Comparative Functional Analysis of DPYD Variants of Potential Clinical Relevance to Dihydropyrimidine Dehydrogenase Activity

Steven M. Offer¹, Croix C. Fossum¹, Natalie J. Wegner¹, Alexander J. Stuflesser¹,², Gabriel L. Butterfield¹, and Robert B. Diasio¹,²

Abstract

Dihydropyrimidine dehydrogenase (DPD) is the initial and rate-limiting enzyme of the uracil catabolic pathway, being critically important for inactivation of the commonly prescribed anti-cancer drug 5-fluorouracil (5-FU). DPD impairment leads to increased exposure to 5-FU and, in turn, increased anabolism of 5-FU to cytotoxic nucleotides, resulting in more severe clinical adverse effects. Numerous variants within the gene coding for DPD, DPYD, have been described, although only a few have been demonstrated to reduce DPD enzyme activity. To identify DPYD variants that alter enzyme function, we expressed 80 protein-coding variants in an isogenic mammalian system and measured their capacities to convert 5-FU to dihydro-fluorouracil, the product of DPD catabolism. The M166V, E828K, K861R, and P1023T variants exhibited significantly higher enzyme activity than wild-type DPD (120%, P = 0.025; 116%, P = 0.049; 130%, P = 0.0077; 138%, P = 0.048, respectively). Consistent with clinical association studies of 5-FU toxicity, the D949V substitution reduced enzyme activity by 41% (P = 0.0031).

Enzyme activity was also significantly reduced for 30 additional variants, 19 of which had <25% activity. None of those 30 variants have been previously reported to associate with 5-FU toxicity in clinical association studies, which have been conducted primarily in populations of European ancestry. Using publicly available genotype databases, we confirmed the rarity of these variants in European populations but showed that they are detected at appreciable frequencies in other populations. These data strongly suggest that testing for the reported deficient DPYD variations could dramatically improve predictive genetic tests for 5-FU sensitivity, especially in individuals of non-European descent. Cancer Res; 74(9); 2545–54. ©2014 AACR.

Introduction

Genetic polymorphisms in the dihydropyrimidine dehydrogenase (DPD) gene (DPYD) have emerged as predictive risk alleles for developing severe toxicity to the commonly prescribed anti-cancer drug 5-fluorouracil (5-FU). Three DPYD variants have consistently been reported to be associated with 5-FU toxicity and impaired DPD enzyme activity. The most studied of the DPYD variants, *2A (rs3918290; also known as 1VS1+1G>A) interrupts a splice acceptor sequence and causes the in-frame deletion of amino acids corresponding to exon 14 (1). Carriers of *2A have significantly reduced DPD enzyme levels, resulting in prolonged clearance times for 5-FU (2) and, as such, are more likely to develop adverse toxicity following treatment with the drug (3, 4). The second well-accepted DPD deficiency–associated variant, I560S (rs55886062), is exceptionally rare in the general population but has been consistently linked to reduced DPD activity (5) and increased incidence of 5-FU toxicity (6, 7). Clinical studies have also consistently shown association between a third variant, D949V, and severe toxicity following chemotherapy that included 5-FU (4, 7).

More than 100 additional missense variants have been reported for DPYD, many from large-scale sequencing efforts using individuals from various racial groups (8, 9). Few of these additional variants have been evaluated in case–control studies of 5-FU toxicity and those that have been studied have yielded unclear or conflicting results. For instance, the M166V variant was shown to strongly associate with grade III and IV toxicity in a cohort of patients with breast, gastroesophageal, or colorectal cancer treated with 5-FU–based therapy (10); however, additional reports failed to confirm a link between the variant and 5-FU toxicity (4, 11). Further studies have suggested that M166V might be protective against specific 5-FU–related toxicities in women (12).

To aid in developing predictive genetic tests of 5-FU toxicity, we determined the contributions of many of these additional DPYD variants to DPD activity. Our laboratory previously demonstrated the use of a recombinant system of protein
expression to measure the enzyme activity of a small set of DPD protein variants using human cells (13). We hypothesized that additional DPD variants may contribute to DPD deficiency, especially in populations not of European ancestry that have been under-represented in large case–control clinical association studies of 5-FU toxicity. In all, 80 DPD variants were expressed in mammalian cells and the enzyme activity of each variant measured. Thirteen variants (9 missense, 2 stop-gained, 1 frame-shift, and 1 in-frame insertion) had less than 12.5% enzymatic activity and were classified as “2A-like.” Six variants had enzyme activities similar to D949V (>25%, but significantly lower than wild type). Four variants showed enzyme activities that were significantly higher than wild type, similar to our previous findings for C29R and S534N (13). Consistent with our hypothesis, these newly classified DPD deficiency variants were present at higher frequencies in non-European populations.

Materials and Methods

In silico functional prediction

A list of missense DPD variants was compiled using the NCBI dbSNP (14), the 1000 Genomes Project (9), and the NIH Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project (ESP; ref. 8) databases. The PolyPhen-2 web server version 2.2.2 was used to predict the impact of amino acid changes on protein function (15) using the translated product of DPD transcript ENST00000370192 from UniProtKB/UniRef100 release 2011_12 as the reference protein sequence. PolyPhen-2 predictions rely upon a naïve Bayes classifier model trained using machine-learning algorithms applied to publicly available datasets. In silico data presented in Fig. 1 were determined using the prediction model trained with the HimDiv dataset (15). The estimated false-positive rate (FPR) for each variant is calculated by the software as the fraction of benign variants incorrectly classified as damaging for a given threshold of naïve Bayes probabilistic scores (15). Qualitative predictors (“benign,” “possibly damaging,” and “probably damaging”) are reported by the software based on the thresholds determined from estimated FPR values. Complete details regarding the algorithms and outputs of the PolyPhen-2 software have been detailed by the software’s developers (15, 16). Additional predictions were performed using PolyPhen-2 trained with the HumVar dataset, PROVEAN version 1.1.3 Mutation Assessor web server (17), SIFT version 4.0.3 (18), and the SNAP webserver (19) using the default settings.

Vector construction

Human DPD variant expression vectors were prepared as previously described (13) and confirmed by the Mayo Clinic Gene Analysis Shared Resource (Rochester, MN). Site-directed mutagenesis primers are listed in Supplementary Table S1.

Figure 1. Predicted impact of missense DPD variations on DPD protein function. A, the probability of each variation being damaging to protein function or structure was determined using PolyPhen-2. B, for each probability reported in A, the estimated FPR was calculated by the software. Each mark represents an individual amino acid substitution. Shading indicates the qualitative classifier predicted by PolyPhen-2 (black, benign; gray, possibly damaging; white, probably damaging).
Experimental design

DPYD variants were randomly divided into groups of 6 for functional evaluation. Each experiment consisted of a group of 6 variants tested in parallel with a positive control (wild-type DPYD) and a negative control (exon 14 deletion mimicking the +2A transcript). Experimental groups were tested in triplicate (technical replicates) with all 3 replicates being transfected, processed, and assayed in parallel. The results for all 3 technical replicates were pooled to constitute a biologic replicate. At least 3 independent biologic replicates were performed for each variant.

Cells

Low-passage HEK293T/c17 cells (culture CRL-11268) were obtained from the American Type Culture Collection and cultured as previously described (13). Aliquots of low-passage cell stocks were prepared within 2 weeks of receipt. Cells were maintained in culture for no more than 10 passages or 2 months. Cell lines were periodically monitored for mycoplasma by Hoechst staining (Sigma-Aldrich). Culture identity and health were monitored by microscopy. Population doubling times were determined by cell counting for each biologic replicate were centered and standardized scores (Z-scores) calculated by subtracting the mean of all data points in the replicate from each individual data point and dividing by the SD of all data points in the replicate. Unpaired 2-tailed Student t tests (assuming equal variance) were used to compare results for a given variant with the matched positive control (wild-type). For data presentation purposes, normalized results for each experiment were re-scaled such that the mean of the negative control was equal to 0 and that of the positive control was equal to 1.

Amino acid alignment and protein modeling

Amino acid sequences for DPYD were retrieved from NCBI for human (*Homo sapiens*, NP_000101.2), pig (*Sus scrofa*, NP_999209.1), mouse (*Mus musculus*, NP_740748.1), chicken (*Gallus gallus*, XP_426639.3), and zebrafish (*Danio rerio*, NP_998058.1). Multiple sequence alignment was performed using Clustal Omega version 1.2.0. Percent identities between amino acid sequences were calculated using NCBI BLASTP version 2.2.28+. Protein modeling was performed using UCSF Chimera version 1.8 (20) and Modeller version 9.12 (21). For homology modeling, the human DPYD protein (NP_000101.2) was used as a query sequence and the pig crystal structure corresponding to substrate-bound DPYD with closed active site loop (PDB ID 1GTH) was used as the template structure. All 3-dimensional protein images were prepared using UCSF Chimera.

Allele frequency data were obtained from the NHLBI ESP (8) and the 1000 Genomes Project (9) databases. The NHLBI ESP dataset is composed of whole exome sequence data from 2,203 African American and 4,300 European American individuals. 1000 Genomes Project data were from 246 individuals of West African ancestry (AFR), 181 of American ancestry (AMR), 286 of East Asian ancestry (ASN), and 379 of European ancestry (EUR). Additional details about the specific racial groups

---

**Screening of DPYD Variants for Altered 5-FU Catabolism**

DPYD enzyme activity assay

Enzyme activity was determined using a method described earlier by our laboratory (13). Briefly, lysates were incubated with 200 μmol/L NADPH (Sigma-Aldrich) and 82 μmol/L [6-C14]-5-fluorouracil ([6-C14]-5-FU; Moravek Biochemicals) for 30 minutes at 37°C. Conversion of [6-C14]-5-FU to [6-C14]-5-dihydrofluorouracil ([6-C14]-5-DHFU) was determined using a reverse-phase C18 HPLC column (Grace) connected to a PerkinElmer Radiomatic 62STR flow scintillation analyzer. DPYD activity was calculated by measuring the percent region of interest as the area under the curve for ([6-C14]-5-DHFU)/([6-C14]-5-FU) + [6-C14]-5-DHFU) using ProFSA software (PerkinElmer). Validation that ([6-C14]-5-FU) and NADPH were not limiting in reactions is presented in Supplementary Fig. S1.

Calculation of enzyme activity for DPYD variants

DPYD enzyme activity for a given variant was normalized by relative input amount of DPYD as measured by dot blot. Data for each biologic replicate were centered and standardized scores (Z-scores) calculated by subtracting the mean of all data points in the replicate from each individual data point and dividing by the SD of all data points in the replicate. Unpaired 2-tailed Student t tests (assuming equal variance) were used to compare results for a given variant with the matched positive control (wild-type). For data presentation purposes, normalized results for each experiment were re-scaled such that the mean of the negative control was equal to 0 and that of the positive control was equal to 1.

**Dot blotting**

Protein lysates were mixed 5:1 with reducing buffer (62.5 mmol/L Tris, pH 6.8, 2% SDS, 5% 2-mercaptoethanol) in a final volume of 50 μL and incubated at 98°C for 10 minutes. Serial dilutions were blotted to nitrocellulose using a BioDot microfiltration apparatus (BioRad). Membranes were blocked using Odyssey blocking buffer (LI-COR). Blots were probed with primary antibodies against DPYD (ab54797, AbCam) and α-tubulin (ab4074, AbCam) and subsequent secondary IRDye800CW-conjugated goat anti-mouse (92632210, LI-COR) and IRDye680-conjugated goat anti-rabbit antibodies (96268071, LI-COR). Blots were scanned and dot intensities quantified using the LI-COR Odyssey Infrared Imaging system. The background signal was determined by repeating quantification using a larger diameter for each dot and subtracting the original intensity value. Relative intensity was calculated as 

\[
\frac{d_{\text{Int}}}{d_{\text{Area}}} - \frac{d_{\text{Int}}}{d_{\text{Area}}} = \frac{[d_{\text{Int}} - d_{\text{Int}}]}{d_{\text{Area}} - d_{\text{Area}}},
\]

where \(d_{\text{Int}}\) is the total intensity of the small dot, \(d_{\text{Area}}\) is the area of the small dot, \(L_{d_{\text{Int}}}\) is the total intensity of the large dot, and \(L_{d_{\text{Area}}}\) is the area of the large dot.

---

**Experimental design**

DPYD variants were randomly divided into groups of 6 for functional evaluation. Each experiment consisted of a group of 6 variants tested in parallel with a positive control (wild-type DPYD) and a negative control (exon 14 deletion mimicking the +2A transcript). Experimental groups were tested in triplicate (technical replicates) with all 3 replicates being transfected, processed, and assayed in parallel. The results for all 3 technical replicates were pooled to constitute a biologic replicate. At least 3 independent biologic replicates were performed for each variant.

**Cells**

Low-passage HEK293T/c17 cells (culture CRL-11268) were obtained from the American Type Culture Collection and cultured as previously described (13). Aliquots of low-passage cell stocks were prepared within 2 weeks of receipt. Cells were maintained in culture for no more than 10 passages or 2 months. Cell lines were periodically monitored for mycoplasma by Hoechst staining (Sigma-Aldrich). Culture identity and health were monitored by microscopy. Population doubling times were determined by cell counting for each biologic replicate were centered and standardized scores (Z-scores) calculated by subtracting the mean of all data points in the replicate from each individual data point and dividing by the SD of all data points in the replicate. Unpaired 2-tailed Student t tests (assuming equal variance) were used to compare results for a given variant with the matched positive control (wild-type). For data presentation purposes, normalized results for each experiment were re-scaled such that the mean of the negative control was equal to 0 and that of the positive control was equal to 1.

**Amino acid alignment and protein modeling**

Amino acid sequences for DPYD were retrieved from NCBI for human (*Homo sapiens*, NP_000101.2), pig (*Sus scrofa*, NP_999209.1), mouse (*Mus musculus*, NP_740748.1), chicken (*Gallus gallus*, XP_426639.3), and zebrafish (*Danio rerio*, NP_998058.1). Multiple sequence alignment was performed using Clustal Omega version 1.2.0. Percent identities between amino acid sequences were calculated using NCBI BLASTP version 2.2.28+. Protein modeling was performed using UCSF Chimera version 1.8 (20) and Modeller version 9.12 (21). For homology modeling, the human DPYD protein (NP_000101.2) was used as a query sequence and the pig crystal structure corresponding to substrate-bound DPYD with closed active site loop (PDB ID 1GTH) was used as the template structure. All 3-dimensional protein images were prepared using UCSF Chimera.

**Allele frequency data were obtained from the NHLBI ESP (8) and the 1000 Genomes Project (9) databases. The NHLBI ESP dataset is composed of whole exome sequence data from 2,203 African American and 4,300 European American individuals. 1000 Genomes Project data were from 246 individuals of West African ancestry (AFR), 181 of American ancestry (AMR), 286 of East Asian ancestry (ASN), and 379 of European ancestry (EUR). Additional details about the specific racial groups
included in 1000 Genomes Project populations can be found on the project’s website (9).

Statistical tests
All data analyses and transformations were performed using JMP 9.0.3 (SAS Institute Inc.), unless otherwise noted. Additional tests and software algorithms used are described in relevant sections above.

Results
In silico prediction of DPYD variant effects
A total of 128 missense DPYD variants were compiled from NCBI dbSNP (14), 1000 Genomes Project (9), and NHLBI ESP (8) databases. A complete list of variants and allele frequencies is presented in Supplementary Table S2. To ascertain which variants were most likely to affect enzyme activity, the PolyPhen-2 software program (15) was used to predict the impact of each amino acid substitution on the structure and function of the DPYD protein. The predicted probability that a given variant is damaging to the protein is presented in Fig. 1A. The estimated FPR for each variant as calculated by PolyPhen-2 is presented in Fig. 1B. Of the 128 missense DPYD variants, 63 were predicted to be probably damaging (<5% estimated FPR), 15 possibly damaging (5%–10% estimated FPR), and 50 benign (>10% estimated FPR). Qualitative predictions for each variant are presented in Supplementary Table S2.

Functional study of variations
Functional studies were performed to directly measure the enzyme activity of 75 transgenically expressed missense DPYD variants using a mammalian system of protein production that we described previously (13). The selected variant pool was enriched for those that had been previously reported in clinical case–control studies and/or case reports relating to 5-FU toxicity or DPD deficiency. Thirty-seven of the selected variants were predicted to be probably damaging, 7 were predicted to be possibly damaging, and 31 were predicted to be benign. Five additional variants (2 truncation, 1 frame-shift, 1 in-frame insertion, and 1 somatic missense mutation) were also studied. Enzyme activity was corrected for differences in expression as determined by quantitative dot blot (Fig. 2A and B), which yielded comparable results to Western blotting (data not shown). A comparative summary of DPYD enzyme activities for selected variants is presented in Fig. 2C–F; numerical values and statistics for missense variants are also detailed in Supplementary Table S2.

Within the probably damaging group of missense variants, 19 (51%) showed significantly decreased enzyme activity compared with wild-type DPD (Fig. 2C). Of these, 8 variants had little to no residual enzymatic activity (<12.5%); including K958E (P = 1.3 × 10⁻⁵), S201R (P = 3.0 × 10⁻⁷), V995F (P = 6.9 × 10⁻⁵), G593R (P = 3.9 × 10⁻⁵), G880V (P = 3.5 × 10⁻⁶), G764D (P = 5.7 × 10⁻⁹), H978R (P = 9.6 × 10⁻⁴), and...
R592W ($P = 4.0 \times 10^{-5}$) and were thus classified as "2A-like" (Fig. 2C). Dramatic reductions in activity (12.5%–25% of wild-type, similar to our previous report for I560S; ref. 13) were noted for 6 variants: R235W ($P = 9.2 \times 10^{-5}$), Y211C ($P = 0.0021$), D495Q ($P = 0.0017$), R592Q ($P = 2.6 \times 10^{-5}$), D342N ($P = 1.9 \times 10^{-5}$), and S492L ($P = 2.3 \times 10^{-5}$; Fig. 2C). D949V, which had previously not been tested using this assay, showed significantly decreased enzyme function compared with wild-type (59% activity, $P = 0.0031$; Fig. 2C). Four additional variants had enzyme activities that were significantly lower than wild-type but greater than 25%: T760I ($P = 0.0035$), P92A ($P = 0.045$), Y304H ($P = 0.026$), and Y186C ($P = 0.027$; Fig. 2C). One variant, M166V, had significantly higher enzyme activity than wild-type DPD (120% activity, $P = 0.0025$; Fig. 2C). The highest average activity in this group was noted for L310S (133%); however, this result was not significantly different from wild-type due to high variability ($P = 0.076$; Fig. 2C).

Fewer variants showed significant reductions in enzyme activity in the possibly damaging and benign prediction groups. Of the 7 possibly damaging variants, 2 (29%) had significantly impaired enzyme function (Fig. 2D). F438L retained 47% of enzymatic activity ($P = 0.0042$) and D687A was 75% active ($P = 0.011$; Fig. 2D). Of the 31 amino acid changes predicted to not affect function, 5 (16%) had enzyme activities significantly lower than wild-type (Fig. 2E). R353C completely lacked enzymatic activity ($P = 2.0 \times 10^{-5}$; Fig. 2E). The enzyme activities of K290E, T983I, L352V, and N893S were 45%, 73%, 77%, and 80%, respectively ($P = 0.0019$, $P = 0.023$, $P = 0.0085$, and $P = 0.018$; Fig. 2E). Three variants in the predicted benign group showed significantly higher enzyme activities than wild-type DPD. E828K was 16% hyperactive ($P = 0.049$), K861R was 30% hyperactive ($P = 0.0077$), and P1023T was 38% hyperactive ($P = 0.048$; Fig. 2E).

We tested 4 additional variants that affected more than 1 amino acid. These included the truncation variants R21X (rs72549310) and E386X (rs78060119), the P633[FS] [rs72549303), and the F100[FS] in-frame 3-nucleotide insertion (rs72549301). In addition, G252V (NM_000101.2-c.755G>T, COSM74340), a somatic missense mutation contained in the Catalogue of Somatic Mutations in Cancer database (22, 23) that was originally identified in an ovarian tumor specimen analyzed as part of The Cancer Genome Atlas project (24), was also assayed. As expected, neither truncation variant, R21X or E386X, yielded detectable enzyme activity (Fig. 2F). Protein fragments corresponding to the predicted sizes were detectable for both variants by Western blotting using an anti-DPD antibody; however, both were expressed at lower levels than wild-type DPD (data not shown). Both P633[FS] and F100[FS] also lacked detectable enzyme activity (Fig. 2F). The enzyme activity of the somatic G252V mutation was significantly lower than wild-type DPD (51% activity, $P = 1.2 \times 10^{-5}$; Fig. 2F), suggesting that de novo tumor mutations can affect catabolism of 5-FU within cancerous cells.

**Performance of in silico prediction tools**

To assess the use of in silico tools for predicting the effects of missense *DPYD* variations on enzyme activity, functional predictions were compared with the actual enzyme activity results for the *DPYD* variants tested (Supplementary Table S3). Polyphen-2 trained with the HumDiv dataset showed a sensitivity of 81% and a specificity of 53%. Tradeoffs between sensitivity and specificity, as well as between negative predictive value (NPV) and positive predictive value (PPV), are noted when low confidence possibly damaging predictions were treated as benign predictions and/or the threshold used to classify variants based on enzyme activity was adjusted (Supplementary Table S3). Overall, balanced accuracy was highest when low confidence predictions were excluded from the deleterious set and only variants with <25% activity were considered deficient. It is notable that while these adjustments to the prediction criteria generally increased sensitivity, specificity, and NPV, as a consequence, PPV was reduced to 38%. Similar performance was noted for PolyPhen-2 trained with the HumanVar dataset, and a comparison with predictions from the PROVEAN (17), SIFT (18), and SNAP (19) software programs is also presented (Supplemental Table S3).

**Modeling the human DPD protein structure**

DPD is highly conserved through vertebrates, with 93% identity between the pig and human amino acid sequences (Fig. 3). The crystal structure of human DPD has not been solved; however, the structure of pig DPD has been reported (25). To identify structural elements that may be disrupted in dysfunctional DPD variants, we generated a theoretical homology model of human DPD using the crystal structure of pig DPD with NADPH and 5-iodouracil as a template. The predicted human structure contained 47 alpha helices and 31 beta sheets (Fig. 3). Of the 27 variants with reduced enzyme activity (Fig. 2C–F), 14 were located in predicted secondary structural elements (Fig. 3). Eleven deficient variants were located in alpha helical domains (Y186C, S201R, D342N, L352V, R353C, S492L, D495G, G880V, N893S, K958E, and V995F) and 3 deficient variants were in beta sheets (Y304H, R592Q, and R592W).

**Location of deficient variants on the DPD structure**

The DPD monomer consists of 5 distinct structural domains, each of which contains a subset of the prosthetic groups and cofactors necessary for enzyme function (26). Domain I contains 2 iron–sulfur (Fe-S) clusters (residues 27–172). The P92A variant was shown to be located at a conserved residue adjacent to the Fe–S coordinating cysteine at position 91 (Fig. 3) and constituted the only deleterious variant located in this domain. Domain II (residues 173–286 and 424–524) and domain III (residues 287–411) are closely intertwined and bind FAD and NADPH, respectively (Fig. 4B). Thirteen of 29 variations tested in the combined domains II and III significantly impaired enzyme function. R235W disrupts a residue important for FAD binding (Fig. 3). D342N and F438L are both located within NADPH coordinating regions (Figs. 3 and 4B). Domain IV contains the uracil-binding site and consists of residues 525–847 (Fig. 4C). G674D is located within the active loop structure, and D687A is located adjacent to the loop (Fig. 3). In all, 4 of 18 variants tested in this region had significantly reduced activity. Domain V (residues 1–26 and...
848–1025) contains 2 additional Fe-S clusters (Fig. 4D). Seven of 16 variants in this region were deficient. Notably, K958E, H978R, and V995F are all located within Fe-S cluster coordinating domains (Fig. 3).

Allele frequencies of deficient alleles
Allele frequencies for deficient variants were obtained from the NHLBI ESP (8) and the 1000 Genomes Project (9) databases and are summarized in Fig. 5 (full data are presented in Supplementary Table S2). Within the NHLBI ESP dataset, the allele frequencies for the 5-FU toxicity-associated DPYD variants *C3_2A, I560S, and D949V were 0.09%, 0.00%, and 0.09%, respectively, in African American individuals. Collectively, variants showing less than 12.5% enzyme activity (denoted as *C3_2A-like) had an allele frequency of 0.43% in the African American population. The additive allele frequencies for I560S-like and D949V-like variants were 0.02% and 2.22%, respectively, suggesting that the newly classified variants may be significant contributors to DPD deficiency in African Americans. Within the European American cohort of the NHLBI ESP dataset, the additive allele frequency of newly classified deficient variants was 0.08%, whereas the frequency of *2A, I560S, and D949V was 0.58%, 0.06%, and 0.54%, respectively.

While the 1000 Genomes Project database has information from fewer individuals than the NHLBI ESP dataset, it contains genotyping information from additional racial groups. Within the AFR 1000 Genomes population, the addition of newly classified variants increases the cumulative allele frequency of deleterious variants to 2.64%, up from 0.41% when only *C3_2A, I560S, and D949V were considered (Fig. 5). The allele frequency

Figure 4. Location of deficient variants relative to major structural domains of DPD. The position of damaging amino acid variants are depicted within the N-terminal Fe-S cluster containing alpha helical domain I (A), the FAD-binding domain II and NADPH-binding domain III (B), the pyrimidine-binding domain IV (C), and the C terminal Fe-S cluster containing domain V (D). Deteriorous amino acid changes are colored as described for Fig. 2. The position of IS605S, which was not evaluated in this study, is shown in C for the benefit of the reader.

Figure 3. Location of variants relative to structural features of DPD. An alignment of vertebrate DPD amino acid sequences was prepared using Clustal Omega. Sequence consensus is denoted below the alignment (asterisk, fully conserved residue; colon, strongly similar properties; period, weakly similar properties; no symbol, dissimilar amino acids). Predicted secondary structural features are indicated by colored bars above the alignment (red, alpha helices; blue, beta turns). Differences between the predicted human structure and the published pig structure are denoted by horizontal (not present in predicted structure) and vertical (not present in published structure) white lines overlaid on the secondary structural prediction. Boxes indicate functional elements (4Fe-4S, iron-sulfur coordinating domains; FAD, FAD-binding domain; NADP, NADP-binding domain; FMN, FMN-binding domain; UBC, uracil-binding domain; ASL, active site loop; F, FAD-interacting residue; N, NADP-interacting residue; P2, FMN-interacting residue; PA, pyrimidine-associated residue). Amino acid changes that significantly reduced enzyme activity are colored as described for Fig. 2. In instances where multiple amino acid variations have been reported for a given residue (i.e., R592W and R592Q), the color indicates the more severe phenotype.
of deleterious variants in the AMR population doubled from 0.28% to 0.56% when newly classified variants were added. '2A, I560S, or D949V were not detected in the ASN population; however, the allele frequencies of newly classified '2A-like and D949V-like variants were 0.17% and 0.52%, respectively. Finally, it was noted that the EUR population did not carry any of the newly identified deficient DPYD variants.

Discussion

A total of 128 missense variants have been reported within the DPYD gene, which alter 119 (12%) of the 1,025 amino acids that comprise DPYD (in 9 instances, variations affect the same codon). Relatively few of these variants have been evaluated in the clinical context and far fewer have been directly studied using functional tests. Three DPYD variants ('2A, I560S, and D949V) are generally considered to be deficiency-associated alleles, whereas many of the remaining variants have been dismissed as being too low penetrance to be of pharmacologic importance. Contrary to this opinion, our results indicate that rare DPYD variants that perturb function of the translated DPD protein are collectively present at sufficiently high frequencies in non-European populations to be considered as candidate risk alleles for developing severe adverse toxicity to 5-FU-based treatments.

The extensively studied '2A, I560S, and D949V variants are most likely to be detected in individuals of European ancestry but are far less penetrant in other racial groups (Fig. 5). Using a system of in vitro DPD enzyme activity measurement, we systematically classified 80 reported variations in DPYD that alter the amino acid sequence of the encoded protein. We show that the enzyme activity of D949V is significantly impaired but not to the extent observed previously for '2A or I560S (13). Thirty of the additional variants tested in the present study were shown to be deficient; 19 of which were severely deficient with enzyme activities similar to those of '2A or I560S (<25% of wild type; Fig. 3). On the basis of publicly available allele frequency data, these deficient variants are expected to be exceedingly rare or nonexistent in European populations but are collectively carried by an appreciable fraction of non-European individuals (Fig. 5). For instance, the cumulative allele frequency for '2A, I560S, and D949V in the ESP African American population was less than 0.2%. In contrast, the cumulative allele frequency of the newly classified deficient variants was approximately 2.7% in that population. In addition, within the ASN population of the 1000 Genomes Project, no individuals carried '2A, I560S, or D949V; however, newly classified deficient variants had an additive allele frequency of 0.7% (Fig. 5). While the population sizes for racial groups in the 1000 Genomes Project are relatively small, these findings suggest that the newly classified deficient variants may account for a significant fraction of DPD deficiency, particularly in individuals with non-European ancestry. Additional studies to ascertain the allele frequencies of DPYD variants in other under-represented populations are underway in our laboratory.

Recently, we examined the enzyme activity of common DPYD variants, C29R, E828K, K861R, and V732I (13). None of these variants showed decreased enzymatic activity relative to wild-type DPD. Surprisingly although, C29R and S534N showed increased 5-FU catabolism, establishing a new class of hyperactive DPYD variations. In the present study, 4 additional variants had significantly higher enzyme activity than wild-type: M166V, E928K, K861R, and P1023T. M166V was initially reported in 2 individuals with partial DPD deficiencies, one of whom also carried D949V (27). A subsequent study of M166V in an extended family containing DPD deficiency showed that the variant did not contribute to the disorder (6). One clinical case–control study with a limited number of patients suggested that M166V may contribute to 5-FU–related toxicities in gastroesophageal and breast cancers, but not in colorectal cancer (10). Three additional clinical association studies failed to establish any link between M166V and 5-FU–related toxicities (4, 11, 28). A fourth study suggested that M166V may protect against hematologic toxicity and neutropenia following administration of 5-FU; however, the association was observed only in women (12). The P1023T variant is rare in European populations and is more common in individuals of African ancestry (Supplementary Table S2). The hyperactive phenotype of P1023T is supported by a previous study from our laboratory in which individuals carrying P1023T, but not the deficiency allele Y186C, had 22% higher DPD activity than those that did not carry P1023T (5). To our knowledge, the hyperactive variants E928K and K861R have not been reported in any clinical or case studies.

Numerous case reports support our findings for many of the variants classified as deficient in this study. In early studies of severe DPD deficiency in children with delayed motor skills development, van Kuilenburg and colleagues identified H978R (29) and subsequently S201R, Y211C, and S492L (30) as potential contributors to DPD deficiency. The truncation variant E386X was reported in a DPD-deficient Japanese individual (31), and R592W was detected in a
Korean individual who experienced grade IV toxicity after receiving 5-FU (32). R21X was detected in an individual also carrying *2A who experienced severe toxicity and died 21 days following 5-FU treatment (33). R235W and the frameshift variant P633[FS] were detected in a compound heterozygous state in a severely DPD-deficient individual (34). V995F was likewise reported in an individual with DPD deficiency (35).

In a previous report, we showed that in a cohort of healthy African American individuals, DPD enzyme activity in peripheral blood mononuclear cells (PBMC) from Y186C carriers was 46% lower than that in non-carriers ($P = 2.9 \times 10^{-4}$), suggesting that the amino acid change impaired the catalytic activity of the enzyme (5). In the present article, we observed that DPD containing the Y186C amino acid substitution was reduced by approximately 15% relative to wild-type. This reduction is less than we expected based on the data from healthy volunteers but is similar to a previously reported study focused on this variant (36). We hypothesized that Y186C may affect dimerization, as the residue is located on the monomer surface, and previous reports have suggested that tyrosine to cysteine substitutions could cause aberrant dimer crosslinking (37). Multiple additional missense variants have been detected in individuals carrying Y186C (5); the potential implication of co-expressed DPD variants to overall DPD activity needs further clarification.

To date, predictive tests of 5-FU toxicity have had limited value, as carriers of known DPD deficiency–associated alleles (*2A, I560S, and D949V) constitute a relatively small percentage of toxicity cases (38). In a previous study, we confirmed that *2A and I560S were deleterious DPD variants with dramatically impaired DPD enzyme activity (13). In the present study, we confirmed D949V as a deleterious variant and presented comparative data showing that at least 30 additional DPD variants impair DPD function. Some of these variants have been documented in case reports; however, the rarity of these alleles has prevented the determination of their statistical significance as predictive markers of 5-FU toxicity. The results presented in this article should address that gap and provide guidance for the individualization of 5-FU therapy for carriers of rare, damaging DPD variants. Our findings also highlight the importance of performing genetic analyses that are unbiased by previous studies conducted in populations of limited diversity.

**Disclosure of Potential Conflicts of Interest**
No potential conflicts of interest were disclosed.

**Authors’ Contributions**
Conception and design: S.M. Offer, C.C. Fossum, A.J. Stuflesser, R.B. Diasio
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): S.M. Offer, C.C. Fossum, N.J. Wegner, A.J. Stuflesser, G.L. Butterfield
Analysis and interpretation of data (e.g., statistical analysis, bioinformatics, computational analysis): S.M. Offer, C.C. Fossum, N.J. Wegner, A.J. Stuflesser
Writing, review, and/or revision of the manuscript: S.M. Offer, C.C. Fossum, A.J. Stuflesser, R.B. Diasio
Study supervision: S.M. Offer, R.B. Diasio

Acknowledgments
The authors thank the Mayo Clinic Center for Clinical and Translational Science for statistical support and the Mayo Clinic Cancer Center Gene Analysis Shared Resource for sequencing services.

**Grant Support**
This study was funded by NIH grant CA116964 (R.B. Diasio). S.M. Offer was financially supported in part by a grant from the Mayo Clinic Center for Individualized Medicine. This study was also supported by the Mayo Clinic Center for Clinical and Translational Science (NCATS U1L TR000135) and the Mayo Clinic Cancer Center Gene Analysis Shared Resource (NCI 5P30 CA15083-37).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received August 30, 2013; revised February 28, 2014; accepted March 6, 2014; published Online First March 19, 2014.

**References**


Comparative Functional Analysis of DPYD Variants of Potential Clinical Relevance to Dihydropyrimidine Dehydrogenase Activity

Steven M. Offer, Croix C. Fossum, Natalie J. Wegner, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-13-2482

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2014/03/20/0008-5472.CAN-13-2482.DC1

Cited articles
This article cites 33 articles, 11 of which you can access for free at:
http://cancerres.aacrjournals.org/content/74/9/2545.full#ref-list-1

Citing articles
This article has been cited by 5 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/74/9/2545.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.