation from snpEff and consolidated classification used in Table 1; Table S7 shows all the SNVs with their raw annotations). The snpEff sequence ontology designation was used in the filtering steps to classify variants generally as noncoding or coding (Table S4).

Identification of Drug Specific Genes. First, we excluded all variants in non-coding regions. Second, we excluded all non-functional variants, retaining only variants with a snpEff definition of HIGH or MODERATE impact (missense, stop lost, stop gain, and structural interaction variants). Finally, we selected only the variants with high allele frequency ($AF > 0.85$) and genes with multiple independent amino acid changes found in the same drug as the final list of candidates. The potential candidate variants were evaluated through Integrative Genomics

Viewer (IGV)(Robinson et al. 2011) to ensure coverage and allele fractions of the mutation positions. The top genes for each drug were included in Table 2 and Table S8.

Somatic Copy Number Variations (CNVs) Analysis. Copy number regions for WGS and WES were called by ControlFreeC⁴⁷ using the default settings for WGS and WES data. Background parental clone samples for each drug served as the control. Recurrent CNV regions were defined as regions observed in more than 1 sample, but not in all of clones from the tested drugs (as these are more likely to indicate potential sequencing artifacts).

Gene knockdowns. shRNAs targeting *TOP2A* (Cat# sc-36695-V), *DCK* (Cat# sc-60509-V), *SLCO3A1* (Cat# sc-62713-V), *SLC13A4* (Cat# sc-89760-V), *KLF-1* (Cat# sc-37831-V), *WWOX* (Cat# sc-44193-V), *WDR33* (Cat# sc-94735-V) and the non-coding control (Cat# sc-108080) were obtained in pLKO.1-Puro from Santa Cruz Biotechnology. *RRM1* (clone ID NM_001033.2-476s1c1) and *CYP1B1* (clone ID NM_000104.2-1176s1c1) were obtained in pLKO.1-Puro-CMV-tGFP from Sigma Aldrich.

Gene expression was knocked down using either a shRNA pool (Santa Cruz Biotechnology) containing between three and five expression constructs each encoding target-specific 19-25 shRNAs or a single shRNA (Sigma Aldrich). HAP-1 cells were plated at 120,000 cells per well (~40% confluency) in a 24-well plate 24 hours prior to viral transduction. On the day of transduction, complete media was replaced with serum-free media and 7µg/mL Polybrene* (Cat# sc-134220) and virus was added to the cells at a multiplicity of infection of 0.5 and cells were incubated overnight at 37°C. The following day, media was replaced with complete media without Polybrene and cells were incubated at 37°C overnight. Cells were then split 1:3 and incubated for 24 hours more and finally stable clones expressing the shRNA were selected using complete media with 2µg/mL puromycin. After 7 days of selection with puromycin, knockdown efficiency was

confirmed by western blot. Cells transduced with shRNAs containing fluorescent tags, were trypsinized (TrypLE™ Express; Cat# 12605-010, Gibco) after puromycin selection, washed twice with DPBS (1X) (Gibco) and sorted by flow cytometry.

Knockout of USP47. *USP47* was knocked out (Cat# HSPD0000092816) using a single plasmid CRISPR-Cas9 system, using as lentivirus backbone the LV01 U6-gRNA:ef1a-puro-2A-Cas9-2A-tGFP targeting *USP47* (Sigma Aldrich). The target sequence (5'-3') was CAATGGGGCTTCTACTAGG. Transduction was as described above. HAP-1 cells were plated at 120,000 cells per well (~40% confluency) in a 24-well plate 24 hours prior to viral transduction. On the day of transduction, complete media was replaced with serum-free media and 7µg/mL Polybrene® (Cat# sc-134220), virus was added to the cells at a multiplicity of infection of 0.5 and cells were incubated overnight at 37°C. The following day, media was replaced with complete media without Polybrene and cells were incubated at 37°C overnight. Cells were then split 1:3 for 24 hours more and stable clones expressing the CRISPR/Cas9 sequence were selected using complete media with 2µg/mL puromycin. After 14 days of selection with puromycin and propagation as required, cells were trypsinized (TrypLE™ Express; Cat# 12605-010, Gibco), washed twice with DPBS (1X) (Gibco) and sorted by flow cytometry using the GFP fluorochrome which is expressed with Cas9. GFP positive cells were plated at an average density of 0.5 cells per well in a 96-well plate (previously treated with poly-L-Lysine (Sigma #P4707-50ml) to improve cell adhesion) in presence of 2µg/mL puromycin (limiting dilution cloning). Cell growth was monitored via microscopy during 25 days to select those wells which were observed to contain single colonies and *USP47* knockout was confirmed in those monoclonal HAP-1 cell lines first via PCR and then reconfirmed by western blot using the *USP47* rabbit polyclonal antibody (Abcam, Cat# ab97835).

Immunoblotting. HAP-1 cells (at least 5×10^6) were trypsinized, washed twice with cold 1X DPBS and then lysed in 500 μ L Pierce™ RIPA Buffer (Thermo Scientific) containing 1:100 protease inhibitor (Halt™ Protease & Phosphatase Inhibitor Cocktail, Thermo Scientific) and 1:100 0.5M EDTA Solution (Thermo Scientific). Total protein concentration was measured using the DC Protein Assay (Bio-Rad). Equal amounts of proteins were resolved by SDS-PAGE and transferred to nitrocellulose membranes (Bio-Rad #1704271), blocked in PBS with 5% (w/v) Blotting-Grade Blocker (Bio-Rad #170-6404) and 0.1% (v/v) Tween20 for 1h and probed. As secondary antibodies, HRP-linked anti-mouse or anti-rabbit (Cell Signaling Technology) were used and the HRP signal was visualized with SuperSignal® West Pico Chemiluminescent Substrate (Thermo Scientific #34080) using Syngene G-Box imager. Protein enrichment was calculated relative to *vinculin*, *γ -tubulin* or *β -actin*. Primary antibodies are listed below.

Antibodies. *TOP2A* (Sigma #SAB4502997), *USP47* (Abcam #ab97835), *WDR33* (Abcam #ab72115), *DCK* (Abcam #ab151966), *β -actin* (Cell Signaling #3700S), *γ -tubulin* (Cell Signaling #4285S), *Vinculin* (Invitrogen #700062), *SLC13A4/SUT-1* (Abcam #ab236619), *WWOX* (Abcam #ab137726), *EKLF/KLF-1* (Abcam #175372), *SLCO3A1/OATP-A* (Santa Cruz #sc-365007), *TOP1* (Proteintech #20705-1-AP), CRISPR/Cas9 (Sigma #SAB4200701), *RRM1* (Abcam #ab133690), *CYP1B1* (Abcam #ab137562), *SPG7* (Sigma #SAB1406470 and Abcam #ab96213), goat anti-mouse (Invitrogen #G21040), goat anti-rabbit (Invitrogen #G21234).

RNA isolation, RT-PCR analysis and sequencing of TOP1 (*His81*). TPT-resistant cells and TPT-WT (1×10^6 cells) were dissociated from plates by the addition of 2mL of TrypLE (Cat #12605-010, Gibco), washed and total RNA was isolated and purified using a Qiagen RNeasy® Mini Kit (Cat #74104, Qiagen) according to manufacturer's instructions. cDNA was synthesized from 1 μ g of total RNA using the Superscript™ First-Strand Synthesis System for RT-PCR Kit

(Invitrogen #11904-018) and random hexamers. The primers used to amplify the region containing *His81* were FWD: GATCGAGAACACCGGCAC and REV: TCAGCATCATCCTCATCTCGAG. DNA from PCR product was extracted, using the QIAquick® Gel Extraction Kit (Qiagen #28706) following the manufacturer's instruction, measured using the Qubit® DNA Assay Kit with the Qubit® 2.0 Fluorometer (Life Technologies), and sequenced.

3.6 Figures

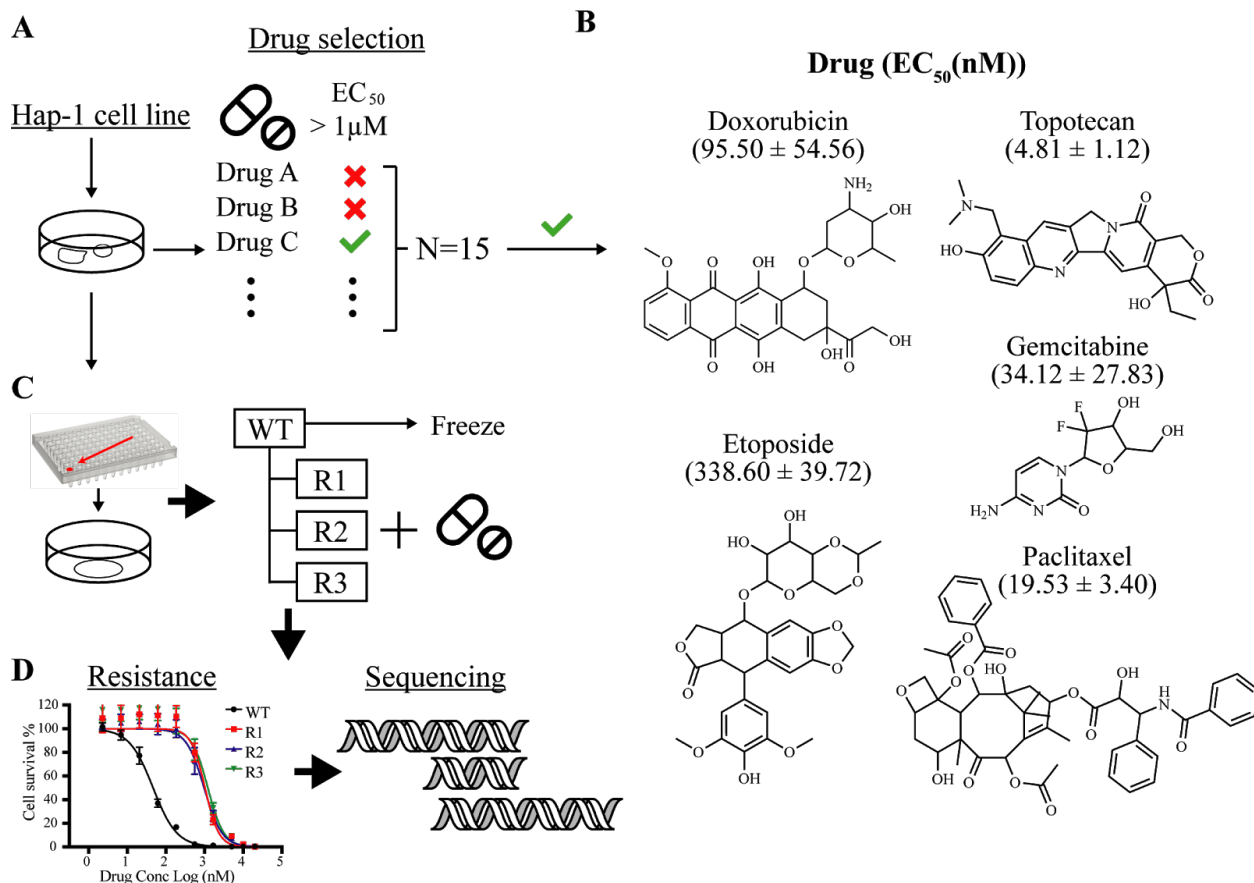


Figure 3.1. Workflow for experimental design.

A. Drug selection step for chemotherapy drugs. Dose-response assays by EC₅₀ determination and bioluminescence quantification were first performed on 15 different chemotherapeutic agents (Table S1). Only drugs to which HAP-1 were sensitive (EC₅₀ value below 1 μM) were considered for IVIEWGA. **B.** Chemical structures of the sensitive chemotherapy agents used in the study. They all present diverse structures with different mechanisms of action. EC₅₀ values are presented as mean ± s.e.m., n=3 with individual biological replicates overlaid and n=8 with individual technical replicates per concentration point. **C.** Clone selection. To ensure the background of the cell line, limiting dilution cloning was used to isolate individual cells and grow the colonies. For each drug selected, three independent selections and one parental line were tested in dose-response drug sensitivity assays to quantify level of drug resistance. **D.** Drug resistance development - cell resistance was achieved in 7 to 30 weeks approximately (49 and 210 generations). The parental cell line and the drug resistant lines were then sequenced.

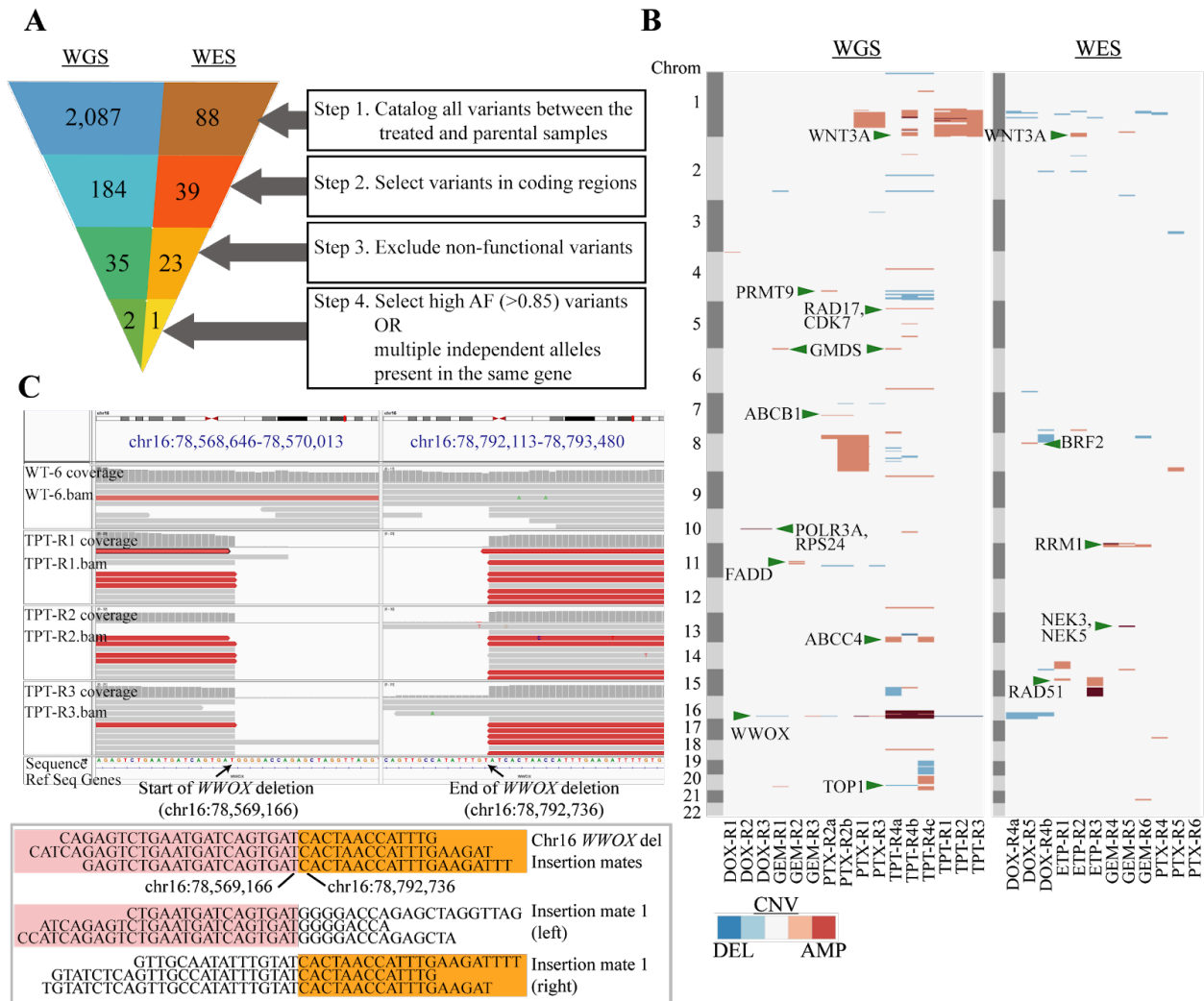


Figure 3.2. Sex- and age-specific MHC-based mutation selection.

A. Variant filtering pipeline and the average number of mutations for WGS and WES samples. **B.** CNV events across samples in each chromosome for WGS and WES samples. Events less than 0.05% ($5E-04$) of the chromosome length and events from chromosome X and Y are not shown. The full list of CNV events can be found in Table S7. Candidate genes in small CNVs were identified by examining protein coding genes in minimal amplified intervals to look for classes of genes known to be associated with resistance (ABC transporters, DNA damage response and cell cycle, cytochrome p450s (Mansoori et al. 2017)) and performing literature searches on potential candidates. **C.** IGV screen view of the *WWOX* deletion event showing the start (left) and end (right) of the deletion as well as a histogram of coverage at each base. Reads with an abnormal paired-end insert size ($\sim 223,000$ versus ~ 200 bases) are highlighted in red. Portions of reads covering the start and end of the deletion events are shown below and allow the identification of the exact breakpoint.

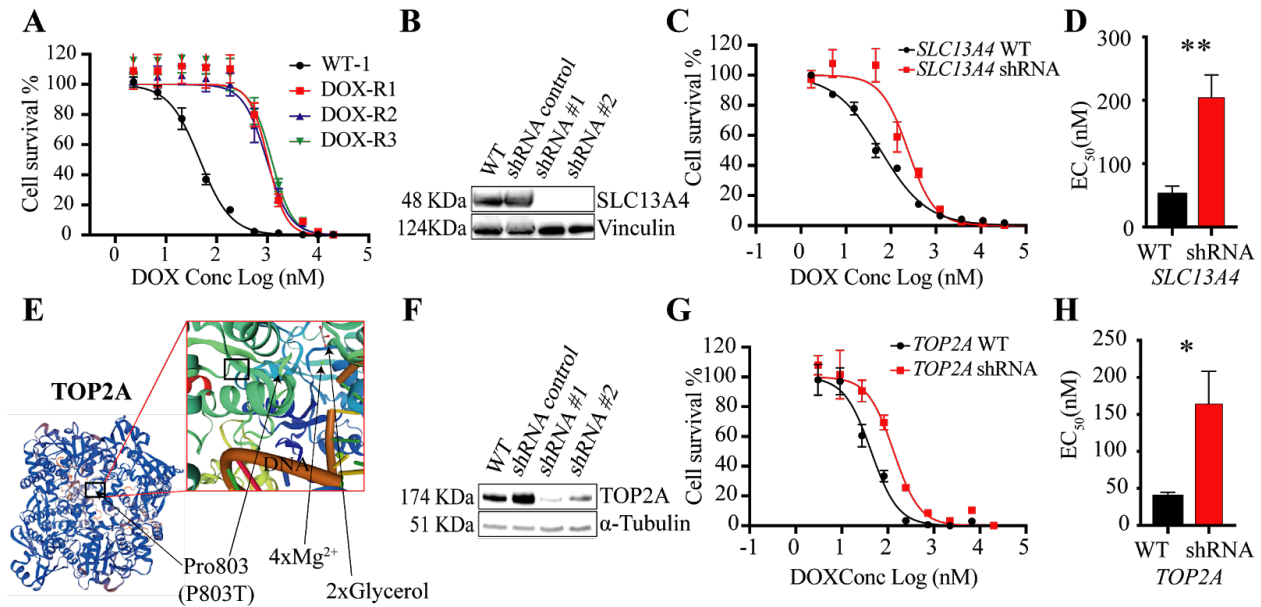


Figure 3.3. DOX-specific analysis and results.

A. EC₅₀ curves for initial screening of DOX resistance **B.** Crystal structure of *TOP2A* highlighting (black box) the allele Pro803Thr next to the principal drug-binding DNA locus, Tyr805 (structure taken from PDBe 4FM9 (Wendorff et al. 2012)). **C.** Western blot confirming that shRNA gene depletion downregulates protein levels for *TOP2A*. **D.** EC₅₀ curves of the WT and shRNA knockdown cell lines for *TOP2A*. **E.** Barplot of the WT and shRNA knockdown cell lines for *TOP2A*. **F.** Western blot confirming that shRNA gene depletion downregulates protein levels for *SLC13A4*. **G.** EC₅₀ curves of the WT and shRNA knockdown cell lines for *SLC13A4*. **H.** Barplot of the WT and shRNA knockdown cell lines for *SLC13A4*. Data is represented by mean ± s.e.m. with n=3 biological replicates; and n=8 (*TOP2A*) and n=4 (*SLC13A4*) technical replicates for every concentration point. * = p value < 0.05, ** = p < 0.01. p values determined by two-tailed *t* test.

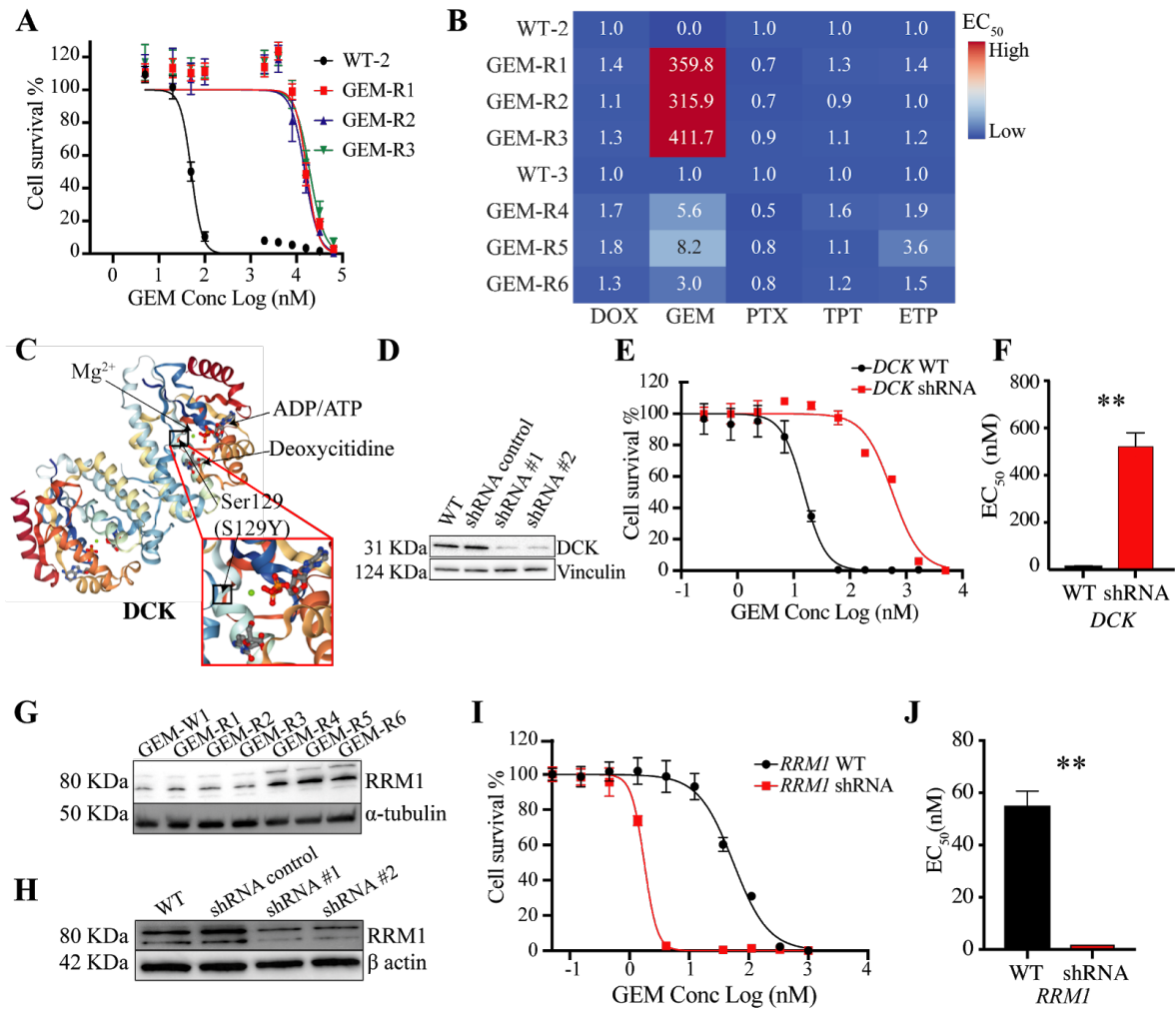


Figure 3.4. GEM-specific analysis and results.

A. EC_{50} curves for initial screening of GEM resistance using $n=8$ technical replicates per concentration point **B.** EC_{50} ratio matrix showing absence of multidrug resistance pathways of GEM resistant lines. **C.** Crystal structure of deoxycytidine kinase (*DCK*) showing the location of the Ser129Tyr (marked by a black box) allele near the deoxycytidine and ATP/ADP binding pocket (structure taken from PDBe 2A7Q (Y. Zhang, Secrist, and Ealick 2006)). **D.** Western blot confirming that shRNA gene depletion downregulates protein levels for *DCK*. **E.** EC_{50} curves of the WT and shRNA knockdown cell lines for *DCK*. $n=8$ with individual technical replicates overlaid for every concentration point. **F.** Barplot of the WT and shRNA knockdown cell lines for *DCK*. **G.** Western blot for *RRM1* across all GEM samples showing overexpression pattern of *RRM1* in GEM-R4-6 resistant clones. γ -tubulin is used as a loading control. **H.** Western blot confirming that shRNA gene depletion downregulates protein levels for *RRM1*. β -actin is used as a loading control. **I.** EC_{50} curves of the WT and shRNA knockdown cell lines for *RRM1*. $n=4$ with individual technical replicates overlaid for every concentration point. **J.** Barplot of the WT and shRNA knockdown cell lines for *RRM1*. All data is represented by mean \pm s.e.m. with $n=3$ with individual biological replicates overlaid. ** = p value < 0.01 . p values determined by two-tailed t test.

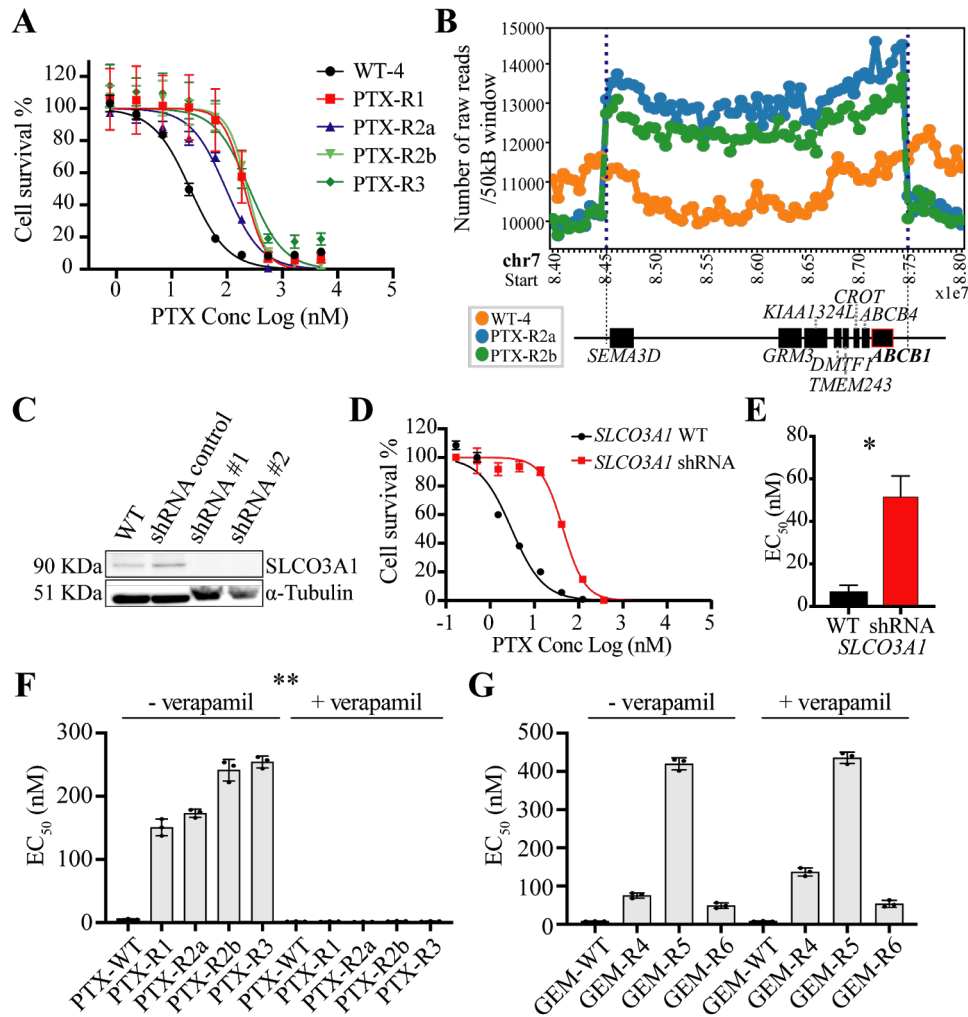
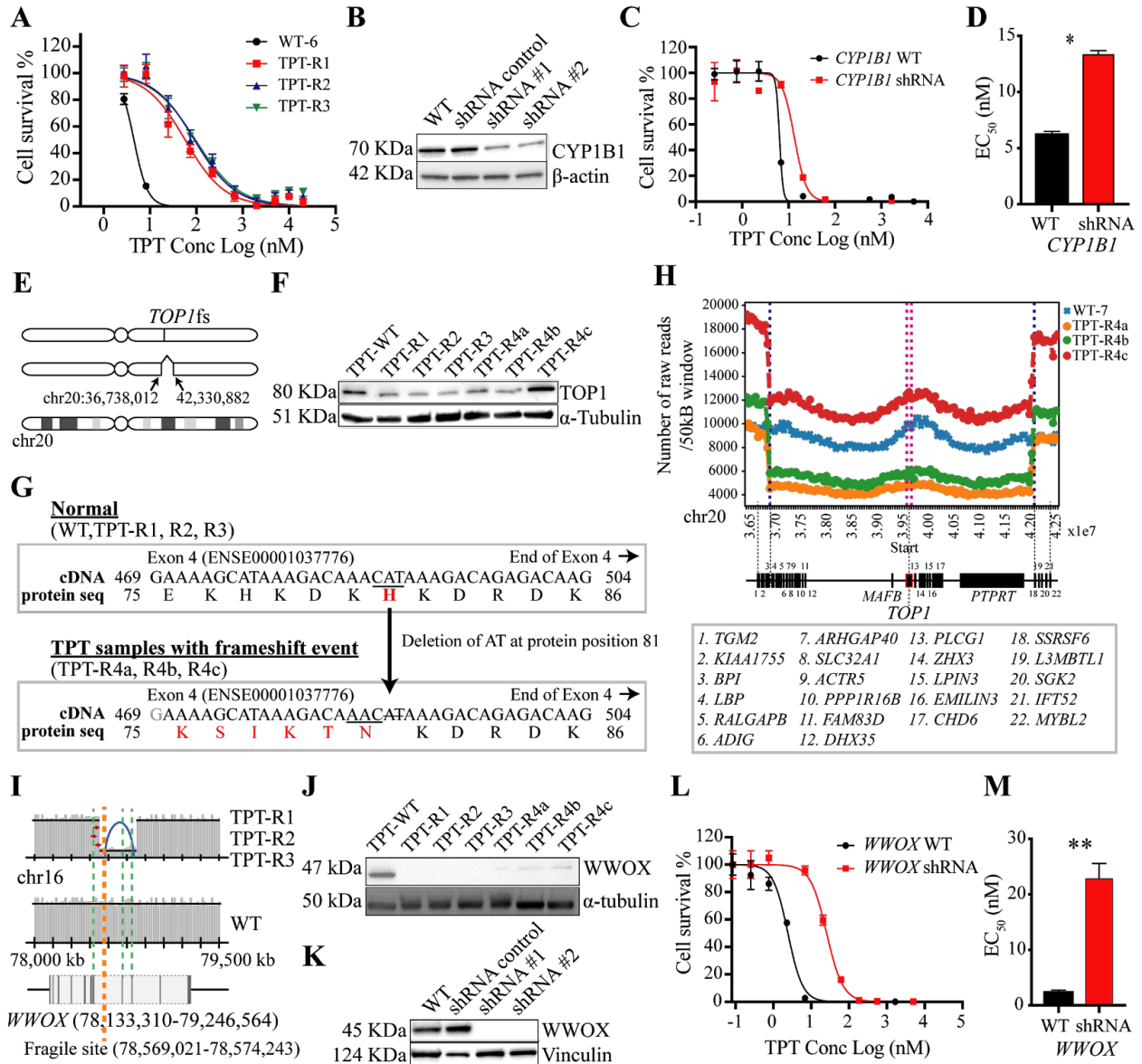


Figure 3.5. PTX-specific analysis and results.

A. EC_{50} curves for initial screening of PTX resistance with $n=8$ technical replicates. Values are given in Table S2. **B.** Raw copy number profile for the amplification event containing protein coding genes including ABC transporters (*ABCB1/ABCB4*) at PTX cell lines. The amplification region (chr7:84,500,000-87,300,000) had a higher number of raw reads (labeled with blue dash lines) with default window size of 50K bp. Genes associated with the CNV event are depicted by black boxes underneath according to their position and sizes. *ABCB1* is highlighted with red outline. **C.** Western blot confirming that shRNA gene depletion downregulates protein levels for *SLCO3A1*. **D.** EC_{50} curves of the WT and shRNA knock-down cell lines. **E.** Barplot of the WT and shRNA knockdown cell lines for *SLCO3A1*. **F.** Barplot of EC_{50} of the PTX treated cell lines with and without verapamil and verapamil alone showing sensitization in presence of verapamil as ABC inhibitor ($n=4$ technical replicates). **G.** Barplot of EC_{50} of the GEM cell lines \pm verapamil showing no EC_{50} shift for GEM cell lines when co-treated with verapamil with $n=4$ technical replicates. All data is represented by mean \pm s.e.m. with $n=3$ with individual biological replicates overlaid. ** = p value < 0.01 . p values determined by two-tailed t test.

Figure 3.6. TPT-specific analysis and results.

A. EC₅₀ curves for initial screening of TPT resistance with n=8 technical replicates. **B.** Western blot confirming that shRNA gene depletion downregulates protein levels for *CYP1B1*. **C.** EC₅₀ curves of the WT and shRNA knock-down cell lines. **D.** Barplot of the WT and shRNA knockdown cell lines for *CYP1B1*. **E.** Complex alteration for regions surrounding *TOP1* gene. Schematic of chr20 reads around *TOP1* showing the frameshift event (top), deletion (middle), and the WT with characteristic banding. **F.** Western blot with their gene depletion downregulates protein levels for *TOP1*. **G.** Transcript sequence (cDNA) and protein sequence for TOP1 transcript for normal (top) and altered (bottom) cases. The figure shows only part of the cDNA (position 469-504) and protein sequence (position 75-86) for TOP1 at the affected exon (Exon 4, ENSE00001037776). The frameshift deletion of nucleotides ‘CAT’ to ‘C’ observed in TPT samples (TPT-R4a, R4b, and R4c) is predicted to give a frameshift at amino acid 81 (His, red highlight in normal). Amino acids affected by the frameshift deletion are highlighted in red. **H.** Raw copy number profiles of *TOP1* deletion at TPT cell lines. The deleted regions (hg19, chr20: 36,738,012-42,330,882) have lower numbers of raw reads (labeled with blue dash lines) than WT-7 with default window size of 50K bp. *TOP1* is labeled with magenta dashed lines. **I.** Schematic of chr16 reads around *WWOX* for TPT-R1, R2, and R3 compares to the WT chromosome 16 parental cell line. Blue arch represents a deleted region and red arch represents an amplified region. Red arrows show insertion events. *WWOX* below shows the exonic (black lines) and intronic (white box) regions of the gene. The start of the deletion event is also close to a known fragile site (orange dashed line). **J.** Western blot showing *WWOX* protein levels in TPT resistant clones. **K.** Western blot shows downregulation of protein levels for *WWOX* in shRNA samples compared to WT and scrambled control. **L.** EC₅₀ curves of the WT and shRNA knock-down cell lines with n=4 technical replicates per concentration point. **M.** Barplot of the WT and shRNA knockdown cell lines for *WWOX*. Data is represented by mean ± s.e.m. with n=3 biological replicates. ** = p < 0.01. p values determined by two-tailed *t* test.



3.7 Tables

Table 3.1. Summary of average number of mutations.

Number of selections performed with the drug is given by n. SNVs and Indels were grouped according to snpEff sequence ontology annotations (Methods, Table S3.4), and detailed counts per clone can be found in Table S3.5.

	WGS					WES			
	DOX (n=3)	GEM (n=3)	PTX (n=4)	TPT (n=3)	TPT (n=3)	DOX (n=3)	ETP (n=3)	GEM (n=3)	PTX (n=3)
<i>Indels</i>									
<i>Disruptive inframe ins.</i>	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00
<i>Frameshift</i>	0.00	1.00	1.33	2.33	0.00	1.00	1.00	1.00	1.00
<i>Frameshift, stop-gained</i>	0.00	0.00	0.00	0.33	0.00	0.00	1.00	0.00	0.00
<i>Inframe insertion</i>	0.00	0.00	1.00	0.33	0.00	0.00	1.00	0.00	0.00
<i>Intergenic</i>	27.67	43.00	26.75	24.67	47.67	2.00	3.50	1.00	1.00
<i>Intragenic</i>	10.00	5.67	9.00	9.00	14.33	1.00	2.00	1.00	1.00
<i>Intron</i>	12.00	20.33	16.25	15.67	32.33	1.00	1.00	4.00	1.50
<i>Splice region plus intron</i>	0.00	0.00	0.00	103.33	0.00	0.00	0.00	1.00	0.00
<i>SNVs</i>									
<i>Disruptive inframe del.</i>	0.00	1.00	1.00	0.67	0.00	1.00	0.00	1.00	1.33
<i>Frameshift</i>	1.00	2.00	1.00	22.33	2.00	3.00	1.67	1.00	3.00
<i>Inframe deletion</i>	0.00	0.00	1.00	0.00	0.00	0.00	2.00	0.00	0.00
<i>Intergenic</i>	898.33	1303.3	1416.6	258.67	2635.7	25.00	19.00	17.00	21.33
<i>Intragenic</i>	272.00	403.33	389.25	77.67	834.67	15.33	12.33	10.33	11.33
<i>Intron</i>	448.67	701.33	764.00	128.33	1358.33	28.00	27.00	33.00	22.67
<i>Missense</i>	16.00	14.00	12.75	7.00	34.33	15.00	19.67	21.67	15.67
<i>Others</i>	1.00	1.00	0.00	0.00	1.33	1.00	1.00	1.50	1.00
<i>Splice region plus intron</i>	1.00	1.67	2.00	30.00	3.33	1.67	1.50	1.00	1.33
<i>Start lost</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00
<i>Stop gained</i>	0.00	2.50	1.00	0.00	1.67	1.50	2.00	2.67	0.00
<i>Stop lost</i>	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Synonymous</i>	3.00	7.33	4.00	6.33	7.67	6.67	5.33	7.67	5.33

Table 3.2. Validation (knockdown) results for selected genes.

CNV, copy-number variant. MS, missense. FS, frameshift variant, KO/KD/CI, knockout, knockdown, chemical inhibition (verapamil, ABCB1). ND: No data: gene knockdowns were attempted but could not be achieved. NE: Not expressed (protein not detected by Western blot). EC_{50} WT and EC_{50} KO/KD/CI are from matched pairs for the given drug and represent the mean \pm s.e.m. (n=3 biological replicates).

Drug	Sample	Gene	Amino Acid Change	Type	AF	EC_{50} (nM) WT	EC_{50} (nM) KD/KO/CI
DOX	R1, R2, R3	TOP2A	Pro803Thr	MS	0.89, 0.87	38.6 \pm 4.3	164.3 \pm 43.9
	R4b	SLC13A4	Gly165His	MS	1	52.9 \pm 11.6	204.3 \pm 35.7
	R5	SPG7	Lys593Asn	MS	1	NE	
PTX	R1, R2a, R3	WWOX	16q23.1	CNV	-	5.8 \pm 2.5	42.9 \pm 11.7
	R2a, R2b	ABCB1	7q21.12	CNV	-	252 \pm 38	1.3 \pm 0.1
	R2b, R6	SLCO3A1	Ile587Asn (R2b), Ala263Thr (R6)	MS		218 \pm 14	1.3 \pm 0.1
GEM	R1, R2, R3	DCK	Ser129Tyr (R1, R2), Asn80fs (R1, R3); Asn113fs (R2); Thr184fs (R3)	MS, FS		6.6 \pm 3.3	51.5 \pm 9.9
	R4, R5, R6	RRM1	11p15.4	CNV	-	14.3 \pm 1.7	521.7 \pm 58.3
ETP	R2, R3	WNT3A	1q42.13 (R2)	CNV	-	54.9 \pm 5.8	1.8 \pm 0.1
	R3	WDR33	P622T	MS	1	ND	-
TPT	R1, R4a, R4b, R4c	CYP1B1	Val432Leu; Asp217Glu (R4a,b,c)	MS	0.13, 0.40, 0.43, 0.42	241.5 \pm 31.0	821.6 \pm 226.9
	R1, R2, R3	WWOX	16q23.1	CNV	-	6.3 \pm 0.2	13.3 \pm 0.3
	R4a, R4b, R4c	TOP1	His81fs; 20q12	FS; CNV	1	2.4 \pm 0.3	22.8 \pm 2.7
	R4a, R4b, R4c	USP47	Arg408*	Stop	0.38, 0.57, 0.58	ND	-
						3.05 \pm 0.2	1.07 \pm 0.07

3.8 Supplemental Tables and Figures

Table S3.3. Sequencing sample characteristics and statistics.

Table S3.5. Summary of average number of mutations for each individual clone.

Table S3.7. All SNVs called from the 28 samples.

Table S3.8. All CNVs called from the 28 samples.

Table S3.1. Summary of all drugs tested for IVIEWGA.

15 different anticancer FDA-approved drugs were tested to determine their EC₅₀. ATP levels were measured via bioluminescence (CellTiterGlo) to determine drug sensitivity in a dose-response assay for 48h with serial dilutions of the drug (10μM max, 1:3 serial dilutions, 8 technical replicates per concentration point). Those drugs showing EC₅₀ values below 1μM were considered for IVIEWGA. Puromycin was used as a positive control.

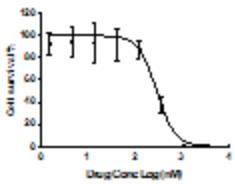
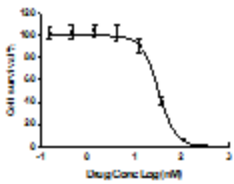
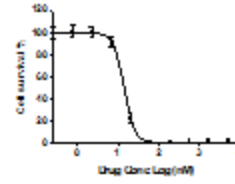
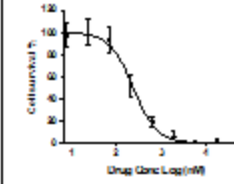
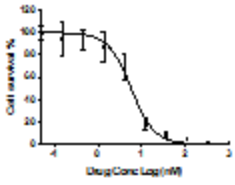
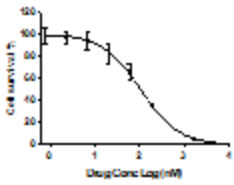
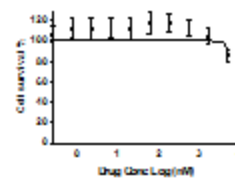
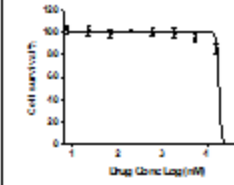
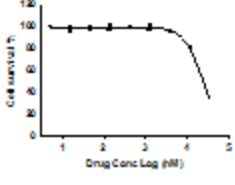
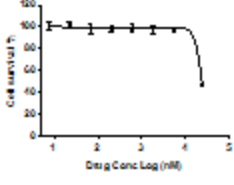
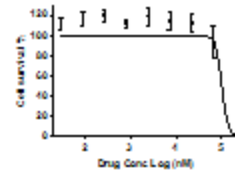
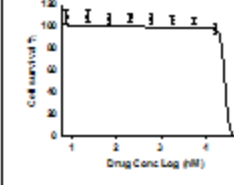
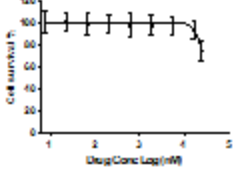
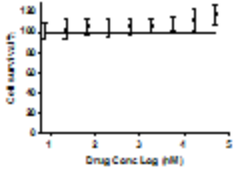
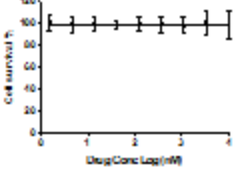
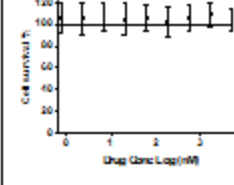
EC ₅₀ curve				
Drug / EC ₅₀ value	Puromycin 288.44 nM (positive control)	Paclitaxel 19.43 nM	Gemcitabine 34.12 nM	Doxorubicin 95.50 nM
EC ₅₀ curve				
Drug / EC ₅₀ value	Topotecan 4.81 nM	Etoposide 338.60 nM	Sorafenib > 5 μM	Tamoxifen > 10 μM
EC ₅₀ curve				
Drug / EC ₅₀ value	Linsitinib > 10 μM	Everolimus > 10 μM	Imatinib > 10 μM	Rapamycin > 10 μM
EC ₅₀ curve				
Drug / EC ₅₀ value	Pterostilbene > 10 μM	Thalidomide N/A	Carboplatin N/A	Imatinib mesylate N/A

Table S3.2. EC₅₀ values for all drug-specific resistant lines and their isogenic parental cell line (WT).

EC₅₀ averaged values (AVRG) are presented as mean with individual biological replicates (BR) overlaid. Doxorubicin (DOX), Gemcitabine (GEM), Paclitaxel (PTX), Topotecan (TPT) and Etoposide (ETP) were used to generate resistant lines.

		BR1	BR2	BR3	AVRG	STDEV
DOXORUBICIN	DOX-WT1	45.2	51.0	41.9	46.1	4.6
	DOX-R1	840.9	1144	945	976.6	154
	DOX-R2	833.7	1106	957.8	965.8	136.3
	DOX-R3	1132	1292	1225	1216.3	80.4
	DOX-WT5	150	150.6	134.2	144.9	9.3
	DOX-R4a	1373	1403	1162	1312.7	131.3
	DOX-R4b	503.4	504.5	466.6	491.5	21.6
	DOX-R5	1510	1513	1455	1492.7	32.7
GEMCITABINE	GEM-WT2	59.3	61.5	57.8	59.5	1.9
	GEM-R1	21774	22069	20377	21406.7	903.8
	GEM-R2	19756	19917	16721	18798	1800.5
	GEM-R3	27722	25236	20529	24495.7	3653.2
	GEM-WT3	9.4	8.8	8.1	8.7	0.7
	GEM-R4	60.1	58.9	45.3	54.8	8.2
	GEM-R5	82.1	78.3	66.1	75.5	8.4
	GEM-R6	33.6	28.1	24.3	28.7	4.6
PACLITAXEL	PTX-WT4	22.8	20.6	21.2	21.5	1.2
	PTX-R1	120.6	231.5	251.4	201.2	70.5
	PTX-R2a	89.4	97.9	102.9	96.7	6.8
	PTX-R2b	152.9	255	215.3	207.7	51.5
	PTX-R3	208.9	283.9	221	237.9	40.3
	PTX-WT5	20.4	19.3	13	17.5	4

Table S3.2. EC₅₀ values for all drug-specific resistant lines and their isogenic parental cell line (WT) (continued).

		BR1	BR2	BR3	AVRG	STDEV
PACLITAXEL	PTX-WT5	20.4	19.3	13	17.5	4
	PTX-R4	855.3	959.1	843.9	886.1	63.5
	PTX-R5	824.8	763.6	659.3	749.2	83.7
	PTX-R6	1018	1073	981.7	1024.2	46
TOPOTECAN	TPT-WT6	5.6	5.7	5.6	5.6	0.1
	TPT-R1	140.8	150.4	79.9	123.7	38.2
	TPT-R2	165.7	299	169.7	211.5	75.8
	TPT-R3	211.2	291.2	177.2	226.5	58.5
	TPT-WT7	3.4	3.7	3.6	3.6	0.2
	TPT-R4a	34.8	23.9	29.4	29.4	7.7
	TPT-R4b	59.2	42.1	51.5	50.9	12.1
	TPT-R4c	68.9	64.9	66.9	66.9	2.8
ETOPOSIDE	ETP-WT3	292.8	363.6	359.4	338.6	39.7
	ETP-R4	5876	9851	8235	7987.3	1999
	ETP-R5	4230	5496	4010	4578.7	802
	ETP-R6	3742	3022	2715	3159.7	527.2

Table S3.4. Summary of effect classification (as referred in the manuscript) of the SNVs and Indels from snpEff annotations.

SnpEff Annotation	Effect Classification
3_prime_UTR_variant	Intergenic
5_prime_UTR_premature_start_codon_gain_variant	Intergenic
5_prime_UTR_variant	Intergenic
TF binding site variant	Intergenic
Conservative inframe deletion	Inframe deletion
Conservative inframe insertion	Inframe insertion
Disruptive inframe deletion	Disruptive inframe deletion
Disruptive inframe insertion	Disruptive inframe insertion
Downstream gene variant	Intergenic
Frameshift variant	Frameshift
Frameshift variant & splice acceptor variant & splice donor variant & splice region variant & intron variant	Frameshift
Frameshift variant & splice acceptor variant & splice region variant & intron variant	Frameshift
Frameshift variant & splice region variant	Frameshift
Frameshift variant & stop gained	Frameshift plus stop-gained
Intergenic region	Intergenic
Intragenic variant	Intragenic
Intron variant	Intron
Missense variant	Missense
Missense variant & splice region variant	Missense
Protein-protein contact	Other nonsynonymous coding
Sequence feature	Intergenic
Splice acceptor variant & intron variant	Splice region plus intron variant
Splice acceptor variant & splice donor variant & splice region variant & 5 prime UTR variant & intron variant	Splice region plus intron variant
Splice acceptor variant & splice donor variant & splice region variant & intron variant	Splice region plus intron variant
Splice acceptor variant & splice region variant & conservative inframe deletion & intron variant	Splice region plus intron variant
Splice acceptor variant & splice region variant & disruptive inframe deletion & intron variant	Splice region plus intron variant
Splice donor variant & intron variant	Splice region plus intron variant
Splice region variant	Splice region plus intron variant
Splice region variant & intron variant	Splice region plus intron variant
Splice region variant & synonymous variant	Splice region plus intron variant
Start lost	Start lost
Stop gained	Stop gained
Stop lost	Stop lost
Structural interaction variant	Other nonsynonymous coding
Synonymous variant	Synonymous

Table S3.6. Summary of the number of mutations along the steps of the filtering pipeline for the five candidate drugs.

Drug	Seq	all	coding	functional	AF and recurrence
DOX	WES	92.00	37.67	20.33	1.67
ETP	WES	87.33	38.33	24.67	2.00
GEM	WES	96.00	46.00	27.00	1.00
PTX	WES	75.33	33.67	20.00	1.00
TPT	WGS	3900.70	308.33	40.33	3.00
DOX	WGS	1512.30	114.33	16.67	5.00
GEM	WGS	2162.70	163.00	17.67	1.00
PTX	WGS	2283.30	167.00	15.75	1.00
TPT	WGS	577.00	168.00	84.67	1.00

Table S3.9. Potential target genes fitting the filtering criteria.

Listed genes are the completed set of mutations that presented with AF > 0.85, and modified the coding potential, or were mutated multiple times at independent samples. AF, allele frequency; CNV, copy number variant; MS, missense; SIV, structural variant; FS, frameshift.

Gene	Sample	Type	Amino acid change	AF
<i>AC091801.1</i>	DOX-R4a	MS	His13Asn	1
<i>TOP2A</i>	DOX-R1, DOX-R2, DOX-R3	MS	Pro803Thr	0.89, 0.87
<i>EDNRA</i>	DOX-R2, DOX-R3	MS	Ala143Ser	0.86, 0.88
<i>ITGB3</i>	DOX-R2, DOX-R3	MS	Cys627Phe	0.88, 0.96
<i>PRAMEF11</i>	DOX-R2, DOX-R3	MS	Gln377Lys	0.91, 0.90
<i>NCAN</i>	DOX-R3	MS	G939Ser	0.97
<i>TRIM45</i>	DOX-R3	MS	Gln305Lys	0.89
<i>SLC13A4</i>	DOX-R4b	MS	Gly165His	1
<i>KCNC3</i>	DOX-R5	MS	Ser379Ile	0.86
<i>SPG7</i>	DOX-R5	MS	Lys593Asn	1
<i>WDR87</i>	DOX-R5	MS	Arg2418Met	0.89
<i>WDR33</i>	ETP-R3	MS	Pro622Thr	1
<i>STARD9</i>	GEM-R4, GEM-R5, GEM-R6	MS	His1021Tyr (R4, R5, R6), Ser1330Ile (R5)	0.48, 0.47, 0.45, 0.33
<i>SLCO3A1</i>	PTX-R2b, PTX-R6	MS	Ile587Asn (R2b), Ala263Thr (R6)	0.23, 0.07
<i>CYP1B1</i>	TPT-R1, TPT-R4a, TPT-R4b, TPT-R4c	MS	Val432Leu, Asp217Glu (R4a, R4b, R4c)	0.13, 0.40, 0.43, 0.42
<i>TOP1</i>	TPT-R4a, TPT-R4b, TPT-R4c	FS	His81fs	1
<i>CEACAM1</i>	TPT-R4c	MS	Asp479Glu	1
<i>VAMP7</i>	TPT-R4c	MS	Gln57Lys	1
<i>ZFP36</i>	TPT-R4c	MS	Asp260Tyr	1
<i>DCK</i>	GEM-R1, GEM-R2, GEM-R3	MS, FS	Ser129Tyr (R1, R2), Asn80fs (R1, R3), Asn113fs (R2), Thr184fs (R3)	0.78, 0.25; 0.28, 0.43; 0.09, 0.14
<i>GRM3</i>	ETP-R3	SIV	-	1

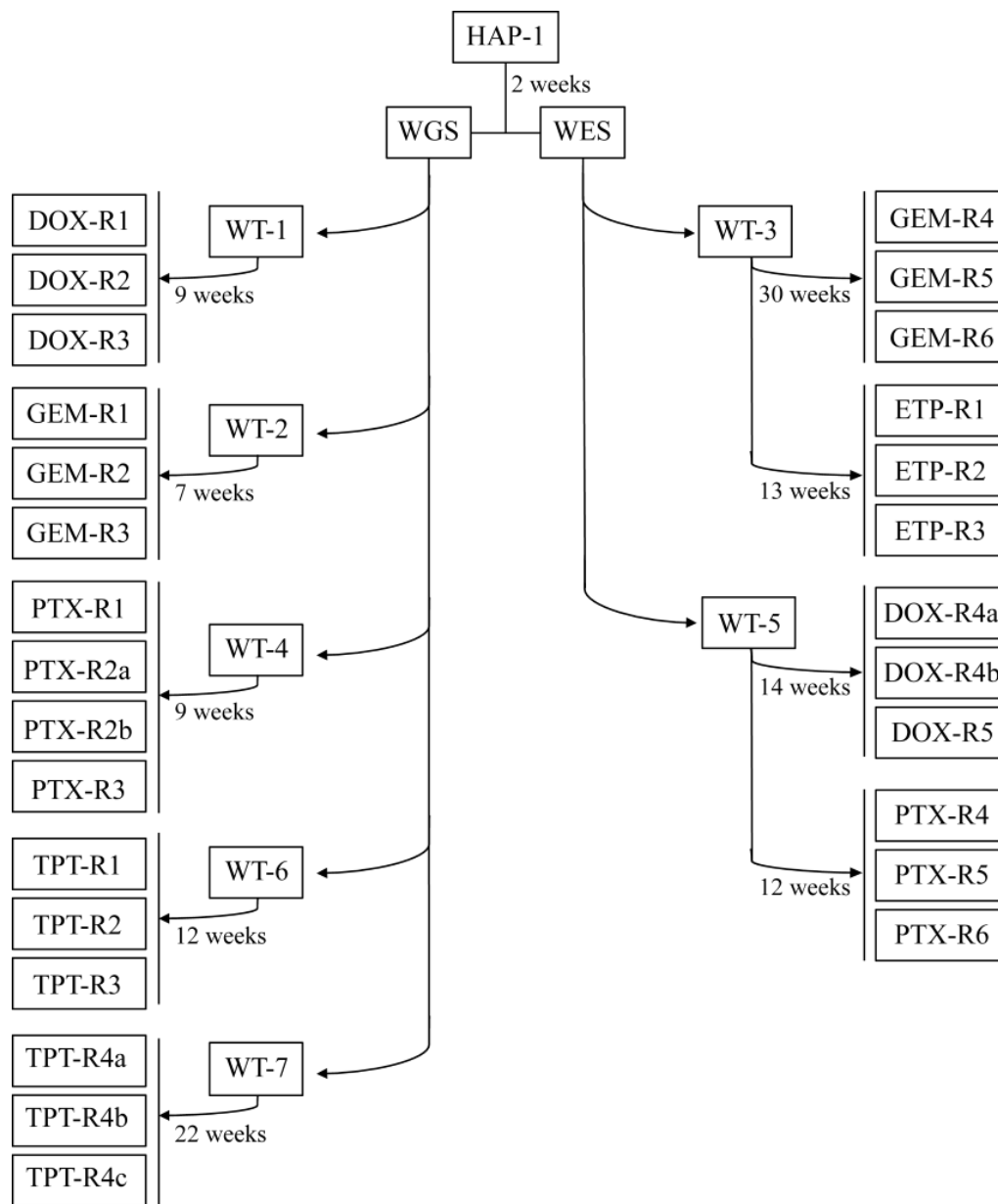


Figure S3.1. Hierarchical tree of the parent clones and the different drug-specific replicates.

At least 3 independent HAP-1 drug-resistant clones were generated directly from their isogenic wildtype parents (WT) and their DNA was whole genome (WGS) or whole exome (WES) sequenced. The independent HAP-1 clones were subjected to increasing sublethal concentrations of the drugs based on the starting EC_{50} values for each chemotherapeutic agent until they acquired resistance. The anticancer drugs used for the study were doxorubicin (DOX), gemcitabine (GEM), paclitaxel (PTX), topotecan (TPT) and etoposide (ETP). The time required to generate drug-resistant clones varied depending on the drug from 7 weeks up to 30 weeks (49 to 210 generations). In a few cases, independent selections could not be achieved and dependent clones with a shared lineage were collected.

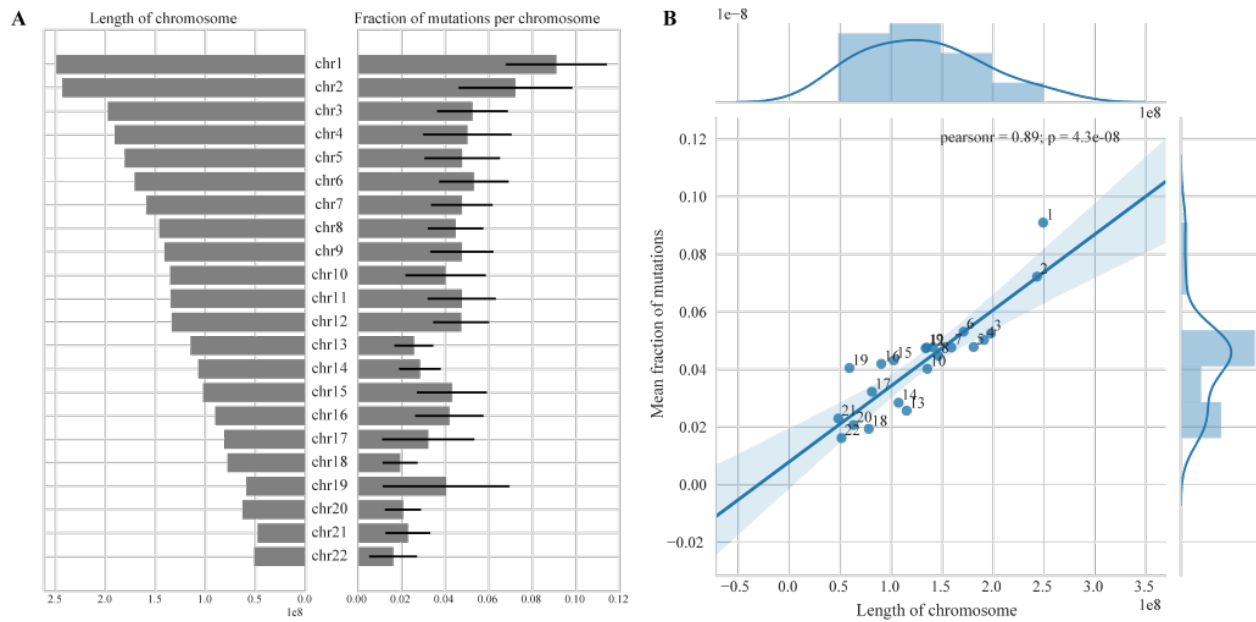


Figure S3.2. Near-normal distribution of the mutations in respect to each chromosome.
A. A bar plot showing length of chromosomes (left) and the fraction of mutations (right) across all samples for each chromosome. Error bars show the standard deviation between samples for their fraction of mutations. **B.** The length of chromosome (x-axis) shows a 0.89 correlation to the mean fraction of mutations (y-axis) for the samples

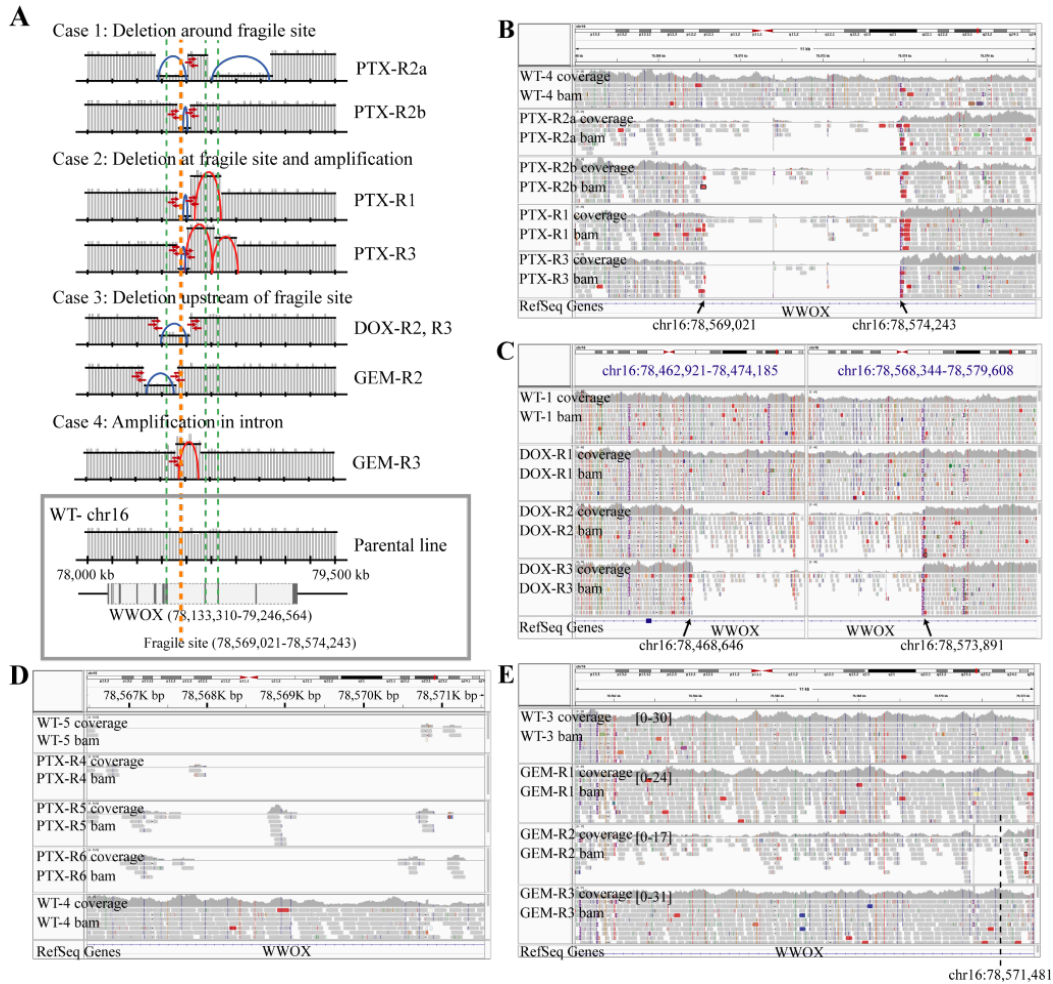


Figure S3.3. Complex alteration for regions surrounding WWOX gene across all samples with WWOX-CNV events.

A. Schematic of chr16 reads around *WWOX*, blue arch represents a deletion region and red arch represents an amplification region. Red arrows show insertion events. Dark bands in the genes represent the exons whereas the light regions represent the introns. Arches are drawn based on the IGV reads as well as CNV calling results (Table S6). **B.** IGVs for the PTX samples are shown with the fragile sites. The reads dropped at the start of a known fragile site (chr16:78,569,021) across the PTX samples. Reads are lower upstream of the fragile site for PTX-R2a, confirming the CNV call results of a deletion event. Reads are higher downstream of the fragile site for PTX-R1 and PTX-R3, confirming a potential amplification event. **C.** IGVs for the DOX samples are shown with the fragile sites. Reads are lower upstream of the fragile site for DOX-R2 and DOX-R3, confirming the CNV call results of a deletion event. **D.** IGVs for the PTX WES and WGS samples around the fragile sites. The lanes of WES samples (WT-5, PTX-R4, R5, and R6) show poor read coverage compare to the WGS sample (WT-4). **E.** IGVs for the GEM samples are shown with the fragile sites. Reads are lower upstream of the fragile site for GEM-R2, confirming the CNV call results of a deletion event. Reads are higher downstream of the fragile site for GEM-R3, confirming a potential amplification event around the intronic region with less effect to the protein.

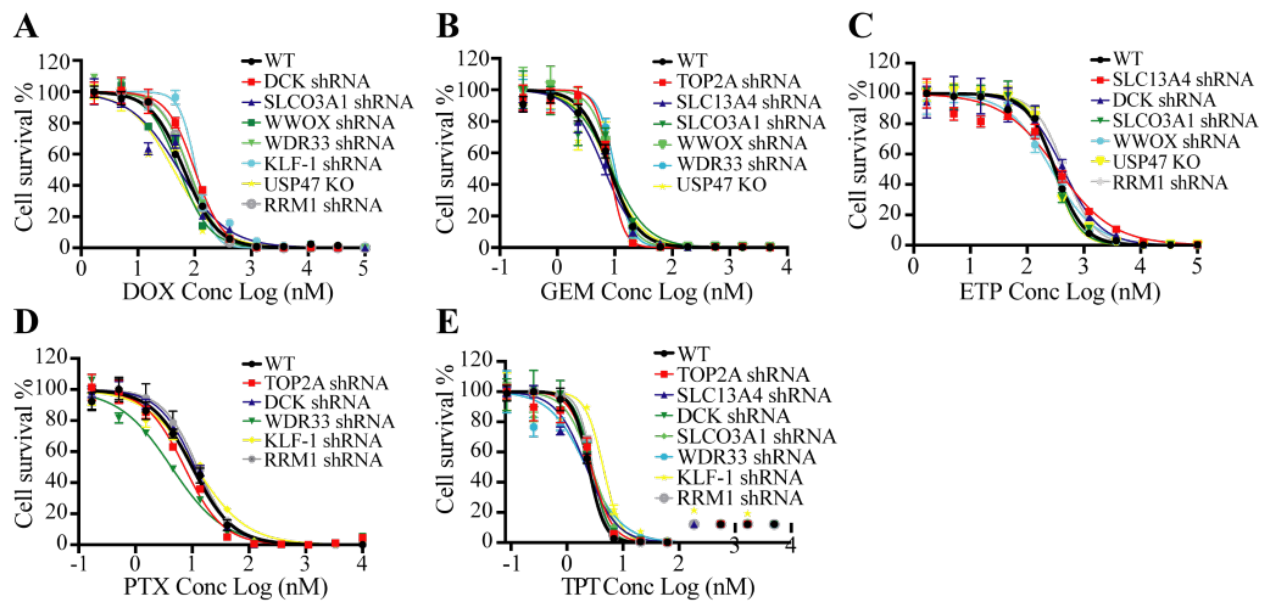


Figure S3.4. Cross-drug resistance EC₅₀ curves.

EC₅₀ curves for all validated gene candidates that showed absence of multidrug resistance (MDR) pathways against the other drugs used in the study. Target gene candidates tested against: **A.** DOX, **B.** GEM, **C.** ETP, **D.** PTX, and **E.** TPT. The results show no cross-drug resistance between drug targets.

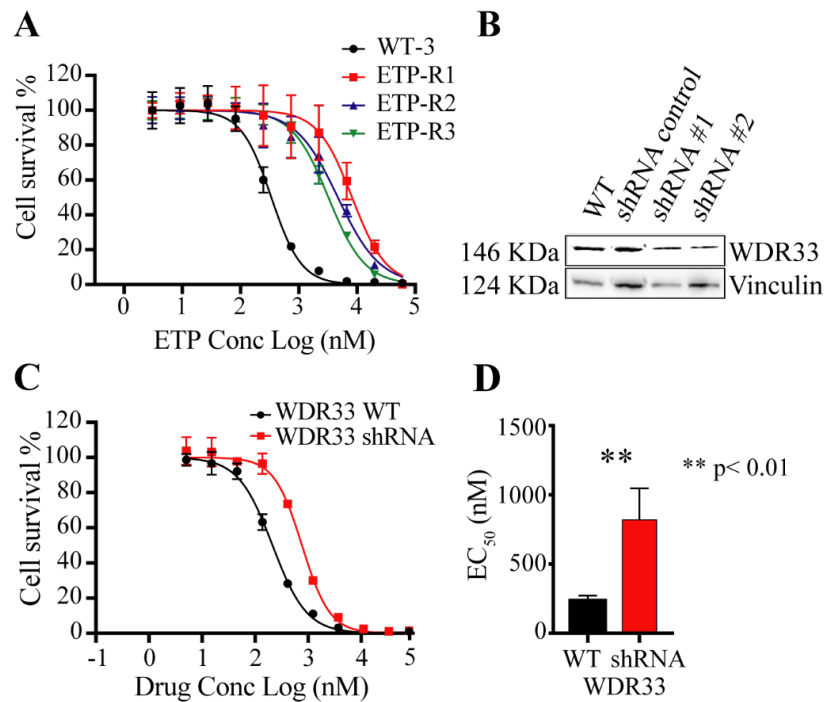


Figure S3.5. ETP target genes and validation results.

A. EC₅₀ curves for initial screening ETP resistance. **B.** Western blot with their gene depletion downregulates protein levels for *WDR33*. **C.** EC₅₀ curves of the WT and shRNA knock-down cell lines. **D.** Boxplot of the WT and shRNA knockdown cell lines for *WDR33*. Data is represented by mean \pm s.e.m. with n=3 biological replicates overlaid and n=4-8 technical replicates. ** = p value < 0.01. p values determined by two-tailed *t* test.

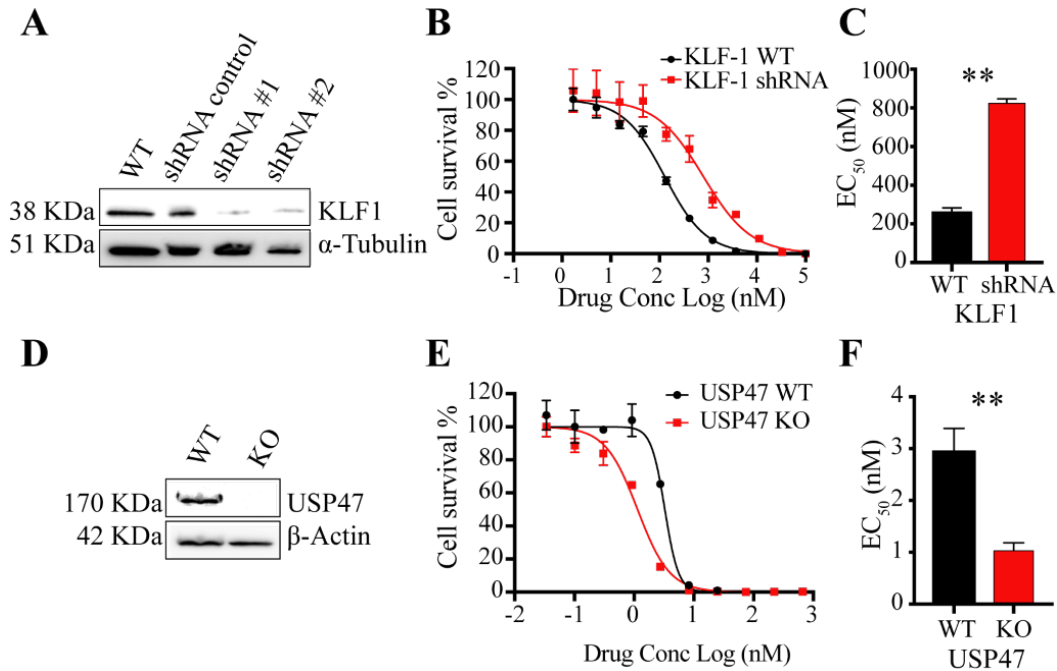


Figure S3.6. Additional mutations likely contribute to drug resistance levels.

A. Western blot with their gene depletion downregulates protein levels for *KLF1*. **B.** EC₅₀ curves of the WT and shRNA knock-down cell lines for *KLF1*. **C.** Boxplot of the WT and shRNA knockdown cell lines for *KLF1*. **D.** Western blot with their gene depletion downregulates protein levels for *USP47*. **E.** EC₅₀ curves of the WT and shRNA knock-down cell lines for *USP47*. **F.** Boxplot of the WT and shRNA knockdown cell lines for *USP47*. Data is represented by mean \pm s.e.m. with n=3 biological replicates overlaid and n=4-8 technical replicates. ** = p value < 0.01. p values determined by two-tailed *t* test.

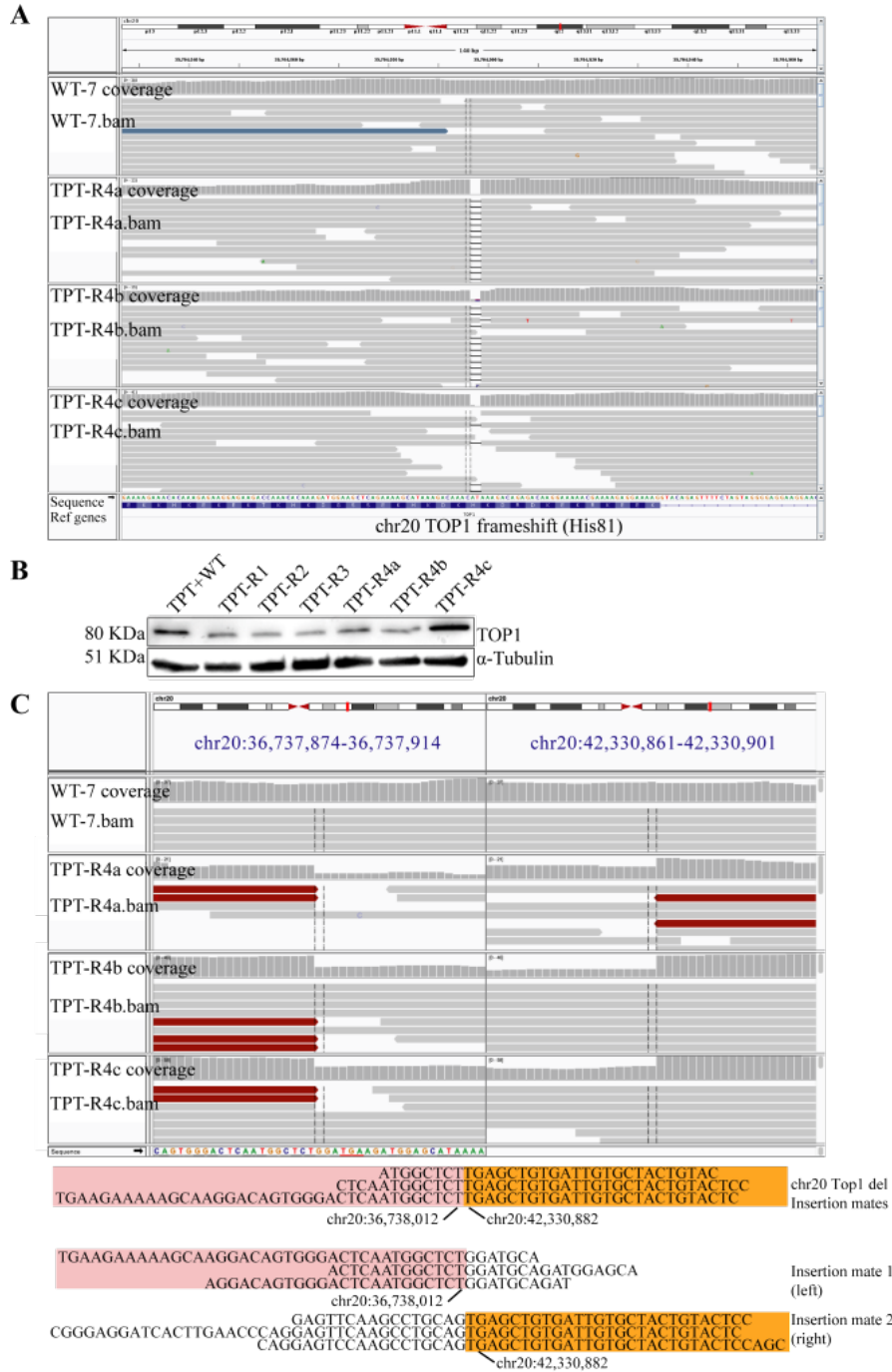


Figure S3.7. IGV views for the TPT target genes.

A. IGV view of *TOP1* frameshift (His81) at TPT cell lines. **B.** Western blot with their gene depletion downregulates protein levels for *TOP1*. **C.** IGV view of the start of the deletion event that includes *TOP1*, where the reads dropped at TPT-R4a, TPT-R4b, and TPT-R4c cell lines compared to WT-7. Left shows the start of the deletion event and right shows the end. At the drop of the reads, there are also insertion events in the replicates but not in the WT.

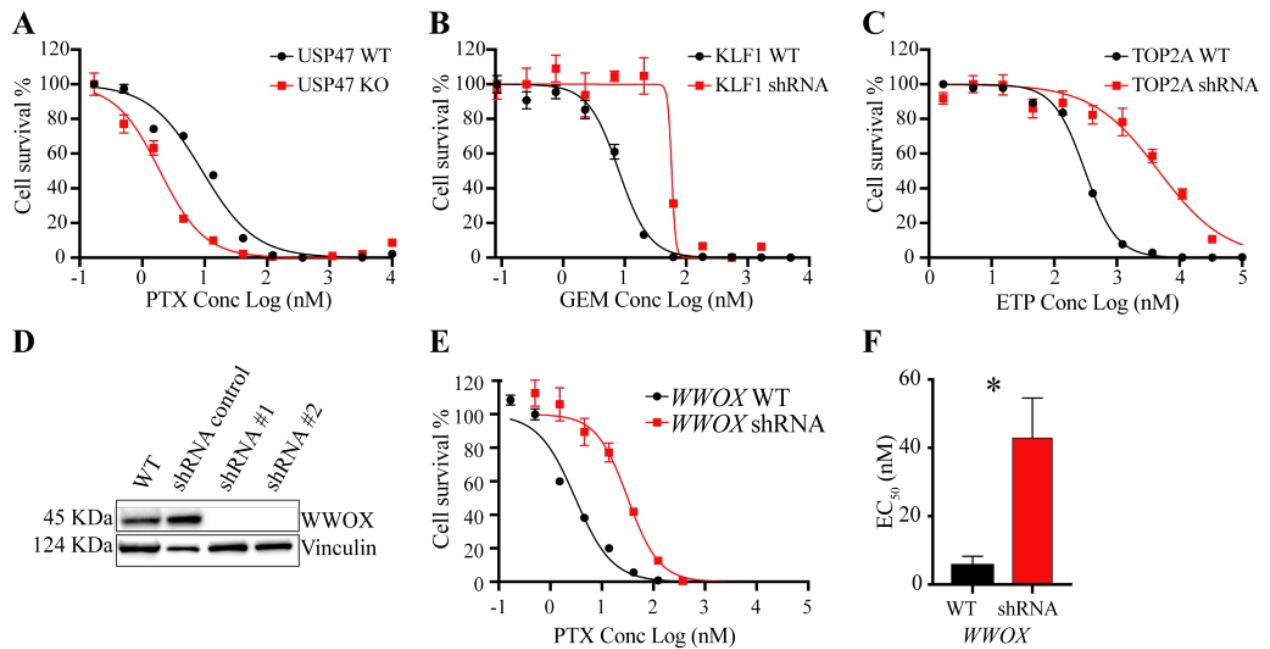


Figure S3.8. Multidrug resistance EC₅₀ curves.

EC₅₀ curves for the validated gene candidates that showed multidrug resistance (MDR) pathways against the other drugs used in the study. **A.** USP47 for PTX, **B.** KLF1 for GEM, **C.** TOP2A for ETP. **D.** Western blot confirming that shRNA gene depletion downregulates protein levels for *WWOX*. **E.** EC₅₀ curves of the WT and shRNA knock-down cell lines. **F.** Boxplot of the WT and shRNA knockdown cell lines for *WWOX*. All data is represented by mean \pm s.e.m. with n=3 biological replicates. Knockdowns are presented with n=4 technical replicates. * = p value < 0.05. p values determined by two-tailed *t* test.

3.9 Author Contributions

Author contributions: JCJ performed selections, validation experiments and wrote the manuscript, MD performed all computational analyses, and assembled figures and tables and wrote the manuscript, GF assisted with CRISPR-Cas9 experiments, HC provided advice and management and wrote the manuscript. EAW performed data analysis, provided advice, obtained funding and wrote the manuscript. TI obtained funding and provided advice.

3.10 Acknowledgements

This work was supported by the National Institute of Health to EAW, HC, and TI (GM085764) and the San Diego Center for Systems Biology and UC San Diego Health Science fellowship to JCJ, and NIH National Library of Medicine Training Grant T15LM011271 to MD. EAW is also supported by grants from the Bill and Melinda Gates Foundation and the Medicines for Malaria Venture. The authors declare no conflicts of interest.

Chapter 3, in full, is a reformatted presentation of the material currently being prepared for submission for publication as “Discovering biomarkers of chemotherapy resistance using *in vitro* evolution in haploid human cells” by Michelle Dow, Juan Carlos Jado, Krypton Carolino, Gregory J. Fonseca, Trey Ideker, Hannah Carter, and Elizabeth A. Winzeler. The dissertation author was the primary investigator and author of this material.

3.11 References

- Agarwal, R., and Kaye, S.B. (2003). Ovarian cancer: strategies for overcoming resistance to chemotherapy. *Nature Reviews Cancer* 3, 502–516.
- Alvarellos, M.L., Lamba, J., Sangkuhl, K., Thorn, C.F., Wang, L., Klein, D.J., Altman, R.B., and Klein, T.E. (2014). PharmGKB summary: gemcitabine pathway. *Pharmacogenet. Genomics* 24, 564–574.
- Amrutkar, M., and Gladhaug, I. (2017). Pancreatic Cancer Chemoresistance to Gemcitabine. *Cancers* 9, 157.
- Andersson, B.S., Collins, V.P., Kurzrock, R., Larkin, D.W., Childs, C., Ost, A., Cork, A., Trujillo, J.M., Freireich, E.J., and Siciliano, M.J. (1995). KBM-7, a human myeloid leukemia cell line with double Philadelphia chromosomes lacking normal c-ABL and BCR transcripts. *Leukemia* 9, 2100–2108.
- Andries, K., Verhasselt, P., Guillemont, J., Göhlmann, H.W.H., Neefs, J.-M., Winkler, H., Van Gestel, J., Timmerman, P., Zhu, M., Lee, E., Williams, P., de Chaffoy, D., Huitric, E., Hoffner, S., Cambau, E., Truffot-Pernot, C., Lounis, N., and Jarlier, V. (2005). A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. *Science* 307, 223–227.
- Antonova-Koch, Y., Meister, S., Abraham, M., Luth, M.R., Otilie, S., Lukens, A.K., Sakata-Kato, T., Vanaerschot, M., Owen, E., Jado, J.C., Maher, S. P., Calla, J., Plouffe, D., Zhong, Y., Chen, K., Chaumeau, V., Conway, A. J., McNamara, C. W., Ibanez, M., and Winzeler, E. A. (2018). Open-source discovery of chemical leads for next-generation chemoprotective antimalarials. *Science* 362.
- Ariey, F., Witkowski, B., Amaratunga, C., Beghain, J., Langlois, A.-C., Khim, N., Kim, S., Duru, V., Bouchier, C., Ma, L., Lim, P., Leang, R., Duong, S., Sreng, S., Suon, S., Chuor, C. M., Bout, D. M., Ménard, S., Rogers, W. O., and Ménard, D. (2014). A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria. *Nature* 505, 50–55.
- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., Reddy, A., Liu, M., Murray, L., Berger, M. F., Monahan, J. E., Morais, P., Meltzer, J., Korejwa, A., Jané-Valbuena, J., and Garraway, L. A. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483, 603–607.
- Bepler, G., Kusmartseva, I., Sharma, S., Gautam, A., Cantor, A., Sharma, A., and Simon, G. (2006). RRM1 modulated in vitro and in vivo efficacy of gemcitabine and platinum in non-small-cell lung cancer. *J. Clin. Oncol.* 24, 4731–4737.
- Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappo, J., Schleiermacher, G., Janoueix-Lerosey, I., Delattre, O., and Barillot, E. (2012). Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* 28, 423–425.

- Brown, R., Curry, E., Magnani, L., Wilhelm-Benartzi, C.S., and Borley, J. (2014). Poised epigenetic states and acquired drug resistance in cancer. *Nat. Rev. Cancer* *14*, 747–753.
- Bürkstümmer, T., Banning, C., Hainzl, P., Schobesberger, R., Kerzendorfer, C., Pauler, F.M., Chen, D., Them, N., Schischlik, F., Rebsamen, M., Smida, M., Fece de la Cruz, F., Lapao, A., Liszt, M., Eizinger, B., Guenzl, P. M., Blomen, V. A., Konopka, T., Gapp, B., and Nijman, S. M. B. (2013). A reversible gene trap collection empowers haploid genetics in human cells. *Nat. Methods* *10*, 965–971.
- Burgess, D.J., Doles, J., Zender, L., Xue, W., Ma, B., McCombie, W.R., Hannon, G.J., Lowe, S.W., and Hemann, M.T. (2008). Topoisomerase levels determine chemotherapy response in vitro and in vivo. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 9053–9058.
- Calle, A.S., Kawamura, Y., Yamamoto, Y., Takeshita, F., and Ochiya, T. (2018). Emerging roles of long non-coding RNA in cancer. *Cancer Science* *109*, 2093–2100.
- Carette, J.E., Raaben, M., Wong, A.C., Herbert, A.S., Obernosterer, G., Mulherkar, N., Kuehne, A.I., Kranzusch, P.J., Griffin, A.M., Ruthel, G., Dal Cin, P., Dye, J. M., Whelan, S. P., Chandran, K., & Brummelkamp, T. R. (2011). Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* *477*, 340–343.
- Chan, S.L., Huppertz, I., Yao, C., Weng, L., Moresco, J.J., Yates, J.R., 3rd, Ule, J., Manley, J.L., and Shi, Y. (2014). CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. *Genes Dev.* *28*, 2370–2380.
- Christowitz, C., Davis, T., Isaacs, A., van Niekerk, G., Hattingh, S., and Engelbrecht, A.-M. (2019). Mechanisms of doxorubicin-induced drug resistance and drug resistant tumour growth in a murine breast tumour model. *BMC Cancer* *19*.
- Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* *6*, 80–92.
- Corey, V.C., Lukens, A.K., Istvan, E.S., Lee, M.C.S., Franco, V., Magistrado, P., Coburn-Flynn, O., Sakata-Kato, T., Fuchs, O., Gnädig, N.F., Goldgof, G., Linares, M., Gomez-Lorenzo, M. G., De Cózar, C., Lafuente-Monasterio, M. J., Prats, S., Meister, S., Tanaseichuk, O., Wree, M., and Winzeler, E.A. (2016). A broad analysis of resistance development in the malaria parasite. *Nat. Commun.* *7*, 11901.
- Cowell, A.N., Istvan, E.S., Lukens, A.K., Gomez-Lorenzo, M.G., Vanaerschot, M., Sakata-Kato, T., Flannery, E.L., Magistrado, P., Owen, E., Abraham, M., and Winzeler, E.A. (2018). Mapping the malaria parasite druggable genome by using in vitro evolution and chemogenomics. *Science* *359*, 191–199.
- Cruz-Muñoz, W., Di Desidero, T., Man, S., Xu, P., Jaramillo, M.L., Hashimoto, K., Collins, C., Banville, M., O'Connor-McCourt, M.D., and Kerbel, R.S. (2014). Analysis of acquired resistance

to metronomic oral topotecan chemotherapy plus pazopanib after prolonged preclinical potent responsiveness in advanced ovarian cancer. *Angiogenesis* 17, 661–673.

De la Casa-Fages, B., Fernández-Eulate, G., Gamez, J., Barahona-Hernando, R., Morís, G., García-Barcina, M., Infante, J., Zulaica, M., Fernández-Pelayo, U., Muñoz-Oreja, M., and López de Munaín, A. (2019). Parkinsonism and spastic paraplegia type 7: Expanding the spectrum of mitochondrial Parkinsonism. *Mov. Disord.* 34, 1547–1561.

Drenberg, C.D., Shelat, A., Dang, J., Cotton, A., Orwick, S.J., Li, M., Jeon, J.Y., Fu, Q., Buelow, D.R., Pioso, M., and Baker, S.D. (2019). A high-throughput screen indicates gemcitabine and JAK inhibitors may be useful for treating pediatric AML. *Nat. Commun.* 10, 2189.

Ghisoni, E., Maggiorotto, F., Borella, F., Mittica, G., Genta, S., Giannone, G., Katsaros, D., Sciarrillo, A., Ferrero, A., Sarotto, I., and Valabrega, G. (2019). TOP2A as marker of response to pegylated liposomal doxorubicin (PLD) in epithelial ovarian cancers. *J. Ovarian Res.* 12, 17.

Greshock, J., Bachman, K.E., Degenhardt, Y.Y., Jing, J., Wen, Y.H., Eastman, S., McNeil, E., Moy, C., Wegrzyn, R., Auger, K., and Wooster, R. (2010). Molecular target class is predictive of in vitro response profile. *Cancer Res.* 70, 3677–3686.

Hansen, L.T., Lundin, C., Spang-Thomsen, M., Petersen, L.N., and Helleday, T. (2003). The role of RAD51 in etoposide (VP16) resistance in small cell lung cancer. *Int. J. Cancer* 105, 472–479.

Hayashi, M.T., and Karlseder, J. (2013). DNA damage associated with mitosis and cytokinesis failure. *Oncogene* 32, 4593–4601.

Huang, F., and Mazin, A.V. (2014). A small molecule inhibitor of human RAD51 potentiates breast cancer cell killing by therapeutic agents in mouse xenografts. *PLoS One* 9, e100993.

Huang, X., Taeb, S., Jahangiri, S., Emmenegger, U., Tran, E., Bruce, J., Mesci, A., Korpela, E., Vesprini, D., Wong, C.S., and Liu, S.K. (2013). miRNA-95 mediates radioresistance in tumors by targeting the sphingolipid phosphatase SGPP1. *Cancer Res.* 73, 6972–6986.

Januchowski, R., Sterzyńska, K., Zawierucha, P., Ruciński, M., Świerczewska, M., Partyka, M., Bednarek-Rajewska, K., Brązert, M., Nowicki, M., Zabel, M., and Klejewski, A. (2017). Microarray-based detection and expression analysis of new genes associated with drug resistance in ovarian cancer cell lines. *Oncotarget* 8, 49944–49958.

Jarrold, J., and Davies, C.C. (2019). PRMTs and Arginine Methylation: Cancer’s Best-Kept Secret? *Trends in Molecular Medicine* 25, 993–1009.

Jeon, K.-H., Yu, H.V., and Kwon, Y. (2018). Hyperactivated m-calpain affects acquisition of doxorubicin resistance in breast cancer cells. *Biochim. Biophys. Acta Gen. Subj.* 1862, 1126–1133.

Jiang, C., Jiang, L., Li, Q., Liu, X., Zhang, T., Yang, G., Zhang, C., Wang, N., Sun, X., and Jiang, L. (2019). Pyrroloquinoline quinone ameliorates doxorubicin-induced autophagy-dependent apoptosis via lysosomal-mitochondrial axis in vascular endothelial cells. *Toxicology* 425, 152238.

Ju, T., Jin, H., Ying, R., Xie, Q., Zhou, C., and Gao, D. (2017). Overexpression of NAC1 confers drug resistance via HOXA9 in colorectal carcinoma cells. *Mol. Med. Rep.* *16*, 3194–3200.

Ko, D.-H., Lee, E.J., Lee, S.-K., Kim, H.-S., Shin, S.Y., Hyun, J., Kim, J.-S., Song, W., and Kim, H.S. (2019). Application of next-generation sequencing to detect variants of drug-resistant *Mycobacterium tuberculosis*: genotype-phenotype correlation. *Ann. Clin. Microbiol. Antimicrob.* *18*, 2.

Krummel, K.A., Roberts, L.R., Kawakami, M., Glover, T.W., and Smith, D.I. (2000). The characterization of the common fragile site FRA16D and its involvement in multiple myeloma translocations. *Genomics* *69*, 37–46.

Leggas, M., Adachi, M., Scheffer, G.L., Sun, D., Wielinga, P., Du, G., Mercer, K.E., Zhuang, Y., Panetta, J.C., Johnston, B., and Schuetz, J.D. (2004). Mrp4 confers resistance to topotecan and protects the brain from chemotherapy. *Mol. Cell. Biol.* *24*, 7612–7621.

Lim, M.Y.-X., LaMonte, G., Lee, M.C.S., Reimer, C., Tan, B.H., Corey, V., Tjahjadi, B.F., Chua, A., Nachon, M., Wintjens, R., and Bifani, P. (2016). UDP-galactose and acetyl-CoA transporters as *Plasmodium* multidrug resistance genes. *Nat Microbiol* *1*, 16166.

Luth, M.R., Gupta, P., Otilie, S., and Winzeler, E.A. (2018). Using in Vitro Evolution and Whole Genome Analysis To Discover Next Generation Targets for Antimalarial Drug Discovery. *ACS Infect Dis* *4*, 301–314.

Mandell, J.B., Lu, F., Fisch, M., Beumer, J.H., Guo, J., Watters, R.J., and Weiss, K.R. (2019). Combination Therapy with Disulfiram, Copper, and Doxorubicin for Osteosarcoma: In Vitro Support for a Novel Drug Repurposing Strategy. *Sarcoma* *2019*, 1–9.

Mansoori, B., Mohammadi, A., Davudian, S., Shirjang, S., and Baradaran, B. (2017). The Different Mechanisms of Cancer Drug Resistance: A Brief Review. *Adv Pharm Bull* *7*, 339–348.

Mini, E., Nobili, S., Caciagli, B., Landini, I., and Mazzei, T. (2006). Cellular pharmacology of gemcitabine. *Ann. Oncol.* *17 Suppl 5*, v7–v12.

Naghavi, L., Schwalbe, M., Ghanem, A., and Naumann, M. (2018). Deubiquitinylase USP47 Promotes RelA Phosphorylation and Survival in Gastric Cancer Cells. *Biomedicines* *6*.

Otilie, S., Goldgof, G.M., Calvet, C.M., Jennings, G.K., LaMonte, G., Schenken, J., Vigil, E., Kumar, P., McCall, L.-I., Lopes, E.S.C., Durrant, J.D., and Winzeler, E.A. (2017). Rapid Chagas Disease Drug Target Discovery Using Directed Evolution in Drug-Sensitive Yeast. *ACS Chem. Biol.* *12*, 422–434.

Pan, P., Li, Y., Yu, H., Sun, H., and Hou, T. (2013). Molecular principle of topotecan resistance by topoisomerase I mutations through molecular modeling approaches. *J. Chem. Inf. Model.* *53*, 997–1006.

Pfister, T.D., Reinhold, W.C., Agama, K., Gupta, S., Khin, S.A., Kinders, R.J., Parchment, R.E., Tomaszewski, J.E., Doroshov, J.H., and Pommier, Y. (2009). Topoisomerase I levels in the NCI-

60 cancer cell line panel determined by validated ELISA and microarray analysis and correlation with indenoisoquinoline sensitivity. *Mol. Cancer Ther.* 8, 1878–1884.

Pommier, Y., Leo, E., Zhang, H., and Marchand, C. (2010). DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chem. Biol.* 17, 421–433.

Raposo, A.E., and Piller, S.C. (2018). Protein arginine methylation: an emerging regulator of the cell cycle. *Cell Division* 13.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26.

Saraswathy, M., and Gong, S. (2013). Different strategies to overcome multidrug resistance in cancer. *Biotechnology Advances* 31, 1397–1407.

Shanmughapriya, S., Rajan, S., Hoffman, N.E., Higgins, A.M., Tomar, D., Nemani, N., Hines, K.J., Smith, D.J., Eguchi, A., Vallem, S., and Madesh, M. (2015). SPG7 Is an Essential and Conserved Component of the Mitochondrial Permeability Transition Pore. *Mol. Cell* 60, 47–62.

Si, W., Shen, J., Zheng, H., and Fan, W. (2019). The role and mechanisms of action of microRNAs in cancer drug resistance. *Clin. Epigenetics* 11, 25.

Srivastava, S., and Panda, D. (2018). A centrosomal protein STARD9 promotes microtubule stability and regulates spindle microtubule dynamics. *Cell Cycle* 17, 2052–2068.

Sur, I., and Taipale, J. (2016). The role of enhancers in cancer. *Nat. Rev. Cancer* 16, 483–493.

Teloni, F., Michelena, J., Lezaja, A., Kilic, S., Ambrosi, C., Menon, S., Dobrovolna, J., Imhof, R., Janscak, P., Baubec, T., and Altmeyer, M. (2019). Efficient Pre-mRNA Cleavage Prevents Replication-Stress-Associated Genome Instability. *Mol. Cell* 73, 670–683.e12.

Tong, M., Zheng, W., Lu, X., Ao, L., Li, X., Guan, Q., Cai, H., Li, M., Yan, H., Guo, Y., Chi, P., and Guo, Z. (2015). Identifying clinically relevant drug resistance genes in drug-induced resistant cancer cell lines and post-chemotherapy tissues. *Oncotarget* 6, 41216–41227.

Turner, J.G., Dawson, J.L., Grant, S., Shain, K.H., Dalton, W.S., Dai, Y., Meads, M., Baz, R., Kauffman, M., Shacham, S., and Sullivan, D.M. (2016). Treatment of acquired drug resistance in multiple myeloma by combination therapy with XPO1 and topoisomerase II inhibitors. *J. Hematol. Oncol.* 9, 73.

Wang, S., Konorev, E.A., Kotamraju, S., Joseph, J., Kalivendi, S., and Kalyanaraman, B. (2004). Doxorubicin induces apoptosis in normal and tumor cells via distinctly different mechanisms. Intermediacy of H(2)O(2)- and p53-dependent pathways. *J. Biol. Chem.* 279, 25535–25543.

Wang, X., Rusin, A., Walkey, C.J., Lin, J.J., and Johnson, D.L. (2019). The RNA polymerase III repressor MAF1 is regulated by ubiquitin-dependent proteasome degradation and modulates cancer drug resistance and apoptosis. *J. Biol. Chem.* 294, 19255–19268.

Weinstein, J.N., Myers, T.G., O'Connor, P.M., Friend, S.H., Fornace, A.J., Jr, Kohn, K.W., Fojo, T., Bates, S.E., Rubinstein, L.V., Anderson, N.L., and Paull, K.D. (1997). An information-intensive approach to the molecular pharmacology of cancer. *Science* 275, 343–349.

Weinstock, J., Wu, J., Cao, P., Kingsbury, W.D., McDermott, J.L., Kodrasov, M.P., McKelvey, D.M., Suresh Kumar, K.G., Goldenberg, S.J., Mattern, M.R., and Nicholson, B. (2012). Selective Dual Inhibitors of the Cancer-Related Deubiquitylating Proteases USP7 and USP47. *ACS Med. Chem. Lett.* 3, 789–792.

Wendorff, T.J., Schmidt, B.H., Heslop, P., Austin, C.A., and Berger, J.M. (2012). The structure of DNA-bound human topoisomerase II alpha: conformational mechanisms for coordinating inter-subunit interactions with DNA cleavage. *J. Mol. Biol.* 424, 109–124.

Wright, C.M., van der Merwe, M., DeBrot, A.H., and Bjornsti, M.-A. (2015). DNA topoisomerase I domain interactions impact enzyme activity and sensitivity to camptothecin. *J. Biol. Chem.* 290, 12068–12078.

Yang, Y., Guan, D., Lei, L., Lu, J., Liu, J.Q., Yang, G., Yan, C., Zhai, R., Tian, J., Bi, Y., Fu, F., and Wang, H. (2018). H6, a novel hederagenin derivative, reverses multidrug resistance in vitro and in vivo. *Toxicol. Appl. Pharmacol.* 341, 98–105.

Yang, Z., Fu, B., Zhou, L., Xu, J., Hao, P., and Fang, Z. (2019). RRM1 predicts clinical outcome of high-and intermediate-risk non-muscle-invasive bladder cancer patients treated with intravesical gemcitabine monotherapy. *BMC Urol.* 19, 69.

Yu, L., Dong, L., Wang, Y., Liu, L., Long, H., Li, H., Li, J., Yang, X., Liu, Z., Duan, G., Dai, X., and Lin, Z. (2019). Reversible regulation of SATB1 ubiquitination by USP47 and SMURF2 mediates colon cancer cell proliferation and tumor progression. *Cancer Lett.* 448, 40–51.

Zeng, C., Fan, W., and Zhang, X. (2015). RRM1 expression is associated with the outcome of gemcitabine-based treatment of non-small cell lung cancer patients--a short report. *Cell. Oncol.* 38, 319–325.

Zhang, Y., Secrist, J.A., 3rd, and Ealick, S.E. (2006). The structure of human deoxycytidine kinase in complex with clofarabine reveals key interactions for prodrug activation. *Acta Crystallogr. D Biol. Crystallogr.* 62, 133–139.

Zhang, Z., Cheng, L., Li, J., Farah, E., Atallah, N.M., Pascuzzi, P.E., Gupta, S., and Liu, X. (2018). Inhibition of the Wnt/ β -Catenin Pathway Overcomes Resistance to Enzalutamide in Castration-Resistant Prostate Cancer. *Cancer Res.* 78, 3147–3162.

Zubovych, I.O., Straud, S., and Roth, M.G. (2010). Mitochondrial dysfunction confers resistance to multiple drugs in *Caenorhabditis elegans*. *Mol. Biol. Cell* 21, 956–969.

EPILOGUE

Limitations and Future Directions

This dissertation provides frameworks to better characterize model organisms used to study cancer and systematically assesses the impact of the genetic variation on tumor formation and response to antitumor treatments. While this effort represents a significant step forward for the field, the approach also has significant limitations and requires further study.

Though I have uncovered very compelling results, the characterization of molecular features for precision medicine has substantial room for improvement. In chapter one, I provided a framework to better characterize model organisms used to study HCC (Dow et al., 2018). In the study for the four mouse models for HCC, there are two specific areas of improvement. Firstly, the study design and samples collected for sequencing can be improved. We utilized the RNAseq data to identify differentially expressed genes (DEGs) between TCGA human subgroups. Using the DEGs, we were able to categorize the mouse models as valid for the study of different stages of human HCC. However, without normal mouse samples for RNAseq, we could not perform a DEGs analysis for tumor development in individual mouse models. While we can identify DEGs between the four mouse models to find the relative DEGs, the normal samples will be essential in understanding how gene expression changes from the normal liver tissues to various stages of HCC. The up- and down- regulated genes in normal versus tumor samples would provide a more accurate set of DEGs and further reveal the developmental differences between the four HCC models (Liang and Pardee, 2003). Tumor-specific transcriptomic profiles for the samples provide clues of potential target genes involved in the four models. Secondly, we could not identify specific immune differences between the mouse models from bulk RNAseq using known immune markers (Davoli et al., 2017). Recent studies suggest that single cell RNAseq (scRNAseq) can provide

better resolutions for specific immune cell expressions (Gomes et al., 2019). Thus, there is room for improvement in the study of HCC transcriptomics to further understand the similarity and differences to between mouse models and human disease.

While chapter two focused on the transcriptomic and chromatin structure of responsive genes; mutations and large genome alteration are known to drive cancer and have potential to reorganized by the immune system (Schmidt et al., 2017). Thus, expanding the study to include genomic alterations such as single nucleotide variants or copy number variations is vital to achieving a more comprehensive picture of the immune landscape in mono- and combined therapies. In addition, expanding the study to alternative pathways of NF- κ B could improve the overall understanding of the response mechanisms.

Expanding analyses to include non-coding genomic alterations would add information for a more comprehensive understanding for results found in chapter one and three. Even though whole genome sequencing data was not available for chapter one, we identified some elements in non-coding regions in chapter three that could be further explored. For example, we found non-coding RNAs such as *EGRF-ASI* and activating cis elements such as enhancers which have previously been implicated in evasion of drug response (Calle et al., 2018; Si et al., 2019; Sur and Taipale, 2016). These intergenic mutations with high allele frequency are present in our provided datasets and provide opportunity for reanalysis or for querying by those interested in a specific noncoding RNA or enhancer.

Finally, I also acknowledge that genetics alterations capture only a small slice of the whole picture of tumor-treatment interactions. In order to probe the whole picture, more samples and more diverse model organisms will be of great value. My framework for studying HCC mouse models and humans from chapter one and my pipeline for identifying potential resistance genes

from chapter three onto different tumor types provide a foundation that could enable the development of improved methods.

Conclusions

The majority of anti-cancer therapies are effective in only a subset of the patient population, thus there is a high urgency to understand the patient heterogeneity leading to the spectrum of outcomes. As the cost for generating NGS data drops, multi-omics analysis can be increasingly used to help characterize individual differences and tailor treatments to specific tissues, genes, and transcriptomic factors relevant to each case of cancer (Krzyszczyk et al., 2018). These approaches could also be utilized to minimize the damage of cancer therapies to healthy cells, improving treatments safety and effectiveness. Lastly, many studies have found that genomic variation plays a tremendous role in shaping tumor phenotypes.

In this dissertation, I furthered efforts to understand the role of tumor heterogeneity on treatment outcomes. Firstly, I systematically compared the HCC model organisms to patients to better categorize the models. I identified mouse models suitable for studying early initiation of HCC (TAK1 and MUP), more aggressive HCC (STAM), and one that should be avoided for HCC studies (DEN). Second, to better understand the effect of mono- and combined therapies on HCC, I studied the different conditions in the mouse models and identified the key response-related mechanisms. The contribution of MHC-I machinery genes was anticipated due to its influence on other tumor phenotypes; however, MHC-I machinery genes' interactions with CREB and P300 regulatory mechanisms show potential novel mechanisms for drug targeting. Finally, I explored the genetics of acquired resistance. I presented evidence that using genomic information, we can identify potential alleles or genes that evolve during drug resistance. Though the models are unable

to capture the full complexity of tumor-treatment interactions, the approach outlined in this dissertation lays a foundation for understanding and capturing the heterogeneity of cancer. This improved understanding has the potential to aid in better design and development of more personalized precision medicine for cancer.

References

- Calle, A.S., Kawamura, Y., Yamamoto, Y., Takeshita, F., and Ochiya, T. (2018). Emerging roles of long non-coding RNA in cancer. *Cancer Science* *109*, 2093–2100.
- Davoli, T., Uno, H., Wooten, E.C., and Elledge, S.J. (2017). Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science* *355*.
- Dow, M., Pyke, R.M., Tsui, B.Y., Alexandrov, L.B., Nakagawa, H., Taniguchi, K., Seki, E., Harismendy, O., Shalapour, S., Karin, M., Font-Burgada, J., and Carter, H. (2018). Integrative genomic analysis of mouse and human hepatocellular carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* *115*, E9879–E9888.
- Gomes, T., Teichmann, S.A., and Talavera-López, C. (2019). Immunology Driven by Large-Scale Single-Cell Sequencing. *Trends Immunol.* *40*, 1011–1021.
- Krzyszczuk, P., Acevedo, A., Davidoff, E.J., Timmins, L.M., Marrero-Berrios, I., Patel, M., White, C., Lowe, C., Sherba, J.J., Hartmanshenn, C., and Yarmush, M.L. (2018). The growing role of precision and personalized medicine for cancer treatment. *Technology* *6*, 79–100.
- Liang, P., and Pardee, A.B. (2003). Analysing differential gene expression in cancer. *Nature Reviews Cancer* *3*, 869–876.
- Schmidt, R.E., Grimbacher, B., and Witte, T. (2017). Autoimmunity and primary immunodeficiency: two sides of the same coin? *Nat. Rev. Rheumatol.* *14*, 7–18.
- Si, W., Shen, J., Zheng, H., and Fan, W. (2019). The role and mechanisms of action of microRNAs in cancer drug resistance. *Clin. Epigenetics* *11*, 25.
- Sur, I., and Taipale, J. (2016). The role of enhancers in cancer. *Nat. Rev. Cancer* *16*, 483–493.