Published in final edited form as:

Clin Pharmacol Ther. 2017 March ; 101(3): 382–390. doi:10.1002/cpt.531.

Novel deleterious dihydropyrimidine dehydrogenase variants may contribute to 5-fluorouracil sensitivity in an East African population

Tarig Elraiyah¹, Calvin R. Jerde^{1,2}, Shikshya Shrestha¹, Rentian Wu¹, Qian Nie¹, Nasra H. Giamma³, Vivekananda Sarangi⁴, Lewis R. Roberts³, Steven M. Offer^{1,2}, and Robert B. Diasio^{1,2,*}

¹Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN

²Mayo Clinic Cancer Center, Rochester, MN

³Division of Gastroenterology and Hepatology, Department of Internal Medicine, Mayo Clinic, Rochester, MN

⁴Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester MN

Abstract

Clinical studies have identified specific genetic variants in dihydropyrimidine dehydrogenase (DPD, *DPYD* gene) as predictors of severe adverse toxicity to the commonly used chemotherapeutic 5-fluorouracil; however, these studies have focused on European and European-American populations. Our lab recently demonstrated that additional variants in non-European haplotypes are predictive of 5-FU toxicity. The objective of this study was to identify potential risk variants in an under-studied East African population relevant to our institution's catchment area. The *DPYD* protein-coding region was sequenced in 588 individuals of Somali or Kenyan ancestry living in central/southeast Minnesota. Twelve novel non-synonymous variants were identified, seven of which significantly decreased DPD activity *in vitro*. The commonly reported toxicity-associated variants *2A, D949V, and I560S were not detected in any individuals. Overall, this study demonstrates a critical limitation in our knowledge of pharmacogenetic predictors of 5-FU toxicity, which has been based on clinical studies conducted in populations of limited diversity.

*Correspondence Robert B. Diasio, MD, Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, 200 First St SW, Rochester, MN 55905, diasio.robert@mayo.edu.

Conflict of Interest/Disclosure

TE, CRJ, SS, RW, QN, NHG, VS, LRR, SMO and RBD have nothing to disclose.

Author Contributions

Conception and design: T. Elraiyah, S.M. Offer, R. B. Diasio

Acquisition of data: T. Elraiyah, C. R. Jerde, S. Shrestha, R. Wu, Q. Nie, N.H. Giamma, L. R. Roberts

Analysis and interpretation of data: T. Elraiyah, C. R. Jerde, V. Sarangi, S. M. Offer

Writing, review, and/or revision of the manuscript: T. Elraiyah, S.M. Offer, R.B. Diasio

Study supervision: S.M. Offer, R.B. Diasio

Keywords

Pharmacogenetics; Fluorouracil; Dihydropyrimidine Dehydrogenase Deficiency; African Americans; Kenya; Somalia; Uracil; Genotype

Introduction

Since it was introduced in 1957, 5-fluorouracil (5-FU) has become the cornerstone of chemotherapeutic regimens for a variety of solid tumors, primarily gastrointestinal, breast, and skin cancers. 5-FU has a narrow therapeutic index that contributes to the risk of dose limiting severe adverse toxicity, which affects 31-34% of treated patients (1). The uracil catabolic pathway is responsible for degrading approximately 85% of administered 5-FU (2). Dihydropyrimidine dehydrogenase (DPD, encoded by *DPYD*) is the initial and rate-limiting enzyme of the uracil catabolism, and, as such, variability in DPD enzyme activity was shown to affect 5-FU metabolism and adverse toxicity (3). A number of genetic variations in *DPYD* have been linked to 5-FU toxicity in several clinical studies (4-7). *2A (rs3918290) has been associated with impaired DPD function and increased incidence of severe toxicity and death (8, 9). Two additional variants p.I560S (rs55886062) and p.D949V (rs67376798) have been consistently associated with toxicity (9, 10). The intronic variant c.1129-5923C>G (rs750187182) has been linked to a toxicity-associated haplotype and was suggested to contribute to 5-FU toxicity through non-obligate alternative splicing of *DPYD* (11, 12). Importantly, the clinical studies that identified and characterized these variants were comprised almost entirely of European or European American patients.

Using a cohort of healthy African-American volunteers, our lab identified p.Y186C (rs115232898), as a potential risk variant for 5-FU-related toxicity. Carriers of p.Y186C had significantly reduced DPD enzyme activity in PBMCs (13). The variant was subsequently detected in patients with 5-FU-related severe toxicity (14), one potentially linked to lethality (15). Notably, none of the toxic p.Y186C carriers presented with any other known toxicity-related variants. An analysis of publically available genotype databases suggested that p.Y186C may be carried as haplotypes associated with West African populations (13).

Recent genetic studies have suggested that Eastern and Saharan Africans have genetic components that are distinguished from population groups in western, central, and southern Africa (16, 17). Therefore, we hypothesized that East African individuals may carry a distinct repertoire of *DPYD* variants that may confer risk of 5-FU toxicity. As such, targeted genotyping for currently accepted 5-FU risk variants would not be expected to provide useful predictive information in individuals with East African ancestry. Our institution's catchment area houses some of the world's largest groups of people with East African origin outside Africa. According to the American Community Survey Data on the Foreign-Born Population conducted by the United States Census Bureau between 2008-2012, Minnesota is home to the largest Somali population in the United States (17,320), and the second largest Kenyan (5,610), and Ethiopian (12,905) populations (18).

Given the sizable number of East Africans in our catchment area, and the putative difference in *DPYD* risk alleles from other African populations, the objectives of the current study

were to assess the distribution of known toxicity-associated variants within the under-studied East African population, and to identify any novel deleterious *DPYD* variants in that group.

Results

Performance and optimization of targeted sequencing strategy

Targeted sequencing of the coding region was performed in two sets. The first set comprised of 270 individuals of Somali origin, while the second set included 278 Somali and 40 Kenyan participants (**Table 1**). The first set had an average coverage of 4000X, with 98.48% of target region covered. Data from the first set was used to improve sequence quality in low-quality areas and to increase the number of specimens being multiplexed per high throughput sequencing lane. The average coverage for the second set was 202X for Somalis and 180X for Kenyans. The target region was covered at 97.64% and 98.64% for the Somalis and Kenyans respectively.

Identification of coding region variants in the *DPYD* gene

A total of 588 individuals (548 Somalis and 40 Kenyans) participated in this study (**Table 1**). In all, 152 variants were identified. A summary of the reported variants is depicted in **Figure 1A**. Thirty-four variants were located in exons, 78 were located in exon-flanking regions, and 40 were located in the 5' and 3' untranslated regions (UTRs). Of the exon variants, 16 were not present in the National Center for Biotechnology Information (NCBI) database of genetic variation (dbSNP) and were thus considered novel (19). These novel variants included 12 non-synonymous variants that are predicted to change one or more amino acids in DPD and 4 synonymous variants. Overall 24 non-synonymous variants were identified in Somalis and 9 in Kenyans (**Figure 1B**). Novel non-synonymous variants included the frameshift variant p.I971N[FS] and the 9 nucleotide in-frame duplication/insertion p.S14-S16[dup], as well as 10 single amino acid substitutions (**Table 2**). The locations and genotype frequencies of novel non-synonymous and synonymous variants are presented in **Table 2**; detected synonymous and non-synonymous variants that have been previously reported in dbSNP are presented in **Supplementary Table S1**. Detected intron and UTR variants that have been previously reported in dbSNP are presented in **Supplementary Table S2**; novel intron and UTR variants that were detected in this study population are presented in **Supplementary Table S3**. Notably, none of the known 5-FU toxicity-associated *DPYD* variants *2A (rs3918290; also known as IVS14+1G>A), p.I560S (rs55886062; also known as *13), or p.D949V (rs67376798) were detected in this study population (**Table 2**). No variants were detected within *DPYD* splice donor or acceptor sequences (data not shown). The genotype frequencies for all detected variants were found to be in Hardy-Weinberg equilibrium ($P>0.05$). All the novel variants have minor allele frequencies of less than 1% (**Table 2**).

Effects of non-synonymous variants on DPD enzyme activity

To assess the potential contribution of detected non-synonymous variants to risk of 5-FU toxicity, we measured the *in vitro* DPD enzyme activity of all 12 novel non-synonymous variants using a mammalian system of transgenic expression that was previously reported by our lab (20, 21). Of the 12 variants tested, 7 showed significant reductions in DPD enzyme

activity (**Figure 2**). The frameshift variant p.I971N[FS] completely ablated enzyme activity (activity=0.02, $P=1.9\times 10^{-6}$, relative to wild-type DPD controls). The single amino acid changing variants p.P86L, p.T793I, p.V941A, p.P1023S, p.A513V, and p.P237L all resulted in significantly reduced DPD activity (activities=0.03, 0.25, 0.65, 0.79, 0.82, and 0.85 respectively; $P=2.0\times 10^{-7}$, 1.9×10^{-5} , 3.7×10^{-4} , and 7.7×10^{-3} , 1.9×10^{-2} , and 4.8×10^{-2} respectively). The 5 additional novel nonsynonymous SNPs detected in the study population (p.R394Q, p.L523F, p.I528V, p.R699T and p.S14-S16[dup]) did not significantly alter DPD function (**Figure 2**).

Our lab has previously determined the impact on DPD enzyme activity for many reported *DPYD* variants (20, 21). A summary of the functional effects for variants detected in this study is included in **Supplementary Table S1**. Notably, 2 of these detected variants (p.Y186C and p.D342N) resulted in significant reductions in enzyme activity.

Structural considerations for novel variants

To investigate the mechanism(s) by which deleterious variants impaired DPD enzyme activity, the location of novel variants was modeled onto the pig DPD crystal structure (**Figure 3**). An overview of the monomeric structural domains is shown in **Figure 3A**. The novel deleterious variant p.P86L is located within domain I (residues 27–172), near to one of the four 4Fe-4S clusters (**Figure 3B**). Domain II (residues 173–286 and 442–524) contains four benign novel variants (p.P237L, p.A513V, p.S14_S16[dup], and p.L523F; **Figure 3C**). Three variants p.A513V, p.S14_S16[dup], and p.L523F are located in a non-structural loop region; p.P237L is located at the beginning of a predicted alpha helical amino acid stretch. A single benign variant (p.R394Q) is located near the middle of a beta turn in domain III (residues 287–441; **Figure 3D**). The deleterious p.T793I variant substitutes a polar uncharged threonine with a non-polar, hydrophobic isoleucine near the uracil binding site within domain IV which consists of residues 525–847 (**Figure 3E**). Two additional benign variants (p.I528V and p.R699T) variants are located in Domain IV (**Figure 3E**). The deleterious missense variant p.V941A is located 3 amino acids upstream from a beta turn that is in close proximity to the pair of 4Fe-4S clusters located within domain V which consist of residues 1–26 and 848–1025 (**Figure 3F**). Likewise, sequences downstream of isoleucine 971, which are altered due to the p.I971[FS] variant, are also in close proximity to the 4Fe-4S clusters in domain V (**Figure 3F**). Finally the moderately deleterious missense variant p.P1023T is located within a likely nonstructural domain near the C-terminal end of DPD, which has not been resolved by crystallography (**Figure 3F**).

Discussion

Numerous pre-clinical and clinical studies have established that certain *DPYD* variants can dramatically increase risk of severe toxicity to 5-FU. The most intensively studied *DPYD* variant *2A disrupts the splice donor site at the 5' end of intron 14, leading to alternative splicing. This alternatively spliced product lacks exon 14 (an in-frame deletion) and functionally inactivates the enzyme (22). The decreased levels of functional DPD lead to reduced 5-FU catabolism and increased 5-FU availability for conversion into active metabolites via anabolism (23). In turn, carriers of deleterious variants such as *2A are at

increased risk for adverse toxicity to 5-FU (8, 24). Two additional variants p.I560S and p.D949V have also been reported to be associated with toxicity (8, 9, 24). Recently, a large clinical study was conducted to evaluate the association of these three variants with clinical adverse events and their predictive value in detecting toxicity; the study demonstrated a statistically significant association between two of the variants and the incidence of adverse toxicity events (grade 3 or more) (7). The third variant, p.I560S, is rare and the study lacked adequate power to conclusively assess this variant. These three variants are mainly detected in individuals of European ancestry, and less likely to be found in individuals of other racial groups (**Figure 4A**). The Clinical Pharmacogenetics Implementation Consortium (CPIC) has recently issued a set of guidelines to provide dosing recommendations for 5-FU chemotherapy based on *DPYD* genotype limited to these aforementioned variants (25).

In the current study we identified and characterized 12 novel non-synonymous variants in individuals of East African ancestry. Seven of these variants were found to be deleterious by *in vitro* measurement of recombinant DPD activity. Interestingly, the three known toxicity associated variants *2A, p.I560S and p.D949V were not detected in the study populations, which is consistent with allele frequency data obtained from 1000 genome project (**Figure 4A**). Benign non-synonymous *DPYD* variants (**Figure 4B**) and synonymous *DPYD* variants (**Figure 4C**) were detected at frequencies broadly consistent with the reported frequencies in the 1000 genome African super-population.

Our lab previously identified the p.Y186C variant in an African American population (13). Data from the 1000 Genomes Project (26), confirms that this variant is mainly found in African populations, including Americans of African Ancestry in Southwest USA (ASW), Yoruba in Ibadan, Nigeria (YRI) and Luhya in Webuye, Kenya (LWK) populations. In the present study, this variant was detected in a single Somali individual (0.2% allele frequency) and in three Kenyans (7.5% allele frequency). Ancestry mapping studies have suggested that Kenyan haplotypes contain predominantly Niger-Congo components (an ancestral component for some of West African haplotypes as well), while Somali from Somalia and Ethiopia haplotypes mainly belong to the Lowland East Cushitic ancestral component (which is more closely related to Middle eastern and European ancestral components) (17). These distinguishing ancestral components potentially explain the elevated frequency of p.Y186C in the Kenyan compared to Somali populations reported herein. The distinction in ancestral haplotypes may also partially explain the unique variants that were detected within the Kenyan (i.e., p. A721T) and Somali (i.e., p.P86L, p.P237L, p.R394Q, p.A513V, p.S14_S16dup, p.L523F, p.I528V, p.R699T, p.T793I, p.V941A, p.I971Nfs, p.K259E, p.G258S, p.D342N, p.S534N and p.V812I) populations (**Figure 1B**).

In addition to p.Y186C (discussed above), our lab previously determined that p.D342N, which was detected as a heterozygous allele in 4 Somali individuals in the present study, significantly impaired DPD function (17% relative DPD activity) (20). The variant p.V732I was detected in 64 East African individuals (MAF 5.7%; **Supplementary Table S1**). A previous *in vitro* study showed that this variant did not alter the enzyme function when compared to wild-type DPD (21). Nevertheless, another study reported that African Americans who carried p.V732I had significant reduction in DPD enzyme activity but only in conjunction with p.Y186C (13). This observation is further supported by the finding of a

study that aimed to detect the correlation of this variant to 5-FU toxicity in a cohort of colorectal cancer patients (5). A recent study reported a statistically significant association of p.V732I with grade 3 or greater toxicities in a group of European patients (27). Collectively, these findings may suggest that an additional variant in linkage disequilibrium (LD) with p.V732I may be contributing to DPD deficiency. Additional study of this variant and any variants in LD with p.V732I will be needed to more clearly understand its contribution to 5-FU toxicity.

Of all variants detected in this study, the benign variant p.C29R was the most common with a minor allele frequency of 43.9% (T allele; **Figure 4B** and **Supplementary Table S1**). This frequency is similar to that reported for the 1000 Genome Project AFR super-population (T allele; 54.9%). Interestingly, a previous study reported that p.C29R was found in a significantly higher frequency in low-toxicity compared to high toxicity groups of colorectal cancer patients, indicating a putative protective effect of this variant (5). Likewise, p.M166V, which was also detected in the East African cohort, showed what appeared to be a similar protective effect with a reduction in overall hematological adverse events and neutropenia (5). However, the protective effect of p.M166V was questioned by another study that described an increased incidence of grade 3-4 5-FU toxicity in patients with gastro-esophageal and breast cancer (4).

Previous studies suggested that genetic polymorphisms that are located outside of *DPYD* coding region or splice sites might also contribute 5-FU toxicity. Among those is the deep intronic variant c.1129–5923C>G, which was found to create a cryptic splice donor site that resulted in an insertion of 44 bp fragment into the mature DPD mRNA (12). An earlier study reported an association of a haplotype (HapB3), consisting of three intronic variants and a synonymous SNP, with severe 5-FU toxicity (11). Notably, individuals with the deep intronic variant were not detected in the present study (data not shown), consistent with previous studies that suggested that this variant is mainly restricted to European haplotypes (6, 12). Similarly, the complete HapB3 haplotype was not detected in any individuals in the present study.

Structurally, we found that the deleterious variants that severely affected enzyme function were generally located within domains that are important for protein function. The complete ablation of the enzyme function by the I971Nfs (located in domain V, **Figure 3F**) could be explained by the altered structure that resulted from shifting of the reading frame. P86L was located close to one of the N-terminal Fe–S sites at domain I (**Figure 3B**); therefore, it may disrupt function by interfering with electron shuttling. T793I changes the first of the 3 amino acids that constitute the binding site for FMN at domain IV (**Figure 3E**). This domain also includes the uracil-binding site. Altering the protein structure at this sensitive location could interfere with the binding of these nucleotides, and thus ablate enzyme function. Both V941A and P1023S are located within domain V in the vicinity of the C-terminal Fe–S sites (**Figure 3F**). Their moderate effect on protein function could be explained by changing the protein structure and the microenvironment surrounding the Fe-S clusters. Of note, A513V and P237L were found to have a marginal effect on enzyme function. Their locations relative to the FAD-binding site in domain II could possibly contribute to this minimal effect (**Figure 3C**). Surprisingly, the 3 amino acid duplication S14_S16dup did not significantly affect

enzyme function. This unexpected finding might be explained by its location in a flexible loop region at the periphery of the protein structure, which is distal from sites critical for enzyme function (**Figure 3C**).

While participants in the present study were not first-degree relatives based on individual and community discussions and questionnaires, we cannot rule out relatedness as a potential contributor to the allele frequencies noted for some variants. Regardless, the findings of the present study highlight the need for further evaluation of the distribution and frequency of *DPYD* variants in diverse populations. Furthermore, clinical studies are needed to evaluate the importance of these variants in predicting the toxicity and adverse events associated with 5-FU therapy.

In conclusion, this study presents a set of novel *DPYD* variants that were detected in East Africans. Some of the identified variants severely affected DPD enzyme function *in vitro* and could potentially contribute to 5-FU toxicity risk. This study also highlights the lack of known toxicity-associated variants (*2A, I560S, and D949V) in this population. The genetic predictive tests of 5-FU toxicity that are recommended for clinical use currently focus only on the aforementioned risk alleles. Based on the findings of this manuscript, these tests are likely of limited value for East African populations. The findings of this study will support the effort of achieving an individualized 5-FU therapy in cancer patients of non-European ancestry.

Materials and methods

Study participants

Samples were collected from healthy volunteers of East African origin with ancestry from Somalia or Kenya who were over 18 years of age. Ancestry was self-declared by study participants. Samples were collected in the Rochester, MN, Mankato, MN, and Minneapolis/St. Paul, MN, metropolitan areas. This study was approved by the Mayo Clinic Institutional Review Board. Individual consent was obtained from each study participant.

Targeted sequencing of *DPYD*

Sequencing libraries were prepared using the Access Array System (Fluidigm Corporation, South San Francisco, CA). Custom primer panels were designed to cover all 23 exons of the *DPYD* gene, including UTR regions, using the Fluidigm D3 design portal. Target amplification was performed in parallel for 48 volunteers on a single Access Array chip. Samples from up to 10 chips were barcoded and pooled according to manufacturer's protocol. Libraries were prepared using two separate iterations of the primer design. The first design (set A) was used to amplify 270 samples (all Somali). Samples were sequenced using an Illumina MiSeq (Illumina, San Diego, CA) by the Mayo Clinic Cancer Center Gene Analysis Shared Resource in the Mayo Clinic Advanced Genomics Technology Center. Sequence quality and depth of coverage was assessed for all 270 samples. Additional tiled primers were designed and added to the primer panels to compensate for regions of low quality sequencing or poor coverage. An additional 318 samples (278 Somali and 40 Kenyan) were amplified using the re-designed primer pools and sequenced. To confirm the

sequence of identified variants, exons containing novel and/or deleterious variants were amplified and sequenced using the Sanger method as previously described (28) by the Mayo Clinic Cancer Center Gene Analysis Shared Resource. Insertion variants were additionally confirmed by cloning the PCR product into the pJET1.2/blunt vector (Thermo Scientific, Waltham, MA) and Sanger sequencing of multiple vector clones.

Expression vector construction

Plasmids with the specific variants were created using site-directed mutagenesis (SDM) as described previously (21). Plasmid sequences were confirmed by the Mayo Clinic Cancer Center Gene Analysis Shared Resource in the Mayo Clinic Advanced Genomics Technology Center. Primers used for SDM are shown in **Supplementary table S4**.

Cell lines and transfection

HEK293T/c17 cells, which do not express detectable DPD (21), were acquired from the American Type Culture Collection and cultured as previously detailed (21). Low passage stocks were prepared after receipt and cells were maintained in culture media for a maximum of 10 total passages after receipt. The cells were regularly monitored by microscopy to confirm cell identity by monitoring cell morphology. Mycoplasma screening was performed every 3 months or whenever a question of contamination arose by using the mycoplasma detection kit Plasmotest (InvivoGen, San Diego, CA). Additionally, cell proliferation was periodically monitored using real-time cell analysis (xCelligence RTCA System, Acea Biosciences, San Diego, CA). For transfection, 1×10^6 cells were seeded in 6-well plates containing Dulbecco's Modified Eagle's Medium (Mediatech, Manassas, VA) supplemented with 10% Fetal Bovine Serum (FBS) (Denville Scientific, Holliston, MA) and 1% L. Glutamine, (Thermo Scientific, Waltham, MA) at 37° C for 24 hours prior to transfection. Cells were transfected using XtremeGENE HP (Roche Diagnostics, Indianapolis, IN) per manufacturers protocol. Media was replaced after 24 hours, and protein lysates were prepared 48 hours after transfection as described previously (21).

Experimental design for DPD enzyme activity assay

Variants to be studied were randomly divided into 3 groups of 4 variants each. Each group was assessed in parallel with vectors encoding wild-type *DPYD* (positive control) and *2A (negative control). Data from 3 technical replicates (i.e., samples transfected and analyzed in parallel) were pooled to form a single biological replicate. Analyses were repeated for 4 independent biological replicates. Statistical analyses presented represent the variability between biological replicates. *In vitro* DPD activity was assessed as detailed previously (20). Briefly, protein lysates were incubated with NADPH and C-14 labeled 5-FU. The conversion of C¹⁴ 5-FU to C¹⁴ 5-dihydrofluorouracil (C¹⁴ 5-DHFU) was assessed using reverse-phase high performance liquid chromatography (HPLC). DPD activity was measured by calculating the area under the curve (AUC) for 5-DHFU compared to the total AUC for the combination of 5-FU and 5-DHFU using Dionex Chromeleon 7.1 Chromatography Data System software (Thermo Fisher Scientific, Waltham, MA). Expression of recombinant DPD variants was assessed as detailed previously (20). To minimize variability due to small

differences in expression, activity was normalized to DPD expression, and relative activity of each variant relative to WT is presented.

***In silico* functional prediction**

The impact of the variants and the subsequent amino acid change on the biological function of DPD protein was predicted using PolyPhen-2 web server version 2.2.2 (29), SIFT version 4.0.3 (30), and Protein Variation Effect Analyzer (PROVEAN) version 1.1.3 (31). PolyPhen-2 software reports the results of the prediction as qualitative terms ("benign," "possibly damaging," and "probably damaging"), based on false positive rate (FPR) estimated for each variant. Sift prediction can either be "Damaging" or "Tolerated", depending on whether the variant affect the protein function or not, while PROVEAN reports either "deleterious" or "neutral."

Protein modeling

UCSF Chimera version 1.10.2 was used to model the protein sequence and generate 3-dimensional images (32). Variants were mapped on the porcine DPD crystal structure (PDBID 1GTH), which has 98% homology to human DPD (33).

Statistical analysis

Statistical analyses were performed using JMP Pro 11.2.1 (SAS Institute Inc., Minneapolis, MN).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors wish to thank Abdirashid Shire, Essa Mohamed, and Hassan Shaleh for facilitating the collection of specimens. This study was funded by Mayo Clinic Cancer Center grant number CA15083 and was supported by NIH training grant number T32 GM008685. The collection of specimens was supported in part by funding from Gilead Sciences. TE presented this study at the ASCPT 2016 meeting and was the recipient of an ASCPT Presidential Trainee Award.

References

- (1). Meta-Analysis Group In, C. et al. Toxicity of fluorouracil in patients with advanced colorectal cancer: effect of administration schedule and prognostic factors. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 1998; 16:3537–41. [PubMed: 9817272]
- (2). Heggie GD, Sommadossi JP, Cross DS, Huster WJ, Diasio RB. Clinical pharmacokinetics of 5-fluorouracil and its metabolites in plasma, urine, and bile. *Cancer Res*. 1987; 47:2203–6. [PubMed: 3829006]
- (3). Diasio RB, Beavers TL, Carpenter JT. Familial deficiency of dihydropyrimidine dehydrogenase. Biochemical basis for familial pyrimidinemia and severe 5-fluorouracil-induced toxicity. *The Journal of clinical investigation*. 1988; 81:47–51. [PubMed: 3335642]
- (4). Gross E, et al. Strong association of a common dihydropyrimidine dehydrogenase gene polymorphism with fluoropyrimidine-related toxicity in cancer patients. *PLoS One*. 2008; 3:e4003. [PubMed: 19104657]

- (5). Kleibl Z, et al. Influence of dihydropyrimidine dehydrogenase gene (DPYD) coding sequence variants on the development of fluoropyrimidine-related toxicity in patients with high-grade toxicity and patients with excellent tolerance of fluoropyrimidine-based chemotherapy. *Neoplasma*. 2009; 56:303–16. [PubMed: 19473056]
- (6). Lee AM, et al. Association between DPYD c.1129-5923 C>G/hapB3 and severe toxicity to 5-fluorouracil-based chemotherapy in stage III colon cancer patients: NCCTG N0147 (Alliance). *Pharmacogenetics and genomics*. 2016; 26:133–7. [PubMed: 26658227]
- (7). Lee AM, et al. DPYD variants as predictors of 5-fluorouracil toxicity in adjuvant colon cancer treatment (NCCTG N0147). *Journal of the National Cancer Institute*. 2014; 106
- (8). Johnson MR, Hageboutros A, Wang K, High L, Smith JB, Diasio RB. Life-threatening toxicity in a dihydropyrimidine dehydrogenase-deficient patient after treatment with topical 5-fluorouracil. *Clin Cancer Res*. 1999; 5:2006–11. [PubMed: 10473079]
- (9). Morel A, et al. Clinical relevance of different dihydropyrimidine dehydrogenase gene single nucleotide polymorphisms on 5-fluorouracil tolerance. *Mol Cancer Ther*. 2006; 5:2895–904. [PubMed: 17121937]
- (10). Johnson MR, Wang K, Diasio RB. Profound dihydropyrimidine dehydrogenase deficiency resulting from a novel compound heterozygote genotype. *Clin Cancer Res*. 2002; 8:768–74. [PubMed: 11895907]
- (11). Amstutz U, Farese S, Aebi S, Largiader CR. Dihydropyrimidine dehydrogenase gene variation and severe 5-fluorouracil toxicity: a haplotype assessment. *Pharmacogenomics*. 2009; 10:931–44. [PubMed: 19530960]
- (12). van Kuilenburg AB, et al. Intragenic deletions and a deep intronic mutation affecting pre-mRNA splicing in the dihydropyrimidine dehydrogenase gene as novel mechanisms causing 5-fluorouracil toxicity. *Hum Genet*. 2010; 128:529–38. [PubMed: 20803296]
- (13). Offer SM, Lee AM, Mattison LK, Fossum C, Wegner NJ, Diasio RB. A DPYD variant (Y186C) in individuals of african ancestry is associated with reduced DPD enzyme activity. *Clinical pharmacology and therapeutics*. 2013; 94:158–66. [PubMed: 23588312]
- (14). Zaanani A, Dumont LM, Lorient MA, Taieb J, Narjoz C. A case of 5-FU-related severe toxicity associated with the p.Y186C DPYD variant. *Clinical pharmacology and therapeutics*. 2014; 95:136. [PubMed: 24037119]
- (15). Saif MW, Lee AM, Offer SM, McConnell K, Relias V, Diasio RB. A DPYD variant (Y186C) specific to individuals of African descent in a patient with life-threatening 5-FU toxic effects: potential for an individualized medicine approach. *Mayo Clinic proceedings*. 2014; 89:131–6. [PubMed: 24388031]
- (16). Tishkoff SA, et al. The genetic structure and history of Africans and African Americans. *Science*. 2009; 324:1035–44. [PubMed: 19407144]
- (17). Shriner D, Tekola-Ayele F, Adeyemo A, Rotimi CN. Genome-wide genotype and sequence-based reconstruction of the 140,000 year history of modern human ancestry. *Scientific reports*. 2014; 4:6055. [PubMed: 25116736]
- (18). Ishikawa Y, et al. Dihydropyrimidine dehydrogenase and messenger RNA levels in gastric cancer: possible predictor for sensitivity to 5-fluorouracil. *Jpn J Cancer Res*. 2000; 91:105–12. [PubMed: 10744051]
- (19). Database of Single Nucleotide Polymorphisms (dbSNP). <http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?showRare=on&chooseRs=coding&go=Go&locusId=1806>. Accessed August 31 2015
- (20). Offer SM, Fossum CC, Wegner NJ, Stuflesser AJ, Butterfield GL, Diasio RB. Comparative Functional Analysis of DPYD Variants of Potential Clinical Relevance to Dihydropyrimidine Dehydrogenase Activity. *Cancer Res*. 2014; 74:2545–54. [PubMed: 24648345]
- (21). Offer SM, Wegner NJ, Fossum C, Wang K, Diasio RB. Phenotypic profiling of DPYD variations relevant to 5-fluorouracil sensitivity using real-time cellular analysis and in vitro measurement of enzyme activity. *Cancer Res*. 2013; 73:1958–68. [PubMed: 23328581]
- (22). Wei X, McLeod HL, McMurrough J, Gonzalez FJ, Fernandez-Salguero P. Molecular basis of the human dihydropyrimidine dehydrogenase deficiency and 5-fluorouracil toxicity. *J Clin Invest*. 1996; 98:610–5. [PubMed: 8698850]

- (23). van Kuilenburg AB, et al. Evaluation of 5-fluorouracil pharmacokinetics in cancer patients with a c.1905+1G>A mutation in DPYD by means of a Bayesian limited sampling strategy. *Clin Pharmacokinet.* 2012; 51:163–74. [PubMed: 22339448]
- (24). 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.* 2008; 26:2131–8. [PubMed: 18299612]
- (25). Caudle KE, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for dihydropyrimidine dehydrogenase genotype and fluoropyrimidine dosing. *Clinical pharmacology and therapeutics.* 2013; 94:640–5. [PubMed: 23988873]
- (26). Vreken P, Van Kuilenburg AB, Meisma R, van Gennip AH. Identification of novel point mutations in the dihydropyrimidine dehydrogenase gene. *J Inherit Metab Dis.* 1997; 20:335–8. [PubMed: 9266349]
- (27). Boige V, et al. DPYD Genotyping to Predict Adverse Events Following Treatment With Fluorouracil-Based Adjuvant Chemotherapy in Patients With Stage III Colon Cancer: A Secondary Analysis of the PETACC-8 Randomized Clinical Trial. *JAMA oncology.* 2016
- (28). Ezzeldin H, Okamoto Y, Johnson MR, Diasio RB. A high-throughput denaturing high-performance liquid chromatography method for the identification of variant alleles associated with dihydropyrimidine dehydrogenase deficiency. *Anal Biochem.* 2002; 306:63–73. [PubMed: 12069415]
- (29). Adzhubei IA, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010; 7:248–9. [PubMed: 20354512]
- (30). Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols.* 2009; 4:1073–81. [PubMed: 19561590]
- (31). Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One.* 2012; 7:e46688. [PubMed: 23056405]
- (32). Pettersen EF, et al. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry.* 2004; 25:1605–12. [PubMed: 15264254]
- (33). Dobritzsch D, Schneider G, Schnackerz KD, Lindqvist Y. Crystal structure of dihydropyrimidine dehydrogenase, a major determinant of the pharmacokinetics of the anti-cancer drug 5-fluorouracil. *EMBO J.* 2001; 20:650–60. [PubMed: 11179210]

Study Highlights

What is the current knowledge on the topic?

Most clinical pharmacogenetic studies of 5-FU toxicity have focused on European and European American populations with only limited information from other racial groups.

What question did this study address?

In this study we examined the occurrence and frequency of *DPYD* variants in an East African population in our catchment area in Central and Southeastern Minnesota.

What this study adds to our knowledge?

The study identified novel *DPYD* variants that cause decreased DPD activity in this East African cohort and are potential risk alleles for 5-FU toxicity.

How this might change clinical pharmacology or translational science?

Data from this study may provide valuable guidance for individualizing 5-FU therapy in East African Cancer patients.

Figure 1.

A. Summary of *DPYD* variants that were identified in this study. **B.** The carrier frequencies of exonic non-synonymous variants in Somali and Kenyan populations.

Figure 2.

Relative DPD enzyme activity was determined *in vitro* for novel *DPYD* variants. For each variant, the horizontal bar represents the mean of four biologic replicates $\pm$ standard deviation (SD). Each biologic replicate is the average of three technical replicates. The mean activity of wild-type *DPYD*, and *2A *DPYD* were used as positive and negative controls, respectively. The horizontal dashed lines represent the SD for wild-type *DPYD*. Numbers represent the *P*-values for the respective deleterious variants compared against matched wildtype controls.

Figure 3.

Protein modeling was performed using UCSF Chimera version 1.10.2. Variants were mapped to the pig crystal structure (PDBID 1GTH). The DPD monomer consists of 5 structural domains that are highlighted with different colors (A). The position of novel *DPYD* variants are shown within the N-terminal 4Fe–4S cluster containing domain I (B), the FAD-binding domain II (C), NADPH-binding domain III (D), the pyrimidine-binding domain IV (E), and the C terminal 4Fe–4S cluster containing domain V (F). P1023S is located in domain V very close to the C-terminus, it is not included as part of the crystal structure.

Figure 4.

The genotype frequencies of (A) previously reported deleterious (intronic and non-synonymous exonic) variants, (B) reported non-synonymous benign variants, and (C) synonymous variants in our East African study population (EA) and 1000 Genomes populations (West African, AFR; Admixed American, AMR; East Asian, ASN; and European, EUR) ancestry are presented. The blue color represents individuals who are homozygous for the reference allele. The red color represents heterozygous individuals. The yellow color represents individuals who are homozygous for the variant allele.

Table 1

Demographic characteristics of study participants

Characteristics	Set A	Set B
Total number	270	318
Median age (range)	47 (20 – 80)	40 (19 – 86)
Gender n (%)		
Male	155 (57.4)	159 (50)
Female	115 (42.6)	159 (50)
Country of origin n (%)		
Somalia	270 (100)	278 (87.4)
Kenya	0 (0)	40 (12.6)

Table 2Description of novel *DPYD* variants identified by NGS and known 5-FU toxicity-associated *DPYD* variants

Variant	Nucleotide Change	Location	Ref/Alt	Geno [*]	MAF	<i>In silico</i> Prediction score [†]
Known risk variants[‡]						
*2A	c.1905+1G>A	In 14	C/T	588/0/0	0	Clinically deleterious
I560S	c.1679T>G	Ex 13	A/C	588/0/0	0	Clinically deleterious
D949V	c.2846A>T	Ex 22	T/A	588/0/0	0	Clinically deleterious
Non-synonymous						
P86L	c.257C>T	Ex 4	G/A	586/2/0	0.0017	3
P237L	c.710 C>T	Ex 7	G/A	585/2/0	0.0017	3
R394Q	c.1181 G>A	Ex 11	C/T	587/1/0	0.0009	1
A513V	c.1538 C>T	Ex 13	G/A	573/11/0	0.0094	0
S14_S16[dup]	c.1540_1548dup9	Ex 13	C/CAGAAACGGA	583/3/0	0.0026	1
L523F	c.1567 C>T	Ex 13	G/A	583/2/0	0.0017	1
I528V	c.1582 A>G	Ex 13	T/C	585/1/0	0.0009	0
R699T	c.2096 G>C	Ex 17	C/G	576/8/0	0.0069	2
T793I	c.2378C>T	Ex 19	G/A	587/1/0	0.0009	2
V941A	c.2822T>C	Ex 22	A/G	587/1/0	0.0009	0
I971Nfs_P1023S	c.2911_2912insA c.3067C>T	Ex 23 Ex 23	A/AT G/A	585/2/0 586/2/0	0.0017 0.0017	1 0
Synonymous						
R78R	c.234 A>G	Ex 4	T/C	582/2/0	0.0017	NA
V535V	c.1605 A>T	Ex 13	T/A	585/1/0	0.0009	NA
G661G	c.1983 A>G	Ex 16	T/C	582/1/0	0.0009	NA
A970A	c.2910 T>C	Ex 23	A/G	584/2/0	0.0017	NA

Abbreviations: In, intron; Ex, exon; Ref: reference (major) allele; Alt, alternative (minor) allele; MAF, Minor allele frequency; NA, not applicable.

^{*} Geno, the frequencies of genotypes presented as (Ref/Ref, Ref/Alt, Alt/Alt)[†] The potential effect of *DPYD* variants on the protein structure or function was predicted by Polymorphism Phenotyping-2 (PolyPhen-2), Sorting Intolerant From Tolerant (SIFT) and Protein Variation Effect Analyzer (PROVEAN). Qualitative predictors (benign, possibly-damaging, and probably-damaging for PolyPhen-2), (damaging and tolerated for Sift) and (deleterious and neutral for PROVEAN) were reported by the software. *In silico* Prediction scores were assigned as follows: **3**, the variant was predicted as damaging/deleterious by all 3 programs; **2**, the variant was predicted as damaging/deleterious by 2 programs; **1**, the variant was predicted as damaging/deleterious by 1 program; **0**, the variant was predicted as benign/neutral/tolerated.[‡] *2A, I560S and D949V are the most commonly reported *DPYD* variants that are known to be associated with 5-FU toxicity. They were not detected in this cohort.