Genetic Markers of 5-Fluorouracil Associated-Toxicity in Colorectal Cancer Patients

By

Chenjie Zeng

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Epidemiology

December 16, 2017

Nashville, Tennessee

Approved:

Wei Zheng, M.D., Ph.D., M.P.H.

Todd L. Edwards, Ph.D.

Bingshan Li, Ph.D.

Chang Yu, Ph.D.

Copyright © 2017 by Chenjie Zeng

All Rights Reserved

ACKNOWLEDGEMENTS

I wish to thank Dr. Wei Zheng, my PhD mentor, for showing me how to be a scientist, sharing me with his insights in cancer, guiding me through those dark days in research, giving me tremendous support and encouragement, teaching me how to write, and showing me how to write a paper during a flight.

I wish to thank Dr. Bingshan Li, my committee member, for sharing me with perspectives of a statistical geneticist on my research in epidemiology and always reminding me positive sides of things. I wish to thank Dr. Todd Edwards, my committee member, for teaching me how to run a GWAS analysis for the first time in my life, encouraging me trying new approaches, and giving me advice in career and life in general. I wish to thank Dr. Chang Yu, my committee member, for sharing me with perspectives of a biostatician and providing valuable suggestions on analyses in my dissertation. I wish to thank Dr. Emily Chan for her advice on evaluating toxicity events in cancer patients and introducing me how oncologists think.

I wish to thank Dr. Qi Liu for helping me understand concepts in bioinformatics and showing me how to perform bioinformatics analyses, Ms. Jing He for teaching me how to program, Dr. Jiajun Shi for performing genotyping for this dissertation. I also wish to thank members of BioVU, particularly Ms. Lisa Price and Dr. Sarah Collier, and members of VICTR.

I also wish to thank my faculty advisor, Dr. Dai Qi for his advice on research projects and career development, and all the teaching faculty members at Vanderbilt Epidemiology for sharing their understandings and knowledge.

In the end, this work would not have been possible without the financial support from Anne Potter Wilson Chair funds from the Vanderbilt University School of Medicine and the National Center for Advancing Translational Sciences of the National Institute of Health under Award Number UL1 TR000445.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS iii

LIST OF TABLES vi

LIST OF FIGURES vii

ABBREVIATIONS viii

Chapter

I. Introduction 1

Significance 1
Specific aims 2

II. Background 4

Chemotherapy in colorectal cancer 4
Pharmacology of 5-FU 17
Predictors of 5-FU associated toxicity 21
Potential roles of regulotary variants in drug-associated toxicity 32

III. Integrative functional genomic analyses identify genetic markers of 5-fluorouracil catabolic pathway activities 37

Introduction 37
Material and Methods 38
Results 42
Discussion 43

IV. Genetic markers of 5- fluorouracil catabolic pathway activities are associated with risk of chemotherapy-associated severe toxicity events in colorectal cancer patients 48

Introduction 48
Methods 49
Results 52
Discussion 54

V. Summary and future directions 63

Summary 63
Future directions 64

Appendix

A. Supplementary information for Chapter III 66

B. Supplementary information for Chapter IV 71

REFERENCES 76

LIST OF TABLES

Table Page

1. AJCC staging -Primary Tumor ... 11
2. AJCC staging -Regional Lymph Nodes (N) ... 12
3. AJCC staging -Distant Metastasis (M) ... 13
4. Anatomic Stage/Prognostic Groups ... 14
5. National Cancer Institute Common Toxicity Criteria for Haematological Adverse Events (CTCAE) v4.0 ... 15
6. National Cancer Institute Common Toxicity Criteria for Gastrointestinal and Dermatologic Adverse Events (CTCAE) v4.0 ... 16
7. Genetic variants in potential regulatory regions predicted the expression of genes in 5-FU catabolic pathway ... 46
8. Integrative analyses of functional genomics data identified potential regulatory variants for genes in the 5-FU catabolic pathway ... 47
9. Selected baseline characteristics of colorectal cancer patients of European ancestry who received 5-FU based chemotherapy at Vanderbilt University Medical Center, 1997-2016 ... 58
10. Association between previously reported variants and new genetic markers that might associated with expression of genes in the 5-fluorouracil catabolic pathway and risk of early-onset severe toxicity events in colorectal cancer patients of European ancestry ... 59
11. Association of genetic markers and risk of subtypes of early-onset severe toxicity events in colorectal cancer patients of European ancestry ... 60
12. Prediction performance of models with genetic markers and non-genetic risk factors in CRC patients ... 61

LIST OF FIGURES

Figure Page

1. Five-FU metabolism pathways 20
2. Flowchart describing the BioVU colorectal cancer cohort with side boxes explaining the reasons for exclusion. 62

ABBREVIATIONS

CRC: Colorectal cancer

5-FU: 5-Fluorouracil

FOLFOX: Leucovorin Calcium+ Fluorouracil+ Oxaliplatin

FOLFIRI: Leucovorin Calcium+ Fluorouracil+ Irinotecan Hydrochloride

DPYD: Dihydropyrimidine dehydrogenase

DPD: Dihydropyrimidine dehydrogenase

DPYS: Dihydropyrimidinase

UPB1: Ureidopropionase, Beta

VUMC: Vanderbilt University Medical Center

EMR: Electronic medical records

MAF: Minor allele frequency

EAF: Effect allele frequency

DHS: DNase I hypersensitivity sites

OR: Odds ratio

HR: Hazard ratio

CI: Confidence interval

AJCC: American Joint Committee on Cancer

CEA: Carcinoembryonic antigen

MSI: Microsatellite instability

KRAS: Kirsten rat sarcoma viral oncogene homolog

LV: Leucovorin

VEGF: Vascular endothelial growth factor

EGF: Epidermal growth factor

EGFR: Epidermal growth factor receptor

BSA: Body surface area

DPD: Dihydropyrimidine dehydrogenase

DHPase: Dihydropyrimidinase

BUP-1: beta-ureidopropionase

FUH2: Dihydrofluorouracil

FUPA: Fluoro-beta-ureidopropionate

TF: Transcription factor

TFBS: Transcription factor binding sites

CHAPTER I

INTRODUCTION

Significance

Colorectal cancer (CRC) is one of most commonly diagnosed cancers in the United States, with an estimated 132,700 new cases in 2015.[1] A steady improvement in survival of CRC has been seen in the past two decades, part of which can be attributed to adjuvant chemotherapy. [2] According to a recent analysis of national data, approximately 70% of the stage III CRC patients and 50% of the stage IV patients diagnosed during 2007 and 2011 received chemotherapy. [3] Despite the dramatic advancement of drug modalities since 1990s, 5-fluorouracil (5-FU) remains the backbone of the chemotherapies for CRC. However, approximately 10~70% of the patients developed at least one severe toxicity event during the treatment, depending on the regimen used.[4] Common 5-FU associated toxicity events include neutropenia, thrombocytopenia, leukopenia, anemia, and gastrointestinal toxicity events as well as hand-foot syndromes, which can lead to dose reduction, poor response, low quality of life, and in extreme cases, deaths. [4, 5] Observational studies showed that patients with 5-FU toxicity events clustered in families, [6,7] suggesting an important role of genetic factors. To identify genetic predictors of toxicity, previous studies investigated extensively protein-coding variants in genes in the metabolism pathways of 5-FU. However, only several coding and splicing variants in the *DPYD* gene that encodes the initial and rate-limiting enzyme of 5-FU degradation showed consistent associations with risk of toxicity in CRC patients. [4,8,9] These variants have low minor allele frequencies (MAFs), among which the most common MAF is 1% in the white population, and thus only explained a small fraction of the genetic variability of what in the patient population. Their predictive values in clinical settings are much debated. To develop a risk assessment model for 5-FU-associated toxicity, additional genetic risk factors should be identified. Previous studies suggested that variants in regulatory regions such as enhancers might be associated with adverse drug reactions through regulating the expression of the drug-metabolizing genes. [10] For example, SNP rs5758550 (Global MAF = 0.25) locating in an enhancer region 115 kb downstream of the gene *CYP2D6* was found to be associated with a more than 2-fold increase in *CYP2D6* expression, [11] and a

CRISPR-mediated deletion of the enhancer region surrounding rs5758550 led to a 70% decreased *CYP2D6* expression. [12] A decreased expression of *CYP2D6* can result in reduced metabolic activities and slow clearance of drugs, [12] and subsequently lead to adverse reactions. [13] However, regulatory variants were understudied in previous pharmacogenetics research of 5-FU associated toxicity. To our knowledge, no such variants in genes in the 5-FU catabolic pathway have been reported in the pharmacogenetics study of 5-FU. In this study, we propose to leverage functional genomic data and gene expression data in liver tissues to identify potential regulatory variants and evaluate their association with risk of 5-FU associated toxicity in CRC patients.

Identification of variants with strong evidence of regulatory functions can help identify causal variants in 5-FU associated toxicity and illuminate the underlying regulatory mechanisms, as discussed above for the *CYP2D6* variant. In addition to the most studied gene *DPYD*, we will also evaluate its downstream genes, *DPYS* and *UPB1*. Previous studies showed that severe 5-FU associated toxicity was found among patients with normal DPD activities, suggesting other genes such as *DPYS* and *UPB1* in the catabolic pathway might play a role. However, previous pharmacogenetics research on these two genes has been limited. [14-16] Our study will provide additional insights into the role of these two genes in 5-FU-associated toxicity. Identification of variants associated with risk of toxicity can facilitate risk assessment and stratification to identify patients at high risk. To date, no risk assessment model for 5-FU toxicity has been established for clinic practice. An important factor hindering clinical implementation is the limited predictive values of the currently known genetic factors. In this study, we aim to identify additional risk variants that explain more genetic variability than the known variants. [17] To improve the predictive performance, I incorporate both genetic and non-genetic risk factors into the prediction models. I anticipate that knowledge gained from this dissertation has the potential for personalized therapeutic management of CRC patients.

Specific aims

In this dissertation, it was hypothesized that regulatory variants for expression of genes in the 5-FU catabolic pathway predicted the risk of 5-FU associated toxicity events.

Aim 1: To identify genetic markers associated with expression of genes in the catabolic pathway of 5-FU including *DPYD*, *DPYS* and *UPB1*, which together have been showed to degrade more than 85% of the administered 5-FU. [23]

Two approaches are proposed. The first approach will focus on predicting expression of genes in the relevant tissues using data from the GTEx projects. The second approach prioritizes potential regulatory variants that regulate gene expression by identifying variants locating at predicted regulatory regions that disrupt the binding of transcription factors.

Aim 2 :To test the association of new genetic markers with risk of severe 5-FU associated toxicity and to build a prediction model using newly identified genetic factors and known non-genetic and genetic factors.

A cohort study of 424 colorectal cancer patients who received chemotherapy between 1997 and 2016 at Vanderbilt University Medical Center is created. The known non-genetic factors include patient characteristics and clinic factors (age, sex, BMI). The known genetic factors include genetic variants that have showed consistent association with the risk of 5-FU-associated toxicity, according to several most recent meta-analyses of genetic markers of 5-FU-associated toxicity. [4,8,9]

CHAPTER II

BACKGROUND

Chemotherapy in colorectal cancer

Epidemiology of colorectal cancer

Colorectal cancer (CRC) is the third most commonly diagnosed cancer in males and the second in females in the world with an estimated 1.4 million new cases in 2012 worldwide. Approximately 50% of the CRC patients will die of this disease. [24] The burden of CRC is expected to increase to more than 2.2 million new cases and 1.1 million deaths worldwide by 2030.[25] Substantial geographic variations in CRC incidence has been observed with incidence rates ranging from less than 5 per 100,000 in some countries in Africa to more than 40 per 100,000 in countries in North America, Europe and Oceania. [25] In the United States, approximately 134,490 new cases of CRC will be diagnosed in 2016, of which 95,270 are colon and the remainder rectal cancers. [26] There are similar incidences of colon cancer in men and women, while there is a higher risk of rectal cancer in men than in women. [26]

Survival of CRC patients has been steadily improved in the developed countries for the last few decades. For example, the 5-year survival rate for patients with CRC in the U.S. increased over the past several decades from 50% in 1970s to 67% in 2008-2013. [26] The improvement in survival can be, in part, attributed to advances in screening, diagnosis and treatment, particularly the development of adjuvant chemotherapy. [2] However, not every subgroup of the population received the survival benefits equally. Racial disparities in CRC survival have been long observed. Multiple studies demonstrated that African Americans had worse overall and stage-specific survival rates than whites. [27,28] In addition, survival disparities across age groups have been increasingly recognized. It was reported that improvements in CRC survival had been much less pronounced among patients aged older than 65 years than among patients who were younger during the last two decades. [29] This disparity was likely due to slower or less adoption of aggressive treatment in the elderly population. [26] There has been a general reluctance in treating older patients with more aggressive chemotherapy (for example, multi-agent regimens), resulting in under-treatment in elderly patients with comparable physical conditions as their younger counterparts.

[30] However, many previous studies have demonstrated that the elderly patients could derive similar benefit from adjuvant chemotherapy as their younger counterparts. [31,32]

The staging, prognosis and management of CRC

The currently recommended staging system for CRC is the TNM system. The TNM system assesses the tumor for its depth of invasion into the bowel wall (T, Table 1), the number of lymph nodes involved (N, Table 2), and the presence of distant metastasis (M, Table 3). CRC patients who are staged by the TNM system can be further grouped into the American Joint Committee on Cancer (AJCC) stages according to their expected prognosis (Table 4). As the AJCC stage increases, the 5-year overall survival decreases dramatically. The 5-year survival rates are 93%, 78%, 64% and nearly 8% for stage I, II, III and IV CRC patients, respectively. [2] Other clinicopathologic features that have been found to be associated with poor prognosis include obstruction, perforation, emergent admission, T4 stage, resection of fewer than 12 lymph nodes, poorly differentiated histology [33] and an increased preoperative carcinoembryonic antigen (CEA) level. [34]

The management for CRC is primarily based on pathological stages. For patients with local or regional diseases (stage I, II or III), which account for 70-80% of all new CRC diagnoses, surgical resection is the cornerstone of curative therapy. In order to remove residual lesions or micro-metastasis after surgery and reduce the risk of recurrence, adjuvant chemotherapy is usually recommended to stage III patients. The role of adjuvant chemotherapy in stage II patients remains unclear. All [35,36] but one [37] clinical trials found no survival benefit of adjuvant chemotherapy in stage II patients who had underwent curative surgery. Even in the only clinical trial showing statistically significant increased survival, the magnitude of the improvement was modest, with an absolute improvement in survival of 3.6% .[37] Due to the large heterogeneity in clinicopathologic features among stage II patients, Moertel *et al.* re-analyzed data from early clinical trials and suggested that stage II patients with poor prognosis features (as described above) might be benefited. [38] However, a population-based study using real-world data of more than 25, 000 stage II CRC patients did not support the use of adjuvant chemotherapy even in patients with poor prognosis features. [39] For patients with metastasis diseases that account for approximately 20% of all

new CRC diagnoses, treatment options include systematic palliative chemotherapy and primary tumor resection. Primary tumor resection is found to be curative in patients with a small number of isolated, organ-confined metastases. [40] However, these patients only account for less than 20% of all patients diagnosed with metastasis diseases. Approximately 80-90% of the stage IV patients are presented with unresectable metastases, for whom systematic palliative chemotherapy is likely the only option. [41,42]

Chemotherapy in colorectal cancer patients

Over the last two decades, tandem advances have been made in the chemotherapy for CRC, which contributed substantially to the improvements in overall survival of cancer patients. [2] Specifically, for patients with resected stage III (node-positive) tumors, an overall survival for 5-fluorouracil (5-FU) based chemotherapy has been well established. For patients with unresectable metastases, systematic palliative chemotherapy has been shown to increase overall survival compared with supportive care.

Adjuvant chemotherapy for localized diseases

The evolution of chemotherapy for CRC is believed to begin with the development of 5-FU in 1957, [43] although early studies failed to demonstrate a survival benefit of 5-FU monotherapy over surgical resection alone. It was in late 1980s that 5-FU gained renewed interest, when Wolmark *et al.* reported that postoperative adjuvant chemotherapy significantly increased disease-free survival in patients with localized diseases in a large scale clinical trial. [44] Following the preclinical findings that reduced folates, such as leucovorin (LV), could enhance the cytotoxicity of 5-FU, [45] Wolmark *et al.* evaluated the effects of LV-modulated 5-FU as adjuvant therapy and found that treatment with 5-FU/LV significantly prolonged disease-free survival and overall survival in Dukes C patients compared with 5-FU only, although at the expense of a higher rate of toxicity events. [46] 5-FU/LV can be administered via infusion or bolus, with varying toxicity profiles. Infusion becomes the preferred route of administration, owing to its superior toxicity profile and response rate. [47,48] Additionally, the oral 5-FU prodrug, capecitabine, which avoids the use of central venous catheters for infusional 5-FU, has showed similar therapeutic efficacy as bolus 5-FU with less severe toxicity profiles. However, data on comparing toxicity profiles between capecitabine and infusional 5-FU has not been available. The next key advance in 5-FU based

chemotherapy is the introduction of oxaliplatin, a platinum-based chemotherapeutic agent with a 1,2-diaminocyclohexane carrier ligand, which forms platinum-DNA adducts and exerts the cytotoxicity by blocking DNA replication. [49] The survival benefit of adding oxaliplatin to postoperative 5-FU based chemotherapy in localized colon cancers has been demonstrated in several randomized trials. [50,51] Furthermore, Sanoff *et al.* pooled data from five population-based databases and showed that the survival benefit of adding oxaliplatin to 5-FU in stage III patients was consistent across different practice settings. [52] For these reasons, the oxaliplatin modified 5-FU/LV regimen (FOLFOX) is considered the current standard approach for adjuvant chemotherapy for patients with localized diseases.

Systematic palliative chemotherapy for metastatic diseases

The concept of systematic therapy for metastatic diseases has been well accepted since Poon *et al.* conducted a pooled analysis of 21 phase II trials in advanced CRC patients and demonstrated that 5-FU/LV significantly increased progress-free survival, compared with 5-FU monotherapy in late 1980s. [53] 5-FU remained the sole active agent for systematic treatment of metastatic CRC for the following decade. Multiple new agents have been introduced in the systematic therapy in the recent two decades, including irinotecan, oxaliplatin, bevacizumab, ramucirumab, cetuximab, panitumumabm, aflibercept, regorafenib trifluridine-tipiracil (TAS-102) and tipiracil. The improvements in response rate and progress free survival of regimens including both oxaliplatin and 5-FU/LV compared with those of 5-FU/LV-only therapy have been demonstrated in several clinical trials. [54-56] The combination of 5-FU/LV and oxaliplatin has become the most commonly used regimen in the first-line therapy of metastasis CRC. [57] The survival benefit of irinotecan plus 5-FU/LV regimens, such as FOLFORI, compared to 5-FU/LV alone was also well recognized, [58,59] although such regimens have been less frequently used than the oxaliplatin-containing regimens in the U.S. [57] Based on results from several clinical trials, multiple biologic agents targeting vascular endothelial growth factor (VEGF) or epidermal growth factor (EGF) and its receptor (EGFR) were approved.[60-63] Although the optimal usage of these agents remains unknown, it was found that a majority of the patients received the biologic agents as part of the 5-FU based chemotherapy regimens in clinical practice. [57]

Despite many advances in cancer treatment over the last two decades, 5-FU remains the backbone for both adjuvant and palliative therapy in CRC patients. Some efforts were taken in order to replace 5-FU during the last few decades, but all of them failed, owing to a lack of survival benefit or a higher financial cost compared with 5-FU. [64-66] 5-FU continues to be the key component in the regimens. However, due to the narrow therapeutic index of 5-FU, toxicity events among CRC patients have been frequently reported.

Five-fluorouracil associated toxicity

Common 5-FU associated toxicities usually involve organs or tissues with proliferating cells, including organs on the gastrointestinal tract, bone marrow and skin. It is thought that 5-FU targets the normal growing cells induce their apoptosis and causes toxicity events. The severity of these toxicity events can be determined according to the NCI CTC grades (National Cancer Institute Common Toxicity Criteria, Table 5-6). The grades range from 0 to 5, with 0 indicating no toxicity, 1 indicating mild toxicity, 2 indicating moderate toxicity, 3 indicating severe toxicity, 4 indicating disabling or life-threatening toxicity, and 5 indicating death.

Gastrointestinal toxicity is characterized by epithelial inflammation and ulceration, manifested as mucositis, diarrhea, nausea, and vomiting. In particular, mucositis and diarrhea are dose-limiting events. Mucositis refers to inflammatory lesions across the continuum of oral and gastrointestinal mucosa, from the mouth to the anus. [67] The frequency of mucositis due to 5-FU based chemotherapy among CRC patients ranged from 20% to 50% while the frequency of severe (grade 3 or more) mucositis ranged from 1.4% to 4.4%, depending on regimens. [67] Mucositis can lead to pain, malnutrition, infections and low quality of life. Diarrhea is another well recognized toxicity events caused by 5-FU. The risk of 5-FU associated diarrhea was estimated to be as high as 50% to 80%, with at least 30% of patients experiencing grade 3 or more diarrhea. [68] Diarrhea can lead to depletion of fluids and electrolytes and malnutrition, both of which can result in cardiovascular compromise, hospitalization and death. Notably, hepatic toxicity is seldom reported for 5-FU based chemotherapy.

Haematological toxicity is characterized by intravascular hemolysis and reduction of circulating blood cells, manifested by leukopenia, anemia, neutropenia, febrile neutropenia and thrombocytopenia, all

of which can be dose-limiting. The incidence of haematological toxicity in cancer patients increased, as more agents such as LV, irinotecan and oxaliplatin, were included in the regimen. For example, the frequency of grade 3 or more haematological toxicity in patients who treated with infusional 5-FU/LV was approximately 5%, while it was 50% in patients treated with the FOLFIRI regimen. [69] These myelosuppression events can lead to infections, fever and sepsis in cancer patients.

Dermatologic toxicity is characterized by decreased pain and temperature sensation in hands and feet with varying severity of pain, tingling, dryness, erythema, scaling, swelling, and vesiculation, manifested as hand foot syndromes, also known as palmar-plantar erythrodysesthesia or palmar-plantar erythema etc. [70] Hand foot syndromes are more common in patients treated with the orally administered 5-FU prodrug-capecitabine. The incidence of hand foot syndromes in patients with capecitabine in clinical trials was reported to be as high as 50%, with more than 17% reporting grade 3 or more toxicity, while the frequency of hand foot syndrome in patients treated with bolus or infusional 5-FU was reported to be 5%, with less than 2% reporting grade 3 or more toxicity. [70,71] Although hand foot syndromes in general are self-limiting and do not involve life-threatening events, it can result in dose reduction, treatment delay and low quality of life in patients.

Other 5-FU associated toxicity events include cardiotoxicity and neurotoxicity. Cardiotoxicity of 5-FU based chemotherapy can be manifested as coronary vasospasms and subsequent calcium antagonist non-responding angina, myocardial infarction, ischemia, dysrhythmia, cardiomyopathy, sinoatrial and atrioventricular nodal dysfunction, QT prolongation with torsades de pointes ventricular tachycardia and cardiac arrest. [72,73] A meta-analysis reported that the frequency of 5-FU associated cardiotoxicity events ranged from 0-35%. [74] The wide variation in incidence is likely due to differences in dose intensity, regimens and prevalence of prior history of cardiovascular diseases in patients across studies. [72] Cardiotoxicity events can lead to treatment delay or discontinuation, hospitalization, and sometimes, death. Saif *et al.* reported an overall death rate of 0.32% for the first cycle of chemotherapy and a death rate of 17% for re-exposure of 5-FU. [75] Neurotoxicity events due to 5-FU based chemotherapy can be manifested as cerebellar syndrome, encephalopathy, subacute multifocal leukoencephalopathy and

seizure. [76,77] The frequency of neurotoxicity events was estimated to be 2-6%. [76-78] Neurotoxicity events are dose-limiting, leading to treatment discontinuation and delay, and sometimes symptoms can persist after 5-FU discontinuation. [79] Unlike common toxicity events that arise in tissues or organs with growing cells, the cellular mechanism of these non-common toxicity events remains unclear. [78]

In summary, common and uncommon 5-FU associated toxicity events remain a major challenge in the clinical management of CRC patients. These toxicity events are likely the consequence of overdose of 5-FU in cancer patients. The standard calculation of 5-FU dose has been based on body surface area (BSA). However, a broad (> 30-fold) range of the plasma 5-FU levels in patients of the same BSA has been observed, [80] suggesting that there are factors other than BSA playing a major role in the metabolism of 5-FU. A comprehensive understanding of the pharmacokinetics and pharmacodynamics of 5-FU will help develop effective strategies for optimal dosage management and preventing severe toxicity events.

Table 1 AJCC staging -Primary Tumor[a]

TX	Primary tumor cannot be assessed.
T0	No evidence of primary tumor.
Tis	Carcinoma *in situ*: intraepithelial or invasion of lamina propria.[b]
T1	Tumor invades submucosa.
T2	Tumor invades muscularis propria.
T3	Tumor invades through the muscularis propria into pericolorectal tissues.
T4a	Tumor penetrates to the surface of the visceral peritoneum.[c]
T4b	Tumor directly invades or is adherent to other organs or structures.[c,d]

[a]*Reprinted with permission from AJCC: Colon and rectum. In: Edge SB, Byrd DR, Compton CC, et al., eds.: AJCC Cancer Staging Manual. 7th ed. New York, NY: Springer, 2010, pp 143-164.*

[b]*Tis includes cancer cells confined within the glandular basement membrane (intraepithelial) or mucosal lamina propria (intramucosal) with no extension through the muscularis mucosae into the submucosa.*

[c]*Direct invasion in T4 includes invasion of other organs or other segments of the colorectum as a result of direct extension through the serosa, as confirmed on microscopic examination (e.g., invasion of the sigmoid colon by a carcinoma of the cecum) or, for cancers in a retroperitoneal or subperitoneal location, direct invasion of other organs or structures by virtue of extension beyond the muscularis propria (i.e., respectively, a tumor on the posterior wall of the descending colon invading the left kidney or lateral abdominal wall; or a mid or distal rectal cancer with invasion of prostate, seminal vesicles, cervix, or vagina).*

[d]*Tumor that is adherent to other organs or structures, grossly, is classified cT4b. However, if no tumor is present in the adhesion, microscopically, the classification should be pT1–4a depending on the anatomical depth of wall invasion. The V and L classifications should be used to identify the presence or absence of vascular or lymphatic invasion whereas the PN site-specific factor should be used for perineural invasion.*

Table 2 AJCC staging -Regional Lymph Nodes (N)[a,b]

NX	Regional lymph nodes cannot be assessed.
N0	No regional lymph node metastasis.
N1	Metastases in 1–3 regional lymph nodes.
N1a	Metastasis in 1 regional lymph node.
N1b	Metastases in 2–3 regional lymph nodes.
N1c	Tumor deposit(s) in the subserosa, mesentery, or nonperitonealized pericolic or perirectal tissues without regional nodal metastasis.
N2	Metastases in ≥4 regional lymph nodes.
N2a	Metastases in 4–6 regional lymph nodes.
N2b	Metastases in ≥7 regional lymph nodes.

[a] *Reprinted with permission from AJCC: Colon and rectum. In: Edge SB, Byrd DR, Compton CC, et al., eds.: AJCC Cancer Staging Manual. 7th ed. New York, NY: Springer, 2010, pp 143-164.*

[b] *A satellite peritumoral nodule in the pericolorectal adipose tissue of a primary carcinoma without histologic evidence of residual lymph node in the nodule may represent discontinuous spread, venous invasion with extravascular spread (V1/2), or a totally replaced lymph node (N1/2). Replaced nodes should be counted separately as positive nodes in the N category, whereas discontinuous spread or venous invasion should be classified and counted in the site-specific factor category Tumor Deposits.*

Table 3 AJCC staging -Distant Metastasis (M)[a]

M0	No distant metastasis.
M1	Distant metastasis.
M1a	Metastasis confined to 1 organ or site (e.g., liver, lung, ovary, nonregional node).
M1b	Metastases in >1 organ/site or the peritoneum.

a Reprinted with permission from AJCC: Colon and rectum. In: Edge SB, Byrd DR, Compton CC, et al., eds.: AJCC Cancer Staging Manual. 7th ed. New York, NY: Springer, 2010, pp 143-164.

Table 4 Anatomic Stage/Prognostic Groups[a,b]

Stage	T	N	M	Dukes[c]	MAC[c]
0	Tis	N0	M0	–	–
I	T1	N0	M0	A	A
	T2	N0	M0	A	B1
IIA	T3	N0	M0	B	B2
IIB	T4a	N0	M0	B	B2
IIC	T4b	N0	M0	B	B3
IIIA	T1–T2	N1/N1c	M0	C	C1
	T1	N2a	M0	C	C1
IIIB	T3–T4a	N1/N1c	M0	C	C2
	T2–T3	N2a	M0	C	C1/C2
	T1–T2	N2b	M0	C	C1
IIIC	T4a	N2a	M0	C	C2
	T3–T4a	N2b	M0	C	C2
	T4b	N1–N2	M0	C	C3
IVA	Any T	Any N	M1a	–	–
IVB	Any T	Any N	M1b	–	–

[a]*Reprinted with permission from AJCC: Colon and rectum. In: Edge SB, Byrd DR, Compton CC, et al., eds.: AJCC Cancer Staging Manual. 7th ed. New York, NY: Springer, 2010, pp 143-164.*

[b]*cTNM is the clinical classification, and pTNM is the pathologic classification. The y prefix is used for those cancers that are classified after neoadjuvant pretreatment (e.g., ypTNM). Patients who have a complete pathologic response (ypT0, N0, cM0) may be similar to stage group 0 or I. The r prefix is to be used for those cancers that have recurred after a disease-free interval (rTNM).*

[c]*Dukes B is a composite of better (T3, N0, M0) and worse (T4, N0, M0) prognostic groups, as is Dukes C (any T, N1, M0 and any T, N2, M0). MAC is the modified Astler-Coller classification.*

Table 5 National Cancer Institute Common Toxicity Criteria for Haematological Adverse Events (CTCAE) v4.0

Scale	Neutrophils	Platelets	Hemoglobin	Lymphocytes (total)	CD4 count	Febrile neutropenia
Grade 1	<LLN to 1500/microL	<LLN to 75,000/microL	<LLN to 10 g/dL	<LLN to 800/microL	<LLN to 500/microL	
Grade 2	1000 to 1500/microL	50,000 to 75,000/microL	8.0 to 10.0 g/dL	500 to 800/microL	200 to 500/microL	
Grade 3	500 to 1000/microL	25,000 to 50,000/microL	<8.0 g/dL	200 to 500/microL	50 to 200/microL	ANC <1000/microL with a single temperature >38.3 °C (100.4°F) or a sustained temperature ≥38°C (100°F) for more than one hour
Grade 4	<500/microL	<25,000/microL	Life-threatening consequences; urgent intervention indicated	<200/microL	<50/microL	Life-threatening consequences; urgent intervention indicated
Grade 5			Death			Death

Table 6 National Cancer Institute Common Toxicity Criteria for Gastrointestinal and Dermatologic Adverse Events (CTCAE) v4.0

Scale	Oral mucositis	Nausea	Vomiting	Diarrhea	Hand-foot syndrome
Grade 1	Asymptomatic or mild symptoms; intervention not indicated	Loss of appetite without alteration in eating habits	One to two episodes (separated by five minutes) in 24 hours	Increase of less than four stools per day over baseline; mild increase in ostomy output compared with baseline	Minimal skin changes or dermatitis (eg, erythema, edema, or hyperkeratosis) without pain
Grade 2	Moderate pain, not interfering with oral intake; modified diet indicated	Oral intake decreased without significant weight loss, dehydration, or malnutrition	Three to five episodes (separated by five minutes) in 24 hours	Increase of four to six stools per day over baseline; moderate increase in ostomy output compared with baseline	Skin changes (eg, peeling, blisters, bleeding, edema, or hyperkeratosis) with pain, limiting instrumental ADL
Grade 3	Severe pain, interfering with oral intake	Inadequate oral caloric or fluid intake; tube feedings, TPN, or hospitalization indicated	≥6 episodes (separated by five minutes) in 24 hours; tube feeding, TPN, or hospitalization indicated	Increase of seven or more stools per day over baseline; incontinence; hospitalization indicated; severe increase in ostomy output compared with baseline; limiting self-care activities of daily living	Severe skin changes (eg, peeling, blisters, bleeding, edema, or hyperkeratosis) with pain, limiting self-care ADL
Grade 4	Life-threatening consequences; urgent intervention indicated		Life-threatening consequences; urgent intervention indicated	Life-threatening consequences; urgent intervention indicated	
Grade 5	Death		Death	Death	

Pharmacology of 5-FU

Based on the observation that some tumors used uracil more rapidly than normal tissues, [81] 5-FU was rationally designed as an analogue of uracil by substituting the hydrogen atom at the C-5 position of uracil with fluorine, which preserves the molecular conformation to maximize the likelihood that 5-FU can be metabolized by the same pathways for uracil. [82,83] After administration, 5-FU is metabolized by two pathways that compete with each other: the catabolic and anabolic pathways. The catabolic pathway degrades 5-FU mainly in liver cells, facilitating its clearance, while the anabolism pathway generates the active metabolites that exert cytotoxicity in cells that utilize uracil actively, such as tumor cells. [84] It is estimated that more than 85% of the administrated 5-FU is deactivated by the catabolic pathway, approximately 1-3% is activated by the anabolism pathway, and 5-10% is eliminated unchanged. [85] Understandings in the metabolism and mechanism of action of 5-FU will provide insights into the underlying mechanisms of 5-FU associated toxicity in cancer patients.

Catabolism

The catabolism pathway include enzymes dihydropyrimidine dehydrogenase (DPD), dihydropyrimidinase (DHPase) and beta-ureidopropionase (BUP-1) (Figure 1). The first enzyme DPD reduces the double-hydro-bonds in the pyrimidine ring of 5-FU and converse it to dihydrofluorouracil (FUH2). The second enzyme DHPase hydrolytically cleaves FUH2 and converts it to fluoro-beta-ureidopropionate (FUPA). The third enzyme BUP-1 further hydrolyses FUPA to FBAL.[86] Notably, the first two steps are reversible, while the last step is irreversible. FBAL is the major metabolite of 5-FU in urine and bile (> 95%).[23] Later studies using more precise analytic methods identified several additional catabolites of 5-FU, including fluoride ion (F−), N-carboxy-α-fluoro-β-alanine (CFBAL), 2-fluoro-3-hydroxypropanoic acid (FHPA) and fluoroacetate (FAC). [86-88]

It is commonly accepted that DPD catalyzes the rate-limiting step in the degradation of 5-FU.[85] However, evidence supporting this conclusion has been limited. Several studies observed a slow clearance of FUH2 and a rapid elimination of 5-FU from plasma in almost every participant of their

studies,[89,90] suggesting that either DHPase or BUP-1could be the rate-limiting enzyme of the catabolism of 5-FU instead of DPD.

Anabolism

While the *DPYD-DPYS-UPB1* pathway is considered the major pathway for 5-FU catabolism in human beings, three pathways are involved in the anabolism of this drug (Figure 1).[84,91] In the first pathway, 5-FU is converted by thymidine phosphorylase (dThdPase) to fluorodeoxyuridine (FUDR), which is subsequenlty phosphorylated by thymidine kinase (TK) to 5-fluoro-2′-deoxyuridine-5′-monophosphate (5-FdUMP). In the other two pathways, 5-FU is converted to 5-fluorouridine-5′-monophosphate (5-FUMP), either directly by orotate phosphoribosyltransferase (OPRT) or indirectly by the uridine phosphorylase (UrdPase) and then uridine kinase (UK).[84] 5-FUMP is then phosphorylated to 5-fluorouridine-5′-diphosphate (5-FUDP) and then 5-fluorouridine-5′-triphosphate (5-FUTP), an analog of uridine-5′-triphosphate (UTP). 5-FUDP and 5-FdUMP can also be converted to 5-fluoro-2′-deoxyuridine-5′-diphosphate (5-FdUDP), which is further phosphorylated to 5-fluoro-2′-deoxyuridine-5′-triphosphate (5-FdUTP), an analog of dUTP.[84,91]

Mechanism of action

The mechanism of action of 5-FU anabolites has been well understood. 5-FU exerts its cytotoxic effects through inhibiting thymidylate synthase (TS) by 5-FdUMP, tincorporating 5-FUTP into RNA and5-FdUTP into DNA [82] (Figure 1).

Inhibition of TS is thought to the primary mechanim for the cytotoxicity of 5-FU anabolites. TS catalyzes the reductive methylation of dUMP to dTMP by transfering a methylene group from a cofactor, CH2H4 folate. [82] dTMP is further phosphorylated to dTTP, which is a precursor for DNA. This reaction provides the only *de novo* source of thymidylate, which is necessary for DNA replication and repair. The anabolite of 5-FU, 5-FdUMP, binds to the nucleotide binding site of TS, forming a stable ternary complex with TS and the cofactor, CH2H4 folate. With the binding, FdUMP competes with dUMP, the natural substrate of TS and inhibits dTMP synthesis, resulting in a decerase of dTTP and an increase in dUTP followed by impaired DNA synthesis and repair. [92] The inhibition of TS can be further enhanced by

leucovorin, the precursor of 5,10-methylene-tetrahydrofolate that stablizes the ternary complex of TS and FdUMP.[82]

Incorporation of 5-FU anabolites into RNA also contributes to the cytotoxic action of 5-FU. 5-FUTP, competing with the natural nucleotide UTP, can be incorporated into all types of RNAs in tumor cells. For example, it has been found that 5-FUTP can be incorporated into snRNA, which inhibits the splicing of pre-mRNAs and the maturation of mRNAs.[93] 5-FUTP also inhibits the synthesis of tRNAs and disrupts post-transcriptional modification. [94] These misincorporations can lead to impaired mRNA and protein synthesis. Interestingly, the incorporation level is though to be dependant on administration routes. The level of incorporation into RNA was found to be higher after bolus administration of 5-FU than that after continuous infusional administration. [95] The precise mechanism remains to be identified.

Another class of 5-FU anabolites, 5-FdUTP, can be incorporated into DNA to exert its cytoxicity. After incorporated into DNA, 5-FdUTP can be excised by uracil–DNA-glycosylase and then cleaved by apurinic–apyrimidinic endonuclease, resulting in DNA strand breaks. Similar to the TS inhibition by 5-FdUMP, misincorporation of 5-FdUTP into DNA can disrupt the balance of intracellular deoxyribonucleotide pool, which subsequently leads to impaired DNA synthesis and repair. [92]

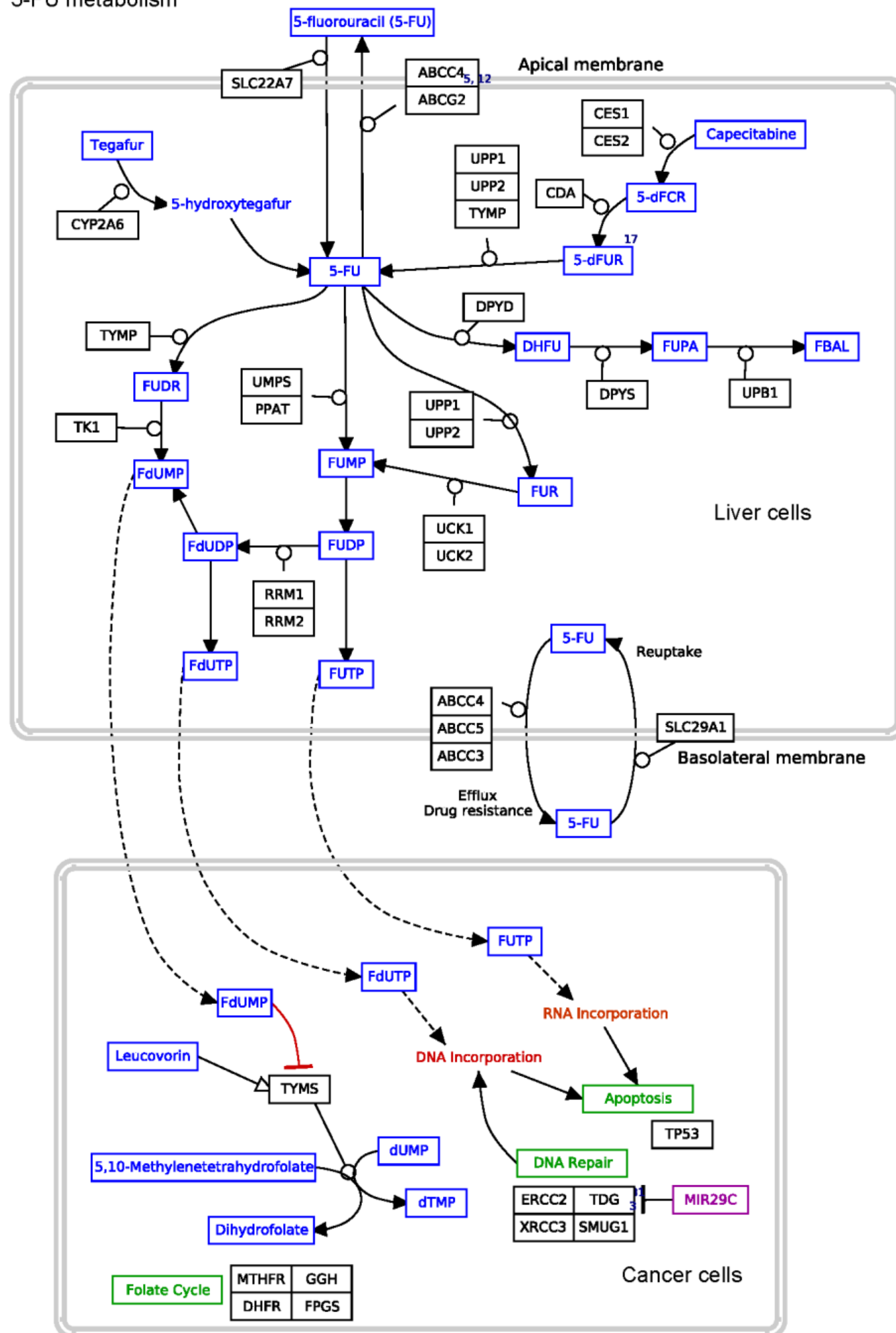

Figure 1 Five-FU metabolism pathways

The figure was generated by the Wikipathways (URL: http://www.wikipathways.org/index.php/Pathway:WP3275).

Predictors of 5-FU associated toxicity

Although the precise molecular mechanisms of 5-FU associated toxicity has not been fully understood, it is believed that a substantial proportion of the interindividual variation in toxicity can be due to differences in drug metabolism. Measurement of activities of metabolic enzymes for 5-FU has thus become appealing for toxicity prediction in cancer patients. However, none of the tests developed so far has been proven to be sufficiently reliable for clinical use. As direct evaluation in liver tissues has been difficult, enzyme activities in blood are usually measured as a surrogate. However, enzymatic activities in blood might not predict the activities in liver cells well. For example, it was reported that the correlation between hepatic – and peripheral blood mononuclear cell (PBMC) – DPD activities was only moderate ($r^2 = 0.31$).[96] Other proposed methods include evaluation of concentrations of metabolites of uracil or 5-FU and close monitoring of pharmacokinetics of 5-FU during the treatment, [97] which are complicated, time-consuming and expensive. Furthermore, there has been no concensus on the optimal methods evaluating enzyme activities, metabolites or pharmacokinetics parameters. All these facotrs prevent health policy makers to make a step towards recommendation of clinical use of phenotyping. To establish clinical utility, markers that are highly predictive and easy to accurately evaluate are needed.

Non-genetic predictors in 5-FU associated toxicity

Multiple non-genetic factors, such as age, sex, race and BMI, have been suggested to influence 5-FU metabolism and the risk of toxicity. To investigate the predictive values of these factors, many previous studies have evaluated their associations with risk of severe 5-FU associated toxicity in cancer patients.

Demographics factors

Age at treatment has been long studied as a potential predictor of 5-FU associated toxicity.[19,98-103] For example, a study using data of 46, 692 patients from population-based cancer registries found that patients aged older than 70 years were more likely to experience gastroenological, heametological toxicities and neurotoxicty from fluoropyrimidine-based chemotherapy than those aged 65–69 years, regardless of the regimens.[101] Futhermore, a recent large prospective cohort study including 1,463 cancer pateints demonstrated that older patients were at least five times more likely to experience lethal toxicity events

than younger patients.[99] Several studies suggested that age-related changes in pharmacokinetics of 5-FU might contribut to the observed age disparities.[104-106]As age increases, the pharmacokinetics profiles of patients can be changed by altered distribution, metabolism and elimination parameters, while alterations in absorption is less likely to lead to substantial change in patients who received infusional 5-FU.[107] For example, older age has been associated with decreased activities of DPD that can lead to slower degradation of 5-FU.[108] However, not all previous data supported this view.[109-111] Other factors have been prosposed included different dose schedule for the elderly[112] and higher prevelence of comorbidities, such as renal and hepatic dysfunctions.[113] For example, several studies showed that there was no difference in pharmcokinetics profiles and risk of severe toxicity between the elderly patients with normal renal functions and their younger counterparts,[71] suggesting that renal functions, instead of age, played a major role. Taken all together, aging is a heterogenous process and age at treatment may represent mulitiple factors that affect the risk of 5-FU associated toxicity. Age alone, although easy to measure and of substantial predictive value, might not be a reproducible predictor of 5-FU associated toxicity across studies. Additional variables are needed to improve the validity of prediction models.

Sex is another potentail risk factor that have been long studied. Previous studies have consistently found that female patients had a 2-3 fold higher risk of grade 3 or more 5-FU associated toxicity than male patients.[19,20,98,100,102,114-117] The sex disparity was often consistent across different regimens,[102,103,117] subtypes of toxicities,[20,102,103,117] modes of administration[114] and cycles of treatment,[114] suggesting a biologic reason underlying the observed differences. It has been consistently found that women tended to have higher 5-FU area under the plasma concentration-time curve (AUC) and lower plasma clearance[106,116,118,119] and plasma half-life of 5-FU than men after first cycle of chemotherapy,[119] all of which supported a slower clearance of 5-FU in women. One possible explanation is that women had a lower DPD activity than men.[120-122] Other factors that have been proposed included a higher percentage of body fat in women that affect drug deposition and clearance[123] and different gut microbiota profiles in women.[124] In summary, sex seems to be a reproducible predictor of 5-FU associated toxicity. Further investigation into the causes of sex disparities in toxicities is needed.

The impact of race in 5-FU associated toxicity remains controversial. Several population-based cohort studies and clinical trials of infusional 5-FU based regimens reported that African American patient had an approximately 40-50% lower risk of severe overall 5-FU associated toxicty than white patients.[103,125,126] Furthermore, it appeared that the lower risk in African Americans were primarily driven by the lower incidences of GI tract toxicities such as diarrhea, nausea, vomitting and stomatitis in this population.[125,126] There was also suggestive evidence for a decreased risk of severe (grade 3 or more) haematological toxicity and an increased risk of mild haematological toxicity (grade 1 or 2) in African American patients compared with their white counterparts.[125,126] One of the possible explanations for the observed discrepancy is the racial difference in the metabolism of 5-FU. However, Mattison *et al.* compared PBMC DPD activities in 149 African Americans and 109 whites and found a higher prevalence of DPD deficiency in African Americans.[122] In a later study, Offer *et al.* suggested that part of the racial differences in DPD activities could be explained by an African American-specific variant Y186C, the minor allele (MAF = 0.03 in African Americans) of which was associated with a 46% reduction of DPD activities in PBMCs.[127] These findings might partially explain the observed higher risk of mild haematological toxicity (grade 1 or 2) in African American patients. However, clear explanations for racial differences in risk of 5-FU associated toxicity, particularly for the lower risk of severe GI tract toxicity in African Americans, remain unavailable.

Clinical factors

Body mass index (BMI) is thought to be an important predictor of 5-FU associated toxicity. In early studies, obesity (defined as a BMI > 30 kg/m^2) has been consistently found to be associated with a reduced risk of grade 3 or more 5-FU-associated toxicity, particularly for haematological toxicity, in cancer patients treated with infusional 5-FU.[128-130] The authors suspected that therapy underdosing for obese patients, which were common in their study populations, might be one of the causes for the lower risk of severe toxicity in these patients.[128,129] To elucidate the role of underdosing, Chambers *et al.* compared toxicity incidences in obese patients fully dosed and those under dosed using data from three clinical trials. They found no difference in rate of severe toxicity events between these two groups (*P* =

0.71).[131] On the other hand, several studies investigated the role of lean body compartments that include metabolic tissues for liver and kidney.[123,132,133] It was hypothesized that a lower proportion of lean body mass, which is more common among obese patients than in normal-weight patients, could result in a higher concerntration of the drug and subsequently increase the risk of toxicity. For example, Prado *et al.* found that low lean body mass was statistically signficantly associated with an increased risk of dose-limiting toxicity in patients treated with 5-FU based chemotherapy.[123] Data on the association of body composition or BMI with pharmacokinetics of 5-FU have been limited. Gusella *et al.* reported that fat-free body mass was associated with 5-FU clearance and distribution, although the association was not statistically significant. [133] This study was conducted in 34 patients, which might have been underpowered. Additional experiments and clinical studies are needed to determine the association of body composition and 5-FU metabolism and toxicity.

Performance status, as an established prognosis factors for CRC patients, has been evaluated for its association with 5-FU associated toxicity in several previous studies.[19,20,134,135] Although early studies showed conflicting results, [19,20,134] a meta-analysis of nine clinical trials with a total of 6,286 CRC patients demonstrated that patients with a poor performance status had an approximatly 2-fold increased risk of grade 3 or more nausea. The association was independent of treatment arms and age. [135] However, it remained unknown whether patients with a poor performance status but still recruited in clinical trials represents those with a poor performance status in the general CRC populations. Therefore, population-based studies are needed to determine whether this finding can be generalized.

Treatment factors including administration routes and dose schedule in addition to regimens have been found to be associated with risk of toxicity.[20,136,137] For example, a meta-analysis of toxicity profiles of cancer patients treated with 5-FU showed that bolus administration was associated with a higher risk of haematological toxicity (eg, neutropenia) than continuous infusion.[20] One possible explanation for the observed differences is that the short peak plasma concentrations of 5-FU induced by bolus administration increases the exposure of the drug to PBMCs. However, these treatment factors might not be relevent for

patients who were treated in recent decades, because of the consensus on the optimal administration routes and schedules. [138-140]

In conclusion, several non-genetic factors, including age, sex, race, body composition and performance status might have marked predictive values for 5-FU associated toxicity. These factors are easy to accurately measure and with clinically relevant effect size, suggesting that these factors might contribute to prediction models for 5-FU associated toxicity. Neverthless, cautions should be exercised that these variables ususally represent a broad spectrum of known and unknown physiological and social factors. The observed associations could have been confounded by these physiological or social factors, leading to inconsistent association results across studies. Furthermore, even within patients of the same age, race, sex and BMI, great variabilities in response to 5-FU have been observed. More accurate predictors are needed to improve the prediction performance.

Genetic predictors in 5-FU associated toxicity

It has been well known that genetic factors account for much of the interpatient variation since the link of familial pyrimidinemia and severe 5-FU associated toxicity was reported approximately 30 years ago. [7] It was estimated that 26-65% of the variations in 5-FU cytotoxicity could be attributed to genetic factors, depending on the dose intensity. [141] To identify genetic markers for risk of toxicity, two approaches have been used in the previous pharmacogenetics studies of 5-FU: candidate gene and genome-wide association.

Candidate gene approach

Most of the previous research on pharmacogenetics of 5-FU has been focused on genes related to drug absorption, distribution, metabolism and excretion (ADME), based on the hypothesis that genetic variants that affect the acitivities of enzymes will affect the pharmacokinetics and pharmacodynamics of 5-FU. Currently known genes involved in the ADME pathways of 5-FU include *DPYD*, *DPYS*, *UPB1*, thymidine phosphorylase (*TYMP*), thymidine kinase 1 (*TK1*), thymidylate synthetase (TYMS), uridine monophosphate synthetase (*UMPS*), phosphoribosyl pyrophosphate amidotransferase (*PPAT*),), uridine phosphorylase 1 (*UPP1*), *UPP2*, uridine-cytidine kinase 1 (*UCK1*), ribonucleotide reductase M1 (*RRM1*),

RRM2, (ATP-binding cassette, sub-family C (CFTR/MRP), member 2) *ABCC2*, *ABCC3*, *ABCC4*, *ABCC5*, solute carrier family 29 (nucleoside transporters), member 1 (*SLC29A1*), and solute carrier family 22 (organic anion transporter), member 7 (*SLC22A7*).

Among these genes, *DPYD* is thought to play a dominant role in severe 5-FU associated toxicity. Previous studies suggested that DPD deficiency could account for up to 40% of the severe 5-FU assocaited toxicity events in cancer patients.[142] Thus, variants in the gene *DPYD* have been the focus of the pharmacogenetic research of 5-FU.

DPYD

Spanning approximately 1Mb on chromosome 1 with 23 exons, *DPYD* is one of the longest genes discoved so far in human. More than 135,000 genetic variants have been identified in this gene, with 625 of them locating in the transcribed exonic regions. With the exception of 17 variants, all of the protein-coding variants in transcribed exonic regions are extremely rare (MAF < 0.1% in all populations). Approximately 90 of the coding or splicing variants in *DPYD* have been investigated for their impacts on DPD activities in cell lines,[143,144] and only 8 of them showed evidence of alterning DPD activities. More than 50 coding or splicing variants in *DPYD* have been investigated for their association with risk of 5-FU associated toxicity.[145,146] However, only 4 of these variants have showed consistent association, including DPYD*2A (IVS14+1G>A, c.1905+1G>A, or rs3918290), c.2846A>T (D949V or rs67376798), c.1679T>G (I560S, *DPYD**13, or rs55886062) and c.1236G>A (E412E or rs56038477). The association results of these variants have been summarized in three recent meta-analyses. [8,9,145]

The *DPYD**2A and c.2846A>T variants are the most well-studied *DPYD* variants. Both variants are rare in all populations, with an MAF of approximtely 1% for *DPYD**2A and 0.5% for c.2846A>T in European descendants. *DPYD**2A causes a glycine to arginine alternation in the 5′ splicing site of intron 14, which is believed to result in a 165-bp deletion in the mRNA. Previous functional studies in cell lines or patient-derived samples demonstrated that a homozygote genotype of this variant resulted in complete DPD deficiency while a heterozygote genotype resulted in partial deficiency.[144] C.2846A>T is a nonsynonymous missense variant on the C-terminus Fe-S motif of exon 22, which directly interupts the

binding of cofactors and electron transport from FAD to FMN and thus affects the activity of DPD. [147] According to a study that evaluated the correlation of genetic variants and the enzyme activity of DPD in transfected cells, the T allele of C.2846A>T was statistically signficantly associated with a 41% reduction in DPD activities.[143] Early studies of these two variants with risk of 5-FU associated toxicity found inconsistent results, likely due to the small sample size and restrospective design of these studies. [8,9,145] Recent large cohort studies [21,103] and meta-analyses [4,9] have confirmed the association of *DPYD**2A and c.2846A>T with risk of severe 5-FU-associated toxicity. In the largest cohort study (n = 2, 886) to date, Lee *et al.* found that the exon-skipping allele A of *DPYD**2A and the T allele of c.2846A>T were statistically significantly associated with an increased risk of grade 3 or greater 5-FU-associated toxicity, adjusting for demographics, tumor characteristics and treatment factors.[103] c.1679T>G and c.1236G>A are also rare, with an MAF of 0.1% for c.1679T>G and 2% for c.1236G>A in European descendents. A functional study in cell lines showed that the G allele of c.1679T>G was associated with a decreased activity of DPD. [144] c.1236G>A is a synonymous variant in the haplotype B3 (HapB3) of *DPYD*, which is in LD with a splice donor variant c.1129-5923C>G (rs75017182) that was thought to creat an additional 44bps exon region in exon 10 and thus reduce the activity of DPD. Results of previous studies on the association of c.1236G>A with DPD activities have been conflicting,[148,149] the functional impact of this variant remains unclear. The clinical validaty of c.1679T>G and c.1236G>A was demostrated in a recent meta-analysis that evaluated the association of these two variants with the risk of severe 5-FU associated toxicity in a total of 7,365 cancer patients from eight studies. The authors found a statistically significant association of c.1679T>G and c.1236G>A/HapB3 with GI tract toxicity (adjusted risk ratio (RR) 5.72, $P = 0.02$; and 2.04 , $P < 0.0001$, respectively) and haematological toxicity (adjusted RR 9.76, $P = 0.0001$; and 2.07, $P = 0.01$, respectively).[8]

DPYS and UPB1

Although *DPYD* genetic variants were thought to account for a substaintial proportion of genetic variablity in 5-FU degradation, 39% -61% of the patients who experienced severe toxicity events had normal DPD activities,[150] suggesting that other factors may also play important roles. Population and

family-based studies suggested that deficiency of the downstream enzymes of DPD (DPH and BUP) might alter uracil catabolism [151,152]and increase risk of 5-FU-associated toxicity.[153,154] DHP is a cytosolic metalloenzyme composed of four identical subunits, encoded by the gene *DPYS*. BUP is a hydrolase enzyme that cleaves carbon-nitrogen bonds, encoded by the gene *UPB1*. Muhale *et al.* specifically knocked down each gene in the metabolism pathway of 5-FU in multiple colorectal cell lines and found that the knockdown of *DPYS* resulted in a signficant increase in 5-FU-induced cell apoptosis while the change in *DPYD* gene knockdown cell lines was modest. The authors also found that *UPB1* knowndown statistically signficantly increased cell apoptosis in *p53*-mutated cell lines.[155] Taken together, these findings suggested an important role of *DPYS* and *UPB1* in the metabolism of 5-FU and in 5-FU-associated toxicity.

The gene *DPYS* spans 87kb on chromome 8, encompassing 10 exons. More than 700 variants have been identified within this gene, only 12 of which are commom (MAF > 1%). *UPB1* spans 34kb on chromosome 22, encompassing 10 exons. Appromixmately 560 variants have been identified within *UPB1*, only 2 of which are common (MAF > 1%). Four previous studies evaluated the association of variants in *DPYS* and/or *UPB1* with 5-FU associated toxicity in cancer patients.[14,15,156,157] These studies have identified mulitple novel variants showing association with risk of toxicity, such as c.1-1T > C and c.265-58T > C in *DPYS*. [14,154] However, the functional impact of these variants needs to be validated in future studies. The association of known genetic variations in *DPYS* and *UPB1* with 5-FU associated toxicity remains inconclusive.

Other genes

In addtion to variants in catabolic genes, variants in other ADME genes such as drug target genes and transporter genes have also been studied. Among these genes, *TYMS* has been the most commonly studied gene, which enocodes the TS enzyme, the major target of 5-FU anabolites (Figure 1). Early studies showed that three variants, rs16430 (a deletion of 6 bp in 3'-UTR),[158] rs45445694 (a repeat of 28 bp in the promoter region) [159] and rs2853542 (a G>C SNP within the 28bp repeat) [160] were associated with altered expression levels of *TYMS in vitro* or in cancer cell lines. Several population- based studies found

that these variants were associated with risk of toxicity, but results from other studies did not confirm their association. A recent meta-analysis showed that rs45445694 and rs16430 were statistically significantly associated with grade 3 or more toxicity in CRC patients (adjusted OR = 1.36, $P < .001$ for rs45445694, OR = 1.25, $P = 0.02$ for rs16340, respectively), and the association of rs16430 was largely dimished in the condtional analysis including both variants in the same model. [4] No association was found for the SNP rs2853542 in this meta-analysis.[4] Furthermore, in the subquent study that comprehensively evaluated variants in the flanking regions of *TYMS*, [146] the authors found that SNP rs2612091, located at a non-coding region downstream of *TYMS*, fully accounted for the association observed for rs45445694 or rs16430. Although the association of rs2612091 remains to be validated, this finding questioned the role of these two variants in the expression of *TYMS*. In addition, to evalute the functional impact of rs16430 and rs45445694, a recent study used allelic-specific analysis in peripheral blood mononuclear cells (PBMCs) from patients which was supposed to avoid the artificial genetic environment created in luciferase assays widely used in previous functional studies. [158] The authors found that none of these variants was statistically signficantly associated with expression of *TYMS.* [161] Moreover, a recent study evaluated the impact of these two variants on the 5-FU degradation rate in the PBMC samples of 1,010 cancer patients and found that no association of these variants with 5-FU clearance.[162] None of the previous studies have established the association of these variants with the activity of *TYMS*. Additionally, Muhale *et al.* found that specific knockdown of the gene *TYMS* did not substantially affect the clearance of 5-FU in colorectal cell lines.[155] More evidence from functional studies is needed to elucidate the role of *TYMS* in the etiology of 5-FU associated toxicity. Other genes that have been investigated included *MTHFR* [163-165] that encodes the cofactor of TS, and *ABCB*1[165] that encodes transporters for 5-FU metabolites.[166] However, results from previous research on these genes have been inconsistent. Taken together, the clinical relevence of the genes *TYMS*, *MTHFR* and *ABCB1* and their genetic variations remains to be clarified.

Genome-wide association approach

Genome-wide association studies (GWAS) that evaluate the association of genetic variants across the whole genome have been shown a powerful tool for uncovering genetic basis of many traits. Since 2005, GWAS studies have identified more than 20, 000 susceptablility loci, most of which would have been difficult to be identified in the candidate gene approach. Because of this feature, there has been increasing interest on the application of GWAS in pharmacogenomics of drugs.[167-169]

To date, only one GWAS has been conducted to identify genetic risk facotors for 5-FU associated toxicity in CRC patients. Fernandez-Rozadilla *et al.* conducted a GWAS study with 221 CRC patients of European ancestry who had been treated with 5-FU alone or in combination with oxaliplatin as the discovery stage. The replication stage included 791 CRC from the same cohort. The outcomes included anemia, leukopenia, thrombocytopenia, neutropenia, nause/vomitting, diarrhoea and mucositis. The authors also included neuropathy as an outcome, which is a common side effect for oxaliplatin but a rare side effect for 5-FU. None of the genetic variants evaluated in this study was found associated with risk of toxicity at a genome-wide significant level ($P < 5 \times 10^{-8}$).[170] The only consistent association in both stages was with SNP rs10876844 ($P_{combined} = 0.01$), which was found to be an eQTL of the gene *RDH5* encoding an enzyme of the dehydrogenase family. [171] The limitations of this study lie in the small sample size of CRC patients and a broad defination of exposures and outcomes. Low *et al.* performed a GWAS for 5-FU- induced severe neutropenia or leucopenia in 1,460 esophageal cancer patients of East Asian ancestory (177 cases/ 952 controls).[172] No statistically significant association was found after correcting for multiple testing. Addtionally, there was no overlap in genomic loci showing suggestive assocations between these two studies. Taken all together, these studies have highlighted the chanllenges in the GWAS approach: sufficient power and clearly defined phenotypes. To achieve a power of 80% in a GWAS study, at least 1,500 cases and 1,500 controls will be needed to detect variants with a OR of 1.5 and a MAF of 20%. Another challenge is that the uniform assessment of toxicity in patients from different insistues. Future GWAS studies that address these challenges are needed to uncover the genetic architecture of chemotherap-induced toxicity.

In summary, currently known genetic risk factors for 5-FU associated toxicity only explained a small fraction of the interpatient variablity of 5-FU response. Although assocations of the majority of the coding variants of the catabolic genes with risk of toxicity remain to be investigated, most of these variants are of extremely low MAFs (less than 0.1%), which seems unlikely to explain the high prevelance of severe 5-FU associated toxicity (up to 30%) in CRC patients. More genetic risk factors remain to be identified. Genetic variants that regulates expression of genes in the catabolic pathways represent promissing candiates for futher investigation.

Potential roles of regulotary variants in drug-associated toxicity

Regulatory variants

It is now believed that noncoding variants play a pivotal role in the genetics of complex diseases through their roles in gene expression regulation. It was estimated that SNPs in regions of potential regulatory elements, such as DNase I hypersensitivity sites (DHS), accounted for the majority of the heritability of multiple common diseases. [17,173] Understanding structures and components of regulatory elements for gene expression is critical for uncovering the causal variants and the etiology of complex diseases.

Regulatory elements include both *cis*- and *trans*-regulatory elements. *Cis*-regulatory elements are sequences that regulate their nearby genes, while *trans*-regulatory elements refer to sequences that regulate target genes by changing the structure, function or expression of a diffusible factor. [22] For example, trans-regulatory elements can alter the expression of transcription factors of the target gene that can be distal to the trans-regulatory elements or even on a different chromosome. Motivated by evidence from GWAS studies that most observed associations are likely due to genetic variants in *cis*-regulatory elements, a large number of the recent studies have focused on *cis*-regulatory elements, contributing to a better understanding of gene expression regulation. [22] Less is known for *trans*-regulatory elements, owing to the complexity of the regulatory network involved. The major focus in this proposal is *cis*-regulatory mechanisms.

Cis-regulatory elements typically include several hundreds of base pairs of nucleotides with multiple occurrences of transcription factor binding sites (TFBS).[22,174] According to their diverse roles in the regulatory network, *cis*-regulatory elements can be categorized into promoters, enhancers, insulators and silencers. Promoters locate near the transcription start site (TSS) of genes, or distal to the TSS in rarer cases, serving as a scaffold for transcription factors and coactivators such as histone acetyltransferases to initiate transcription. Enhancers, which interact with promoters, can locate either upstream or downstream of their target genes, and sometimes in another gene. Although distal to their target promoters linearly on the chromosome, enhancers can interact with their target promoters by chromatin looping. [175] Enhancers increase gene expression by recruiting transcription factors and their

coactivators and increasing the access of the transcription machinery to promoters. [176] In contrast to enhancers, insulators negatively regulate gene expression by creating boundaries of regulatory domains and prevent the spread of heterochromatin and the activation of promoters. [177] Similar to enhancers, insulators can locate in either downstream or upstream of their target genes. Insulators decrease gene expression by recruiting repression factors and mediating silencing of promoters by interacting with enhancers or other part of the transcriptional machinery. [178] These regulatory elements are characterized by specific chromatin modifications. [179] Additionally, these effects of *cis*-regulatory elements are often context-specific, depending on the cell types and states (activated or steady). [180]

Variations in the *cis*-regulatory regions can cause phenotype changes through altering the transcription regulation of genes. [180] Thus, identifying *cis*-regulatory elements and variants that trigger the change in the regulatory machinery is an important goal in studies that investigate genetic basis of complex traits. [181] Although systematic characterization of regulatory variants across the human genome is still in its infancy, recent advances in genomics, epigenomics and bioinformatics have begun to make identifying potential functional variants possible. Multiple approaches have been proposed to uncover the potential functional variants. In general, these approaches can be categorized into two groups: expression quantitative trait loci (eQTLs) mapping and regulatory mechanism mapping. [182]

The eQTLs are the genetic loci with variants that affect the expression levels of genes, which can be identified through evaluating the correlations between genotypes and mRNA levels of genes in a genetically diverse population. *cis*-eQTLs are common across the human genome. Battle et al. reported that at least one *cis*-eQTL could be identified for approximately 80% of the expressed gene in whole blood. [183] *Trans*-eQTLs also account for a substantial heritability of gene expression across the genome. [184] However, reliable detection of *trans*-eQTLs has been difficult, due to the large sample size required. eQTL mapping has become an important tool for investigating regulatory variants, their target genes and the associated traits. For example, Lappalainen *et al.* performed genome-wide eQTL analyses and identified candidate causal regulatory variants for gene expression across the genome, some of which

were validated in experiments. By combining with results from GWAS studies, the authors also identified potential causal variants for multiple disease-associated loci. [185]

Despite a powerful method that directly evaluates genetic variants and expression levels of genes, the eQTL mapping approach suffers from several limitations. It typically identifies associated loci rather than the regulatory variants of gene expression. The best eQTL variants may not be the causal variants owing to noises in the measurements in genotypes and phenotypes. Moreover, eQTL mapping does not provide information on the regulatory mechanism, and thus functional investigation is usually needed to establish a causal relationship.

Based on the premise that altering any step of the transcription can affect the expression level of genes, regulatory mechanism mapping that identifies variants that affect regulatory elements has become an attractive alternative of eQTL mapping. [182] Regulatory mechanism mapping studies usually used sequencing-based assays such as chromatin immunoprecipitation sequencing (ChIP-seq) and DNase I hypersensitive site sequencing (DNase-seq) [186] or transposase-accessible chromatin with sequencing (ATAC-seq) [187,188] to explore the genetic basis of variations in regulatory elements. However, these assays are expensive, labor-intensive, and sometimes, require a large number of cells. Most of these studies were conducted with a small sample size ($N < 100$), limiting their power to detect variants with a moderate effect size.

Recent studies suggested that the alteration in transcription factor binding was the central event initiating concerted changes in regulatory mechanisms. [182] For example, integrated analyses of data of histone modification, transcription binding sites and gene expression found that many variants that affect transcription factor binding were also associated with variations in chromatin states and expression of genes, [189-191] suggesting that differential transcription factor binding might affect chromatin states and subsequently affect gene expression. Furthermore, Jolma *et al.* characterized the DNA sequences that recognized pairs of transcription factors and found that the binding of transcription factors was dependent on the spacing and orientation of their binding motifs instead of any other known regulatory mechanism. [192] These results suggested that variants that alter transcription factor binding were important in

transcriptional regulation. Multiple *in vivo* and *in silico* approaches have been developed to identify genetic variations in transcription factor binding. [186,193,194]

Regulatory variants in drug-associated toxicity

Although no regulatory variants have been reported for the genes in the catabolic pathway of 5-FU, several functional variants that affect expression of drug-metabolizing enzymes have been identified and characterized. In addition, more than 96% of the GWAS-identified variants for drug response to date located in the non-coding regions. [195]Although causal variants underlying the observed associations remain to be identified, these findings suggested an important role of regulatory variants in drug response.

The most commonly studied regulotary variants in pharmacogenetics of drug response are promoter variants. For example, the promoter variant of the *UGT1A1* gene, *UGT1A1*28,* was found to be associated with risk of diarrahea in patients receiving irinotecan. This variant locates 39 bp upstream of the transcription start site of the *UGT1A1* gene and creates an additional TA repeat in the TATATATATATA sequence of the promoter. *UGT1A1*28* has been found to be significantly associated with a decreased expression level of *UGT1A1,* which is thought to delay the clearance of the metabolite of irinotecan, SN-38 and thus cause severe toxicity events.[196,197] Other promoter variants that have been well investigated include rs9923231 in *VKORC1,*[198] rs717620 in *ABCC2*[199] and rs2413775 in *CNT2*,[200] suggesting a role of promoter variants in drug response.

Enhancer variants have also been implicated in pharmacogenetic studies. The most studied enhancer variants for drug reponse so far are those for genes in the cytochromes P450 (CYP) family, which encode major enzymes for drug metabolism. In addtion to the enhancer variant for *CYP2D6* described in Chapter I, enhancer variants or regions have been identified for other members in the CYP family, including *CYP1A1*,[201] *CYP1B1,*[202] *CYP2E1,*[203] *CYP2B6,*[204] *CYP2C19,*[205] *CYP2E1*[203]and *CYP3A4*.[206] Ehancer variants also play an important role in transcriptional regulation of drug transporters. For example, a study that characterized enhancers for liver membrane transporters identified multiple variants with functional impacts in the potential enhancer regions of the gene *SLCO1A2*.[207] Addtionally, several recent studies that systematically characterized drug-induced regulotary elements across the genome using RNA-

seq and ChIP-seq suggested that enhancers might predominantly contribute to the variation in exprssion of drug metabolism genes. [140]

In conclusion, regulotary variants have important roles in influencing drug response through regulating expression of drug metabolism and transportation genes. Recent advances in functional genomics has allowed us to predict putative regulotary elements and variants.

CHAPTER III

INTEGRATIVE FUNCTIONAL GENOMIC ANALYSES IDENTIFY GENETIC MARKERS OF 5-FLUOROURACIL CATABOLIC PATHWAY ACTIVITIES

Introduction

Genetic variations are thought to contribute a substantial proportion of the interpatient variation in severe toxicity events due to 5-fluorouracil (5-FU) based chemotherapy. Previous studies have investigated multiple coding and splicing variants in the genes in the 5-FU metabolism pathways. However, only four of these variants, all of which were located in the *DPYD* gene, showed consistent association with 5-FU associated toxicity, all of which are rare (minor allele frequencies range from 0.005 to 0.02 in all 1000 Genomes populations). [8,103] These variants together explained a small fraction of the genetic variability of 5-FU associated toxicities. A recent fine mapping study of genetic variants in and around genes in the 5-FU metabolism pathways suggested that non-coding common variants might play an important role in the chemotherapy-associated toxicities. [21] It is hypothesized that non-coding variants mediate the risk of severe 5-FU associated toxicity by regulating the expression of genes in the 5-FU metabolism pathway. However, none of the regulatory variants for genes in the 5-FU metabolism pathways has been identified.

Although there has been an increasing understanding in how genetic variants regulate gene expressions and phenotype variability, [22] systematic identification of regulatory elements and the causal variants that contribute to gene expression variation remains a major challenge. One of the principal hurdles is that capturing regulatory activities directly has been difficult. [208] Instead of direction assessment, recent research has focused on characterization of chromatin states that are indicative of regulatory activities. Systematic characterization can be done by integrating epigenomics data of DNase I hypersensitive sites (DHS), histone modifications and transcription factor binding sites (TFBS). The most well-known resources of these data include the Encyclopedia of DNA Elements (ENCODE),[209] the Roadmap Epigenomics Mapping Consortium (REMC) [210] and the Functional Annotation of the Mammalian Genome 5 (FANTOM5) projects. [176] Another focus in recent research is the expression quantitative trait loci (eQTLs) mapping, [22] which have been demonstrated useful in uncovering genetic

loci that regulate the expression of their target genes. [185] Typical eQTL analyses usually evaluate the association of expression with each genetic variant at a given locus and rarely consider the combinatorial effects of multiple regulatory genetic variants. However, recent studies have suggested that expression of genes were likely to be regulated by multiple genetic variants at the locus. [185] [211]

To identify variants that predict expression levels of genes in the 5-FU catabolic pathway (*DPYD*, *DPYS* and *UPB1*), two approaches were developed in this study. Both approaches started with identifying candidate variants in the predicted regulatory regions near the target genes. In the first approach, based on the hypothesis that multiple *cis*-variants co-regulate the expression of genes, variable-selection algorithms such as LASSO [212,213] and elastic net [214] that allowed for identifying multiple variants at the same region were used to identify variants predicting the expression of genes in relevant tissues. In the second approach, based on the hypothesis that genetic variants in regulatory regions that alter the binding of relevant transcription factors are more likely to regulate the expression of genes, potential regulatory variants were prioritized according to their predicted functional potentiality.

Material and Methods

Predicted functional regions and candidate variants for the genes *DPYD*, *DPYS* and *UPB1*

Data on the enhancer-like (predicted from DNase hypersensitivity and histone modification H3K27ac signals), promoter-like regions (predicted from DNase hypersensitivity and histone modification H3K4me3 signals) and chromatin state (18 states) for 350 samples from the REMC and ENCODE projects, corresponding to 60 normal tissues and 74 noncancerous cell types, were obtained from the ENCODE webportal. [209] Data on enhancers and promoters predicted by Cap Analysis of Gene Expression (CAGE) across different cell types were downloaded from the web portal for the FANTOM5 project. [176] In addition, transcription factor binding sites (TFBSs) derived from ChIP-seq experiments across multiple cell lines by the ENCODE project [215] were obtained from the UCSC genome browser data portal. [216] The details of data processing and the algorithms of predicting enhancers and promoters in the ENCODE, REMC and FANTOM5 projects have been described previously. [176,209-211] We selected regions of

predicted enhancers and promoters, regions of TFBSs and regions with a chromatin state of active transcription start site (TSS), flanking active TSS, genic enhancers or enhancers as predicted regulatory regions of interest.

To identify the likely genomic boundaries of the regulatory landscape for each gene in the 5-FU catabolic pathway, the topologically associated domain (TAD) [217] flanking each gene was identified from the Hi-C interaction matrices in a human embryonic stem cell line (H1-hESC) and a human adult normal liver cells (liver_STL011) obtained from the ENCODE webportal. The genetic interval for the TAD flanking each gene was identical in both cell types. The TAD spans 1.65 Mb on chromosome 1 (chr1: 97350000-99000000, hg19), 0.8 Mb on chromosome 8 (chr8: 104825000-105625000, hg19), and 0.35 Mb on chromosome 22 (chr22: 24650000-25000000, hg19), respectively, for the genes *DPYD*, *DPYS*, and *UPB1*.

Because the gene *DPYD* is ubiquitously expressed across tissue and cell types (Appendix A, Figures S1), the predicted regulatory regions across all cell types in ENCODE, REMC and FANTOM5 that overlap with the TAD for *DPYD* were defined as the potential regulatory regions for this gene. The expression of *DPYS* and *UPB1* is largely confined in liver tissues (Appendix A, Figures S1 and S2), which was consistent with the previous estimate that more than 85% of 5-FU were degraded in liver.[23] Therefore, the predicted regulatory regions in normal liver cells and tissues (namely, hepatocytes derived from H9 cells and liver tissues from a healthy adult donor) that overlap with the TADs for *DPYS* and *UPB1* were defined as the potential regulatory regions for these two genes. Genetic variants with a minor allele frequency (MAF) larger than 0.01 in 1000 Genomes populations were defined as the candidate variants for further evaluation. A total of 899, 154, and 300 genetic variants were remained, respectively, for *DPYD*, *DPYS* and *UPB1*.

Approach I: Predicting gene expression by potential functional variants using data from the GTEx project

The mapped RNA-seq data from 7,051 tissue samples obtained from 450 donors, the genotype data and demographic characteristics (age, sex and race) data from these donors were downloaded from

the web portal of GTEx.[218] The genotype and RNA-seq data were filtered and processed using the protocols by the GTEx consortium. [218] To control for potential hidden confounding factors in the RNA-seq data, probabilistic estimation of expression residuals (PEER) analysis were performed to derive PEER factors. [219] To increase the coverage of the genotype data, imputation of genotypes using the 1000 Genomes project (Phase 3) as the reference panel was performed. The top three principal components were derived from the genotype data as described previously.

Prediction models for the expression level of each gene were built using the GTEx data. For the ubiquitously expressed gene *DPYD* (Figure 2), its expression levels were decomposed into a component that is common across all tissues (termed as the cross-tissue component) and tissue-specific components using a mixed effects model as previously described by H. Wheeler *et al*. [220] Thus, the outcomes in the prediction models for *DPYD* included the cross-tissue component (n=449, , including all ancestral groups)) and the liver-tissue-specific component (n= 97, including all ancestral groups) for this gene. For the genes *DPYS* and *UPB1*, the outcomes were the expression levels in liver tissues (n = 97, including all ancestral groups). Prediction models for expression levels were developed using the predictor selection algorithms LASSO and elastic-net (α=0.5) in the glmnet R package. [221] The predictive performance of the selected variants were measured by the Pearson's R^2. The LASSO slightly outperformed the elastic-net selection, so the results from the LASSO were used in the downstream analyses. To evaluate whether the selection by LASSO was consistent, bootstrap resampling with replacement (B=500) in GTEx samples were used. In each bootstrap dataset, the same prediction model using LASSO was performed and the selected variants were recorded. The genetic variants selected in the primary dataset remained the top selected variants across all bootstrap datasets, indicating that the selection was relatively consistent. Due to the relatively small sample size of the GTEx data and the random errors in measurements of gene expression and genotypes in participants, not all the potential regulatory variants could not be identified with this approach. Additional methods were needed to identify the additional variants.

Approach II: identifying potential regulatory variants by evaluating the disruptiveness of the variants on transcription factor binding sites

To identify potential regulatory variants that were missed in the approach I, another approach searching for variants affecting gene expression was developed. Previous fine mapping studies of causal variants in diseases suggested that a proportion of functional variants regulated expression by disrupting or creating a binding motif for transcription factors (TFs) for the target genes. [222-224] Therefore, it is hypothesized that functional variants can be identified by evaluating the disruptiveness of the variants on the binding of the relevant TFs for genes in the 5-FU catabolic pathway.

To predict the disruptiveness of transcription factor binding of candidate variants, position weight matrices (PWMs) of over 2800 motifs were obtained from the following databases, JASPAR,[225] TRANSFAC (commercial version),[226] Homer,[227] Factorbook,[228] ENCODE,[229] HOCOMOCO,[230] and Hi-SELEX.[231] In addition, a literature review was performed to identify transcription factors with evidence from functional assays. The R package motifbreakR [232] was used to estimate the binding affinities of the reference and alternative alleles of each candidate variant. The details of the algorithms were described previously. [232] Only variants with a PWM match *P* value less than 1×10^{-6} and effects on disrupting transcription factor bindings predicted to be strong by motifbreakR were remained for further analyses.

To better characterize each variants additional data of epigenetic features were obtained from the aforementioned functional genomic databases. It was hypothesized that variants locating at the enhancers that interacted with promoters of the genes were more likely to be functional. For *DPYD*, data on cross-tissue correlation of DNase I hypersensitivity (DHS) regions that predicted regulatory interaction between promoters and enhancers [233] were obtained from the ENCODE project data portal, [215] and data on the cross-tissue enhancer-promoter link were obtained from the FANTOM5 project. [176] For *DPYS* and *UPB1,* of which the expression was largely confined in liver cells, no data on the liver-specific enhancer-promoter interactions were publicly available. Therefore, no enhancer regions could be prioritized for these two genes under the hypothesis mentioned above. Instead, we hypothesized that liver-specific regulatory regions were more likely to be functional for *DPYS* and *UPB1*. Data on enhancers and

promoters that were specifically expressed in normal liver cells were obtained from the FANTOM5 project.[176]

Combining data on epigenetics features and predicted disruptiveness of transcription factor binding, a functional potential score for each candidate variant was created as follows,

$$S = W_{reg} \times \log\left(\frac{TAD}{D} \times TFS_{ref} \times \left|\left(TFS_{ref} - TFS_{alt}\right)\right|\right)$$

Where W_{reg} represents the weight for the predicted regulatory region where the candidate variant is located, for candidate variants of *DPYS* and *UPB1*, the weight for liver-specific promoter regions of the gene is 5, for liver-specific enhancers is 3, and for other enhancer regions is 1, respectively. For candidate variants of *DPYD*, the weight for promoter regions of the gene is five, the strong enhancers for which expression were correlated with the expression of promoters is 3*(1 + correlation coefficient (R^2)), and other enhancers is one, respectively. D represents the distance to the transcription start site of the gene in kb. TFS_{ref} and TFS_{alt} represent the predicted binding affinity scores of the reference and the alternative allele of the candidate variant, respectively. If more than one transcription factors were predicted, the one with the highest binding affinity of either allele was selected.

Results

The genetic variants that predicted the expression of genes in the 5-FU catabolic pathway in GTEx datasets were showed in Table 1. Twelve variants in predicted regulatory regions were identified for the gene *DPYD*, which together explained 12% of the variation of the cross-tissue component of the *DPYD* expression. Notably, the variant rs56038477, which is one of the four established genetic variants for 5-FU toxicity [8] as previously described was identified as one of the predictors. No genetic variants were identified for the liver-specific expression of *DPYD*. Four highly correlated variants locating in the active enhancer regions in liver cells were identified for the gene *UPB1*, which together explained 7% of the variation of the expression of *UPB1* in the liver tissues. No genetic variants were identified for predicting the expression of *DPYS* in liver tissues.

By integrating data on epigenetic features and predicted disruptiveness of transcription factor binding of candidate variants, sixteen potential regulatory variants were prioritized (Table 2 and Table S2 in Appendix A). For *DPYD*, three common variants locating at the enhancer regions that were predicted to interact with the promoter regions of *DPYD* were found (Appendix A, Table S2), but none of these variants passed the PWM match score criteria. Three variants locating at other enhancers were prioritized instead. For *UPB1*, two common variants, rs2070475 and rs2032116, were identified at the predicted liver-specific promoter near this gene, which were also predicted to disrupt TF binding. No common variants were identified at the predicted liver-specific enhancers or promoters for the gene *DPYS*. According to the functional potentiality scores of candidate variants, eleven variants were prioritized for *DPYS* and *UPB1*. Notably, the prioritized promoter variant rs2070474 for *UPB1* was found to be associated with a seven-fold increased risk of severe 5-FU associated toxicity in cancer patients in a previous study. [15] In addition, several variants that were correlated with rs2070474 were also prioritized for *UPB1*, including rs2032116, rs2298383 and rs5760447. To test if the prioritized variants predicted the expression of their target genes, we evaluated the association of these variants with expressions of genes in relevant tissues using the GTEx data (Appendix A, Table S2). With the exception of rs74450569, no statistically significant association was found for the prioritized variants. Similarly, no association in liver tissues were found for the established variants that had been found associated with expression of *DPYD*,[144,149] suggesting that the GTEx data were underpowered. Datasets with a larger sample size of liver tissues were needed to validate the association of the prioritized variants with expression of genes.

Discussion

Two approaches have been developed to identify genetic variants that may predict the expression of genes in the 5-FU catabolic pathway in this study. In the first approach that integrated epigenomics, genetics and transcriptomics data, 12 variants that together explained 12% of the variations in the cross-tissue expression of *DPYD* and 4 variants that together explained 7% of the variations in the expression of *UPB1* in normal liver tissues were identified. In the second approach, a functional potentiality score was created for each common variant in potential regulatory regions, and 18 variants were prioritized as

potential regulatory variants that may predict the expression of *DPYD*, *DPYS*, and *UPB1* in relevant tissues.

Although replication efforts and functional assays are needed to confirm findings from both approaches developed in this study, multiple lines of evidence suggested that these approaches were effective. In the first approach, the only variant with a MAF larger than 1% in the four established variants for 5-FU associated toxicity (rs56038477) was identified as one of the predictors for the cross-tissue component of *DPYD* in the GTEx data. In the second approach, a variant near *UPB*1 (rs2070474) that had previously been reported to be associated with the risk of severe 5-FU associated toxicity [15] was prioritized as a potential regulatory variant. Additionally, although the prioritized variant rs75570956 was not associated with expression of *DPYD* in the GTEx data (Appendix A, Table S2), it was found to be highly statistically significantly associated with the expression of this gene in whole blood cells in a larger dataset, according to a recent eQTL analysis in approximately 5,000 participants (beta for one G allele (predicted to decrease binding affinity) = -0.202, $P = 4.1 \times 10^{-10}$). [234] These consistencies with previous findings suggested that approaches developed in this study could be effective in identifying genetic variants predicting expression of genes. Additionally, under the assumption that genetic control of transcription is similar across genes, these two approaches can also be applied to other genes of which the expressions are at least partially regulated by *cis* genetic variants, which enables the genome-wide identification of variants predicting gene expression.

Approaches developed in this study were unique in searching for genetic variants in predicted regulatory regions and exploring the transcriptional variations in relevant tissues only. By limiting analyses to variants with evidence of regulatory roles in relevant tissues, we were able to detect variants predicting expression of target genes in a small sample size of transcription data and prioritize variants that were likely to be functional among tens of thousands of genetic variants. Several previous studies proposed to identify genetic components of expression of genes. [235,236] However, few studies had integrated functional genomics data to inform the analyses. To demonstrate that focusing on variants with evidence of regulatory roles improved the power of detecting potential variants, we performed analyses using all

common variants in the flanking region (±1M) in the same GTEx data as proposed by one of the previous studies, [235] but no variants were identified in the relevant tissues for genes in the 5-FU catabolic pathway, which was consistent with the results provided by this study. [235]

There are several limitations in this study. First, no validation analyses were performed to evaluate the association of the variants identified in the first approach. This was due to the lack of data on both genotype and gene expression in normal liver tissues. The Cancer Genome Atlas (TCGA) project represents a commonly used resource for genotype and phenotype analyses. However, *DPYD*, *DPYS* and *UPB1* together were mutated in approximately 50% of the liver cancer patients in TCGA, which could introduce substantial noises in the analyses using data derived from the adjacent normal liver tissues. [237] Second, no *trans* effects of genetic variants on gene expression were evaluated. This was due to that current knowledge in the transcription regulation of these genes was limited and no previous studies had identified the relevant *trans* regulatory regions and transcription factors. Future work on the transcription regulation of these genes was needed. Third, the epigenomics data in normal tissues from the REMC and FANTOM5 projects were usually generated in cells and tissues from one or a few healthy donors, which might not represent the general population of cancer patients and could not fully account for the inter-individual variabilities in epigenetic features. Additionally, the approached developed in this study could not account for many other factors, such as translation post-translational modification of the genes, which can also affect the activities of genes. Additional studies are needed to capture the genetic components of gene activities comprehensively.

In summary, this study has provided a framework that integrate epigenomics, transcriptomics and genomics data to identify *cis*- genetic markers that can predict expression of genes of interest. With this framework, genetic markers that might predict the 5-FU catabolic pathway activities were identified. Studies in cancer patients are needed to evaluate the association of these genetic markers with the risk of severe 5-FU associated toxicity events.

Table 7 Genetic variants in potential regulatory regions predicted the expression of genes in 5-FU catabolic pathway

Genes	Variants	Position (hg19)	Alleles[a]	EAF	Potential regulatory elements[b]
DPYD***					
	rs11165845	chr1:97819405	C/T	0.28	strong enhancer
	rs12032384	chr1:97843647	C/T	0.25	strong enhancer
	rs1415683	chr1:97845053	G/T	0.25	strong enhancer
	rs72728442	chr1:97849858	G/A	0.20	strong enhancer
	rs72728443	chr1:97849910	T/A	0.20	strong enhancer
	rs1356919	chr1:97852376	T/A	0.20	promoter
	rs115358442	chr1:98002678	A/C	0.005	strong enhancer
	rs116772342	chr1:98037830	C/T	0.02	promoter
	rs56038477	chr1:98039419	T/C	0.02	strong enhancer
	rs78944474	chr1:98049321	C/T	0.02	strong enhancer
	rs114806143	chr1:98093710	G/T	0.01	strong enhancer
	rs78593303	chr1:98133440	T/G	0.02	strong enhancer
UPB1***					
	rs1892721	chr22:24919329	C/T	0.16	strong enhancer
	rs5996713	chr22:24921558	C/T	0.16	strong enhancer
	rs12159862	chr22:24920322	T/C	0.16	strong enhancer
	rs5996712	chr22:24921214	A/G	0.16	strong enhancer

Abbreviations: EAF: effect allele frequency

[a] Effect allele/ reference allele in the GTEX data.

[b] The variants identified for *DPYD* were located in predicted regulatory regions in normal liver cells or other noncancerous cells in the REMC and ENCODE datasets. The variants identified were located in predicted enhancer regions in normal liver cells in the REMC datasets.

[*] The12 variants explained 12% of the variations in the cross-tissue component of *DPYD* expression while the four variants explained 7% of the variations in the expression of *UPB1* in the liver tissues in the GTEx project.

Table 8 Integrative analyses of functional genomics data identified potential regulatory variants for genes in the 5-FU catabolic pathway

Genes	Variants	Position (hg19)	Potential regulatory regions [a]	Distance to TSS (kb)	Ref	Alt	Transcription factor [b]	TFS*$_{ref}$	TFS*$_{atl}$
DPYD	rs74450569	1:98372645	strong enhancer	13.9	T	C	NF-E2	15.6	13.9
	rs75570956	1:98377149	active promoter	9.5	A	G	ERV1	12.5	11.2
DPYS									
	rs78426610	8:105479302	active promoter	0.02	G	C	RORA	19.9	22.2
	rs182332679	8:105479134	active promoter	0.1	C	T	PAX5	10.0	8.9
	rs2298840	8:105478933	active promoter	0.3	G	A	PU.1	14.6	2.6
	rs3793354	8:105424774	strong enhancer	54.5	G	A	TFAP2	11.0	9.1
	rs13274374	8:105291325	strong enhancer	250	T	C	RBP-Jkappa	10.9	11.9
UPB1									
	rs2070474	22: 24891292	active promoter	0.01	C	G	ZFX	12.8	11.8
	rs2070475	22:24891355	active promoter	0.1	A	T	GR	5.4	-6.6
	rs2032116	22: 24888796	active promoter	2.5	A	G	TBX21	22.4	20.8
	rs2298383	22:24825511	strong enhancer	65.8	C	T	ZBT7B	13.6	12.9
	rs5760447	22:24879289	strong enhancer	12.0	A	G	SOX9	23.8	22.3
	rs131455	22:24908536	strong enhancer	17.2	T	C	DMRT2	10.1	11.5
	rs7286246	22:24948759	strong enhancer	57.5	A	G	IRF4	9.7	8.8
	rs62234044	22:24823731	strong enhancer	67.6	C	T	TFAP2	9.0	7.4
	rs77804922	22:24830029	strong enhancer	61.3	G	A	ETS	6.4	5.2
	rs116907149	22:24683497	strong enhancer	207.8	A	G	ETS1	15.5	16.5

Abbreviations: TSS: transcription start site; Ref: reference allele; Alt: alternative allele;

a: predicted regulatory regions for *DPYD* across multiple tissues an cells including liver cells; predicted regulatory regions for *DPYS* and *UPB1* in liver cells.

b: transcription factors in the position weight matrices obtained from databases described in the Materials and methods.

*: S_{ref} and S_{alt} : the predicted binding affinity scores for the reference and alternative alleles of the variant, respectively.

CHAPTER IV

GENETIC MARKERS OF 5- FLUOROURACIL CATABOLIC PATHWAY ACTIVITIES ARE ASSOCIATED WITH RISK OF CHEMOTHERAPY-ASSOCIATED SEVERE TOXICITY EVENTS IN COLORECTAL CANCER PATIENTS

Introduction

Colorectal cancer (CRC) is the second leading cause of cancer-related deaths in men and the third in women in the United States. [238,239] Five-fluorouracil (5-FU) based chemotherapy has been extensively used in the treatment of CRC for both curative and palliative intent, which have been showed to improve survival in many patients. [240-243] However, dose-limiting side effects of chemotherapy remain a major clinical challenge. A recent meta-analyses of eight cohort studies showed that more than 25% of the patients who received 5-FU based chemotherapy developed severe toxicity during the treatment. [8] Severe toxicity events can lead to suboptimal treatment intensity, low quality of life and sometimes, death, highlighting the need of reliable prediction models. [8] Genetic testing of variants in the 5-FU metabolic pathway has showed some promising results. [244] However, the low mutation frequencies of the established variants (MAF ranged from 0.1% to 2%)[8,103,145] did not account for the magnitude of severe toxicity events observed in the population. Additional genetic risk factors remain to be identified. Several recent studies suggested regulatory variants associated with expression of genes in the metabolism pathway might account for a sizable portion of the interpatient variability of 5-FU associated toxicity. [21]

In the previous chapter, genetic variants predicting expression of genes in 5-FU catabolic pathway including and potential regulatory variants for these genes were identified. It was hypothesized that the decreased genetically predicted expression of 5-FU catabolic genes, the alleles predicted to decrease the binding of transcription factor in potential regulatory variants were associated with an increased risk of 5-FU-associated toxicity events in cancer patients. It was also hypothesized that genetic risk scores constructed by combining genetic markers in 5-FU catabolic genes to represent genetically regulated catabolic pathway activities were highly predictive. To test these hypotheses, we evaluated the

association of these genetic markers with the risk of 5-FU associated toxicity in a well-characterized cohort study of more than 400 CRC patients.

Methods

Study population

This study was approved by the Institutional Review Board (IRB) at VUMC. This study utilized resources from the Vanderbilt cancer registry, electronic medical record (EMR) database-Synthetic Derivative (SD) and the DNA biobank at Vanderbilt (BioVU). The details of the cancer registry, SD and BIOVU have been described previously. [245,246] Briefly, the Vanderbilt cancer registry is the official data repository for cancer patients at VUMC, which collects detailed clinical information of patients, including primary tumor site, stage at diagnosis, histology and date of recurrence. The SD is a de-identified data repository of clinical information of patients who received care at VUMC. BioVU is the repository of DNA samples from patients with a linkage to SD. All the CRC patients documented in the cancer registry with records in SD and non-compromised DNA samples available in BioVU were identified. Patients who met each of the following criteria were included: 1) with a primary diagnosis of CRC; 2) received 5-FU based chemotherapy; 3) with detailed records on toxicity events for at least the first four cycles of the treatment (see “outcomes” and Figure 2).

Data on the tumor characteristics (tumor site, stage at diagnosis, grade and histology) and date of diagnosis were extracted from the cancer registry. Data on height and weight at diagnosis for each patient were extracted from the SD. The algorithms for extracting these variables were included in the Appendix. Medical records of each patient were reviewed to obtain the information on chemotherapy regimens, lifestyle factors (smoking and alcohol drinking), and toxicity events during the treatment.

Outcomes

The primary outcome in this study was defined as developing at least one of the severe toxicity events (including hematological toxicity events, gastrointestinal toxicity events, cardio-toxicity events and dermatological toxicity events) in any of the first four cycles. To determine the severity of toxicity events, the NCI Common Toxicity Criteria for Adverse Events (CTCAE) V.4.0 (Table 5 & 6) were used to

evaluate the grade of each haematological, gastrointestinal or dermatological event. All cardiotoxicity events due to 5-FU were included as severe events. Secondary outcomes were defined as patients with any individual toxicity events within the first four cycles, respectively. With the exception of dermatologic events (mainly hand-foot syndromes that are usually due to 5-FU), these toxicity events can be caused by many conditions other than 5-FU. For example, virus infection can cause severe diarrhea and other GI events, and Dilantin taper can cause neutropenia. To distinguish the events caused by chemotherapy from those by other conditions, clinical notes for each patient were explored. The likely etiologies of events were determined according to the judgement of treating physicians documented in the clinical notes. Patients who developed toxicity solely due to conditions other than 5-FU based chemotherapy were included as patients without the events. For example, patient A developed diarrhea leading to hospitalization after the first cycle of chemotherapy that was later found to be caused by foodborne virus, and no severe toxicity events were observed in the other early cycles. Therefore, patient A was included as a patient without early-onset severe toxicity events. Patients who developed severe events but the etiology remained undetermined according to the judgement of treating physicians were excluded. Because no patients with severe early-onset 5-FU-associated toxicity events were identified in the populations of African ancestry or populations of Asian ancestry, only patients of European ancestry were remained in the genetic analyses.

Genotyping

Targeted genotyping was performed using the iPLEX Sequenom MassARRAY platforms. For the 39 SNPs identified in the previous chapter, 3 failed and 12 were replaced by surrogate SNPs (Appendix B, Table S1). To be consistent with the reporting in the previous chapter, the identified SNPs instead of their surrogate SNPs were presented in the main tables. Four negative controls (water) and eight positive quality controls (HapMap or duplicate samples) were included in each 384-well plate. We filtered out participants with a genotype call rate $< 98\%$. We filtered out SNPs with (i) genotype call rate $< 95\%$, (ii) genotyping concordance rate $< 95\%$ in positive control samples, (iii) an unclear genotype call or (iv) P for Hardy-Weinberg equilibrium < 0.01. We calculated the mean concordance rate using data from positive

quality control samples. Each genetic variant was coded into 0, 1, 2 for homozygotes of reference alleles, heterozygotes, and homozygotes of effect alleles, respectively.

Statistical analysis

Differences in the demographic, lifestyle and clinical characteristics of patients with and without early-onset severe toxicity were tested using *t*-tests for continuous variables or χ^2 test for categorical variables. Unconditional logistic regression models were performed to evaluate the association of genetic markers identified in Chapter III with the risk of early-onset chemotherapy-associated severe toxicity events in CRC patients. For genetic markers (genetic component of expression of genes) identified in the first approach (Chapter III), genetic risk scores for *DPYD* and *UPB1* were created using the equation $\sum_{i=1}^{k} \beta_i SNP_i$, where k was the number of SNPs selected, β was the weight derived from the LASSO algorithm using the GTEx data (Appendix A, Table S1). Furthermore, a genetic risk score representing the predicted 5-FU catabolic pathway activities, termed as 5-FU catabolic pathway activities GRS 1, was created by adding up the genetic scores for genes in the pathway. For genetic variants prioritized in the second approach, variants for each gene were tested, assuming an additive effect. For the highly correlated variants, only one variant was selected for reporting. To evaluate the cumulative association of variants in the same gene, a genetic risk score was created for each gene using the equation as follows, $\sum_{i=1}^{k} aSNP_i$, where k was the number of SNPs prioritized for the gene, a =1 if the effect allele is predicted to decrease the binding of transcription factor, and a =-1 if the effect allele is predicted to increase the binding of transcription factor (Appendix A, Table S1). For the highly correlated variants, only one variant was selected to create the score. Similarly, these scores were added up to represent the predicted 5-FU catabolic pathway activities, termed as 5-FU catabolic pathway activity GRS 2. Genetic variants that have been reported in previous studies (see Chapter II) were also evaluated, including the established *DPYD* variants, common variants in *DPYD* that were identified in a previous fine-mapping study [146] and variants in *DPYS* [14] and *UPB1* [15] (Appendix A, Table S1). Adjustment of demographic, lifestyle, and clinical variables did not materially change the estimates of the genetic markers evaluated.

Therefore, the crude odds ratios were reported. The associations with early-onset severe haematological toxicity events, gastrointestinal toxicity events, cardiotoxicity events, and dermatologic toxicity events were also evaluated separately.

To evaluate the prediction values of these genetic markers, prediction models were built. Because of the small sample size, selecting predictors using data in this study was less likely to generate reproducible results. Therefore, pre-specified prediction models were developed instead. As discussed in Chapter II, age at diagnosis, sex and BSA were well-established risk factors for chemotherapy-associated toxicity events. These variables were included as predictors. Genetic risk scores for 5-FU catabolic pathway activities, age and BSA at diagnosis were included as linear continuous variables. Sex, tumor characteristics variables (tumor site, stage) and treatment variables (regimens and dose reduction in the beginning of the treatment) were included as categorical variables. Bootstrap sampling was used to estimate the optimism in the performance of the prediction models. The optimism-corrected *c* statistics was used to evaluate prediction performance. A *c* statistics larger 0.75 was considered of high predictive value, a *c* statistics between 0.60 and 0.75 was considered of some predictive value, whereas a *c* statistics less than 0.6 was considered of minimal predictive value.

Results

A total of 424 colorectal cancer patients of European ancestry who received 5-FU based chemotherapy were included in this study, 24% of whom developed early-onset severe toxicity events. Patient who had early-onset of severe toxicity events were less likely receive a standard dose of 5-FU when initiating treatment and receive FOLFOX-based regimens than patients who did not (Table 9).

Association results of newly identified genetic markers that predicted expression of 5-FU catabolic genes and risk of early-onset severe toxicity events were shown in Table 10. The genetically predicted expressions of 5-FU catabolic genes showed a moderate association (OR per standard deviation (SD) decrease in predicted expression = 1.19, (95% CI: 0.98-1.45, $P = 0.08$) and 1.20 (95% CI: 0.97-1.49, $P = 0.08$), respectively, for *DPYD* and *UPB1*). To evaluate the association of the predicted 5-FU catabolic pathway activity, both predicted *DPYD* and *UPB1* expression were combined. The genetically predicted

5-FU catabolic activity score was statistically significantly associated with risk of early-onset severe toxicity events (OR per one standard deviation (SD) decrease in predicted catabolic activity score = 1.28, (95% CI: 1.05-1.59, *P* = 0.02)). Among genetic variants that were identified as potential regulatory variants of gene expression, a common variant rs2032116 (MAF = 0.43), lying in a strong enhancer region of *UPB1* in normal liver tissues, was statistically significant associated with risk of early-onset severe toxicity events (OR for each effect allele = 1.48, 95% CI: 1.08-2.02, *P* = 0.01). The variant rs75570956, predicted to disrupt the binding of transcription factors in the promoter region of *DPYD,* showed an association with borderline significant (OR for each effect allele = 1.94, 95% CI: 0.94-2.94, *P* = 0.10). We created a genetic risk score for each gene by combining potential regulatory variants (OR per effect allele =1.37 (95% CI: 0.92 - 2.03, *P* = 0.12), 1.20 (95% CI: 0.85-1.68, *P* = 0.30), and 1.44 (95% CI: 1.06 -1.94, *P* = 0.02), respectively, for *DPYD*, *DPYS* and *UPB1*). The genetic risk score of each gene was combined to derive a catabolic pathway activity genetic score, which was found to be statistically associated with risk of early-onset severe toxicity (OR per one allele that predicted to decrease binding of transcription factor = 1.33 (95% CI: 1.09 -1.61, *P* = 0.004)). Additionally, several previously reported variants were also evaluated in this study (Table 10, and Appendix B, Table S2). Among the four established *DPYD* variants, *DPYD**2A and c.1679T>G were invariant in this cohort, and the genotype data of *DPYD* c.2846A>T were not available. Therefore, no association results were available for these three variants. All the tested variants showed allelic associations in the same direction as previously reported. Specifically, the established variant *DPYD* c.1129-5923 C>G (rs75017182), which is thought to change a splice donor site and lead to the retention of an additional 44-nucleotide cryptic region at the exon 10, [149] showed an association of marginal significance with the risk of early-onset severe toxicity events (OR for each effect allele = 1.93, 95% CI: 0.87-4.26, *P* = 0.10). The *UPB1* promoter variant rs2070474 also showed a marginally significant association (OR for each effect allele = 1.19, 95% CI: 0.98-1.44, *P* = 0.07).

Among patients who developed early onset severe toxicity events, the most commonly observed events were gastrointestinal toxicity events (45.6%, Table 11), followed by myelosuppression (36.9%),

and dermatological toxicities (22.3%) and cardiotoxicities (10.7%). No statistically significant differences were found in the associations of the evaluated genetic markers across subtypes of toxicity events (Table 11).

The prediction performance of genetic risk scores of 5-FU catabolic pathway activities was assessed, in combination with the previously reported markers and demographic and clinical factors (age at diagnosis, sex, BSA, chemotherapy regimens and dose reduction at the beginning of treatment) in the BIOVU cohort. After correcting for the optimism, the *c* statistics for the full model including both the new genetic markers and the known genetic markers and non-genetic risk factors was 0.62 while the *c* statistics for the model including the previously reported markers and clinical factors was 0.58 (Table 12).

Discussion

By analyzing genetic markers identified in the framework proposed in the previous chapter in a well-characterized cohort of CRC patients, this study showed that genetically predicted *DPYD* and *UPB1* expression together (5-FU catabolic pathway activity score 1) were associated with the risk of severe 5-FU-associated toxicity in CRC patients. This study also showed that the genetic risk score combining potential regulatory variants for genes in the 5-FU catabolic pathway (5-FU catabolic pathway activity score 2) was statistically significantly associated with the risk of severe 5-FU-associated toxicities. These findings highlighted the importance of common variants that were associated with expression of genes in the risk of 5-FU associated toxicity. These finding also supported the importance of evaluating all gene in the 5-FU catabolic pathway. In combination with known genetic and non-genetic risk factors, these pathway genetic risk scores had moderate predictive value, showing promise for pretreatment risk stratification in cancer patients.

Findings from this study supported that the approaches developed in Chapter III were effective in identifying genetic markers relevant expression of genes and provided insight into the underlying regulatory mechanism. Among all the genetic variants evaluated, the most statistically association was observed with a common variant rs2032116 (MAF=0.43), of which was associated with a 50% increased risk for each effect allele. This variant is located in a liver-specific promoter region of *UPB1* and

predicted to change the binding affinity of the transcription factor *TBX21*, the deficiency of which had been showed to drive the carcinogenesis of colitis-associated CRC. [247] This SNP is in LD with the previous reported variant rs2070474 (R^2= 0.78 in study population, OR = 1.33, P = 0.07). Conditional analyses showed that the association of rs2032116 remain statistically significant (P = 0.04) while the association of rs2070474 diminished in the model including both variants in the same model. Another variant that showed evidence of association is a promoter variant of *DPYD*, rs75570956. This variant was predicted to located in the binding site of ERV1 that is likely to co-occur with P53 binding sites.[248] By evaluating the TFBS near this variant in the TRANFAC database, we found that this variant also disrupted the binding site of P53. A recent study in showed that P53 repressed the expression of *DPYD* in liver cells treated with 5-FU.[249] This observation raised a possibility that certain genetic variants might be associated with the expression of *DPYD* only in the presence of 5-FU, suggesting that gene-environment interaction might be an important factor in the etiology of 5-FU associated toxicity.

The observation that the association of previously reported variants in this study was in the same direction as in the initial reports suggested that the algorithms detecting patients with the primary outcome developed in this study were effective, suggesting that mining clinic notes could be valuable for identifying phenotypes of complicated presentations and etiology. It has been well recognized that identifying patients with specific adverse drug reactions remained one of the challenges in in EMR-based studies. Previous work in phenotyping in EMR mainly relied on ICD codes, medications and lab results, which could be documented in structured forms. [245,250] However, the precision and recall rates of these algorithms were not optimal, likely due to the broad spectrum of symptoms in these side effects, which were likely to be under-documented or inaccurately documented in structured forms. To date, no algorithms for detecting 5-FU associated toxicity events in EMR has been published. It is expected that detecting such events using structured forms would be challenging because of the broad spectrum of clinical presentations of these events. In this study, there were four subtypes of toxicities, each of which included multiple phenotypes that could be linked to different ICD codes. For example, the gastrointestinal toxicity events include diarrhea, vomiting, stomach mucositis etc. In addition, these

events could be missed in structured forms if patients received treatment for these events in other hospitals. Furthermore, a majority of these symptoms could be caused by many other factors than chemotherapy. The consistencies of findings with previous studies suggested that using treating physicians' judgements on outcomes in the unstructured clinical narratives could be effective in phenotyping complex traits in EMR.

No substantial heterogeneity in the associations with subtypes of toxicity events was observed for the new genetic markers or previously reported markers for which differentiated risks by subtypes have been reported. One of the possible explanation was the lack of power due to the small sample size for each subtype. For example, the previously reported variant, rs12022243 , which initially showed a stronger association with the risk of diarrhea compared with hand-foot syndromes in patients who received the oral form of 5-FU, [146] showed some evidence of differences in the effect size of associations between gastrointestinal toxicity and dermatologic toxicities.

This study has several limitations. No independent studies were available to validate the associations of the new genetic markers with the risk of early-onset severe toxicity events. However, these markers were identified through integrating data from multiple external datasets, which could increase the likelihood of reproducibility of the findings in this study. Nevertheless, independent studies are desired to validate the observed associations. Another limitation was that participants in this study were highly selective and might not represent the population of CRC patients receiving treatment at VUMC. Because only patients with detailed records were included, and those who developed severe toxicity events were more likely to have detailed information than those who did not, this lead to a higher percentages (24%) of patients who developed early-onset severe events than expected (approximately 10 to 15%) in this cohort. The prediction models built in this study and their predictive performance were less likely to be generalized to other populations of CRC patients. Additional studies that are presentative of CRC patients who receive 5-FU based chemotherapy in the population are needed to validate the predictive value of these newly identified markers.

In conclusion, this study identified novel genetic markers that were associated with the risk of early-onset severe toxicity events in CRC patients. This study supported that the framework developed in this dissertation were effective in identifying genetic markers that predict gene expression, suggested the importance of common regulatory variants in the etiology of 5-FU-associated toxicity, and provided directions for future pharmacogenetics studies.

Table 9 Selected baseline characteristics of colorectal cancer patients of European ancestry who received 5-FU based chemotherapy at Vanderbilt University Medical Center, 1997-2016

Characteristics	Patients with early-onset severe toxicity events (n=103)		Patients without early-onset severe toxicity events (n=321)		*P* value
	%	Mean(SD)	%	Mean(SD)	
Age at diagnosis, years		57.6 (11.6)		55.2 (12.0)	0.07
Sex (male)	51.4		56.7		0.35
Performance status*					0.81
0-1	72.8		78.8		
2-4	5.8		5.6		
unknown	21.3		15.6		
BMI at diagnosis , continuous		28.5 (5.9)		29.0 (6.2)	0.54
BSA at diagnosis					0.48
<= 2.0	57.3			52.3	
>2.0	42.7			46.7	
Smoking					0.59
Current	17.5		15.9		
Former	26.2		31.4		
Never	56.3		52.6		
Tumor site					0.55
colon	64.1		67.3		
rectum	35.9		32.7		
Stage					0.63
regional	65.0		67.6		
distal	35.0		32.4		
Chemotherapy regimen					0.04
Single agent 5-FU	22.3		15.0		
FOLFOX**	51.4		65.1		
FOLFIRI**	22.3		18.7		
FOLFOXIRI	3.9		1.2		
5-FU dose reduction when initiating the treatment					
Yes	6.8		4.9		0.02
No	85.4		92.8		
Unknown	0.07		0.02		

Abbreviation: 5-FU: 5- fluorouracil; FOLFOX: folinic acid, fluorouracil and oxaliplatin; FOLFIRI: folinic acid, fluorouracil and irinotecan; FOLFOXIRI: folinic acid, fluorouracil, oxaliplatin and irinotecan; SD: standard deviation; BMI: body mass index; BSA: body surface area.

*Performance status were graded according to the Eastern Cooperative Oncology Group (ECOG) performance status grading scale.

**Including regimens with or without biological agents, such as anti-VEGF and anti-EGFR antibodies.

Table 10 Association between previously reported variants and new genetic markers that might associated with expression of genes in the 5-fluorouracil catabolic pathway and risk of early-onset severe toxicity events in colorectal cancer patients of European ancestry

Genes	Genetic markers	Alleles[a]	EAF[b]	Odds ratios (95% CI)[c]	P value
Previously reported genetic markers					
DPYD	rs75017182	C/G	0.03	1.93 (0.87-4.26)	0.10
DPYD	rs12132152	A/G	0.02	1.23 (0.49-3.11)	0.66
DPYD	rs12022243	T/C	0.19	1.18 (0.79-1.78)	0.41
DPYS	rs2959023	A/G	0.44	0.97 (0.71-1.33)	0.86
UPB1	rs2070474	C/G	0.44	1.33 (0.97-1.83)	0.07
grs for reported markers*		-	-	1.19 (0.98-1.44)	0.07
Approach 1: genetically predicted expression					
*DPYD**	See Appendix A	-	-	1.19 (0.98-1.45)	0.08
*UPB1**	See Appendix A	-	-	1.20 (0.97-1.49)	0.08
5-FU catabolic pathway activity grs 1*		-	-	1.28 (1.05-1.59)	0.02
Approach 2: other potential regulatory variants					
DPYD	rs74450569	C/T	0.09	1.07 (0.63-1.83)	0.89
DPYD	rs75570956	G/A	0.10	1.54 (0.94-2.54)	0.09
DPYD grs	-	-	-	1.37 (0.92-2.03)	0.12
DPYS	rs78426610	C/G	0.03	1.35 (0.57-3.18)	0.49
DPYS	rs2298840	T/C	0.20	0.94 (0.64-1.40)	0.77
DPYS	rs3793354	T/C	0.07	0.57 (0.27-1.17)	0.12
DPYS	rs13274374	C/T	0.03	1.41 (0.60-3.35)	0.43
DPYS grs	-	-	-	1.20 (0.85-1.68)	0.30
UPB1	rs2032116	A/G	0.43	1.48 (1.08-2.02)	0.01
UPB1	rs131455	C/T	0.01	0.52 (0.06-4.34)	0.52
UPB1	rs62234044	T/C	0.01	0.44 (0.05-3.61)	0.44
UPB1 grs	-	-	-	1.44 (1.06-1.94)	0.02
5-FU catabolic pathway activity grs 2*		-	-	1.33 (1.09-1.61)	0.004

Abbreviations: GRS: genetic risk score; 5-FU, 5-fluorouracil, EAF: effect allele frequency; OR: odds ratio; CI: confidence interval.

[a]Alleles: effect allele/reference allele.

[b]EAF: effect allele frequency in the BIOVU cohort.

[c]The odds ratios for each SNP was estimated as per one copy of effect allele of genetic variants, the odds ratios for genetically predicted gene expression were estimated as odds ratios per one standard deviation decrease in the genetically predicted expression. The odds ratio for 5-FU catabolic pathway activity grs 2 were estimated as the odds ratio per effect allele in any of the variants included in the score. Adjustment of age at diagnosis, sex and all the available clinical variables did not change the estimates of these genetic markers.

* the genetic risk score for previously reported variants were created by adding all the variants into one genetic risk score. The genetic risk score for genetically predicted expression was constructed using the variants and weights estimated in approach 1(see Chapter III) so that a decrease in this score was predicted to be associated with a decreased expression level of the target gene. The genetic risk score using potential regulatory variant for each gene was constructed by summing each allele predicted to decrease binding of transcription factor for variants prioritized in approach 2 (see Chapter III) so that increase in this score was predicted to be associated with a decreased expression level of the target gene.

Table 11 Association of genetic markers and risk of subtypes of early-onset severe toxicity events in colorectal cancer patients of European ancestry

Gene	Genetic markers	GI tract* (n=47) OR (95% CI)	Hematologic* (n=38) OR (95% CI)	Dermatologic* (n=23) OR (95% CI)	Cardiotoxicity* (n=11) OR (95% CI)	P heterogeneity
Previously reported genetic markers						
DPYD	rs75017182	0.86 (0.21-3.59)	2.56 (0.94-6.98)	2.49 (0.74-8.37)	1.74 (0.25-12.1)	0.63
DPYD	rs12132152	1.76 (0.60-5.16)	1.11 (0.27-4.63)	0.93 (0.13-6.57)	NA	0.93
DPYD	rs12022243	1.37 (0.81-2.34)	1.41 (0.78-2.56)	1.08 (0.50-2.33)	0.69 (0.19-2.43)	0.73
DPYS	rs2959023	0.97 (0.63-1.50)	0.96 (0.60-1.55)	0.88 (0.48-1.62)	0.58 (0.23-1.46)	0.79
UPB1	rs2070474	1.48 (0.96-2.28)	1.77 (1.09-2.87)	1.25 (0.68-2.28)	1.13 (0.48-2.66)	0.75
GRS for reported markers		1.25 (0.97-1.61)	1.38 (1.04-1.82)	1.11 (0.78-1.58)	0.80 (0.47-1.37)	0.35
Approach 1: genetically predicted expression						
DPYD predicted expression		0.96 (0.68-1.37)	1.33 (1.04-1.69)	1.23 (0.80-1.64)	1.09 (0.65-1.85)	0.51
UPB1 predicted expression		1.22 (0.92-1.61)	1.30 (0.97-1.75)	1.10 (0.74-1.64)	1.05 (0.60-1.85)	0.88
5-FU catabolic pathway grs 1		1.10 (0.91-1.52)	1.45 (1.05-1.89)	1.20 (0.90-1.69)	1.12 (0.81-1.89)	0.59
Approach 2: other potential regulatory variants						
DPYD GRS		0.96 (0.53-1.74)	1.28 (0.70-2.36)	1.63 (0.81-3.27)	0.44 (0.10-1.97)	0.39
DPYS GRS		1.15 (0.72-1.82)	1.12 (0.67-1.85)	1.27 (0.68-2.37)	0.98 (0.39-2.44)	0.65
UPB1 GRS		1.64 (1.08-2.50)	1.73 (1.10-2.72)	1.47 (0.83-2.60)	1.24 (0.56-2.73)	0.89
5-FU catabolic pathway grs 2		1.28 (0.98-1.66)	1.40 (1.04-1.87)	1.39 (0.99-1.96)	0.95 (0.55-1.62)	0.63

Abbreviations: GI: gastrointestinal tract; GRS: genetic risk score; 5-FU, 5-fluorouracil, OR: odds ratio; CI: confidence interval, NA: not applicable
*GI tract toxicity included diarrhea, vomiting, nausea, stomatitis and mucositis; Hematologic toxicity included neutropenia, thrombocytopenia and leukopenia; Dermatologic toxicity included hand foot syndromes; cardiotoxicity included coronary vasospasm, myocardial infarction and any other ECG abnormalities.

Table 12 Prediction performance of models with genetic markers and non-genetic risk factors in CRC patients

Prediction models	Predictors included in the model	Uncorrected *c* statistics	Corrected *c* statistics
Model 1	5-FU catabolic pathway GRS 1 and 2	0.62	0.61
Model 2	Clinical factors and previously reported genetic markers*	0.63	0.58
Model 3	Clinical factors only*	0.61	0.57
Model 4	Previously reported genetic markers	0.57	0.56
Full model	5-FU catabolic pathway GRS 1 and 2, previously reported markers, and clinical factors	0.66	0.62

*Clinical factors including age at diagnosis, sex, body surface area, chemotherapy regimen and dose reduction when initiating treatment.
**Previously reported genetic markers included rs75017182, rs12132152, rs12022243, rs2959023, and rs2070474. SNP rs56038477 were excluded due to its high LD with rs75017182 in this cohort.

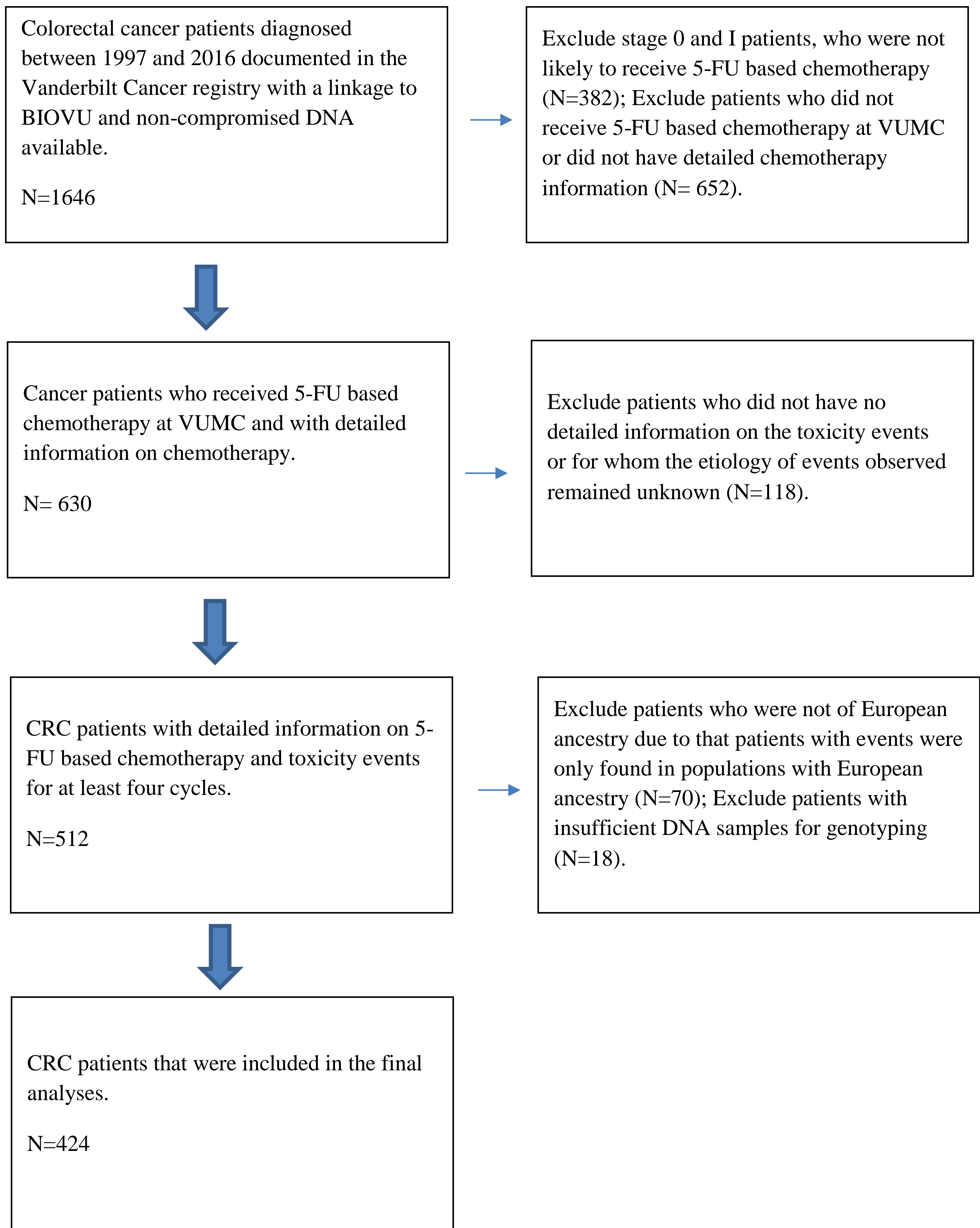

Figure 2 Flowchart describing the BioVU colorectal cancer cohort with side boxes explaining the reasons for exclusion.

CHAPTER V

SUMMARY AND FUTURE DIRECTIONS

Summary

This dissertation sought to identify new genetic markers for 5-FU-associated toxicity events in cancer patients. To achieve this goal, I first identified genetic markers that might predict expression of 5-FU catabolic genes in relevant tissues by integrative analyses of epigenetic, genetic and transcriptomic data. I next built a retrospective cohort study of CRC patients who received 5-FU based chemotherapy at VUMC in the last two decades. By evaluating the association of genetic markers of expression of 5-FU catabolic genes with the risk of severe 5-FU associated toxicity events in these patients, I found that predicted expression of the catabolic genes *DPYD* and *UPB1* by common variants and the combined genetic risk score of potential regulatory variants for *DPYD*, *DPYS* and *UPB1* were associated with the risk of severe 5-FU associated toxicity events. Together with establish genetic and non-genetic risk factors, these new markers provided moderate predictive value for risk stratification in CRC patients. These findings support my hypothesis that regulatory variants play a critical role in the catabolism of 5-FU and can be important markers for 5-FU-associated toxicity in cancer patients. To my knowledge, this dissertation is the first study that systematically evaluated common variants that might regulate expressions of genes in the 5-FU catabolic pathway in cancer patients. If findings from this study were validated in independent cohorts, this dissertation would provide a new direction for future pharmacogenetics studies.

This dissertation attempted to address the current knowledge gap in the role of regulatory variants in the pharmacogenetics of 5-FU-associated toxicity events by exploring the regulatory landscape of genes in the 5-FU catabolic pathway and evaluating potential regulatory variants or variants associated with expression of genes in a well-characterized cohort of CRC patients. Specifically, this dissertation has tried to address several questions and raised new ones. First, this dissertation proposed a framework of identifying potential regulatory variants through integrative analyses of multiple functional genomics datasets. Although functional studies are needed to validate the regulatory role of these variants, association results of these variants with risk of 5-FU-associated toxicity suggest that these variants might

be functional or in LD with functional variants for 5-FU catabolic genes. However, due to the small sample size of the GTEx project, I am not able to estimate the heritability of expressions of these genes confidently and how much of the heritability these newly identified genetic variants together account for. It is likely that this framework still miss a substantial percentage of regulatory variants for these genes due to the limited knowledge in the regulatory mechanisms of gene expression. Nevertheless, the framework proposed in this dissertation provides a direction of identifying potential regulatory variants using existing functional genomics data and enable the identification of potential regulatory variants across the genome. Second, this study, along with previous studies, supports the potential of using EMR as reliable resources for pharmacogenetics studies of side effects in therapy. It has been well recognized that measuring side effects, such as toxicity events in cancer patients in this study, due to the broad spectrum of relevant symptoms and the lack of structured records for these events. Through manual reviews of all patients included in the study, I find that the judgement of treating physicians on toxicity events documented in the narrative notes are important for identifying patients with events and determining the severity and nature of the toxicity events. This enables identification of patients with events of interest through natural language processing algorithms in studies of large sample size.

Future directions

Future studies proposed here will be aimed at addressing the limitations of this dissertation and extending research strategies explored in this study to address current hurdles to implementing pretreatment genetic testing for risk stratification in cancer patients and further understand the regulatory mechanism of 5-FU catabolic genes.

The major limitation of this study lies in that no replication studies were available to validate the observed association of the new genetic markers and the prediction models with known risk factors and these markers. To validate the association of these markers with toxicity events, I propose to evaluate these new genetic markers in studies with accurately measured treatment and outcome data, for example, clinical trials. To evaluate the predictive performance, I propose to use studies that are representative of the general population of CRC patients, for example, population-based studies.

If findings from this dissertation were validated, studies that evaluate the plausibility of incorporating these new genetic markers into genetic testing for 5-FU-associated toxicity would be needed. Early detection of patients at a high risk of severe toxicity, which allows for modifications of regimens and/or before the treatment, is desirable to the personalized management of CRC patients who are treated with chemotherapy. Several pharmacogenetics kits have been developed to identify patients at risk of severe toxicity, which include different combinations of coding and/or splicing variants in genes in the metabolism pathways of 5-FU, such as *DPYD* and *TYMS*. [4,146] However, none of these pharmacogenetics kits or any other pharmacogenetics tests have been routinely used in clinical practice. One of the primary factors that prevent their implementation is the low sensitivity of these tests. It was estimated that the sensitivity of these kits ranged from 10% to 30%, which does not meet the need in clinical practice. [4] The new genetic makers are composed of common variants, which are likely to explain a larger proportion of genetic variability in the susceptibility of severe toxicity and increase the sensitivity of the pharmacogenetics tests.

Functional studies are needed to elucidate the underlying regulatory mechanism of 5-FU catabolic genes and identify causal variants that affect the expression of these genes. Although it is not the major focus of this dissertation, the approaches developed to identify genetic component of gene expression and prioritize potential regulatory variants has also provided valuable insight into the regulatory mechanisms of 5-FU catabolic genes. For example, this study suggested that the variant rs75570956 might be associated with expression of *DPYD* through disrupting a P53 binding site, which was consistent with a recent finding that P53 is an important regulator of *DPYD* expression in liver cells.[249]

In summary, these proposed studies will address different aspects of questions that this dissertation has raised, inform future studies in pretreatment risk stratification in cancer patients, and provide insight into the regulatory mechanism of 5-FU catabolic genes. The research strategies presented in this dissertation has laid a foundation that these proposed studies could be built on. The translational nature of this dissertation project will hopefully accelerate the personalized management of chemotherapy in cancer patients.

APPENDIX

A. Supplementary information for Chapter III

Table S1 Genetic variants predicting the cross-tissue component of expression of *DPYD* and the expression of *UPB1* in the liver tissues in the GTEx data

Gene	SNP	Position	Ref	Alt	Predicted regulatory regions	Weight
DPYD	rs6691565	chr1:97827246	T	C	active enhancer	-0.00050038
DPYD	rs12032384	chr1:97843647	T	C	active enhancer	-0.00004739
DPYD	rs1415683	chr1:97845053	T	G	active enhancer	-0.00000107
DPYD	rs72728442	chr1:97849858	A	G	active enhancer	-0.01648624
DPYD	rs72728443	chr1:97849910	A	T	active enhancer	-0.00012756
DPYD	rs1356919	chr1:97852376	A	T	promoter	-0.00081639
DPYD	rs115358442	chr1:98002678	C	A	active enhancer	-0.03802596
DPYD	rs116772342	chr1:98037830	T	C	promoter	-0.02247810
DPYD	rs56038477	chr1:98039419	C	T	active enhancer	-0.00028277
DPYD	rs78944474	chr1:98049321	T	C	active enhancer	-0.00740645
DPYD	rs114806143	chr1:98093710	T	G	active enhancer	-0.06322717
DPYD	rs78593303	chr1:98133440	G	T	active enhancer	-0.06941193
UPB1	rs1892721	chr22: 24919329	T	C	active enhancer	-0.027427060
UPB1	rs5996713	chr22: 24921558	C	T	active enhancer	-0.000005764
UPB1	rs12159862	chr22: 24920322	T	G	active enhancer	-0.000363967
UPB1	rs5996712	chr22: 24921214	T	C	active enhancer	-0.000097857

Table S2. The association of variants prioritized in the second approach with expression of their target genes in the relevant tissues in the GTEx project

Gene	SNP	Position (hg19)	Ref	Alt	predicted regulatory region	Liver specific region	interacting with promoters	PMW match	eQTL analysis in GTEx datasets		
									Tissues	beta*	*P* value*
DPYD	rs4294451	chr1:98395714	T	A	Enhancer	No	Yes	No	Cross-tissues	0.002	0.88
DPYD	rs6676451	chr1:98291840	T	A	Enhancer	No	Yes	No	Cross-tissues	-0.013	0.29
DPYD	rs6682525	chr1: 98403398	T	G	Enhancer	No	Yes	No	Cross-tissues	-0.06	0.57
DPYD	rs74450569	chr1:98372645	T	C	Enhancer	No	No	Yes	Cross-tissues	-0.07	0.0004
DPYD	rs75570956	chr1:98377149	A	G	Enhancer	No	No	Yes	Cross-tissues	0	0.78
DPYD	rs78334244	chr1: 98386000	A	G	Promoter	No	No	Yes	Cross-tissues	0.05	0.82
DPYS	rs3793354	chr8:105424774	G	A	Promoter	No	N/A	Yes	Liver	-0.03	0.81
DPYS	rs2298840	chr8:105478933	G	A	Promoter	No	N/A	Yes	Liver	-0.02	0.82
DPYS	rs182332679	chr8:105479134	C	T	Promoter	No	N/A	Yes	Liver	0.24	0.57
DPYS	rs78426610	chr8:105479302	G	C	Enhancer	No	N/A	Yes	Liver	0.24	0.57
DPYS	rs13274374	chr8:105291325	T	C	Enhancer	No	N/A	Yes	Liver	0.45	0.04
UPB1	rs116907149	chr22:24683497	A	G	Enhancer	No	N/A	Yes	Liver	0.31	0.27
UPB1	rs62234044	chr22:24823731	C	T	Enhancer	No	N/A	Yes	Liver	0.28	0.41
UPB1	rs2298383	chr22:24825511	C	T	Enhancer	No	N/A	Yes	Liver	-0.03	0.62
UPB1	rs77804922	chr22:24830029	G	A	Enhancer	No	N/A	Yes	Liver	0.28	0.41
UPB1	rs5760447	chr22:24879289	A	G	Enhancer	No	N/A	Yes	Liver	0	0.98
UPB1	rs2070475	chr22:24891355	A	T	Promoter	Yes	N/A	Yes	Liver	0.17	0.10
UPB1	rs131455	chr22:24908536	T	C	Enhancer	No	N/A	Yes	Liver	0.08	0.72
UPB1	rs7286246	chr22: 24948759	A	G	Enhancer	No	N/A	Yes	Liver	-0.02	0.90
UPB1	rs2070474	chr22:24891292	C	G	Promoter	No	N/A	Yes	Liver	-0.03	0.66

*beta and P-value was estimated with the alternative allele as the effect allele in the linear regression analysis.

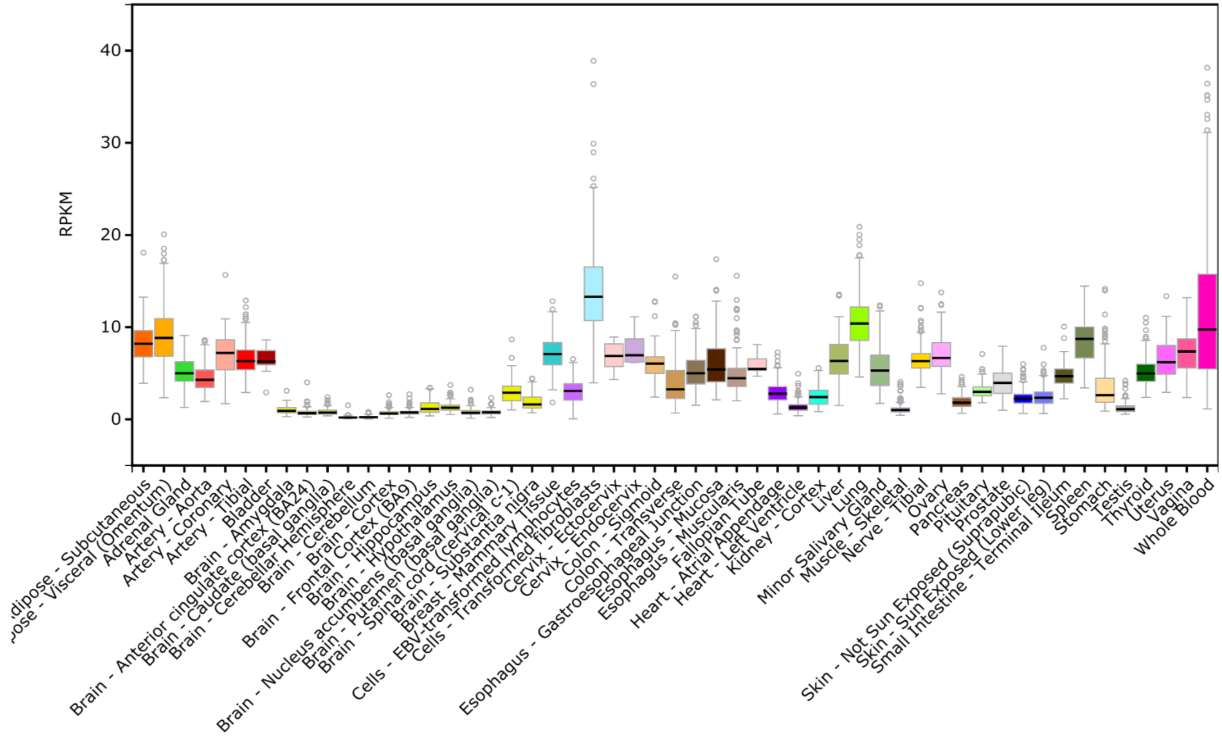

Figure S1 The mRNA expression levels of DPYD across normal tissues from the Genotype-Tissue Expression (GTEx) project (Version 6)

DPYS Gene Expression

Figure S2 The mRNA expression levels of *DPYS* across normal tissues from the Genotype-Tissue Expression (GTEx) project.

UPB1 Gene Expression

Figure S3 The mRNA expression levels of UPB1 across normal tissues from the Genotype-Tissue Expression (GTEx) project.

B. Supplementary information for Chapter IV

Table S1 The genetic markers identified in Chapter III and the association result of each variant with the risk of early-onset severe 5-FU associated toxicity events in the BioVU cohort

Gene	SNP identified in Chapter III	surrogate SNP	Allele	Minor	MAF	HWE.*P*-value	Beta*	se	*P*-value
Approach 1									
DPYD	rs11165845		G/A	A	0.28	0.54	0.25	0.18	0.16
DPYD	rs12032384		T/C	C	0.21	0.71	0.23	0.19	0.23
DPYD	rs1415683		T/G	G	0.21	0.81	0.24	0.19	0.20
DPYD	rs72728442		A/G	G	0.17	0.48	0.17	0.21	0.42
DPYD	rs72728443		A/T	T	0.17	0.48	0.17	0.21	0.42
DPYD	rs1356919		A/T	T	0.17	0.48	0.17	0.21	0.42
DPYD	rs116772342		T/C	C	0.03	0.31	0.66	0.40	0.10
DPYD	rs56038477	rs116566349	C/T	T	0.02	0.30	0.62	0.42	0.14
DPYD	rs78944474	rs114170368	C/T	T	0.02	0.28	0.43	0.43	0.32
DPYD	rs78593303		G/T	T	0.02	0.28	0.78	0.41	0.06
DPYD	rs114806143		T/G	G	0.00	1.00	NA	NA	NA
DPYD	rs115358442		T/G	G	0.01	1.00	NA	NA	NA
UPB1	rs1892721		T/C	C	0.17	0.66	0.34	0.20	0.08
UPB1	rs5996713	rs12159862	T/C	C	0.17	0.66	0.34	0.20	0.08
UPB1	rs5996712		T/C	C	0.17	0.66	0.34	0.20	0.08
UPB1	rs12159862		T/C	C	0.17	0.66	0.34	0.20	0.08
Approach 2									
DPYD	rs74450569		T/C	C	0.08	0.58	0.07	0.27	0.80
DPYD	rs75570956		A/G	G	0.10	0.82	0.43	0.25	0.09
DPYS	rs78426610		G/C	C	0.03	1.00	0.30	0.44	0.49
DPYS	rs2298840	rs3750187	G/A	A	0.21	0.62	-0.06	0.20	0.77
DPYS	rs3793354		G/A	A	0.06	0.73	-0.57	0.37	0.12
DPYS	rs13274374	rs36090760	C/T	T	0.03	1.00	0.35	0.44	0.43

UPB1	rs2070475	Fail to genotype			-	-	NA	NA	NA
UPB1	rs2070474		G/C	C	0.43	1.00	0.29	0.16	0.07
UPB1	rs2032116	rs4822500	C/T	T	0.43	0.93	0.39	0.16	0.01
UPB1	rs2298383		T/C	C	0.47	0.32	0.35	0.16	0.03
UPB1	rs5760447		G/A	A	0.42	0.87	0.37	0.16	0.02
UPB1	rs131455		T/C	C	0.01	1.00	-0.66	1.09	0.54
UPB1	rs116907149	rs62233071	T/G	G	0.01	1.00	-0.82	1.08	0.44
UPB1	rs62234044	rs62231882	C/T	T	0.01	1.00	-0.82	1.08	0.44
UPB1	rs77804922		G/A	A	0.01	1.00	-0.81	1.08	0.45
Established markers									
DPYD	rs3918290	rs189653741	T/T	-	-	-	NA	NA	NA
DPYD	rs55886062	rs192732997	A/A	-	-	-	NA	NA	NA
DPYD	rs67376798	Fail to genotype			-	-			
DPYD	rs75017182		G/C	C	0.03	0.31	0.66	0.40	0.10
Previously reported common variants for 5-FU associated toxicity									
DPYD	rs12132152 [21]		G/A	A	0.03	0.08	0.21	0.47	0.66
DPYD	rs12022243 [21]	rs12040763	T/C	C	0.17	0.56	0.17	0.21	0.42
DPYS	rs2959023 [14]	rs11783979	C/A	A	0.43	0.32	-0.03	0.16	0.86
UPB1	rs2070474 [156]		G/C	C	0.43	1.00	0.29	0.16	0.07

*beta, se and *P* value were derived from the unconditional logistic regression with the risk of early-onset severe 5-FU associated toxicity events as the outcome.

BIOVU Programming

Study Population: All Colorectal Cancer Patients in the tumor registry receiving 5-FU based chemotherapy with non-comprised DNA samples available.

Include:

1. All CRC cancer patients available in **tumor registry**, identified by the following codes

ICD-O-3 codes: C180, C181, C182, C183, C184, C185, C186, C187, C188, C189, C199, C209

2. **Chemotherapy use** using billing codes or clinic notes
   ICD-9 procedure code: 9925
   ICD-9-CM codes: E0781, E9331, V58.1, v58.11, V66.2, or V67.2
   ICD-10 procedure code: 3E03305, 3E04305

   Or

   ICD-10-CM codes: Z51.11,, Z08

   Or

   CPT codes: 96400–96549, J8510, J8520, J8521, J8530–J8999, J9000–J9999, or Q0083–Q0085

   Or

   Chemo, chemotherapy, adjuvant chemotherapy, 5FU, 5-Fu, 5-FU, leucovorin by the deGramont LV5FU2, 5FU+LV, 5-FU + leucovorin, 5FU + leucovorin, FOLFOX, FOLFRI, oxaliplatin, capecitabine, XEOLDA, cape or CAPE, appeared at least once in the clinic notes

Define *baseline* as t_0 = the date of first diagnosis of colorectal cancer

Variables extracted at first diagnosis t_0

1. All variables included in **tumor registry**
2. **Height , weight at each time t (height_0 = height at t0; …); continuous variables, in m, kg, and kg/m2**

Adjuvant chemotherapy after diagnosis

A. Create time variables Date of each cycle of chemotherapy (chemotherapy_date1= date of first cycle of chemotherapy…. chemotherapy_date12, etc.)

At **each cycle** create the following variables: (i = 1, 2, 3, 4, .the number of last cycle) the maximum of I is usually 12.

A1. Height, weight, or BMI at each time t (height_1 = height at chemotherapy_date1; weight_1=weight at chemotherapy_date1 …); continuous variables, in m, kg, and kg/m2

A2. Create a set of variables: ECOG_PS_i: ECOG performance status (0, 1, 2, or 3), this can be extracted from clinic narratives, most of the time abbreviation was used: ps or PS (ECOG_PS_1 AT chemotherapy cycle 1…)

A3. Regimen of the chemotherapy create a set of variables regimen_i (categorical variables, description of the regimen used)

Extract from clinic narratives or CPT Codes:
5FU, or 5-Fu, 5-FU/leucovorin by the deGramont (LV5FU2), 5FU+LV, 5-FU + leucovorin, 5FU + leucovorin, FOLFOX, FOLFRI, oxaliplatin, capecitabine, XEOLDA, cape or CAPE, Aflibercep, bevacizumab, cetuximab,
Avastin or avastin, Gemcitabine or gemcitabline, Irinotecan or irinotecan,Rituximabm, erbitux ,IMO-2055, panitumumab

A4. CEA level reported at each cycle

Create a set of variables: CEA_i

A5. Blood test result at first cycle of chemo: baseline_wbc_i (i = the number of this cycle), baseline_ Hgb _i, baseline_ PCV _i, baseline_ Plt-Ct _i, baseline_ Neut _i, baseline_ NTauto _i, baseline_ Lym _i, baseline_ Monocy_i, baseline_ Eos_i, baseline_ Baso_i,

Note: Blood test components

WBC, Hgb , PCV, Plt-Ct , Neut, NTAuto, Lym , Monocy, Eos , Baso

B. Toxicities check

Create time variables: Date of each toxicity check (toxicitycheck_date1= date of first check…. toxicitychek_date12, etc.);
At **each check, create series of variables for blood test result for each blood cell count at each check:** toxicitycheck_wbc_i (i = the number of this cycle), toxicitycheck_ Hgb _i, toxicitycheck_ PCV _i, toxicitycheck_ Plt-Ct _i, toxicitycheck_ Neut _i, toxicitycheck_ NTauto _i, toxicitycheck_ Lym _i, toxicitycheck_ Monocy_i, toxicitycheck_ Eos_i, toxicitycheck_ Baso_i,

Death

Create a variable: date _of_death, extract from tumor registry if available.

Date of last follow-up

Create a variable: date _of_lastfollowup, extract from tumor registry if available, or the last date of clinic visit to VUMC.

Other covariates

Date of birth

Sex

REFERENCES

1. Siegel, R.L., Miller, K.D. & Jemal, A. Cancer statistics, 2015. *CA Cancer J Clin* **65**, 5-29 (2015).
2. Siegel, R., Desantis, C. & Jemal, A. Colorectal cancer statistics, 2014. *CA Cancer J Clin* **64**, 104-17 (2014).
3. Bradley, C. *et al.* Trends in chemotherapy-related treatment of advanced-stage colorectal (CRC) cancer. *Journal of Clinical Oncology* **31**(2013).
4. Rosmarin, D. *et al.* Genetic Markers of Toxicity From Capecitabine and Other Fluorouracil-Based Regimens: Investigation in the QUASAR2 Study, Systematic Review, and Meta-Analysis. *J Clin Oncol* **32**, 1031-9 (2014).
5. Cortejoso, L. & Lopez-Fernandez, L.A. Pharmacogenetic markers of toxicity for chemotherapy in colorectal cancer patients. *Pharmacogenomics* **13**, 1173-91 (2012).
6. Tuchman, M. *et al.* Familial pyrimidinemia and pyrimidinuria associated with severe fluorouracil toxicity. *N Engl J Med* **313**, 245-9 (1985).
7. Diasio, R.B., Beavers, T.L. & Carpenter, J.T. Familial deficiency of dihydropyrimidine dehydrogenase. Biochemical basis for familial pyrimidinemia and severe 5-fluorouracil-induced toxicity. *J Clin Invest* **81**, 47-51 (1988).
8. Meulendijks, D. *et al.* Clinical relevance of DPYD variants c.1679T>G, c.1236G>A/HapB3, and c.1601G>A as predictors of severe fluoropyrimidine-associated toxicity: a systematic review and meta-analysis of individual patient data. *Lancet Oncol* **16**, 1639-50 (2015).
9. Terrazzino, S. *et al.* DPYD IVS14+1G>A and 2846A>T genotyping for the prediction of severe fluoropyrimidine-related toxicity: a meta-analysis. *Pharmacogenomics* **14**, 1255-72 (2013).
10. Yang, X. *et al.* Systematic genetic and genomic analysis of cytochrome P450 enzyme activities in human liver. *Genome Res* **20**, 1020-36 (2010).
11. Wang, D. *et al.* Common CYP2D6 polymorphisms affecting alternative splicing and transcription: long-range haplotypes with two regulatory variants modulate CYP2D6 activity. *Hum Mol Genet* **23**, 268-78 (2014).
12. Wang, D., Papp, A.C. & Sun, X. Functional characterization of CYP2D6 enhancer polymorphisms. *Hum Mol Genet* **24**, 1556-62 (2015).
13. Meyer, U.A. Pharmacogenetics and adverse drug reactions. *Lancet* **356**, 1667-71 (2000).
14. Kummer, D. *et al.* Dihydropyrimidinase and beta-ureidopropionase gene variation and severe fluoropyrimidine-related toxicity. *Pharmacogenomics* **16**, 1367-1377 (2015).
15. Fidlerova, J., Kleiblova, P., Kormunda, S., Novotny, J. & Kleibl, Z. Contribution of the beta-ureidopropionase (UPB1) gene alterations to the development of fluoropyrimidine-related toxicity. *Pharmacological Reports* **64**, 1234-1242 (2012).
16. Hiratsuka, M. *et al.* Genetic Polymorphisms of Dihydropyrimidinase in a Japanese Patient with Capecitabine-Induced Toxicity. *Plos One* **10**(2015).
17. Gusev, A. *et al.* Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am J Hum Genet* **95**, 535-52 (2014).
18. Eng, C. Toxic effects and their management: daily clinical challenges in the treatment of colorectal cancer. *Nat Rev Clin Oncol* **6**, 207-18 (2009).
19. Stein, B.N. *et al.* Age and sex are independent predictors of 5-fluorouracil toxicity. Analysis of a large scale phase III trial. *Cancer* **75**, 11-7 (1995).
20. Meta-Analysis Group In, C. *et al.* Toxicity of fluorouracil in patients with advanced colorectal cancer: effect of administration schedule and prognostic factors. *J Clin Oncol* **16**, 3537-41 (1998).
21. Rosmarin, D. *et al.* A candidate gene study of capecitabine-related toxicity in colorectal cancer identifies new toxicity variants at DPYD and a putative role for ENOSF1 rather than TYMS. *Gut* **64**, 111-20 (2015).
22. Albert, F.W. & Kruglyak, L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet* **16**, 197-212 (2015).

23. Heggie, G.D., Sommadossi, J.P., Cross, D.S., Huster, W.J. & Diasio, R.B. Clinical pharmacokinetics of 5-fluorouracil and its metabolites in plasma, urine, and bile. *Cancer Res* **47**, 2203-6 (1987).
24. Torre, L.A. *et al.* Global cancer statistics, 2012. *CA Cancer J Clin* **65**, 87-108 (2015).
25. Arnold, M. *et al.* Global patterns and trends in colorectal cancer incidence and mortality. *Gut* (2016).
26. Siegel, R.L., Miller, K.D. & Jemal, A. Cancer statistics, 2016. *CA Cancer J Clin* **66**, 7-30 (2016).
27. DeSantis, C.E. *et al.* Cancer statistics for African Americans, 2016: Progress and opportunities in reducing racial disparities. *CA Cancer J Clin* **66**, 290-308 (2016).
28. Robbins, A.S., Siegel, R.L. & Jemal, A. Racial disparities in stage-specific colorectal cancer mortality rates from 1985 to 2008. *J Clin Oncol* **30**, 401-5 (2012).
29. Zeng, C. *et al.* Disparities by Race, Age, and Sex in the Improvement of Survival for Major Cancers: Results From the National Cancer Institute Surveillance, Epidemiology, and End Results (SEER) Program in the United States, 1990 to 2010. *JAMA Oncol* **1**, 88-96 (2015).
30. Keating, N.L. *et al.* Adjuvant chemotherapy for stage III colon cancer: Do physicians agree about the importance of patient age and comorbidity? *Journal of Clinical Oncology* **26**, 2532-2537 (2008).
31. Sanoff, H.K. *et al.* Effect of adjuvant chemotherapy on survival of patients with stage III colon cancer diagnosed after age 75 years. *J Clin Oncol* **30**, 2624-34 (2012).
32. Sargent, D.J. *et al.* A pooled analysis of adjuvant chemotherapy for resected colon cancer in elderly patients. *N Engl J Med* **345**, 1091-7 (2001).
33. Compton, C., Fenoglio-Preiser, C.M., Pettigrew, N. & Fielding, L.P. American Joint Committee on Cancer prognostic factors consensus conference - Colorectal working group. *Cancer* **88**, 1739-1757 (2000).
34. Wanebo, H.J. *et al.* Preoperative Carcinoembryonic Antigen Level as a Prognostic Indicator in Colorectal Cancer. *New England Journal of Medicine* **299**, 448-451 (1978).
35. Erlichman, C. *et al.* Efficacy of adjuvant fluorouracil and folinic acid in B2 colon cancer. *Journal of Clinical Oncology* **17**, 1356-1363 (1999).
36. Tournigand, C. *et al.* Adjuvant Therapy With Fluorouracil and Oxaliplatin in Stage II and Elderly Patients (between ages 70 and 75 years) With Colon Cancer: Subgroup Analyses of the Multicenter International Study of Oxaliplatin, Fluorouracil, and Leucovorin in the Adjuvant Treatment of Colon Cancer Trial. *Journal of Clinical Oncology* **30**, 3353-3360 (2012).
37. Gray, R. *et al.* Adjuvant chemotherapy versus observation in patients with colorectal cancer: a randomised study. *Lancet* **370**, 2020-2029 (2007).
38. Moertel, C.G. *et al.* Intergroup Study of Fluorouracil Plus Levamisole as Adjuvant Therapy for Stage-Ii Dukes B2 Colon-Cancer. *Journal of Clinical Oncology* **13**, 2936-2943 (1995).
39. O'Connor, E.S. *et al.* Adjuvant Chemotherapy for Stage II Colon Cancer With Poor Prognostic Features. *Journal of Clinical Oncology* **29**, 3381-3388 (2011).
40. Tomlinson, J.S. *et al.* Actual 10-year survival after resection of colorectal liver metastases defines cure. *J Clin Oncol* **25**, 4575-80 (2007).
41. Scoggins, C.R., Meszoely, I.M., Blanke, C.D., Beauchamp, R.D. & Leach, S.D. Nonoperative management of primary colorectal cancer in patients with stage IV disease. *Ann Surg Oncol* **6**, 651-7 (1999).
42. Adam, R. Chemotherapy and surgery: new perspectives on the treatment of unresectable liver metastases. *Ann Oncol* **14 Suppl 2**, ii13-6 (2003).
43. Heidelberger, C. *et al.* Fluorinated pyrimidines, a new class of tumour-inhibitory compounds. *Nature* **179**, 663-6 (1957).
44. Wolmark, N. *et al.* Postoperative adjuvant chemotherapy or BCG for colon cancer: results from NSABP protocol C-01. *J Natl Cancer Inst* **80**, 30-6 (1988).
45. Mini, E., Trave, F., Rustum, Y.M. & Bertino, J.R. Enhancement of the Antitumor Effects of 5-Fluorouracil by Folinic Acid. *Pharmacology & Therapeutics* **47**, 1-19 (1990).

46. Wolmark, N. *et al.* The Benefit of Leucovorin-Modulated Fluorouracil as Postoperative Adjuvant Therapy for Primary Colon-Cancer - Results from National Surgical Adjuvant Breast and Bowel Project Protocol C-03. *Journal of Clinical Oncology* **11**, 1879-1887 (1993).
47. Rodel, C. Efficacy and toxicity of intravenous continuous infusion of fluorouracil compared with bolus administration in advanced colorectal cancer. *Strahlentherapie Und Onkologie* **175**, 294-295 (1999).
48. Wolmark, N. *et al.* Efficacy of intravenous continuous infusion of fluorouracil compared with bolus administration in advanced colorectal cancer. *Journal of Clinical Oncology* **16**, 301-308 (1998).
49. Raymond, E., Faivre, S., Woynarowski, J.M. & Chaney, S.G. Oxaliplatin: Mechanism of action and antineoplastic activity. *Seminars in Oncology* **25**, 4-12 (1998).
50. Andre, T. *et al.* Improved Overall Survival With Oxaliplatin, Fluorouracil, and Leucovorin As Adjuvant Treatment in Stage II or III Colon Cancer in the MOSAIC Trial. *Journal of Clinical Oncology* **27**, 3109-3116 (2009).
51. Kuebler, J.P. *et al.* Oxaliplatin combined with weekly bolus fluorouracil and leucovorin as surgical adjuvant chemotherapy for stage II and III colon cancer: Results from NSABP C-07. *Journal of Clinical Oncology* **25**, 2198-2204 (2007).
52. Sanoff, H.K. *et al.* Comparative Effectiveness of Oxaliplatin vs Non-Oxaliplatin-containing Adjuvant Chemotherapy for Stage III Colon Cancer. *Jnci-Journal of the National Cancer Institute* **104**, 211-227 (2012).
53. Poon, M.A. *et al.* Biochemical Modulation of Fluorouracil with Leucovorin - Confirmatory Evidence of Improved Therapeutic Efficacy in Advanced Colorectal-Cancer. *Journal of Clinical Oncology* **9**, 1967-1972 (1991).
54. de Gramont, A. *et al.* Leucovorin and fluorouracil with or without oxaliplatin as first-line treatment in advanced colorectal cancer. *J Clin Oncol* **18**, 2938-47 (2000).
55. Ducreux, M. *et al.* Sequential versus combination chemotherapy for the treatment of advanced colorectal cancer (FFCD 2000-05): an open-label, randomised, phase 3 trial. *Lancet Oncol* **12**, 1032-44 (2011).
56. Giacchetti, S. *et al.* Phase III multicenter randomized trial of oxaliplatin added to chronomodulated fluorouracil-leucovorin as first-line treatment of metastatic colorectal cancer. *Journal of Clinical Oncology* **18**, 136-147 (2000).
57. Abrams, T.A. *et al.* Chemotherapy Usage Patterns in a US-Wide Cohort of Patients With Metastatic Colorectal Cancer. *Jnci-Journal of the National Cancer Institute* **106**(2014).
58. Douillard, J.Y. *et al.* Irinotecan combined with fluorouracil compared with fluorouracil alone as first-line treatment for metastatic colorectal cancer: a multicentre randomised trial. *Lancet* **355**, 1041-1047 (2000).
59. Saltz, L.B. *et al.* Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. *New England Journal of Medicine* **343**, 905-914 (2000).
60. Hurwitz, H.I. *et al.* Efficacy and Safety of Bevacizumab in Metastatic Colorectal Cancer: Pooled Analysis From Seven Randomized Controlled Trials. *Oncologist* **18**, 1004-1012 (2013).
61. Tabernero, J. *et al.* Aflibercept versus placebo in combination with fluorouracil, leucovorin and irinotecan in the treatment of previously treated metastatic colorectal cancer: Prespecified subgroup analyses from the VELOUR trial. *European Journal of Cancer* **50**, 320-331 (2014).
62. Vaccaro, M. *et al.* Efficacy, tolerability and impact on quality of life of clindamycin phosphate and benzoyl peroxide for the treatment of cetuximab-associated acneiform eruption in patients with metastatic colorectal cancer. *Journal of Dermatological Treatment* **27**, 148-152 (2016).
63. Maughan, T.S. *et al.* Addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. *Lancet* **377**, 2103-2114 (2011).

64. Skelton, L.A., Ormerod, M.G., Titley, J.C. & Jackman, A.L. Cell cycle effects of CB30865, a lipophilic quinazoline-based analogue of the antifolate thymidylate synthase inhibitor ICI 198583 with an undefined mechanism of action. *Cytometry* **33**, 56-66 (1998).
65. Calvert, A.H. *et al.* A Phase-I Evaluation of the Quinazoline Antifolate Thymidylate Synthase Inhibitor, N-10-Propargyl-5,8-Dideazafolic Acid, Cb3717. *Journal of Clinical Oncology* **4**, 1245-1252 (1986).
66. Wilson, K.S. & Taylor, S.C.M. Raltitrexed: optimism and reality. *Expert Opinion on Drug Metabolism & Toxicology* **5**, 1447-1454 (2009).
67. Peterson, D.E., Bensadoun, R.J., Roila, F. & Grp, E.G.W. Management of oral and gastrointestinal mucositis: ESMO Clinical Recommendations. *Annals of Oncology* **20**, 174-177 (2009).
68. Benson, A.B. *et al.* Recommended guidelines for the treatment of cancer treatment-induced diarrhea. *Journal of Clinical Oncology* **22**, 2918-2926 (2004).
69. Braun, M.S. & Seymour, M.T. Balancing the efficacy and toxicity of chemotherapy in colorectal cancer. *Ther Adv Med Oncol* **3**, 43-52 (2011).
70. Walko, C.M. & Lindley, C. Capecitabine: a review. *Clin Ther* **27**, 23-44 (2005).
71. Cassidy, J. *et al.* First-line oral capecitabine therapy in metastatic colorectal cancer: a favorable safety profile compared with intravenous 5-fluorouracil/leucovorin. *Annals of Oncology* **13**, 566-575 (2002).
72. Layoun, M.E., Wickramasinghe, C.D., Peralta, M.V. & Yang, E.H. Fluoropyrimidine-Induced Cardiotoxicity: Manifestations, Mechanisms, and Management. *Curr Oncol Rep* **18**, 35 (2016).
73. Focaccetti, C. *et al.* Effects of 5-fluorouracil on morphology, cell cycle, proliferation, apoptosis, autophagy and ROS production in endothelial cells and cardiomyocytes. *PLoS One* **10**, e0115686 (2015).
74. Polk, A., Vistisen, K., Vaage-Nilsen, M. & Nielsen, D.L. A systematic review of the pathophysiology of 5-fluorouracil-induced cardiotoxicity. *BMC Pharmacol Toxicol* **15**, 47 (2014).
75. Saif, M.W., Shah, M.M. & Shah, A.R. Fluoropyrimidine-associated cardiotoxicity: revisited. *Expert Opin Drug Saf* **8**, 191-202 (2009).
76. Moore, D.H., Fowler, W.C., Jr. & Crumpler, L.S. 5-Fluorouracil neurotoxicity. *Gynecol Oncol* **36**, 152-4 (1990).
77. Pirzada, N.A., Ali, II & Dafer, R.M. Fluorouracil-induced neurotoxicity. *Ann Pharmacother* **34**, 35-8 (2000).
78. Lukaschek, J. *et al.* Cardiotoxicity and neurotoxicity of high-dose continuous fluorouracil as a result of degradation compounds in the drug vials. *J Clin Oncol* **22**, 5022-5 (2004).
79. Davis, S.T., Joyner, S.S., Baccanari, D.P. & Spector, T. 5-Ethynyluracil (776C85): protection from 5-fluorouracil-induced neurotoxicity in dogs. *Biochem Pharmacol* **48**, 233-6 (1994).
80. Saam, J. *et al.* Body Surface Area-based Dosing of 5-Fluoruracil Results in Extensive Interindividual Variability in 5-Fluorouracil Exposure in Colorectal Cancer Patients on FOLFOX Regimens. *Clinical Colorectal Cancer* **10**, 203-206 (2011).
81. Rutman, R.J., Cantarow, A. & Paschkis, K.E. The catabolism of uracil in vivo and in vitro. *J Biol Chem* **210**, 321-9 (1954).
82. Longley, D.B., Harkin, D.P. & Johnston, P.G. 5-Fluorouracil: Mechanisms of action and clinical strategies. *Nature Reviews Cancer* **3**, 330-338 (2003).
83. Heidelberger, C. Fluorinated pyrimidines. *Prog Nucleic Acid Res Mol Biol* **4**, 1-50 (1965).
84. Diasio, R.B. & Harris, B.E. Clinical-Pharmacology of 5-Fluorouracil. *Clinical Pharmacokinetics* **16**, 215-237 (1989).
85. Heggie, G.D., Sommadossi, J.P., Cross, D.S., Huster, W.J. & Diasio, R.B. Clinical Pharmacokinetics of 5-Fluorouracil and Its Metabolites in Plasma, Urine, and Bile. *Cancer Research* **47**, 2203-2206 (1987).

86. Heggie, G.D., Sommadossi, J.P., Cross, D.A., Huster, W.J. & Diasio, R.B. Catabolism of 5-Fluorouracil (Fu) in Man - Evaluation of Dihydrofluorouracil (Fuh2) Formation at Different Doses of Fu with Evidence of a Novel Metabolite in Bile. *Proceedings of the American Association for Cancer Research* **27**, 174-174 (1986).
87. Malet-Martino, M. & Martino, R. Clinical studies of three oral prodrugs of 5-fluorouracil (capecitabine, UFT, S-1): a review. *Oncologist* **7**, 288-323 (2002).
88. Sommadossi, J.P., Cross, D.S., Goldman, I.D. & Diasio, R.B. Modulation of 5-Fluorouracil (Fu) Catabolism (Cat) in Liver-Cells by Thymine (T) and Uracil (U) - Therapeutic Implications and Evidence for Induced Formation of a Novel Fu Metabolite. *Proceedings of the American Association for Cancer Research* **24**, 302-302 (1983).
89. Casale, F. *et al.* Plasma concentrations of 5-fluorouracil and its metabolites in colon cancer patients. *Pharmacological Research* **50**, 173-179 (2004).
90. Ito, S. *et al.* Physiologically based pharmacokinetic modelling of the three-step metabolism of pyrimidine using C-13-uracil as an in vivo probe. *British Journal of Clinical Pharmacology* **60**, 584-593 (2005).
91. Miura, K. *et al.* 5-fu metabolism in cancer and orally-administrable 5-fu drugs. *Cancers (Basel)* **2**, 1717-30 (2010).
92. Noordhuis, P. *et al.* 5-fluorouracil incorporation into RNA and DNA in relation to thymidylate synthase inhibition of human colorectal cancers. *Annals of Oncology* **15**, 1025-1032 (2004).
93. Lenz, H.J., Manno, D.J., Danenberg, K.D. & Danenberg, P.V. Incorporation of 5-fluorouracil into U2 and U6 snRNA inhibits mRNA precursor splicing. *J Biol Chem* **269**, 31962-8 (1994).
94. Santi, D.V. & Hardy, L.W. Catalytic mechanism and inhibition of tRNA (uracil-5-)methyltransferase: evidence for covalent catalysis. *Biochemistry* **26**, 8599-606 (1987).
95. Katsumata, K. *et al.* Correlation between clinicopathologic factors and kinetics of metabolic enzymes for 5-fluorouracil given to patients with colon carcinoma by two different dosage regimens. *Cancer Chemotherapy and Pharmacology* **51**, 155-160 (2003).
96. Chazal, M. *et al.* Link between dihydropyrimidine dehydrogenase activity in peripheral blood mononuclear cells and liver. *Clinical Cancer Research* **2**, 507-510 (1996).
97. Saif, M.W., Choma, A., Salamone, S.J. & Chu, E. Pharmacokinetically Guided Dose Adjustment of 5-Fluorouracil: A Rational Approach to Improving Therapeutic Outcomes. *Journal of the National Cancer Institute* **101**, 1543-1552 (2009).
98. Zalcberg, J., Kerr, D., Seymour, L. & Palmer, M. Haematological and non-haematological toxicity after 5-fluorouracil and leucovorin in patients with advanced colorectal cancer is significantly associated with gender, increasing age and cycle number. Tomudex International Study Group. *Eur J Cancer* **34**, 1871-5 (1998).
99. Meulendijks, D. *et al.* Renal function, body surface area, and age are associated with risk of early-onset fluoropyrimidine-associated toxicity in patients treated with capecitabine-based anticancer regimens in daily clinical care. *European Journal of Cancer* **54**, 120-130 (2016).
100. Schwab, M. *et al.* Role of genetic and nongenetic factors for fluorouracil treatment-related severe toxicity: a prospective clinical trial by the German 5-FU Toxicity Study Group. *J Clin Oncol* **26**, 2131-8 (2008).
101. Cen, P., Liu, C. & Du, X.L. Comparison of toxicity profiles of fluorouracil versus oxaliplatin regimens in a large population-based cohort of elderly patients with colorectal cancer. *Annals of Oncology* **23**, 1503-1511 (2012).
102. Boige, V. *et al.* DPYD Genotyping to Predict Adverse Events Following Treatment With Flourouracil-Based Adjuvant Chemotherapy in Patients With Stage III Colon Cancer: A Secondary Analysis of the PETACC-8 Randomized Clinical Trial. *JAMA Oncol* (2016).
103. Lee, A.M. *et al.* DPYD Variants as Predictors of 5-fluorouracil Toxicity in Adjuvant Colon Cancer Treatment (NCCTG N0147). *Jnci-Journal of the National Cancer Institute* **106**(2014).
104. Denham, J.W. *et al.* Causes for increased myelosuppression with increasing age in patients with oesophageal cancer treated by chemoradiotherapy. *Eur J Cancer* **35**, 921-7 (1999).

105. Louie, S.G. *et al.* Higher capecitabine AUC in elderly patients with advanced colorectal cancer (SWOGS0030). *British Journal of Cancer* **109**, 1744-1749 (2013).
106. Port, R.E., Daniel, B., Ding, R.W. & Herrmann, R. Relative importance of dose, body surface area, sex, and age for 5-fluorouracil clearance. *Oncology* **48**, 277-81 (1991).
107. Klotz, U. Pharmacokinetics and drug metabolism in the elderly. *Drug Metab Rev* **41**, 67-76 (2009).
108. Lu, Z., Zhang, R. & Diasio, R.B. Dihydropyrimidine dehydrogenase activity in human peripheral blood mononuclear cells and liver: population characteristics, newly identified deficient patients, and clinical implication in 5-fluorouracil chemotherapy. *Cancer Res* **53**, 5433-8 (1993).
109. Duffour, J. *et al.* Inter-observer agreement between dermatologists and oncologists in assessing dermatological toxicities in patients with metastatic colorectal cancer treated by cetuximab-based chemotherapies: A pilot comparative study. *European Journal of Cancer* **46**, 3169-3174 (2010).
110. Mueller, F. *et al.* Gender-specific elimination of continuous-infusional 5-fluorouracil in patients with gastrointestinal malignancies: results from a prospective population pharmacokinetic study. *Cancer Chemotherapy and Pharmacology* **71**, 361-370 (2013).
111. Daher Abdi, Z. *et al.* Pharmacokinetics and exposure-effect relationships of capecitabine in elderly patients with breast or colorectal cancer. *Cancer Chemother Pharmacol* **73**, 1285-93 (2014).
112. Lichtman, S.M. *et al.* International Society of Geriatric Oncology Chemotherapy Taskforce: evaluation of chemotherapy in older patients--an analysis of the medical literature. *J Clin Oncol* **25**, 1832-43 (2007).
113. Dees, E.C. *et al.* A prospective pharmacologic evaluation of age-related toxicity of adjuvant chemotherapy in women with breast cancer. *Cancer Investigation* **18**, 521-529 (2000).
114. Sloan, J.A. *et al.* Women experience greater toxicity with fluorouracil-based chemotherapy for colorectal cancer. *J Clin Oncol* **20**, 1491-8 (2002).
115. Hamza, S. *et al.* Toxicity of Oxaliplatin Plus Fluorouracil/Leucovorin Adjuvant Chemotherapy in Elderly Patients with Stage III Colon Cancer: A Population-Based Study. *Annals of Surgical Oncology* **21**, 2636-2641 (2014).
116. Garg, M.B. *et al.* Predicting 5-fluorouracil toxicity in colorectal cancer patients from peripheral blood cell telomere length: a multivariate analysis. *British Journal of Cancer* **107**, 1525-1533 (2012).
117. Chansky, K., Benedetti, J. & Macdonald, J.S. Differences in toxicity between men and women treated with 5-fluorouracil therapy for colorectal carcinoma. *Cancer* **103**, 1165-1171 (2005).
118. Milano, G. *et al.* Influence of sex and age on fluorouracil clearance. *J Clin Oncol* **10**, 1171-5 (1992).
119. Gusella, M. *et al.* Pharmacokinetic and demographic markers of 5-fluorouracil toxicity in 181 patients on adjuvant therapy for colorectal cancer. *Ann Oncol* **17**, 1656-60 (2006).
120. Etienne, M.C. *et al.* Population Study of Dihydropyrimidine Dehydrogenase in Cancer-Patients. *Journal of Clinical Oncology* **12**, 2248-2253 (1994).
121. Lu, Z., Zhang, R., Carpenter, J.T. & Diasio, R.B. Decreased dihydropyrimidine dehydrogenase activity in a population of patients with breast cancer: implication for 5-fluorouracil-based chemotherapy. *Clin Cancer Res* **4**, 325-9 (1998).
122. Mattison, L.K. *et al.* Increased prevalence of dihydropyrimidine dehydrogenase deficiency in African-Americans compared with Caucasians. *Clin Cancer Res* **12**, 5491-5 (2006).
123. Prado, C.M.M. *et al.* Body composition as an independent determinant of 5-fluorouracil-based chemotherapy toxicity. *Clinical Cancer Research* **13**, 3264-3268 (2007).
124. Garrett, W.S. Cancer and the microbiota. *Science* **348**, 80-6 (2015).
125. McCollum, A.D. *et al.* Outcomes and toxicity in african-american and caucasian patients in a randomized adjuvant chemotherapy trial for colon cancer. *J Natl Cancer Inst* **94**, 1160-7 (2002).

126. Sanoff, H.K., Sargent, D.J., Green, E.M., McLeod, H.L. & Goldberg, R.M. Racial differences in advanced colorectal cancer outcomes and pharmacogenetics: a subgroup analysis of a large randomized clinical trial. *J Clin Oncol* **27**, 4109-15 (2009).
127. Offer, S.M. *et al.* A DPYD Variant (Y186C) in Individuals of African Ancestry Is Associated With Reduced DPD Enzyme Activity. *Clinical Pharmacology & Therapeutics* **94**, 158-166 (2013).
128. Meyerhardt, J.A. *et al.* Influence of body mass index on outcomes and treatment-related toxicity in patients with colon carcinoma. *Cancer* **98**, 484-95 (2003).
129. Meyerhardt, J.A. *et al.* Impact of body mass index on outcomes and treatment-related toxicity in patients with stage II and III rectal cancer: findings from Intergroup Trial 0114. *J Clin Oncol* **22**, 648-57 (2004).
130. Farker, K. *et al.* Chronomodulated chemotherapy with oxaliplatin, 5-FU and sodium folinate in metastatic gastrointestinal cancer patients: original analysis of non-hematological toxicity and patient characteristics in a pilot investigation. *Int J Clin Pharmacol Ther* **44**, 31-7 (2006).
131. Chambers, P., Daniels, S.H., Thompson, L.C. & Stephens, R.J. Chemotherapy dose reductions in obese patients with colorectal cancer. *Ann Oncol* **23**, 748-53 (2012).
132. Ali, R. *et al.* Lean body mass as an independent determinant of dose-limiting toxicity and neuropathy in patients with colon cancer treated with FOLFOX regimens. *Cancer Med* **5**, 607-16 (2016).
133. Gusella, M., Toso, S., Ferrazzi, E., Ferrari, M. & Padrini, R. Relationships between body composition parameters and fluorouracil pharmacokinetics. *British Journal of Clinical Pharmacology* **54**, 131-139 (2002).
134. Capitain, O. *et al.* The influence of fluorouracil outcome parameters on tolerance and efficacy in patients with advanced colorectal cancer. *Pharmacogenomics Journal* **8**, 256-267 (2008).
135. Sargent, D.J. *et al.* Pooled Safety and Efficacy Analysis Examining the Effect of Performance Status on Outcomes in Nine First-Line Treatment Trials Using Individual Data From Patients With Metastatic Colorectal Cancer. *Journal of Clinical Oncology* **27**, 1948-1955 (2009).
136. Levy, E. *et al.* Toxicity of fluorouracil in patients with advanced colorectal cancer: Effect of administration schedule and prognostic factors. *Journal of Clinical Oncology* **16**, 3537-3541 (1998).
137. Patel, K. *et al.* Weekly 5-fluorouracil and leucovorin: achieving lower toxicity with higher dose-intensity in adjuvant chemotherapy after colorectal cancer resection. *Annals of Oncology* **15**, 568-573 (2004).
138. Chagpar, R. *et al.* Adherence to stage-specific treatment guidelines for patients with colon cancer. *J Clin Oncol* **30**, 972-9 (2012).
139. Benson, A.B. *et al.* Rectal Cancer, Version 2.2015 Featured Updates to the NCCN Guidelines. *Journal of the National Comprehensive Cancer Network* **13**, 719-728 (2015).
140. Benson, A.B. *et al.* Colon Cancer, Version 3.2014. *Journal of the National Comprehensive Cancer Network* **12**, 1028-1059 (2014).
141. Watters, J.W., Kraja, A., Meucci, M.A., Province, M.A. & McLeod, H.L. Genome-wide discovery of loci influencing chemotherapy cytotoxicity. *Proc Natl Acad Sci U S A* **101**, 11809-14 (2004).
142. Lee, A., Ezzeldin, H., Fourie, J. & Diasio, R. Dihydropyrimidine dehydrogenase deficiency: impact of pharmacogenetics on 5-fluorouracil therapy. *Clin Adv Hematol Oncol* **2**, 527-32 (2004).
143. Offer, S.M. *et al.* Comparative Functional Analysis of DPYD Variants of Potential Clinical Relevance to Dihydropyrimidine Dehydrogenase Activity. *Cancer Research* **74**, 2545-2554 (2014).
144. Offer, S.M., Wegner, N.J., Fossum, C., Wang, K.S. & Diasio, R.B. Phenotypic Profiling of DPYD Variations Relevant to 5-Fluorouracil Sensitivity Using Real-time Cellular Analysis and In Vitro Measurement of Enzyme Activity. *Cancer Research* **73**, 1958-1968 (2013).

145. Rosmarin, D. *et al.* Genetic Markers of Toxicity From Capecitabine and Other Fluorouracil-Based Regimens: Investigation in the QUASAR2 Study, Systematic Review, and Meta-Analysis. *Journal of Clinical Oncology* **32**, 1031-1039 (2014).
146. Rosmarin, D. *et al.* A candidate gene study of capecitabine-related toxicity in colorectal cancer identifies new toxicity variants at DPYD and a putative role for ENOSF1 rather than TYMS. *Gut* **64**, 111-120 (2015).
147. Mattison, L.K., Johnson, M.R. & Diasio, R.B. A comparative analysis of translated dihydropyrimidine dehydrogenase cDNA; conservation of functional domains and relevance to genetic polymorphisms. *Pharmacogenetics* **12**, 133-144 (2002).
148. Meulendijks, D. *et al.* Clinical relevance of DPYD variants c.1679T > G, c.1236G > A/HapB3, and c.1601G > A as predictors of severe fluoropyrimidine-associated toxicity: a systematic review and meta-analysis of individual patient data. *Lancet Oncology* **16**, 1639-1650 (2015).
149. Nie, Q. *et al.* Quantitative Contribution of rs75017182 to Dihydropyrimidine Dehydrogenase mRNA Splicing and Enzyme Activity. *Clin Pharmacol Ther* **102**, 662-670 (2017).
150. van Kuilenburg, A.B. Dihydropyrimidine dehydrogenase and the efficacy and toxicity of 5-fluorouracil. *Eur J Cancer* **40**, 939-50 (2004).
151. Thomas, H.R. *et al.* Genetic regulation of dihydropyrimidinase and its possible implication in altered uracil catabolism. *Pharmacogenetics and Genomics* **17**, 973-987 (2007).
152. Lee, J., Pitt, J., Meijer, J. & van Kuilenburg, A.B.P. beta-ureidopropionase deficiency presenting with congenital anomalies of the urogenital and colorectal systems. *Journal of Inherited Metabolic Disease* **30**, 136-136 (2007).
153. Sumi, S. *et al.* Population and family studies of dihydropyrimidinuria: prevalence, inheritance mode, and risk of fluorouracil toxicity. *Am J Med Genet* **78**, 336-40 (1998).
154. van Kuilenburg, A.B. *et al.* Dihydropyrimidinase deficiency and severe 5-fluorouracil toxicity. *Clin Cancer Res* **9**, 4363-7 (2003).
155. Muhale, F.A., Wetmore, B.A., Thomas, R.S. & McLeod, H.L. Systems pharmacology assessment of the 5-fluorouracil pathway. *Pharmacogenomics* **12**, 341-350 (2011).
156. Fidlerova, J. *et al.* Contribution of dihydropyrimidinase gene alterations to the development of serious toxicity in fluoropyrimidine-treated cancer patients. *Cancer Chemother Pharmacol* **65**, 661-9 (2010).
157. Loganayagam, A. *et al.* Pharmacogenetic variants in the DPYD, TYMS, CDA and MTHFR genes are clinically significant predictors of fluoropyrimidine toxicity. *Br J Cancer* **108**, 2505-15 (2013).
158. Kawakami, K., Omura, K., Kanehira, E. & Watanabe, Y. Polymorphic tandem repeats in the thymidylate synthase gene is associated with its protein expression in human gastrointestinal cancers. *Anticancer Research* **19**, 3249-3252 (1999).
159. Horie, N., Aiba, H., Oguro, K., Hojo, H. & Takeishi, K. Functional analysis and DNA polymorphism of the tandemly repeated sequences in the 5'-terminal regulatory region of the human gene for thymidylate synthase. *Cell Struct Funct* **20**, 191-7 (1995).
160. Mandola, M.V. *et al.* A novel single nucleotide polymorphism within the 5 ' tandem repeat polymorphism of the Thymidylate synthase gene abolishes USF-1 binding and alters transcriptional activity. *Cancer Research* **63**, 2898-2904 (2003).
161. Balboa-Beltran, E., Cruz, R., Carracedo, A. & Barros, F. Delimiting Allelic Imbalance of TYMS by Allele-Specific Analysis. *Medicine* **94**(2015).
162. Mazzuca, F. *et al.* Pre-treatment evaluation of 5-fluorouracil degradation rate: association of poor and ultra-rapid metabolism with severe toxicity in a colorectal cancer patients cohort. *Oncotarget* **7**, 20612-20 (2016).
163. Thomas, F. *et al.* Methylenetetrahydrofolate reductase genetic polymorphisms and toxicity to 5-FU-based chemoradiation in rectal cancer. *British Journal of Cancer* **105**, 1654-1662 (2011).
164. Glimelius, B. *et al.* Prediction of irinotecan and 5-fluorouracil toxicity and response in patients with advanced colorectal cancer. *Pharmacogenomics Journal* **11**, 61-71 (2011).

165. Gusella, M. *et al.* Predictors of survival and toxicity in patients on adjuvant therapy with 5-fluorouracil for colorectal cancer. *British Journal of Cancer* **100**, 1549-1557 (2009).
166. Gonzalez-Haba, E. *et al.* ABCB1 gene polymorphisms are associated with adverse reactions in fluoropyrimidine-treated colorectal cancer patients. *Pharmacogenomics* **11**, 1715-1723 (2010).
167. Ni, X., Zhang, W. & Huang, R.S. Pharmacogenomics discovery and implementation in genome-wide association studies era. *Wiley Interdiscip Rev Syst Biol Med* **5**, 1-9 (2013).
168. Motsinger-Reif, A.A. *et al.* Genome-wide association studies in pharmacogenomics: successes and lessons. *Pharmacogenet Genomics* **23**, 383-94 (2013).
169. Daly, A.K. Genome-wide association studies in pharmacogenomics. *Nat Rev Genet* **11**, 241-6 (2010).
170. Fernandez-Rozadilla, C. *et al.* Pharmacogenomics in colorectal cancer: a genome-wide association study to predict toxicity after 5-fluorouracil or FOLFOX administration. *Pharmacogenomics J* **13**, 209-17 (2013).
171. Westra, H.J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nature Genetics* **45**, 1238-U195 (2013).
172. Low, S.K. *et al.* Genome-wide association study of chemotherapeutic agent-induced severe neutropenia/leucopenia for patients in Biobank Japan. *Cancer Sci* **104**, 1074-82 (2013).
173. Nica, A.C. *et al.* Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet* **6**, e1000895 (2010).
174. Claussnitzer, M. *et al.* Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. *Cell* **156**, 343-58 (2014).
175. Whalen, S., Truty, R.M. & Pollard, K.S. Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat Genet* **48**, 488-96 (2016).
176. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455-61 (2014).
177. Phillips-Cremins, J.E. & Corces, V.G. Chromatin insulators: linking genome organization to cellular function. *Mol Cell* **50**, 461-74 (2013).
178. Li, Y., Chen, C.Y., Kaye, A.M. & Wasserman, W.W. The identification of cis-regulatory elements: A review from a machine learning perspective. *Biosystems* **138**, 6-17 (2015).
179. Sexton, T. & Cavalli, G. The role of chromosome domains in shaping the functional genome. *Cell* **160**, 1049-59 (2015).
180. Mathelier, A., Shi, W. & Wasserman, W.W. Identification of altered cis-regulatory elements in human disease. *Trends Genet* **31**, 67-76 (2015).
181. Khurana, E. *et al.* Role of non-coding sequence variants in cancer. *Nat Rev Genet* **17**, 93-108 (2016).
182. Pai, A.A., Pritchard, J.K. & Gilad, Y. The Genetic and Mechanistic Basis for Variation in Gene Regulation. *Plos Genetics* **11**(2015).
183. Battle, A. *et al.* Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res* **24**, 14-24 (2014).
184. Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nature Genetics* **44**, 1084-+ (2012).
185. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506-11 (2013).
186. Maurano, M.T. *et al.* Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo (vol 47, pg 1393, 2015). *Nature Genetics* **47**(2015).
187. Buenrostro, J.D., Wu, B., Chang, H.Y. & Greenleaf, W.J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**, 21 29 1-9 (2015).
188. Kumasaka, N., Knights, A.J. & Gaffney, D.J. Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nat Genet* **48**, 206-13 (2016).
189. Kasowski, M. *et al.* Extensive Variation in Chromatin States Across Humans. *Science* **342**, 750-752 (2013).

190. McVicker, G. *et al.* Identification of Genetic Variants That Affect Histone Modifications in Human Cells. *Science* **342**, 747-749 (2013).
191. Kilpinen, H. *et al.* Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science* **342**, 744-7 (2013).
192. Jolma, A. *et al.* DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature* **527**, 384-+ (2015).
193. Kasowski, M. *et al.* Variation in Transcription Factor Binding Among Humans. *Science* **328**, 232-235 (2010).
194. Moyerbrailean, G.A. *et al.* Which Genetics Variants in DNase-Seq Footprints Are More Likely to Alter Binding? *Plos Genetics* **12**(2016).
195. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res* **45**, D896-D901 (2017).
196. Biason, P., Masier, S. & Toffoli, G. UGT1A1*28 and other UGT1A polymorphisms as determinants of irinotecan toxicity. *J Chemother* **20**, 158-65 (2008).
197. Schulz, C. *et al.* UGT1A1 gene polymorphism: impact on toxicity and efficacy of irinotecan-based regimens in metastatic colorectal cancer. *World J Gastroenterol* **15**, 5058-66 (2009).
198. Wang, D. *et al.* Regulatory polymorphism in vitamin K epoxide reductase complex subunit 1 (VKORC1) affects gene expression and warfarin dose requirement. *Blood* **112**, 1013-21 (2008).
199. Grover, S. & Kukreti, R. A systematic review and meta-analysis of the role of ABCC2 variants on drug response in patients with epilepsy. *Epilepsia* **54**, 936-45 (2013).
200. Yee, S.W. *et al.* Identification and characterization of proximal promoter polymorphisms in the human concentrative nucleoside transporter 2 (SLC28A2). *J Pharmacol Exp Ther* **328**, 699-707 (2009).
201. Talwar, P. *et al.* Genetic contribution of CYP1A1 variant on treatment outcome in epilepsy patients: a functional and interethnic perspective. *Pharmacogenomics J* (2016).
202. Beedanagari, S.R. *et al.* Role of epigenetic mechanisms in differential regulation of the dioxin-inducible human CYP1A1 and CYP1B1 genes. *Mol Pharmacol* **78**, 608-16 (2010).
203. Shadley, J.D. *et al.* Identification and functional analysis of a novel human CYP2E1 far upstream enhancer. *Mol Pharmacol* **71**, 1630-9 (2007).
204. Wang, H. *et al.* A novel distal enhancer module regulated by pregnane X receptor/constitutive androstane receptor is essential for the maximal induction of CYP2B6 gene expression. *J Biol Chem* **278**, 14146-52 (2003).
205. Sanford, J.C., Guo, Y., Sadee, W. & Wang, D. Regulatory polymorphisms in CYP2C19 affecting hepatic expression. *Drug Metabol Drug Interact* **28**, 23-30 (2013).
206. Matsumura, K. *et al.* Identification of a novel polymorphic enhancer of the human CYP3A4 gene. *Mol Pharmacol* **65**, 326-34 (2004).
207. Kim, M.J. *et al.* Functional characterization of liver enhancers that regulate drug-associated transporters. *Clin Pharmacol Ther* **89**, 571-8 (2011).
208. Nelson, A.C. & Wardle, F.C. Conserved non-coding elements and cis regulation: actions speak louder than words. *Development* **140**, 1385-1395 (2013).
209. Consortium, E.P. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* **9**, e1001046 (2011).
210. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).
211. Gamazon, E.R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nature Genetics* **47**, 1091-+ (2015).
212. Tibshirani, R. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society Series B-Methodological* **58**, 267-288 (1996).
213. Tibshirani, R. Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **73**, 273-282 (2011).

214. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **67**, 301-320 (2005).
215. Gerstein, M.B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91-100 (2012).
216. Rosenbloom, K.R. *et al.* ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res* **41**, D56-63 (2013).
217. Pope, B.D. *et al.* Topologically associating domains are stable units of replication-timing regulation. *Nature* **515**, 402-5 (2014).
218. Keen, J.C. & Moore, H.M. The Genotype-Tissue Expression (GTEx) Project: Linking Clinical Data with Molecular Analysis to Advance Personalized Medicine. *J Pers Med* **5**, 22-9 (2015).
219. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc* **7**, 500-7 (2012).
220. Heather E Wheeler, K.P.S., Jonathon Brenner, Tzintzuni Garcia, Keston Aquino-Michaels, GTEx Consortium, Nancy J Cox, Dan L Nicolae, Hae Kyung Im. Survey of the Heritability and Sparsity of Gene Expression Traits Across Human Tissues. *biorxiv* (2016).
221. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* **33**, 1-22 (2010).
222. Farh, K.K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337-43 (2015).
223. Varshney, A. *et al.* Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proc Natl Acad Sci U S A* **114**, 2301-2306 (2017).
224. Huang, H. *et al.* Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature* **547**, 173-178 (2017).
225. Mathelier, A. *et al.* JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **44**, D110-5 (2016).
226. Wingender, E. The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. *Brief Bioinform* **9**, 326-32 (2008).
227. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-89 (2010).
228. Wang, J. *et al.* Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res* **22**, 1798-812 (2012).
229. Kheradpour, P. & Kellis, M. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Res* **42**, 2976-87 (2014).
230. Kulakovskiy, I.V. *et al.* HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. *Nucleic Acids Res* **41**, D195-202 (2013).
231. Jolma, A. *et al.* DNA-binding specificities of human transcription factors. *Cell* **152**, 327-39 (2013).
232. Coetzee, S.G., Coetzee, G.A. & Hazelett, D.J. motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics* **31**, 3847-9 (2015).
233. Thurman, R.E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75-82 (2012).
234. Jansen, R. *et al.* Conditional eQTL analysis reveals allelic heterogeneity of gene expression. *Hum Mol Genet* **26**, 1444-1451 (2017).
235. Gamazon, E.R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* **47**, 1091-8 (2015).
236. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48**, 245-52 (2016).
237. Huang, X., Stern, D.F. & Zhao, H. Transcriptional Profiles from Paired Normal Samples Offer Complementary Information on Cancer Patient Survival--Evidence from TCGA Pan-Cancer Data. *Sci Rep* **6**, 20567 (2016).

238. Siegel, R.L. *et al.* Colorectal cancer statistics, 2017. *CA Cancer J Clin* **67**, 177-193 (2017).
239. Siegel, R.L., Miller, K.D. & Jemal, A. Cancer Statistics, 2017. *CA Cancer J Clin* **67**, 7-30 (2017).
240. Mitry, E. *et al.* Adjuvant chemotherapy after potentially curative resection of metastases from colorectal cancer: a pooled analysis of two randomized trials. *J Clin Oncol* **26**, 4906-11 (2008).
241. Schmoll, H.J. *et al.* Capecitabine Plus Oxaliplatin Compared With Fluorouracil/Folinic Acid As Adjuvant Therapy for Stage III Colon Cancer: Final Results of the NO16968 Randomized Controlled Phase III Trial. *Journal of Clinical Oncology* **33**, 3733-+ (2015).
242. Yothers, G. *et al.* Oxaliplatin As Adjuvant Therapy for Colon Cancer: Updated Results of NSABP C-07 Trial, Including Survival and Subset Analyses. *Journal of Clinical Oncology* **29**, 3768-3774 (2011).
243. Rodel, C. *et al.* Oxaliplatin added to fluorouracil-based preoperative chemoradiotherapy and postoperative chemotherapy of locally advanced rectal cancer (the German CAO/ARO/AIO-04 study): final results of the multicentre, open-label, randomised, phase 3 trial. *Lancet Oncol* **16**, 979-89 (2015).
244. Deenen, M.J. *et al.* Upfront Genotyping of DPYD*2A to Individualize Fluoropyrimidine Therapy: A Safety and Cost Analysis. *J Clin Oncol* **34**, 227-34 (2016).
245. Ritchie, M.D. *et al.* Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. *Am J Hum Genet* **86**, 560-72 (2010).
246. Pulley, J., Clayton, E., Bernard, G.R., Roden, D.M. & Masys, D.R. Principles of human subjects protections applied in an opt-out, de-identified biobank. *Clin Transl Sci* **3**, 42-8 (2010).
247. Garrett, W.S. *et al.* Colitis-associated colorectal cancer driven by T-bet deficiency in dendritic cells. *Cancer Cell* **16**, 208-19 (2009).
248. Bao, F., LoVerso, P.R., Fisk, J.N., Zhurkin, V.B. & Cui, F. p53 binding sites in normal and cancer cells are characterized by distinct chromatin context. *Cell Cycle*, 0 (2017).
249. Gokare, P. *et al.* P53 represses pyrimidine catabolic gene dihydropyrimidine dehydrogenase (DPYD) expression in response to thymidylate synthase (TS) targeting. *Sci Rep* **7**, 9711 (2017).
250. Sai, K. *et al.* Development of a detection algorithm for statin-induced myopathy using electronic medical records. *Journal of Clinical Pharmacy and Therapeutics* **38**, 230-235 (2013).