

2018

Analyzing Genetic Variants Related to 5-FU and Capecitabine Toxicity

Justin Kritzinger

Western University, jkritzinger24@gmail.com

Follow this and additional works at: https://ir.lib.uwo.ca/undergradawards_2018



Part of the [Medical Sciences Commons](#)

Recommended Citation

Kritzinger, Justin, "Analyzing Genetic Variants Related to 5-FU and Capecitabine Toxicity" (2018). *2018 Undergraduate Awards*. 11.
https://ir.lib.uwo.ca/undergradawards_2018/11

Analyzing genetic variants related to 5-FU and capecitabine toxicity

Justin Kritzing

(Dr. Richard Kim)

The Department of Physiology and Pharmacology, Schulich School of Medicine & Dentistry,
The University of Western Ontario, London, ON, Canada

Abstract

Chemotherapy regimens containing 5-fluorouracil (5-FU) or the oral pro-drug capecitabine are often used to treat colorectal cancer patients. Unfortunately, toxicity resulting from inappropriate dosing occurs in approximately 33% of patients. Currently, select polymorphisms in the dihydropyrimidine dehydrogenase (DPD) gene (*DPYD*) are used to predict the occurrence of toxicity and to guide dose reductions. As a vital enzyme involved in the metabolism of 5-FU to inactive metabolites, DPD has been the focus of studies related to 5-FU toxicity. However, patients lacking these variants still experience toxic reactions to fluoropyrimidine treatments. Here, we examined variants in 12 genes within the metabolic pathway of 5-FU and capecitabine, and investigated potentially deleterious *DPYD* polymorphisms that may contribute to toxicity. These genes include various transporters involved in the efflux of toxic fluoropyrimidines, drug target enzymes, and enzymes involved in the catabolism of 5-FU and capecitabine. Using next generation sequencing, 69 colorectal cancer patients had targeted regions sequenced within the 12 genes. Subjects were initially characterized based on *DPYD* genotypes (those containing the SNPs predicting toxicity, and those without). The cohort lacking the known *DPYD* variants were subsequently further characterized based on those who experienced adverse reactions (ARs) to therapy and those who did not. CADD, Polyphen, and SIFT *in silico* prediction tools were used to identify potentially deleterious variants. More predicted deleterious variants were identified exclusively within the AR cohort than the no reaction cohort. We propose several polymorphisms within multiple genes that could have contributed to toxicity seen within both *DPYD* genotype cohorts. In order to create a more comprehensive screening technique, it is essential to further investigate the role these

deleterious variants may have in the toxic build up of fluoropyrimidines within cells. This may help clinicians improve patient care, and result in less ARs to fluoropyrimidine based treatments.

Keywords: *DPYD*, Metabolism, Polymorphism, 5-FU, Adverse Reaction

Table of Contents

Abstract	2
Introduction	5
Materials & Methods	11
Results	14
Figures and Tables	
Table 1	15
Figure 1	16
Figure 2	17
Figure 3	18
Figure 4	20
Table 2	22
Figure 5	24
Figure 6	26
Discussion	27
Acknowledgements	33
References	34

Introduction

Colorectal cancer affects approximately 27 000 Canadians each year, resulting in roughly 9400 deaths annually, according to the Canadian Cancer Society¹. As the second most commonly diagnosed cancer in Canada, and the second and third leading causes of death from cancer in men and women respectively, this disease exerts a great stress on patients and the health care system¹. Treatment for colorectal cancer typically includes a combination of surgery, radiation, and chemotherapy, which can vary depending on the patient's case². The current method of dosing chemotherapy drugs is based on body surface area. Due to the high degree of variation among patients with regard to their response to treatment, a personalized approach using genetic biomarkers has recently been considered to improve care³. This approach includes screening for particular genetic variants associated with adverse reactions (ARs) prior to the initiation of chemotherapy treatments, and subsequently lowering the administered dose to an appropriate amount if certain polymorphisms are present. Several genes have been identified by our lab as potential markers based on their role in coding transporters, drug targets, or in the metabolism of two chemotherapy drugs commonly used to treat colorectal cancer, 5-fluorouracil (5-FU) and capecitabine. Both 5-FU and capecitabine act by disrupting DNA synthesis and causing cell death primarily in rapidly dividing cells, such as those found in tumours, the gastrointestinal tract, and the skin³. As a result of differences in metabolism largely attributed to varying genetic compositions among patients, toxic metabolites can accumulate leading to an AR³. Thus, by developing a more comprehensive pre-emptive screening technique, an appropriate dosing regimen can be determined based on an

individual's genetic composition. In turn, potentially fewer patients will experience toxicity from their treatment, and clinical care may be improved.

5-FU is a fluorinated pyrimidine analogue commonly used in combinational chemotherapy regimens to treat breast, lung, and colorectal cancers³. Administered as a continuous intravenous infusion, 5-FU is often combined with oxaliplatin, irinotecan, or other chemotherapy drugs used to improve the efficacy of treatment⁴. Upon injection into the blood stream, 5-FU rapidly enters cells following the same transport mechanism as uracil, largely via the SLC22a7 (OAT2) transporter⁴. More than 80% of 5-FU is catabolized within the liver to inactive metabolites by dihydropyrimidine dehydrogenase (DPD), leaving only 1-3% of the original dose to mediate cytotoxic effects on tumor cells and normal tissues^{4,5}. The catabolic pathway thus plays a critical role in determining a patient's response to 5-FU, as reduced enzymatic activity may result in an increased half-life of the drug and an increased risk of dose-dependent severe toxicity⁴. Several metabolites of 5-FU act to disrupt normal cell function in some manner, however the primary mechanism of the cytotoxic effects involves fluorodeoxyuridine monophosphate (FdUMP)⁴. This metabolite functions by inhibiting thymidylate synthase (TS), ultimately disrupting thymidine formation required for DNA synthesis⁴. By forming a ternary complex with TS and 5,10-methylene tetrahydrofolate (CH₂THF), FdUMP inhibits TS from methylating deoxyuridine monophosphate (dUMP) to form deoxythymidine monophosphate (dTMP)⁴. This complex blocks dUMP from accessing the nucleotide-binding site of TS by competitively binding at this site. Due to the fluorinated C-5 of the uridine analog, the complex will bind, yet no reaction will occur. This results in a pool

imbalance of deoxynucleotides, with an increased level of deoxyuridine triphosphate (dUTP) that ultimately disrupts DNA synthesis and repair, leading to DNA damage⁴.

In addition to FdUMP, 5-FU is converted into two other active metabolites: fluorouridine triphosphate (FUTP), which competes with uridine triphosphate (UTP) to be incorporated into RNA; and fluorodeoxyuridine triphosphate (FdUTP), which competes with deoxythymidine triphosphate (dTTP) to be incorporated into DNA⁴. Both FdUTP and FdUMP cause DNA damage while FUTP affects RNA processing and function, all of which result in cell death⁴.

Thymidylate phosphorylase (TYMP) catalyzes the conversion of 5-FU to fluorodeoxyuridine (FdUR), which is subsequently phosphorylated by thymidine kinase (TK) to FdUMP, the main active metabolite of 5-FU⁵. The other two active metabolites, FdUTP and FUTP, are generated from fluorouracil monophosphate (FUMP), a metabolite of 5-FU. FUMP is phosphorylated to fluorouracil diphosphate (FUDP), and either further phosphorylated to the active metabolite FUTP, or converted to fluorodeoxyuridine diphosphate (FdUDP) by ribonucleotide reductase⁴. FdUDP is subsequently either phosphorylated again to FdUTP, or dephosphorylated to FdUMP⁴.

Capecitabine, the oral prodrug of 5-FU, is often used as an alternative due to the benefits it provides patients. Its rapid and nearly complete absorption through the gastrointestinal wall allows for direct intestinal exposure of 5-FU to be largely avoided⁶. In addition, a continuous infusion is not required as in 5-FU therapy in order to maintain sufficient concentrations of the drug for effective treatment⁶. This is due to the prolonged release of the oral prodrug, and increased specificity of action⁶. Capecitabine is then

metabolized to 5-FU through a three-step enzymatic process mediated by carboxyl esterases (CES) 1 and 2, cytidine deaminase (CDA), and finally TYMP⁵.

As capecitabine is administered orally, the risk of thrombosis and infection are eliminated, unlike during a continuous infusion of 5-FU where the rate of such complications are reported to be as high as 20-60% with chronic venous access devices⁶. These complications pose a serious risk to cancer patients and lead to further medical expenses and time in hospital. As a result, capecitabine is often given to older patients, where the benefits of a direct infusion treatment are outweighed by the risk of adverse complications. In addition to the benefits of oral administration, capecitabine generates 5-FU preferentially within tumours due to the increased expression of TYMP within these tissues⁶. Systemic exposure to 5-FU is thereby reduced, potentially improving the efficacy and safety of the drug⁶.

Capecitabine and 5-FU treatments have a narrow therapeutic window and display significant differences in individual responses that frequently result in elevated toxicity⁷. ARs typically result from a buildup of cytotoxic fluoropyrimidine metabolites within tissues, often due to variants resulting in a loss of function within genes associated with the disposition of 5-FU and capecitabine⁴. Tissues such as tumors where cells are constantly undergoing replication and DNA synthesis are primarily affected by these chemotherapy agents⁷. Thus not only does tumor size decrease as a result of cell death, but growth is restricted by targeting cells that rapidly divide⁷. However, other rapidly dividing tissues including the gastrointestinal tract and skin, are also susceptible to damage by 5-FU⁸. Toxicities due to these treatments are primarily manifested in four common ARs: hand-foot syndrome, the reddening, swelling, and desquamation of the palms of the hands and

soles of the feet; neutropenia, an abnormally low level of neutrophils; diarrhea; and mucositis, the inflammation and ulceration of the mucous membranes lining the digestive tract^{8,9}. It has been shown that hand-foot syndrome is more likely to occur in patients treated with capecitabine, and variants in *CES1* and *CDA* were associated with this particular toxicity¹⁰. Furthermore, in a meta-analysis of 1219 colorectal cancer patients receiving 5-FU, it was reported that severe toxicity was encountered in 31-34% of patients, with 0.5% mortality, highlighting the need to examine possible genetic factors resulting in these ARs¹¹.

Currently, variation in the gene *DPYD*, encoding the protein DPD is the strongest predictor of an AR to 5-FU or capecitabine. While DPD activity has been identified within various tissues, the liver is thought to be the primary organ where 5-FU catabolism occurs¹². There is a high degree of variation in DPD function in the population, with an estimated 5% of individuals exhibiting low or deficient DPD activity¹². Patients with low DPD activity are expected to be at a substantially greater risk of experiencing severe and potentially lethal toxicity to standard doses of fluoropyrimidine treatments¹². As demonstrated in a study examining 80 patients with severe 5-FU toxicity, 71% were reported to have reduced DPD activity, suggesting DPD function plays a major role in 5-FU related adverse events¹².

In accordance with the personalized approach to treating colorectal cancer patients, the Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines have suggested three variants of *DPYD* to be screened for as markers predicting toxicity: *DPYD**2A, which results in skipping exon 14 entirely; *DPYD**13, which appears to destabilize the Flavin mononucleotide binding site; and *DPYD* rs67376798, which

substitutes an aspartic acid residue for a valine residue (D949V) , all of which result in a non-functional protein^{12,13}. However, patients lacking these variants may still experience toxicity to fluoropyrimidine treatments, indicating the possible role of novel *DPYD* variants or other deleterious mutations in proteins involved in the disposition of 5-FU and capecitabine. To further improve patient clinical care and reduce adverse events experienced during treatment, a more comprehensive screening technique should be developed. As the body-surface area based chemotherapy dosing method does not account for complex metabolic processes that can vary considerably between individuals, a personalized treatment based on genetic composition can improve clinical care substantially¹¹. By identifying novel deleterious variants in proteins within this metabolic pathway, a more comprehensive predictor of toxicity can be developed, and an appropriate dosing strategy can be implemented¹¹.

The objective of this study is to identify novel germline variants in *DPYD* and other genes within the pathway of 5-FU and capecitabine disposition that are potentially deleterious within the context of fluoropyrimidine treatments. Furthermore, we wish to investigate associations between variants in particular genes with specific types of adverse events. Previous literature has demonstrated that a polymorphism in *CDA* predicts severe capecitabine-induced hand-foot syndrome¹⁴. Our study aims to identify other polymorphisms as predictors of certain adverse events to provide further relevant information when designing a personalized treatment strategy. Due to the numerous proteins and enzymes involved in the metabolism and drug action of 5-FU and capecitabine, we hypothesize there are yet to be discovered deleterious variants that contribute to ARs resulting from these treatments.

Methods and Materials

Patient Population

69 colorectal cancer patients were selected to be genotyped by next generation sequencing. These patients were grouped based on their classification as one of three types of *DPYD* metabolizers: (1) an extensive metabolizer (EM) is classified as lacking any of the three *DPYD* variants that CPIC guidelines suggest be screened for; (2) an intermediate metabolizer (IM) is classified as being heterozygous for one of these variants; (3) and a poor metabolizer (PM) is classified as being either homozygous for one of these variants or heterozygous for two (heterozygous compound). From this population, 56 patients were considered EMs, 12 were IMs, and 1 was a PM. Clinical and demographic information were recorded for the purpose of this study. This included the patient's date of birth, sex, reason for consult, chemotherapy regimen, type of ARs if any, and whether or not they were genotyped pre-emptively or after an adverse event.

DNA Sequencing

Illumina MiSeq next generation sequencing (NGS) (Illumina, San Diego, CA) was conducted for these 69 patients, where targeted sequencing of the exons of genes involved in 5-FU and capecitabine disposition occurred. DNA samples were amplified and sequenced via the sequencing by synthesis technique. This technique of sequencing DNA involves fluorescently labeled reversible terminators that are imaged as each dNTP is added, and then cleaved to allow incorporation of the following base¹⁵. As a result, base-by-base sequencing occurs creating an accurate and reliable method of determining DNA sequences¹⁵.

Genetic Analyzation and Annotation

Subsequent to sequencing the patient's DNA, the samples were analyzed using CLC genomics program (QIAGEN, San Francisco, CA) and annotated using ANNOVAR software. Among other outputs, ANNOVAR provides information about the type of mutation, the location of the gene and where the mutation occurs, reported frequencies according to the 1000 Genomes Project, as well as various *in silico* scores that are used to predict the functional consequences of the mutation (eg. benign or deleterious)¹⁶. In particular, three scores were examined to provide a more robust classification: Combined Annotation Dependent Depletion (CADD); Sorting Intolerant From Tolerant (SIFT); and Polymorphism Phenotyping-2 (PolyPhen-2)¹⁶. Using information regarding the type of mutation (eg. synonymous or non-synonymous), location within the gene (eg. within an exon), and various other parameters, these scores provided *in silico* predictions about how potentially deleterious a variant was.

Genes of Interest

Among the 100 genes that were sequenced, 12 genes of interest were identified: five coding transporters (*SLC22A7*, *ABCG2*, *ABCC3*, *ABCC4*, and *ABCC5*); four coding enzymes involved in converting capecitabine to 5-FU within the cell (*TYMP*, *CDA*, *CES1*, and *CES2*); two coding a protein or compound forming the inhibitory complex (*MTHFR*, and *TYMS*); and *DPYD*. Previous literature has demonstrated how deleterious variants in many of these genes have been associated with the toxic build of fluoropyrimidines that can lead to ARs^{17,18,19}.

Identifying Variants

The NGS data was analyzed first by investigating the EM population. Two groups were created to compare differences in variation within the 12 genes of interest: EM

patients who had experienced an AR; and EM patients who did not. The EM patients who did not have an adverse event served as a control group. Within these two groups, all variants within the 12 genes of interest were characterized by ANNOVAR annotation¹⁶. Variants with a scaled CADD score greater than or equal to 20 were identified as being potentially deleterious. A cut off of 20 was determined arbitrarily as there is no set score to classify a variant as deleterious or not. Rather, a scaled CADD score of 20 corresponds to a variant that is amongst the top 1% of deleterious variants in the human genome²⁰. Upon identifying variants with scaled CADD scores greater than or equal to 20, a Chi-squared test and Fischer's exact test were conducted to compare allele frequencies of polymorphisms found within the EM population who experienced an AR to that reported by the 1000 Genome (Euro). In addition, variants found exclusively within the AR group within the EM population were examined. Based on the *in silico* prediction scores, prevalence of variants within the sample populations, and role of the protein within the metabolic pathway, NGS coverage of select genes was examined to provide confidence in the correct sequencing of identified variants. Finally, the variation within the EM patients who experienced an AR was compared to that of the IM/PM group. This was done as the variants contributing to toxicity found within the EM patients could potentially be contributing to toxicity found within the IM/PM patients, in addition to their *DPYD* genotype presumably being the primary cause.

Results

Cohort Summary and Workflow

The study sample was categorized based on *DPYD* genotype and further described by demographic information, treatment regimen, reason for genotyping, and type of AR that occurred (**Table 1**). The majority of subjects within the sample were categorized as *DPYD* EMs and had a mean age ($\pm$ SD) of 66.23 ($\pm$ 5.95). Capecitabine and 5-FU treatment regimens were given nearly equally within the EM cohort. In contrast, the majority of subjects within the IM/PM cohort received a 5-FU based regimen (approximately 77%). Within both EM and IM/PM cohorts, the majority of subjects were genotyped subsequent to experiencing an AR to their chemotherapy. Diarrhea was the most prevalent AR experienced among the IM/PM cohort (61.5%), whereas hand-foot syndrome was the most prevalent among the EM cohort (25%).

Table 1. Subject demographics and clinical characteristics. The sample population (n=69) was categorized based on *DPYD* genotypes into EMs, IMs, and PMs. Among the EM subgroup, hand-foot syndrome was the most prevalent AR, whereas among the IM/PM subgroup, diarrhea was the most common AR.

Characteristic	<i>DPYD</i> Genotype	
	EM	IM/PM
No. of subjects	56	13
No. of females, (%)	30 (53.6%)	8 (61.5%)
Age, mean $\pm$ SD	66.23 $\pm$ 5.95	
Treatment Regimen		
Capecitabine, (%)	20 (35.7%)	3 (23.1%)
5-FU, (%)	24 (42.9%)	10 (76.9%)
Both, (%)	3 (5.4%)	0
Unknown, (%)	9 (16.1%)	0
Reason for Genotyping		
Adverse Event, (%)	42 (75%)	6 (46.2%)
Pre-emptive genotyping, (%)	9 (16.1%)	0
Pre-emptive genotyping, (%) (subsequent adverse event)	5 (8.9%)	7 (53.8%)
Adverse Reaction		
Diarrhea, (%)	13 (23.2%)	8 (61.5%)
Neutropenia, (%)	11 (19.6%)	7 (53.8%)
Mucositis, (%)	10 (17.9%)	6 (46.2%)
Hand-Foot Syndrome, (%)	14 (25%)	5 (38.5%)

NGS data of targeted exonic regions of the 12 identified genes were analyzed based on the categorization of *DPYD* genotype (EM or IM/PM), and further analyzed based on their prevalence within EM adverse reaction (EM AR) and no adverse reaction (EM no AR) subgroups. IM and PM subjects were analyzed as one group due to the small sample size, and due to the assumption their *DPYD* genotype was the primary reason for experiencing toxicity. *In silico* prediction tools were used to classify potentially deleterious variants (**Figure 1**).

When analyzing the prevalence and *in silico* predictions of variants, potential outcomes of loss of function or decreased expression were considered based on the metabolic pathway of 5-FU and Capecitabine (**Figure 2**).

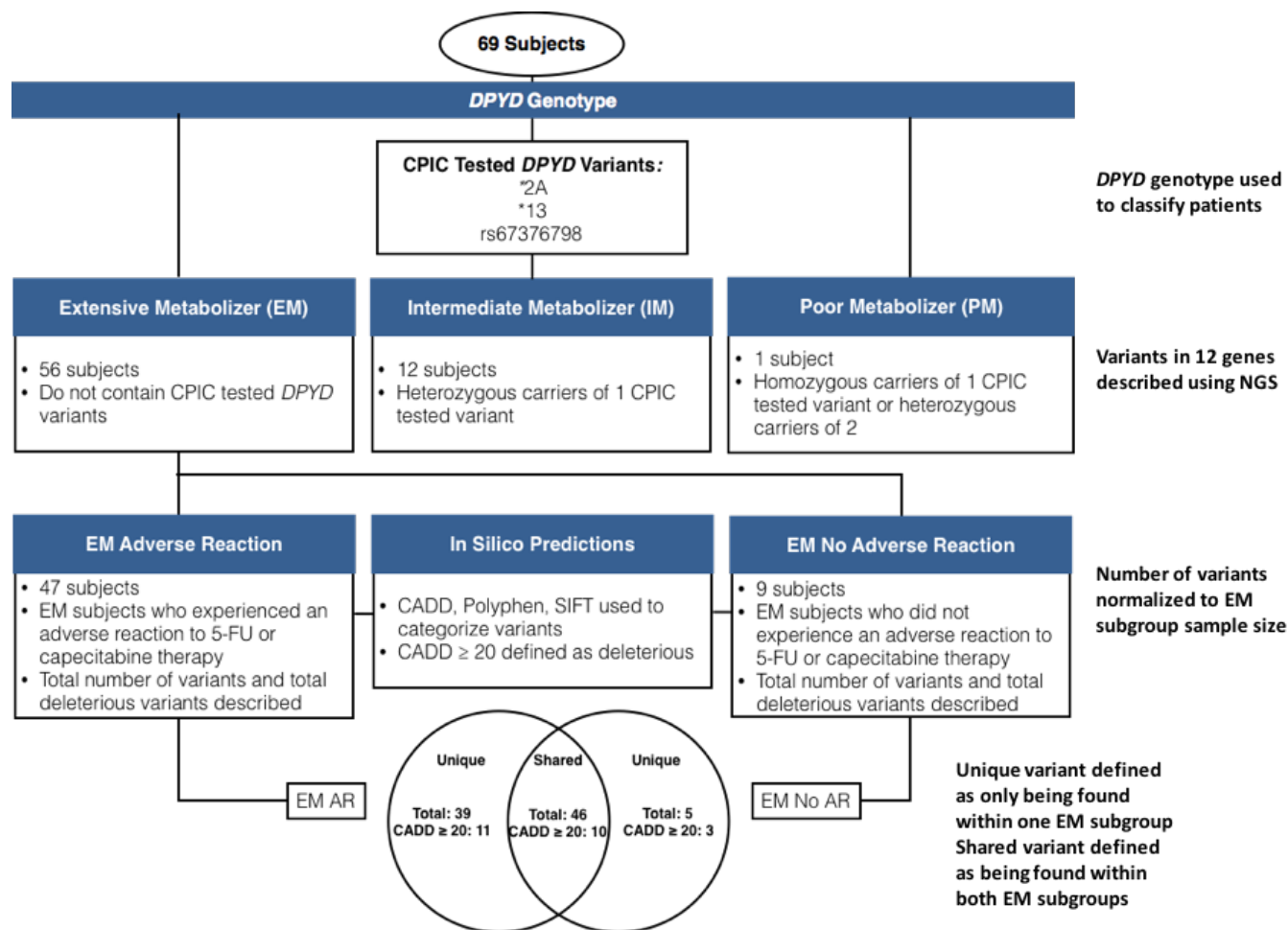


Figure 1. Work flow of subject categorization, variant analysis, and *in silico* predictions. From 56 subjects that were categorized as *DPYD* EMs, 47 subjects experienced ARs to 5-FU or capecitabine therapies, whereas 9 subjects did not experience an AR. The total number of variants, and number of variants predicted *in silico* to be deleterious (CADD score ≥ 20) were compared between EMs who experienced an AR and those who did not. Variants were also classified as unique based on whether they were found exclusive to either subgroup. IM (n=12) and PM (n=1) subjects were analyzed as one group since the sample size was limited. The number of predicted deleterious variants was compared across all 12 genes between the EM AR subgroup and IM/PM group.

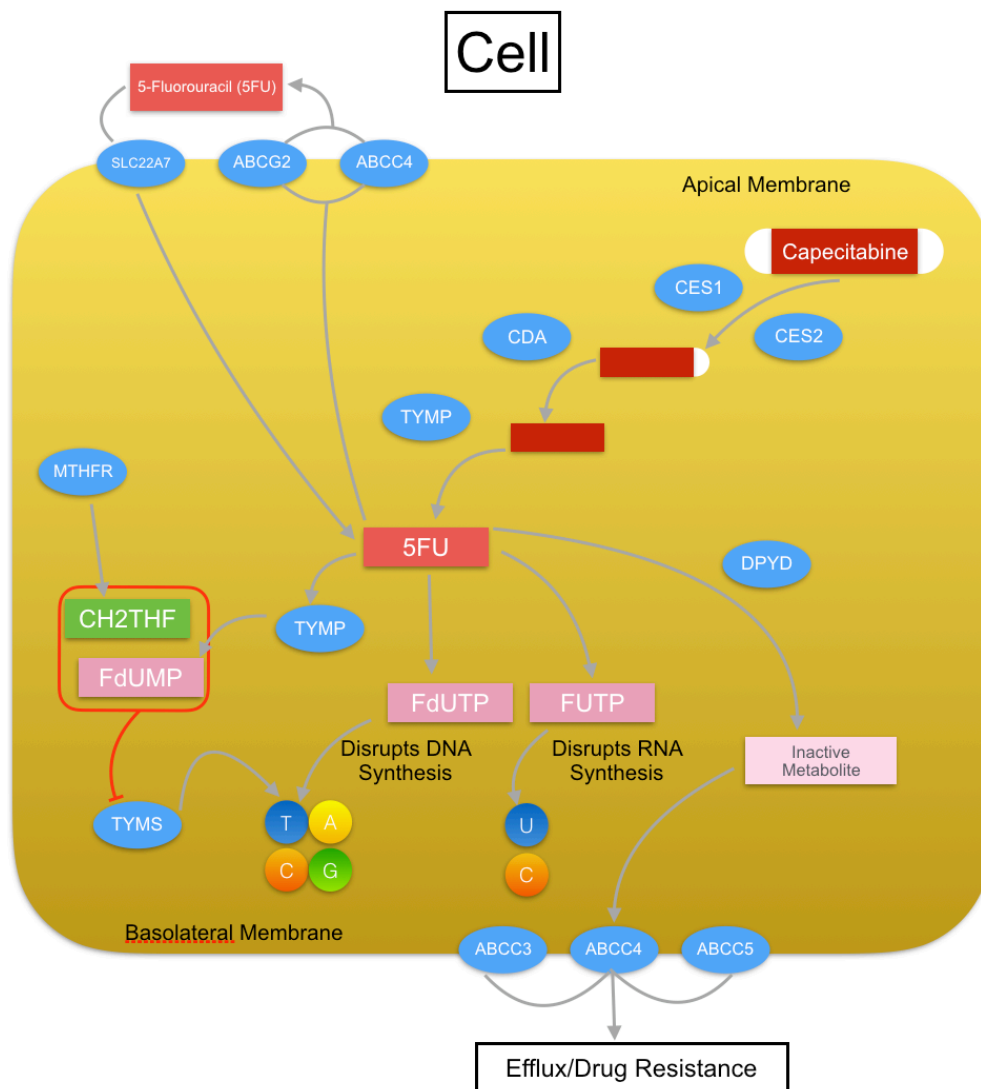


Figure 2. Overview of the capecitabine and 5-FU metabolic pathway within a cell. Transporters include ABCC3/4/5, ABCG2, and SLC22A7. Capecitabine is converted into 5-FU via a three step pathway involving CES1 and 2, CDA, and TYMP. A complex formed by 5,10-CH₂THF and FdUMP inhibits TYMS, while FdUTP and FUTP disrupt DNA and RNA synthesis, respectively. Through a three step enzymatic process beginning with DPYD, 5-FU is catabolized into inactive metabolites.

Genetic Variation Among Total Sample and Targeted Exon Length

The total number of variants, as well as the number of nonsynonymous variants within the sample population were described, with regard to the 12 genes investigated within the metabolic pathway of 5-FU and capecitabine (**Figure 3**). In addition, the length of the targeted exonic regions were characterized. Transporter genes *ABCC3* and *ABCC4*, as well as *CES1* had the greatest number of exonic variants. However, the total length of exonic regions of *CES1* was less than half of those of *ABCC3* and *ABCC4*. *ABCC3* and *CES1* had the greatest number of nonsynonymous variants (7/14) followed by *DPYD* (6/8).

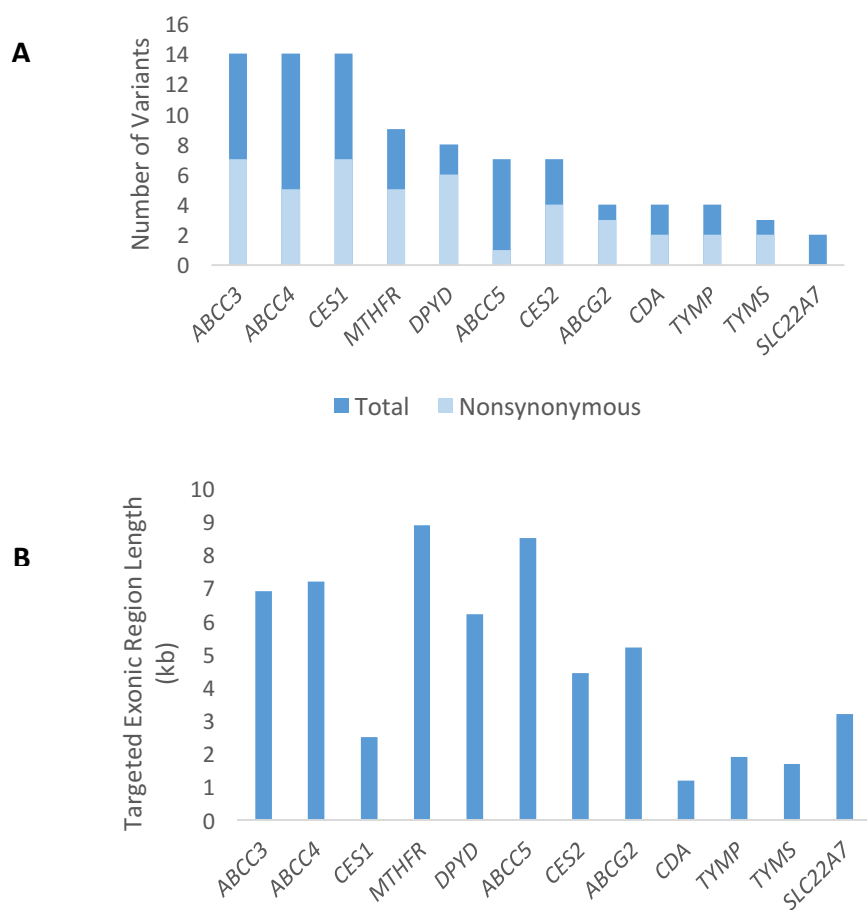


Figure 3. (A) Total number of variants (dark blue) and number of nonsynonymous variants (light blue) identified within targeted exonic regions by next generation sequencing. (B) Targeted exonic region size (kb) of the 12 genes of interest.

Transporters *ABCC3* and *ABCC4*, and esterase *CES1* were found to have the greatest number of exonic variants. The greatest number of nonsynonymous variants were found within *ABCC3* and *CES1*, and *CES1* had the fourth smallest exonic region length. *DPYD* along with *ABCG2* conveyed the greatest proportion of nonsynonymous variants.

Genetic Variation Among DPYD EM Group

Patients classified as *DPYD* EMs still experienced ARs to 5-FU or capecitabine treatment regimens. Thus to investigate the variants that were potentially contributing to toxicity among these patients, the *DPYD* EM group was divided into AR and no AR subgroups. The EM no AR subgroup had greater variation per person within all targeted genes, except for *CES2*, compared to the EM AR subgroup (**Figure 4**). However, within *ABCC3*, *ABCC4*, *CES2*, *MTHFR*, *ABCC5*, and *TYMS*, there were more variants found exclusive to the EM AR subgroup (i.e. unique), than to the EM no AR subgroup. In addition, there were more variants with CADD scores ≥ 20 and found unique to the EM AR subgroup, within *ABCC3*, *ABCC4*, *CES2*, *CES1*, and *MTHFR*, than there were in the same genes within the no AR subgroup.

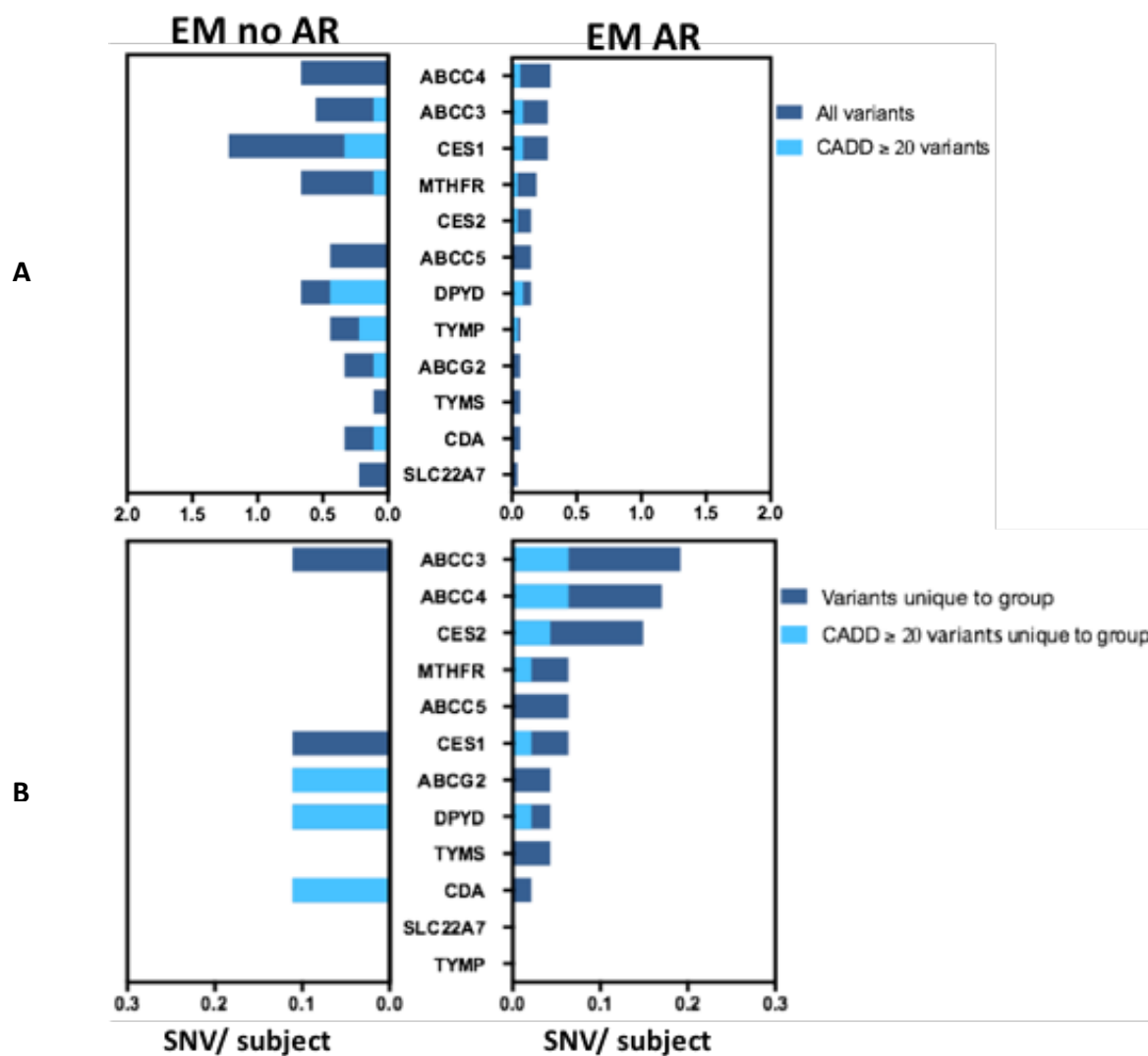


Figure 4. Left: EM no AR subgroup (n=9). Right: EM AR subgroup (n=47). (A) Total number of different single nucleotide variants (SNV) (navy) and different SNVs with CADD scores ≥ 20 (blue) normalized to the subgroup population. (B) Total number of unique SNVs (navy) and number of unique SNVs with CADD scores ≥ 20 (blue), normalized to subgroup population.

Genetic Variation Among the DPYD EM AR Subgroup

After describing the variation broadly within the 12 genes among the EM AR group, specific SNPs were characterized based on allele frequencies within the subgroup population, compared to those found by the 1000G (European). In addition, CADD, SIFT, and Polyphen scores were used to provide *in silico* predictions of whether a variant would likely be deleterious or benign (**Table 2**). Furthermore, CPIC guidelines had identified a *DPYD* variant c.1129-5923C>G, which introduces a cryptic splice site and the partial production of a nonfunctional transcript²¹. Due to the fact that this particular SNP is located within an intron, a CADD score was not assigned as this *in silico* prediction tool applies only to exonic variants. However, exonic SNP *DPYD* rs56038477 has been identified as being in perfect linkage disequilibrium with *DPYD* c.1129-5923C>G (rs75017182), and thus was used as a proxy within **Table 2**. No variant allele frequency was found to be significantly greater within the EM AR population compared to the 1000G (European) reported frequencies. However, it should be noted that several predicted deleterious variants were found exclusively within the EM AR population.

Table 2. Exonic variants found within the EM AR subgroup, predicted *in silico* as deleterious (CADD ≥ 20) or identified in literature as deleterious (*). CADD, SIFT, and Polyphen scores were used to characterize the potential of variants to be deleterious. A Chi-square and Fischer's exact test were performed to identify significant differences between variant allele frequencies within the EM AR subgroup (n=47) and the 1000G (European) reported frequencies. No allele frequency in the EM AR subgroup was significantly greater than the 1000G (European) frequency.

No. of subjects	Gene	Position	dbSNP137	1000G (Euro)	Sample frequency	p value	CADD	SIFT	Polyphen
21	<i>MTHFR</i>	1:11856378	rs1801133	0.35	0.22	0.01	25	0.06 ^a	1.00 ^c
1	<i>MTHFR</i>	1:11861223	NA	NA	0.01	NA	35	0 ^b	0.99 ^c
9	<i>DPYD</i>	1:97770920	rs1801160	0.05	0.10	0.09	25.9	0.04 ^b	0.30 ^b
2	<i>DPYD</i>	1:97981421	rs1801158	0.03	0.02	NA	23.4	0.02 ^b	0.03 ^a
8	<i>DPYD</i>	1:98165091	rs2297595	0.12	0.12	0.40	24.5	0.01 ^b	0.99 ^c
41	<i>DPYD</i>	1:98348885	rs1801265	0.78	0.57	0.00	23.7	0.18 ^a	0.00 ^a
3	<i>DPYD*</i>	1:98039419	rs56038477	0.02	0.03	0.44	14.03	NA	NA
1	<i>ABCC4</i>	13:95705380	rs11568644	0.01	0.01	NA	26.6	0.1 ^a	0.93 ^c
3	<i>ABCC4</i>	13:95859035	rs2274407	0.07	0.03	0.20	20.9	0.01 ^b	0.06 ^a
1	<i>ABCC4</i>	13:95863008	rs11568658	0.03	0.01	NA	33	0 ^b	0.89 ^c
2	<i>CES1</i>	16:55853481	rs202001817	NA	0.02	NA	25.1	0.01 ^b	0.45 ^b
1	<i>CES1</i>	16:55853545	rs115629050	0.04	0.01	NA	24	0.01 ^b	0.59 ^b
31	<i>CES1</i>	16:55862691	rs62028647	0.45	0.33	0.03	26.1	0.01 ^b	0.48 ^b
31	<i>CES1</i>	16:55862883	rs34380375	0.44	0.33	0.05	25	0 ^b	0.75 ^b
1	<i>CES2</i>	16:66971964	rs140461033	0.01	0.01	NA	23.3	0.06 ^b	0.39 ^b
2	<i>CES2</i>	16:66974546	NA	NA	0.02	NA	33	0 ^b	0.99 ^c
1	<i>ABCC3</i>	17:48736618	rs148287642	NA	0.01	NA	32	0.04 ^b	0.48 ^b
2	<i>ABCC3</i>	17:48761053	rs11568591	0.06	0.02	NA	34	0 ^b	1.00 ^c
2	<i>ABCC3</i>	17:48761367	rs141762939	NA	0.02	NA	35	0 ^b	1.00 ^c
1	<i>ABCC3</i>	17:48761397	rs11568588	NA	0.01	NA	20.9	0.12 ^a	0.01 ^a
5	<i>TYMP</i>	22:50964236	rs11479	0.08	0.05	0.43	24.1	0.34 ^a	0.68 ^b
9	<i>TYMP</i>	22:50964255	rs112723255	0.06	0.10	0.18	20.5	0.34 ^a	0.05 ^a

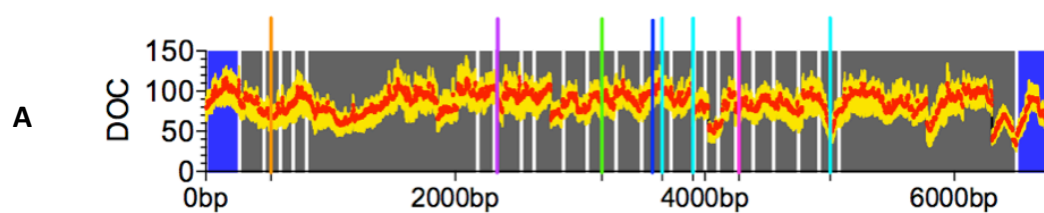
SIFT: ^a = benign ^b = deleterious

Polyphen: ^a = benign ^b = possibly damaging ^c = probably damaging

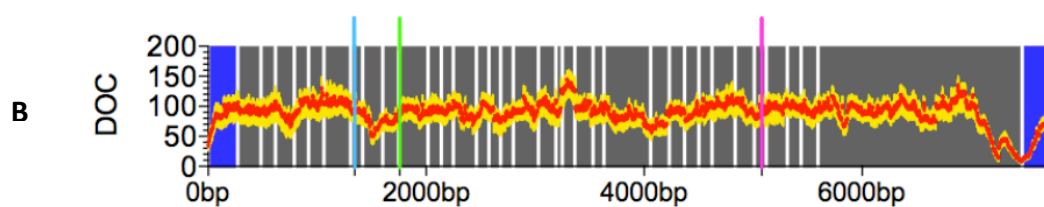
Highlighted variants not found within EM non AR subgroup

Depth of Coverage of ABCC4, ABCC3, TYMP, and DPYD

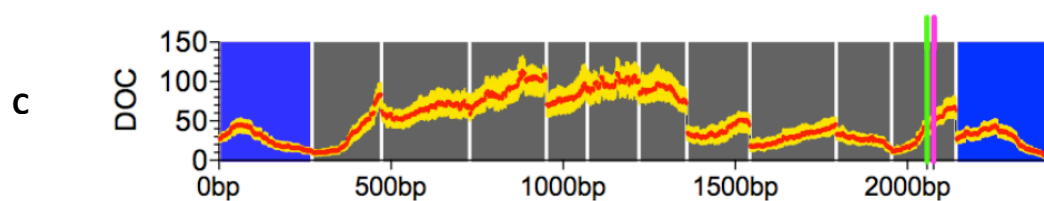
Depth of coverage (DOC)—defined as the number of reads that include a given nucleotide in the reconstructed sequence—was characterized for four genes of interest. These four genes were chosen based on the the *in silico* prediction scores of the variants, the prevalence of these deleterious variants within the EM AR subgroup, as well as the role of each protein within the metabolic pathway of 5-FU and capecitabine (**Figure 5**). The level of coverage required for a sequence to be deemed reliable within this study was set at 30x depth of coverage. Variants with CADD scores ≥ 20 found within the EM AR subgroup are highlighted, indicating their specific position within a particular exon. Based on CPIC guidelines, three *DPYD* SNPs were recommended to be screened for prior to 5-FU or capecitabine treatment, as these genotypes have been linked to toxicity²¹. Five other *DPYD* SNPs were presented on the basis of *in silico* prediction scores or clinical relevance, including *DPYD* rs56038477, which the latest CPIC guidelines suggests be screened for²¹. It should be noted that the latest guidelines became available subsequent to the initiation of this study, and thus *DPYD* rs56038477 was not included in the criteria for classifying *DPYD* IM/PM patients.

DPYD

Targeted Region

*ABCC4*

Targeted Region

*TYMP*

Targeted Region



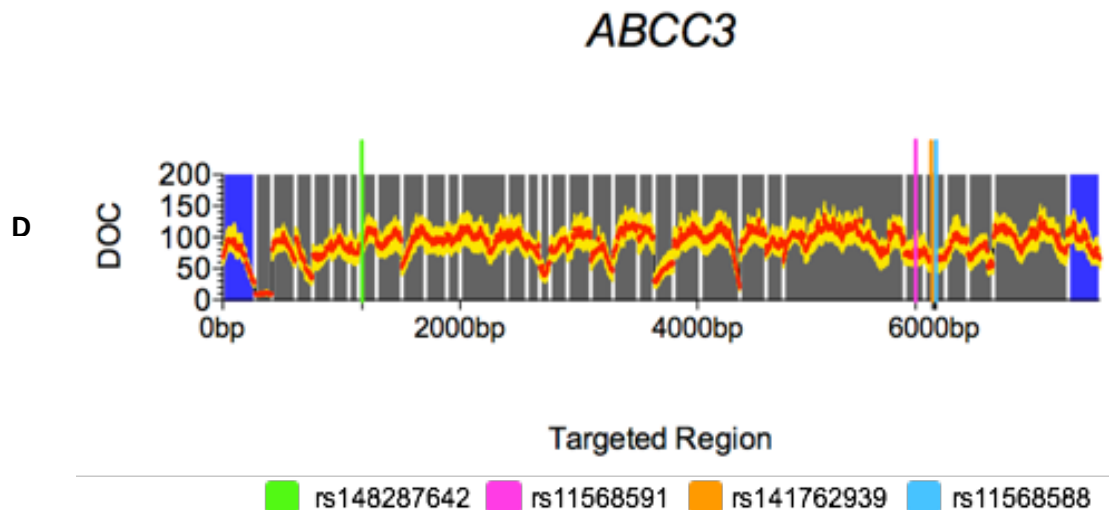


Figure 5. Median (red) and 25th to 75th percentile (yellow) of depth of coverage (DOC) of targeted exonic regions within (A) *DPYD*, (B) *ABCC4*, (C) *TYMP*, and (D) *ABCC3*. White lines indicate division between different exonic regions. Blue shaded areas indicate portions of 5' and 3' UTR sequenced on either side of exonic regions. Predicted deleterious SNPs are highlighted indicating the coverage at a particular position. *DPYD*, *ABCC4*, and *ABCC3* demonstrate reliable coverage (>30x DOC), whereas *TYMP* demonstrates reliable coverage in selective areas.

Comparing Distribution of Predicted Deleterious Variants within DPYD EM AR and IM/PM Populations

Assuming other variants within the 5-FU and capecitabine pathway contributed to the toxicity seen within the EM AR subgroup, those same variants may have been present in the IM/PM population, and thus further contributed to toxicity seen in *DPYD* IM/PM patients. The total number of variants with CADD scores ≥ 20 found within the EM AR and IM/PM population were compared (**Figure 6**). It was found that *DPYD*, *CES1*, *ABCC4*, and *TYMP* had noticeably greater number of variants per person with CADD scores ≥ 20 within the IM/PM population, compared to the EM AR population. *ABCC3* was found to be the only gene studied with more variants per person with CADD scores ≥ 20 in the EM AR population, than the IM/PM population.

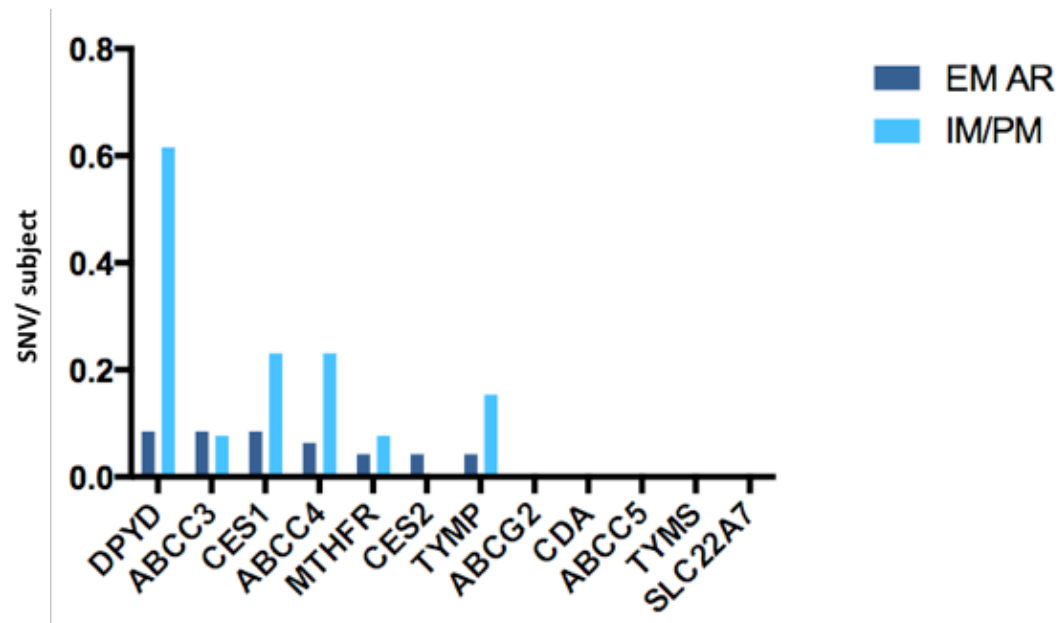


Figure 6. Total number of variants with CADD scores ≥ 20 within the *DPYD* EM AR (n=47) and IM/PM population (n=13), normalized to sample size.

Discussion

NGS data identified variants within exonic regions of 12 identified genes within the metabolic pathway of 5-FU and capecitabine, of 69 colorectal cancer patients. It was shown that among the EM population, there were more variants found exclusively within the EM AR population than exclusively within the EM non AR population, for 50% of the genes analyzed. Furthermore, 22 SNPs within 7 different genes, that have been predicted to be deleterious *in silico* or have been identified in previous literature as such, were found within the EM AR population. Of these, 12 SNPs were found exclusive to the EM AR subgroup among the total EM population. The depth of coverage of the targeted sequenced regions of *DPYD*, *ABCC4*, *ABCC3*, and *TYMP*, were described to provide information on the reliability of coverage for specific SNPs found at particular locations. Finally, the total number of variants with CADD scores ≥ 20 were compared between the IM/PM and the EM AR groups.

DPYD

It was found that *DPYD* along with *ABCG2* contained the highest proportion of variants with CADD scores ≥ 20 (75% of exonic variants within the genes). This lead to the investigation of variants within *DPYD*, as it is well established that SNPs within this gene have been associated with toxicity due to 5-FU and capecitabine therapies. Five *DPYD* variants were identified within the EM AR subgroup based on *in silico* prediction scores or prevalence in current literature predicting toxicity²¹. The *DPYD* rs1801158 SNP was found heterozygous within two patients of our study, and has been identified by Loganayagam et al. as significantly associated with grade 3-4 toxicity¹¹. In addition, *DPYD* rs75017182, found in perfect linkage disequilibrium with *DPYD* rs56038477, was found

heterozygous in three patients. One study conducted by Nie et al. demonstrated DPD enzyme function was reduced by 35% in heterozygous carriers, similar to that of the well studied toxicity-associated variant rs67376798 (31% reduction)²². This variant is now included in current CPIC guidelines and suggested to be screened for prior to clinical treatment²¹. A common *DPYD* variant rs2297595¹² found within eight EM AR patients of this study, has been associated with 5-FU toxicity within two separate studies^{12,23}. Both studies reported a significant association between this variant and grade 3 & 4 5-FU related toxicity within breast and gastroesophageal patients, however the association was not significant in colorectal cancer patients^{12,23}. This was likely due to differences in treatment regimens between the different cancers. In accordance with these findings, Pancyzk et al. indicated this variant was located at a significantly conserved site, and may disrupt electron transport and ultimately decrease DPD activity¹⁹. Additionally, they presented findings regarding decreased expression of *DPYD* in carriers of the rs1801160 variant (*DPYD**6)¹⁹. In an association analysis conducted by Ruzzo et al., *DPYD**6 allele carriers were significantly associated with $\geq$ grade 3 fluoropyrimidine-related ARs, and were further significantly associated with time to neutropenia following a time-to-toxicity analysis²³. This variant was found within nine patients of the EM AR subgroup, three of which developed neutropenia at some point throughout their treatment. Furthermore, this SNP was found at twice the frequency compared to that of the 1000G (Euro). Although not significant ($p=0.09$), more research into the clinical applications of this SNP should be conducted with greater sample populations.

MTHFR

Variants in *MTHFR* were examined due to the role 5,10-CH₂THF plays in the inhibition of TS. Several polymorphisms within *MTHFR* were found unique to the EM AR subgroup, while none were found unique to the EM no AR subgroup. Two SNPs with CADD scores ≥ 20 were identified within the EM AR population: rs1801133, and a novel SNP with no previous reported frequency by 1000G (Euro). In the same study conducted by Loganayagam et al., rs1801133 was identified as a predictor of fluoropyrimidine toxicity¹¹. During a subgroup analysis restricted to patients treated with capecitabine, they found a significant association between this variant genotype and hand-foot syndrome¹¹. In our study, 21 patients in the EM AR subgroup possessed the rs1801133 variant. Nine of these patients were treated with capecitabine, of which 67% developed hand-foot syndrome. This is compared to eight patients within this group who were treated with 5-FU, of which 25% developed hand-foot syndrome. In the case of two of the four remaining patients, it is unclear which treatment they received, while the other two received both, none of which developed hand-foot syndrome. An association appears to be present between the presence of this variant and the development of hand-foot syndrome subsequent to capecitabine treatment, although larger sample sizes are necessary to provide more reliable conclusions. Other studies have demonstrated rs1801133 resulted in decreased enzymatic activity, which may have a significant effect on the pharmacological efficacy of 5-FU^{19,25}. As *MTHFR* converts 5,10-CH₂THF into 5-CH₂THF, a decrease in the activity of *MTHFR* may lead to an accumulation of 5,10-CH₂THF, the key one-carbon donor and co-substrate of the TYMS enzyme during the methylation of dUMP to dTMP¹⁹. As a result, this may contribute to changes in chemosensitivity of cells exposed to 5-FU by increasing the amount and stability of the ternary complex formed, leading to greater inhibition of

DNA synthesis¹⁹. Both *in vitro* and *in vivo* studies further observed that the presence of rs1801133 is responsible for greater chemosensitivity in colon cancer cells, suggesting that this variant may be a pharmacogenetic factor used to assess the effectiveness of 5-FU based chemotherapy²⁶.

TYMP

Based on the role *TYMP* plays in the conversion of capecitabine to 5-FU, as well as the generation of the active metabolite FdUMP, variants within this gene were investigated as to their potential contribution to fluoropyrimidine-related toxicity. This study identified two SNPs that were predicted *in silico* as deleterious: rs11479, and rs112723255. Found within five EM AR patients within this study, rs11479 has shown to be significantly associated with early dose modifications and/or severe adverse events²⁷. However, these variants were identified in areas of low coverage and thus subsequent validation is needed. In a study conducted by Evrard et al., an expression vector containing human *TYMP* cDNA was transfected into human colon carcinoma cells²⁸. They found the cytotoxic effects of 5-FU were higher in transfected cells compared to wild-type cells, and that increased sensitivity to 5-FU by these transfected cells was significantly correlated with an increase in both *TYMP* activity and expression²⁸. Polymorphisms in *TYMP* resulting in increased expression or function may have clinical consequences by increasing the formation of FdUMP, the primary active metabolite of 5-FU. Currently, there is little *in vitro* literature characterizing the functional consequences of the two *TYMP* variants presented, and further investigation is needed to determine whether these SNPs could induce a gain of function.

ABCC4/ABCC3

The transporter ABCC4 plays a vital role in the efflux of toxic fluoropyrimidine metabolites. Thus, a reduction in the expression or function of this membrane protein could potentially lead to a toxic accumulation of fluoropyrimidines within cells. This study identified three variants within the EM AR subgroup that were predicted *in silico* to be deleterious: rs11568644, rs2274407, and rs11568658. All three polymorphisms were found exclusively within the AR subgroup of the EM cohort. Banerjee et al. showed the transport of MMA(GS)₂, a substrate of MRP4, was reduced by 73% and 30% compared to the wild-type allele, in cells carrying rs11568658 and rs2274407, respectively²⁹. Levels of ABCC4 in cells carrying rs2274407 were also 50% of those found in wild-types²⁹. Currently, no literature has examined the effects of these variants with that of toxicity subsequent to 5-FU treatment. However, the results presented by Banerjee et al. provide some insight as to the possible mechanisms by which variants in *ABCC4* may contribute to fluoropyrimidine-related toxicity.

Although limited studies have been conducted investigating the functional characteristics of *ABCC3* variants, Kobayashi et al. described three nonsynonymous SNPs resulting in either an intracellular accumulation of an immature protein, or a loss of their transport activity³⁰. Based on the role *ABCC3* plays in the clearance of fluoropyrimidines, mutations such as these could lead to an intracellular accumulation of toxic metabolites. Variants with CADD scores ≥ 20 found within the EM AR population of this study were mostly novel (3/4 with no reported frequencies by the 1000G), and found exclusive to this subgroup of the EM population (3/4 variants). Furthermore, 75% of the identified variants had CADD scores >30 , corresponding to the top 0.1% of deleterious variants.

Comparing Number of Deleterious Variants Between IM/PM and EM AR Groups

As variants within the EM AR subgroup could also be found within the IM/PM population and contribute to toxicity, the number of deleterious variants between both groups were compared. Of note, there were many more deleterious variants within *DPYD*, *TYMP*, and *ABCC4* in the IM/PM group compared to the EM AR group. When examining the role of these proteins regarding fluoropyrimidine metabolism, deleterious variants could contribute to toxicity as previously mentioned. However, due to differences in sample sizes between the two groups, no absolute conclusions can be drawn.

Limitations and Conclusion

Due to the relatively small sample size of this study, limited statistical significance was able to be drawn, however predictions were made based on *in silico* scores and the given role of a protein within the metabolic pathway. Future studies should focus on *in vitro* characterization of these SNPs, as in many cases little information is available, as well as the clinical implications using greater sample populations.

In conclusion, considering the substantial number of patients classified as *DPYD* EMs who continue to experience ARs to their treatment, it is vital a more comprehensive screening technique be developed, which encompasses other genes that likely contribute to toxicity subsequent to fluoropyrimidine therapies. By identifying key variants of interest, this study provided an area of focus for further studies to examine both the functional and clinical significance of variants within genes pertaining to the metabolic pathway of 5-FU and capecitabine. By developing a screening panel which is more comprehensive along with corresponding dose reduction guidelines, clinicians may be able to provide better care to colorectal cancer patients treated with fluoropyrimidines.

Acknowledgements

I would like to thank Dr. Wendy Teft and Dr. Ute Schwarz for their help and guidance throughout this project. I would like to acknowledge Markus Gulilat's creative assistance in the development and formatting of the figures, as well as his guidance throughout the project. Cam Ross and the personalized medicine department are thanked for their assistance with day-to-day activities. Dr. Richard Kim is thanked for his supervision throughout the project.

References

1. Canadian Cancer Society's Advisory Committee on Cancer Statistics, Canadian Cancer Statistics 2017, <http://www.cancer.ca/en/cancer-information/cancer-type/colorectal/statistics/?region=on> (2017).
2. Colorectal Cancer Canada, Treatments & Side Effects, <http://www.colorectal-cancer.ca/en/treating-cancer/treatment-cancer/> (2017).
3. F. Mir, S. Rafiyath. Medscape, Fluorouracil Toxicity and DPYD, <https://emedicine.medscape.com/article/1746057-overview#a2> (2016).
4. Miura, K. *et al.* 5-FU metabolism in cancer and orally-administrable 5-FU drugs. *Cancers (Basel)*. **2**, 1717–1730 (2010).
5. C.F. Thorn *et al.* PharmGKB, Fluoropyrimidine Pathway, Pharmacokinetics, <https://www.pharmgkb.org/pathway/PA150653776> (2018).
6. Van Cutsem, E. *et al.* Oral capecitabine vs intravenous 5-fluorouracil and leucovorin: Integrated efficacy data and novel analyses from two large, randomised, phase III trials. *Br. J. Cancer* **90**, 1190–1197 (2004).
7. Fang, L. *et al.* Determining the optimal 5-FU therapeutic dosage in the treatment of colorectal cancer patients. *Oncotarget* **7**, 81880–81887 (2016).
8. Wang, B., Walsh, S. J. & Saif, M. W. Pancytopenia and Severe Gastrointestinal Toxicities Associated with 5-Fluorouracil in a Patient with Thymidylate Synthase (TYMS) Polymorphism. *Cureus* **8**, e798 (2016).
9. Inokuchi, M. *et al.* Treatment of capecitabine-induced hand-foot syndrome using a topical retinoid: A case report. *Oncol. Lett.* **7**, 444–448 (2014).
10. Wheeler HE, González-Neira A, Pita G, de la Torre-Montero JC, Alonso R, Lopez-Fernandez LA, *et al.* Identification of genetic variants associated with capecitabine-induced hand-foot syndrome through integration of patient and cell-line genomic analyses. *Pharmacogenet Genomics* **24**, 231–237 (2014).
11. 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**, 2505–2515 (2013).
12. Amstutz, U., Froehlich, T. K. & Largiadèr, C. R. Dihydropyrimidine dehydrogenase gene as a major predictor of severe 5-fluorouracil toxicity. *Pharmacogenomics* **12**, 1321–1336 (2011).
13. Dobritzsch, D., Schneider, G., Schnackerz, K. D. & Lindqvist, Y. Crystal structure of dihydropyrimidine dehydrogenase, a major determinant of the pharmacokinetics of the anti-cancer drug 5-fluorouracil. *EMBO J.* **20**, 650–660 (2001).
14. Caronia, D. *et al.* A polymorphism in the cytidine deaminase promoter predicts severe capecitabine-induced hand-foot syndrome. *Clin. Cancer Res.* **17**, 2006–2013 (2011).
15. Illumina, Introduction to SBS Technology, <https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html> (2015).
16. ANNOVAR Documentation, <http://annovar.openbioinformatics.org/en/latest/> (2017).
17. Ansari, M. *et al.* Polymorphisms in multidrug resistance-associated protein gene 4 is associated with outcome in childhood acute lymphoblastic leukemia. *Blood* **114**, 1383–1386 (2009).

18. Banerjee, M. *et al.* Polymorphic variants of MRP4/ABCC4 differentially modulate the transport of methylated arsenic metabolites and physiological organic anions. *Biochem. Pharmacol.* **120**, 72–82 (2016).
19. Panczyk, M. Pharmacogenetics research on chemotherapy resistance in colorectal cancer over the last 20 years. *World J. Gastroenterol.* **20**, 9775–9827 (2014).
20. Combined Annotation Dependent Depletion (CADD), <http://cadd.gs.washington.edu/info> (2014).
21. Amstutz, U. *et al.* Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Dihydropyrimidine Dehydrogenase Genotype and Fluoropyrimidine Dosing: 2017 Update. *Clin. Pharmacol. Ther.* **103**, 210–216 (2018).
22. Nie, Q. *et al.* Quantitative Contribution of rs75017182 to Dihydropyrimidine Dehydrogenase mRNA Splicing and Enzyme Activity. *Clin. Pharmacol. Ther.* **102**, 662–670 (2017).
23. Gross, E. *et al.* Strong association of a common dihydropyrimidine dehydrogenase gene polymorphism with fluoropyrimidine-related toxicity in cancer patients. *PLoS One* **3**, 1–7 (2008).
24. Ruzzo, A. *et al.* Dihydropyrimidine dehydrogenase pharmacogenetics for predicting fluoropyrimidine-related toxicity in the randomised, phase III adjuvant TOSCA trial in high-risk colon cancer patients. *Br. J. Cancer* **117**, 1269–1277 (2017).
25. Lam, S. W., Guchelaar, H. J. & Boven, E. The role of pharmacogenetics in capecitabine efficacy and toxicity. *Cancer Treat. Rev.* **50**, 9–22 (2016).
26. Sohn, K. J. *et al.* Effect of the methylenetetrahydrofolate reductase C667T polymorphism on chemosensitivity of colon and breast cancer cells to 5-fluorouracil and methotrexate. *J Natl Cancer Inst.* **96**, 134–144 (2004).
27. Jennings, B. A. *et al.* Evaluating Predictive Pharmacogenetic Signatures of Adverse Events in Colorectal Cancer Patients Treated with Fluoropyrimidines. *PLoS One* **8**, 1–9 (2013).
28. Evrard, A., Cuq, P., Ciccolini, J., Vian, L. & Cano, J. P. Increased cytotoxicity and bystander effect of 5-fluorouracil and 5'-deoxy-5-fluorouridine in human colorectal cancer cells transfected with thymidine phosphorylase. *Br. J. Cancer* **80**, 1726–1733 (1999).
29. Banerjee, M. *et al.* Polymorphic variants of MRP4/ABCC4 differentially modulate the transport of methylated arsenic metabolites and physiological organic anions. *Biochem. Pharmacol.* **120**, 72–82 (2016).
30. Kobayashi, K., Ito, K., Takada, T., Sugiyama, Y. & Suzuki, H. Functional analysis of nonsynonymous single nucleotide polymorphism type ATP-binding cassette transmembrane transporter subfamily C member 3. *Pharmacogenet. Genomics* **18**, 823–833 (2008).