Biomedicine & Pharmacotherapy 154 (2022) 113644

Contents lists available at ScienceDirect

Biomedicine & Pharmacotherapy

journal homepage: www.elsevier.com/locate/biopha

Rare genetic variant burden in DPYD predicts severe fluoropyrimidine-related toxicity risk

Elena De Mattia [a], Marco Silvestri [b], Jerry Polesel [c], Fabrizio Ecca [a], Silvia Mezzalira [a], Lucia Scarabel [a], Yitian Zhou [d], Rossana Roncato [a], Volker M. Lauschke [d,e,f], Stefano Calza [g], Michele Spina [h], Fabio Puglisi [h,i], Giuseppe Toffoli [a,*], Erika Cecchin [a,*]

[a] Experimental and Clinical Pharmacology, Centro di Riferimento Oncologico di Aviano (CRO) IRCCS, via Franco Gallini n. 2, 33081 Aviano PN, Italy
[b] Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Department of Applied Research and Technological Development, Via Giacomo Venezian 1, 20133 Milano, Italy
[c] Unit of Cancer Epidemiology, Centro di Riferimento Oncologico di Aviano (CRO) IRCCS, via Franco Gallini n. 2, 33081 Aviano PN, Italy
[d] Department of Physiology and Pharmacology, Karolinska Institutet, 171 77 Stockholm, Sweden
[e] Dr Margarete Fischer-Bosch Institute of Clinical Pharmacology, Auerbachstraße 112, 70376 Stuttgart, Germany
[f] University of Tuebingen, Geschwister-Scholl-Platz, 72074 Tuebingen, Germany
[g] University of Brescia, Department of Molecular and Translational Medicine, Viale Europa 11, 25123 Brescia, Italy
[h] Department of Medical Oncology, Centro di Riferimento Oncologico di Aviano (CRO), IRCSS, via Franco Gallini n. 2, 33081 Aviano PN, Italy
[i] Department of Medicine, University of Udine, Via delle Scienze, 206, 33100 Udine UD, Italy

ARTICLE INFO

Keywords:
Ds
DPYD
Rare variant
Fluoropyrimidine
Toxicity
Next-generation sequencing
Clinical implementation

ABSTRACT

Preemptive targeted pharmacogenetic testing of candidate variations in *DPYD* is currently being used to limit toxicity associated with fluoropyrimidines. The use of innovative next generation sequencing (NGS) approaches could unveil additional rare (minor allele frequency <1%) genetic risk variants. However, their predictive value and management in clinical practice are still controversial, at least partly due to the challenges associated with functional analyses of rare variants. The aim of this study was to define the predictive power of rare *DPYD* variants burden on the risk of severe fluoropyrimidine-related toxicity.

The *DPYD* coding sequence and untranslated regions were analyzed by NGS in 120 patients developing grade 3–5 (NCI-CTC vs3.0) fluoropyrimidine-related toxicity and 104 matched controls (no-toxicity). The functional impact of rare variants was assessed using two different *in silico* predictive tools (i.e., Predict2SNP and ADME Prediction Framework) and structural modeling. Plasma concentrations of uracil (U) and dihydrouracil (UH2) were quantified in carriers of the novel variants.

Here, we demonstrate that the burden of rare variants was significantly higher in patients with toxicity compared to controls ($p = 0.007$, Mann-Whitney test). Carriers of at least one rare missense *DPYD* variant had a 16-fold increased risk in the first cycle and an 11-fold increased risk during the entire course of chemotherapy of developing a severe adverse event compared to controls ($p = 0.013$ and $p = 0.0250$, respectively by multinomial regression model). Quantification of plasmatic U/UH2 metabolites and *in silico* visualization of the encoded protein were consistent with the predicted functional effect for the novel variations.

Analysis and consideration of rare variants by *DPYD*-sequencing could improve prevention of severe toxicity of fluoropyrimidines and improve patients' quality of life.

Abbreviations: ADME, absorption distribution metabolism and excretion; APF, ADME-optimized prediction framework; CI, confidence interval; DEL, deletion; DPD *DYPD*, dihydropyrimidine dehydrogenase; INS, insertions; MAF, minor allele frequency; NGS, next-generation sequencing; OR, odds ratio; SNP, single nucleotide polymorphisms; U, uracil; UH2, dihydrouracil; UTR, untranslated region.

* Corresponding authors.

E-mail addresses: edemattia@cro.it (E. De Mattia), marco.silvestri@istitutotumori.mi.it (M. Silvestri), polesel@cro.it (J. Polesel), ecca.fabrizio@gmail.com (F. Ecca), silvia.mezzalira.23@gmail.com (S. Mezzalira), lucia.scarabel@cro.it (L. Scarabel), yitian.zhou@ki.se (Y. Zhou), rroncato@cro.it (R. Roncato), volker.lauschke@ki.se (V.M. Lauschke), stefano.calza@unibs.it (S. Calza), mspina@cro.it (M. Spina), fabio.puglisi@cro.it (F. Puglisi), gtoffoli@cro.it (G. Toffoli), ececchin@cro.it (E. Cecchin).

https://doi.org/10.1016/j.biopha.2022.113644
Received 20 June 2022; Received in revised form 23 August 2022; Accepted 31 August 2022
Available online 2 September 2022

1. Introduction

Fluoropyrimidines represent the cornerstone of several antineoplastic regimens used to treat a broad spectrum of solid tumors. Although fluoropyrimidines are generally well tolerated, they cause severe toxicity in up to 30 % of patients, leading to patient death in approximately 1 % of cases. [1,2]. Dihydropyrimidine dehydrogenase (DPD, encoded by the *DYPD* gene) is the first and rate-limiting enzyme in the detoxification pathway of fluoropyrimidines and its deficiency could lead to overexposure and increased toxicity. About 3–5 % of Caucasians have partial deficiency of DPD, while 0.2 % of patients have complete deficiency [3]. Four *DPYD* variants (*DPYD*2A,* rs3918290*; DPYD*13,* rs55886062; c.2846A>T, rs67376798; c.1236G>A-HapB3; rs56038477) are associated with complete or partial DPD deficiency and are currently validated for their clinical impact on fluoropyrimidine-related toxicity. DPD loss is associated with impaired conversion of dihydrouracil (UH2) to uracil (U), and analysis of the concentration of these metabolites and their ratio in patients' plasma could represent a surrogate marker of intra-cellular DPD activity [4]. Specific guidelines for drug dose adjustments have been developed for the clinical translation of *DPYD* testing, and pre-treatment analysis is now recommended by regulatory agencies throughout Europe [1,5,6]. These four *DPYD* variants, while useful for identifying at-risk individuals, explain only a fraction (approximately 17 %) of patients who experience severe fluoropyrimidine-related toxicity [7].

Compared with conventional targeted genotyping strategies, high-throughput sequencing technologies have opened up the possibility of obtaining a more complete picture of candidate genes variability by capturing the full spectrum of variation, including rare variants. The Exome Aggregation Consortium has reviewed data from population-scale sequencing programs (e.g., 1000 Genomes Program) showing that most human germline variants are rare (Minor Allele Frequency, MAF, <1 %) [8] and that the probability of a variant being deleterious is inversely related to its frequency [9]. Among clinically tested pharmacogenes, *DPYD* has emerged as one of the pharmacogenes with the largest number of relevant variants not detected by conventional targeted genotyping [10,11], making it a good candidate for more comprehensive profiling. Recent clinical case reports have shown that retrospective next-generation sequencing (NGS) is useful for detecting rare and novel variants in *DPYD* that appear to cause severe toxic reactions to fluoropyrimidine treatment [12–15]. It is therefore likely that an upfront complete sequencing of the *DPYD* gene and the characterization of rare, very rare, and novel *DYPD* variants could intercept a significant percentage of fluoropyrimidine-related toxicities, with critical implications for genotype-driven drug prescribing [11].

However, the added benefit of systematic clinical application of pre-treatment *DPYD* NGS remains questionable due to issues related to the functional interpretation of detected variants that could provide advice for the management of fluoropyrimidine treatment. Recently, it has been reported that the burden of rare genetic variants in pharmacogenes may be an important marker of drug-associated toxicity, regardless of their functional effect, underscoring that frequency may be a criterion for selecting potentially harmful variants [16]. Moreover, innovative bioinformatics tools have recently been developed to analyze *in silico* the functional role of variants related to the absorption, distribution, metabolism, and excretion (ADME) of drugs that could be integrated into clinical diagnostic pipelines [17].

The main objective of our study was to investigate the predictive role of rare variants in *DPYD* on the risk of developing severe fluoropyrimidine-related toxicity. To this end, an integrated approach involving next-generation *DPYD* sequencing in conjunction with a series of *in silico* functional analyses of rare, very rare, and novel genetic variants, was applied to 120 individuals with severe fluoropyrimidine-related toxicity and 104 matched control subjects.

2. Patients and methods

2.1. Patient cohorts and clinical data collection

A group of cancer patients who developed severe fluoropyrimidine-related toxicity was selected from a biobank of clinical cases enrolled in prospective pharmacogenetic studies aimed at defining predictive markers of fluoropyrimidine-related toxicity at the Clinical and Experimental Pharmacology Unit of the Centro di Riferimento Oncologico (CRO), Aviano (PN) [18], Italy, based on the following criteria: (1) histologically confirmed diagnosis of solid cancer; (2) available peripheral blood sample; (3) available detailed clinical data (4); assumption of treatment containing fluoropyrimidines (5-FU or capecitabine); (5) absence of recommended genetic variants (i.e., *DPYD**2A, *DPYD**13, c.2846A>T, c.1236G>A-HapB3); and (6) development of at least one episode of hematological or non-hematological fluoropyrimidine-related toxicity of grade $\geq$ 3 according to the Common Terminology Criteria for Adverse Events (CTCAE) v.3.0. This group is hereinafter referred to as the "toxicity group".

An additional group of patients who did not develop toxicity during treatment ("no-toxicity group") was selected from the same biobank [18], using the same eligibility criteria as above except for item (6). Patients in the "no-toxicity group" were selected to be not statistically different from the "toxicity group" in terms of sex, age, tumor type, treatment setting and chemotherapy regimen.

Clinical and toxicity data were obtained from patients' medical records as previously described [18]. The maximum grade of toxicity experienced by the patients during the entire chemotherapy course was considered in the selection of patients.

Through the selection process, 120 patients were identified for the "toxicity group" and 104 for the "no-toxicity group".

All patients in the study were self-reported to be Caucasian. The study protocol complied with the ethical guidelines of the 1975 Declaration of Helsinki. All patients signed written informed consent approved by the local Ethical Committee before entering the study. All experiments were performed in accordance with the relevant guidelines and regulations of the Centro di Riferimento Oncologico (CRO), Aviano (PN), Italy.

2.2. Library preparation and sequencing

Information on *DPYD* gene variants was extracted from sequencing data obtained by the analysis of a panel comprising 54 fluoropyrimidine-related genes using an NGS-based method.

Genomic DNA was extracted from peripheral blood samples using the BioRobot EZ1 automated extractor, in association with the "EZ1 DNA Blood Kit 350 μl" kit (Qiagen) and stored at + 4 °C. To improve the quality of the extracted DNA and to remove any contaminants that may interfere with library preparation (e.g., EDTA), a DNA purification was also performed using Agencourt AMPure XP Beads (Beckman Coulter). The quality of DNA samples was assessed using the Nanodrop spectrophotometric method (Thermo Scientific) (i.e., 260/280 and 260/230 ratio), while the quantity was determined using the Quantus™ Fluorometer (Promega).

Gene sequencing was performed using a custom hybrid capture-based Roche/NimbleGen assay. The custom design of the panel was carried out by NimbleDesign software based on Genome Build hg19/GRCh37 (February 2009). For the *DPYD* gene, the design captured the genetic variability of all exons and their adjacent splice junctions (approximately 35 bases upstream and downstream of the exon), the 5' and 3' untranslated region (UTR), and the potential proximal promoter region (approximately 3000 bp in 5′ of the first exon) with a total length of 7526 bp. DNA library preparation for all samples was performed with 100 ng of input DNA purified using the KAPA HyperPlus Library Preparation Kit according to the manufacturer's instructions in the SeqCap EZ HyperCap Workflow User's Guide v 2.3 (Roche). Concentration of

single and pooled libraries was assessed using the Quantus™ Fluorometer (Promega), while quality and size distribution were assessed using the 2200 TapeStation system (Agilent). For more details on library preparation, see Supplementary Methods. Pooled libraries were sequenced on a Miseq platform (Illumina), according to the manufacturer's instructions, using a 300-cycle kit with 2 × 150 paired-end read setup.

2.3. Bioinformatic analysis, variants annotation and functional prediction

Raw sequencing data were quality checked using FastQC [19] and aligned to the human reference genome (hg19/GRCh37) using BWA-MEM software [20]. Alignment sequencing data were quality checked using Qualimap [21] and removed for duplicated sequences using picard [22] (http://broadinstitute.github.io/picard). Bedtools2 [23] was used to compute mapped reads for each genomic position reported in the manifest file of the panel. Mutect2 from GATK4.1 was used for calling germline variants and Oncotator for variants annotation. To reduce the number of false-positive calls and obtain a list of confident genetic variations, variant with at least one of the following features were discarded: (1) Variant Classification= "nontranslated intergenic regions (IGR)" or "intron"; (2) read position= "FAIL"; (3) bad haplotype= "FAIL"; (4) base quality= "FAIL"; (5) mapping quality= "FAIL"; (6) strand artifact= "FAIL"; (7) clustered events= "FAIL"; (8) fragment_length= "FAIL"; (9) t_lod= "FAIL"; (10) multiallelic= "FAIIL"; (11) read depth < 20X; (12) variant allele frequency < 0.15 to reduce the risk of false positive calls; (13) variants in genomic regions outside the panel.

MAF from the 1000 Genomes database (European population) was used to classified germline variants into very rare (MAF≤0.1 %), rare (0.1 %<MAF<1 %), common (MAF≥1 %) and novel (MAF not registered; absence of rs ID in dbSNP database, [24]). Only in case of no data available in 1000 Genomes European population database, ExAC database (non-Finnish European population) was considered.

To quantitatively predict the functional impact of missense variants, the ADME-optimized Prediction Framework (APF), which provides normalized quantitative functionality prediction scores in the range from 1 (neutral) to 0 (deleterious), was used [11,17]. For other types of variants, the PredictSNP algorithm was used [25]. For splice site polymorphisms, additional functional annotations were performed using the Loss-Of-Function Transcript Effect Estimator (LOFTEE) and SpliceAI [26]. 3'UTR regions were screened for potential microRNA binding sites using MicroSNiPer [27] and the MirSNP database [28].

2.4. Sanger sequencing

Very rare and novel *DPYD* variants detected by NGS sequencing were validated by Sanger sequencing. Each assay was designed referring to Ensembl ENST00000370192.3 (GRCh37.p13) transcript of the *DPYD* gene (ENSG00000188641). Primers used for amplification of target regions by polymerase chain reaction (PCR) were selected using the online tool Primer3Plus [29]. PCR reactions were performed in an Eppendorf Mastercycler gradient with TaqGold DNA Polymerase (ThermoFisher--Applied Biosystems). Details of primers sequence and PCR conditions are provided in Supplementary Methods. Purified reactions were sequenced using the Big Dye Terminator kit (ThermoFisher-Applied Biosystems) and the ABI PRISM 3130 capillary sequencer. Both reverse and forward primers were used for sequencing target regions. Chromatograms were visualized using the Chromas software version 2.5 to check the quality of the sequencing results. Chromatograms were manually reviewed to identify and validate the genetic variant of interest.

2.5. Structural modeling

The 3D protein structure of DPD was analyzed using Chimera software [30] to visualize the location of the protein residues and to determine the amino acid category resulting from each identified rare genetic variant (including novel ones) and its surrounding regions [31]. The probability to establish clashes and contacts with the nearest residues at a distance of less than 5 Å by the mutant respect to the wild-type residue was analyzed. Structure minimization was also performed. In the absence of a human reference structure for DPD, the structure of Sus scrofa (Pig) (entry 1h7w; resolution of 1.9 Å) was selected from the Protein Data Bank (PDB; [32]) because this protein sequence has more than 90 % similarity to the DPD of Homo sapiens [33]. The bioinformatics tool I-Mutant 3.0 [34] was used to calculate a prediction of the change in Gibb free energy (ΔΔG) between the wild type and mutant missense variants based on the protein structure at a pH of 7 and a temperature of 25 °C, and the changes in protein stability were automatically calculated based on this. The SVM3 support vector machines approach provides the sign of the ΔΔG value and an additional class, Neutral, which refers to small ΔΔG value changes in wild-type/mutant protein variants [35].

2.6. Determination of UH2/U ratio

DPD activity was indirectly assessed by quantifying the endogenous DPD substrate uracil (U) and its metabolite dihydrouracil (UH2) in human plasma using an ultra-performance liquid chromatography–tandem mass spectrometry (UPLC–MS/MS) assay described previously [36]. The UH2/U ratio was also calculated. These analyses were performed in the Pharmacy Department Laboratory of the Nederlands Kanker Institut (Amsterdam, The Netherlands). Based on published data, the uracil threshold to distinguish between normal and deficient DPD activity was set at 16 ng/ml (DPD deficiency with U concentration >16 ng/ml) [1]. The UH2/U ratio was also used to confirm DPD deficiency (with UH2/U <6), as described in the literature [37].

2.7. Statistical analysis

Socio-demographic and clinical characteristics were expressed as percentages, and differences between the "toxicity" and "no-toxicity" groups were evaluated using Fisher's exact test. The number of variants was counted for each patient, distinguished by variant type and MAF frequency (i.e., <1 %, <0.1 %, novel). For each group, the mean number of variants was calculated as the ratio between the total number of variants and the number of patients. Considering the low frequency, the mean number of variants was expressed per 100 patients by multiplying it by 100; differences between groups were assessed by the Mann-Whitney test. Toxicity risk was evaluated through odds ratio (OR) and corresponding 95 % confidence intervals (CI) estimated by means of unconditional logistic regression model, adjusting for sex, age, cancer site, treatment setting, and fluoropyrimidines.

Further, the association between common *DPYD* polymorphisms and risk of toxicity was examined. For each polymorphism, OR was calculated by an unconditional logistic regression model with the covariates listed above. Dominant, recessive, and additive genetic models were considered for each polymorphism by combining heterozygous with homozygous genotypes; the best-fitting genetic model was selected according to the Wald chi-squared test. Association ($p < 0.05$) were further tested for robustness by a bootstrap procedure with 1000 re-sampling.

3. Results

3.1. Sequencing and patient characteristics

A total of 224 fluoropyrimidine-treated patients were sequenced by NGS, of which 120 developed grade 3–5 toxicity and 104 had no

fluoropyrimidine toxicity. After quality control, 11/224 samples were excluded from analysis because they did not pass the quality check of sequencing data by fastQC. All of these samples were characterized by a low-number of produced paired-end reads (number of reads < 1000000) and a sequence length > 220 bp versus the expected 150–200 bp, indicating a failure in the sequencing step.

Therefore, the final "toxicity group" included 109 patients, whereas the "no-toxicity group" included 104 patients. Overall, the coverage analysis showed a median number of 542162 mapped reads (range: 61074–1563800) with a median percentage of reads at 1x and 10x (depth of coverage) of 86 % and 2 %, respectively.

Looking at coverage at the genomic level, the median depth at the gene level was homogeneous between the "toxicity" and "no-toxicity" groups, allowing comparison of the two cohorts.

The main demographic and clinical characteristics of the two study populations are reported in Table 1. Of the 109 patients in the "toxicity group", 62 (56.9 %) developed grade 3 toxicity, 46 (42.2 %) developed grade 4 toxicity, and 1 (0.9 %) developed grade 5 toxicity as the maximum grade of hematological or non-hematological toxicity that patients experienced during the entire treatment. Sixty-five of 109 patients (59.6 %) developed grade ≥ 3 hematological toxicity, with neutropenia being the most common adverse event (50/65; 76.9 %). Seventy-four of 109 patients (67.9 %) developed grade ≥ 3 non-hematological toxicity, with diarrhea being the most common adverse event (36/74; 48.6 %). Forty-nine of 109 patients (45.0 %) experienced severe hematological or non-hematological toxicity within the first cycle of treatment (acute toxicity, cycle ≤1), and sixty patients (55 %) after the first cycle.

Table 1
Socio-demographic and clinical characteristic of solid cancer patients enrolled in the study.

	"toxicity" group (n = 109)		"no-toxicity" group (n = 104)		
	n	(%)	n	(%)	Fisher's Exact Test
Sex					
Female	59	(54.1)	50	(48.1)	
Male	50	(45.9)	54	(51.9)	*P = 0.4121*
Age, years (median, range)	63	(30–82)	64	(26–98)	
Cancer Type					
Colon	64	(58.7)	72	(69.2)	
Rectum	22	(20.2)	19	(18.3)	
Breast	6	(5.5)	5	(4.8)	
Stomach	5	(4.6)	1	(1.0)	
Head and neck	2	(1.8)	3	(2.9)	
Pancreas	1	(0.9)	3	(2.9)	
Others	3	(2.8)	1	(1.0)	
Unknown	6	(5.5)	–		*P = 0.0999*
Chemotherapy					
Fluoropyrimidines					
5-Fluorouracil	93	(85.3)	89	(85.6)	
Capecitabine	16	(14.7)	15	(14.4)	*P = 1.0000*
Monotherapy	8	(7.3)	7	(6.7)	
Association with oxaliplatin	39	(35.8)	43	(41.3)	
Association with irinotecan	38	(34.9)	38	(36.6)	
Association with other drugs	24	(22.0)	16	(15.4)	*P = 0.6323*
Therapy setting					
Neo-adjuvant	5	(4.6)	8	(7.7)	
Adjuvant	45	(41.3)	47	(45.2)	
First-line or more	58	(53.2)	49	(47.1)	
Unknown	1	(0.9)	–		*P = 0.5010*
Max Toxicity Grade*					
3	62	(56.9)	–		
4	46	(42.2)	–		
5	1	(0.9)	–		

* Maximum grade of hematological or non-hematological toxicity experienced by the patients.

3.2. Identified variations in DPYD

A total of 375 and 349 germline variants were called against the reference genome in the "toxicity" and "no-toxicity" groups, respectively. The mean coverage (read depth) of the identified genetic variants was 103 (range: 20–797) and 78 (range: 21–186) for the "toxicity" and "no-toxicity" group, respectively.

Overall, in the "toxicity group" 34 unique genetic variants were identified (30 single nucleotide polymorphisms [SNP], 2 deletions [DEL], 2 insertions [INS]). Of these, 17/34 (50.0 %) were common (MAF≥1 %) and 17/34 (50.0 %) were rare/very rare (MAF<1 %) or novel variants. Nine of 17 (52.9 %) rare/very rare or novel variants were classified as deleterious by functional prediction (Fig. 1a). Most of the identified variants were missense (15/34, 44.1 %), and two of them also fell on a canonical splice site. Of the remaining variants, 1/34 (2.9 %) were synonymous, 11/34 (32.4 %) were in the 5' region, and 7/34 (20.6 %) were in the 3' UTR (Fig. 1b).

In the "no-toxicity group" 28 unique genetic variants were identified (26 SNP, 2 DEL, 2 INS). Of these 18/28 (64.3 %) were common (MAF≥1 %) and 10/28 (35.7 %) were rare/very rare (MAF<1 %) or novel variants. Four of 10 (40.0 %) rare/very rare or novel variants were classified as deleterious by functional prediction (Fig. 1c). Most of the identified variants were located in the 5' flaking (10/28, 35.7 %) or 3'UTR (9/28, 32.1 %) region. Of the remaining variants, 6/28 (21.4 %) were missense, 2/28 (7.2 %) were synonymous, and 1/28 (3.6 %) were located at a canonical splice site (Fig. 1b).

All very rare (MAF≤0.1 %) and novel variants identified by NGS analysis in the "toxicity" and "no-toxicity" groups were validated by Sanger sequencing with a concordance rate of 100 %.

3.3. Common DPYD variants and risk of toxicity

The toxicity risk for each common (MAF≥1 %) polymorphism identified by NGS analysis between the "toxicity" and "no-toxicity" groups is summarized in Table 2. Of the 18 genetic markers tested, only the *DPYD* rs72981743 variant was significantly more frequent in the "toxicity group", and the polymorphic T-allele was predictive of an increased risk of severe toxicity (OR=3.19, p = 0.019; Bootstrap p = 0.020). The same effect was observed for the risk of severe acute toxicity (OR=4.22, p = 0.018). For *DPYD* rs56160474, only a non-significant trend was observed, with the polymorphic GG genotype associated with a lower risk of severe toxicity (OR=0.22, p = 0.074; Bootstrap p = 0.081).

To investigate the functional role of rs72981743, bio-informatic analysis was performed using HaploReg v4.1 [38], RegulomeDB v2.0 [39] and Ensembl's Variant Effect Predictor (VEP) Ensembl release 105, December 2021 [40]. The methods and detailed results are summarized in Supplementary Table 1. Briefly, the rs72981743 polymorphism could have a moderate impact on gene functionality and/or expression, as it broadly alters regulatory chromatin state (i.e., 24 promoter histone marks, 52 DNAse items), proteins bound (i.e., 5 hits), and motifs (i.e., 13 changed motifs). This effect was summarized by a RegulomeDB rank score of 2b (i.e., TF binding + any motif + DNase Footprint + DNase peak) and a probability score of 0.58. The VEP tool showed a CADD score of 4.829. HaploReg detected four additional polymorphisms in the *DPYD*-rs72981743 haploblock (r2 > 0.8).

3.4. Rare, very rare and novel DPYD variants burden and risk of toxicity

The complete list of rare (MAF<1 %) and very rare (MAF≤0.1 %) variants identified in the *DPYD* gene in the "toxicity" and "no-toxicity" groups is shown in the Supplementary Table 2. Previously unidentified genetic variants (i.e., novel) that are expected to be very rare in

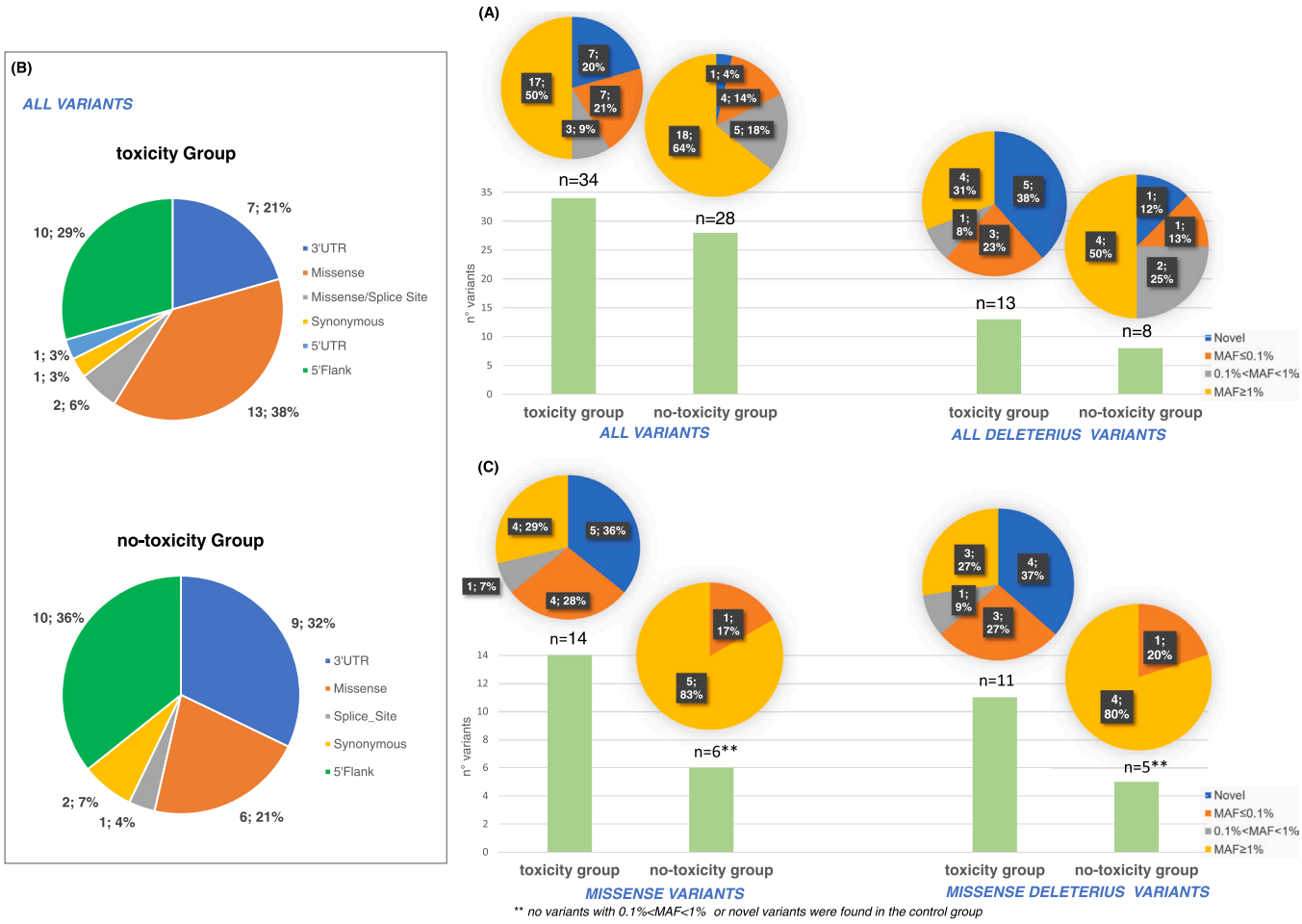

Fig. 1. (A) Pie chart visualizing the typology of all *DPYD* genetic variants identified in the "toxicity" and "no-toxicity" group. (B) Distribution of all identified *DPYD* genetic variants based on minor allele frequency (MAF) in the "toxicity" and "no-toxicity" group (number of patients and percentage are reported). (C) Focus on *DPYD* missense variants and their distribution based on MAF in the "toxicity" and "no-toxicity" group (number of patients and percentage are reported).

frequency are also reported.

An enrichment of variants with a MAF< 1 % (including novel variants) was detected in the toxicity group compared to the "no-toxicity" group (50.0 % versus 35.7 %). Similarly, variants with a MAF< 1 % and a deleterious impact were more common in the "toxicity group" (52.9 % versus 40.0 %). This difference becomes even more proninced when only the very rare variants (including novel ones) are considered. In the "toxicity group" 14 (14/34, 41.2 %) of the identified variants were classified as very rare, including 8 (57.1 %) with a putative deleterious impact on the *DPYD* gene. In the control group, only 5 of 28 (17.9 %) variants were very rare, of which 2 (40.0 %) were predicted as deleterious (Fig. 1a).

When considering the variants burden and the risk of developing severe fluoropyrimidine toxicity (Table 3A), a borderline significant association was observed for very rare variants (including novel ones). Specifically, the mean number of very rare variants per 100 patients was 13.8 and 5.8 in the "toxicity" and "no-toxicity" groups, respectively ($p = 0.051$). The number of patients with at least one very rare variant was 15 (13.8 %) and 6 (5.8 %) in the "toxicity" and "no-toxicity" groups, respectively ($p = 0.065$). Carrying at least one very rare variant increased the risk of developing severe fluoropyrimidine toxicity in multivariate analysis (OR=2.53, $p = 0.071$). The strength of the association increased further when only the risk of acute toxicity was considered (OR=4.01, $p = 0.017$). Furthermore, when focusing only on novel variants, it appeared that harboring at least one novel variant resulted in an increased risk of developing severe acute toxicity (OR=6.13, $p = 0.036$).

When considering missense variants, we found a remarkable enrichment of variants with a MAF< 1 % (including novel variants) in the "toxicity group", either when considering all (10/14, 71.4 % versus 1/6, 16.7 %) or only the deleterious variants (8/11, 72.7 % versus 1/5, 20.0 %) (Fig. 1c). The mean number of variants per 100 patients was significantly higher in the "toxicity" group than in the "no-toxicity" group, considering all groupings (MAF<1 %, $p = 0.007$; MAF≤0.1 %, $p = 0.012$; novel variants $p = 0.027$). Similar results were obtained when only missense variants with a predicted deleterious impact were considered (Table 3B). The number of patients with at least one rare variant (including novel variants) differed significantly between the two groups (9.2 % versus 1.0 %, $p = 0.010$). Carrying at least one rare variant was associated with an approximately 11-fold higher risk of severe toxicity ($p = 0.025$) and a 16-fold higher risk of acute severe toxicity ($p = 0.013$) in multivariate analysis. Similar significant associations were found when only variants with a predicted deleterious impact on DPD were considered. The number of patients with at least one very rare variant (including novel variants) was 8.3 % and 1.0 % in the "toxicity" and "no-toxicity" groups, respectively, $p = 0.019$). Carrying at least one very rare variant was associated with an approximately 9-fold higher risk of developing severe toxicity ($p = 0.037$) and a 16-fold higher risk of acute toxicity ($p = 0.014$) in multivariate analysis. The same trend was observed when only the deleterious very rare variants were considered, although this was not statistically significant due to the smaller number of events.

No significant results were obtained considering only the variants located in the 3'UTR or 5'UTR regions (data not shown).

Table 2
Risk of severe fluoropyrimidine-related toxicity according to identified *DPYD* common polymorphisms. The risk of acute toxicity (cycle ≤ 1) was also considered. Associations with *P*-value < 0.05 are in bold.

SNP	Base change	Type	"toxicity" group (n = 109)			"no-toxicity" group (n = 104)			Model	OR (95 % CI)[a]	p-value	Cycles (≤1) (n = 49)	
			AA	Aa	aa	AA	Aa	aa				OR (95 % CI)[a]	p-value
rs41285690	T > C	3'UTR	0.973	0.027	0.000	0.971	0.029	0.000	Dominant	1.00 (0.19–5.24)	0.996	1.80 (0.27–11.80)	0.543
rs17470762	A>G	3'UTR	0.927	0.073	0.000	0.942	0.058	0.000	Dominant	1.46 (0.46–4.58)	0.518	1.30 (0.30–5.58)	0.722
rs291593	G>A	3'UTR	0.596	0.339	0.064	0.644	0.250	0.106	Recessive	0.69 (0.25–1.89)	0.469	0.51 (0.10–2.51)	0.404
rs291592	C>T	3'UTR	0.440	0.413	0.147	0.404	0.414	0.183	Additive	0.85 (0.58–1.25)	0.407	1.15 (0.69–1.91)	0.598
rs1042482	C>T	3'UTR	0.881	0.110	0.009	0.904	0.096	0.000	Additive	1.27 (0.55–2.96)	0.574	1.15 (0.42–3.21)	0.784
rs56160474	A>G	3'UTR	0.716	0.266	0.018	0.702	0.231	0.067	Recessive	0.22 (0.04–1.16)	0.074	–	–
rs1801160	C>T	Missense	0.852	0.139	0.009	0.865	0.135	0.000	Additive	1.34 (0.62–2.88)	0.460	0.90 (0.29–2.82)	0.860
rs17376848	A>G	Silent	0.936	0.064	0.000	0.971	0.029	0.000	Dominant	2.27 (0.56–9.22)	0.254	3.32 (0.67–16.47)	0.141
rs1801159	T > C	Missense	0.679	0.294	0.028	0.673	0.298	0.029	Recessive	0.90 (0.17–4.67)	0.897	2.38 (0.43–13.14)	0.319
rs1801158	C>T	Missense	0.926	0.074	0.000	0.933	0.067	0.000	Dominant	1.27 (0.42–3.81)	0.671	2.65 (0.77–9.15)	0.122
rs56038477	C>T	Silent	1.000	0.000	0.000	0.962	0.038	0.000	–	–	–	–	–
rs2297595	T > C	Missense	0.771	0.229	0.000	0.750	0.231	0.019	Additive	0.84 (0.45–1.55)	0.576	0.86 (0.39–1.90)	0.712
rs1801265	A>G	Missense	0.633	0.321	0.046	0.635	0.298	0.067	Recessive	0.71 (0.22–2.35)	0.579	0.32 (0.04–2.73)	0.296
rs72981743	**C>T**	**5'Flank**	**0.807**	**0.193**	**0.000**	**0.924**	**0.076**	**0.000**	**Dominant**	**3.19 (1.21–8.4)**	**0.019**	**4.22 (1.28–13.88)**	**0.018**
rs61787828	A>C	5'Flank	0.881	0.119	0.000	0.817	0.164	0.019	Additive	0.58 (0.28–1.21)	0.145	0.58 (0.23–1.52)	0.270
rs1471548772	Ins/Del	5'Flank	0.914	0.086	0.000	0.856	0.135	0.010	Additive	0.58 (0.25–1.36)	0.210	0.79 (0.27–2.28)	0.658
rs57862948a	Ins/Del	5'Flank	0.789	0.173	0.039	0.808	0.115	0.077	Recessive	0.53 (0.15–1.87)	0.325	0.27 (0.03–2.30)	0.230
rs57862948b	Ins/Del	5'Flank	0.991	0.009	0.000	0.990	0.010	0.000	Dominant	0.91 (0.05–16.19)	0.949	2.01 (0.10–39.98)	0.648

[a] Odds ratio (OR) and corresponding 95 % confidence intervals (CI) were estimated from unconditional logistic regression model, adjusting for sex, age, cancer site, treatment setting, and Fluoropyrimidines.

3.5. Structural analysis of DPYD variants

Chimera was used to visualize the protein domain localization of each identified missense variant, as well as the corresponding chemical shift of the encoded amino acids and the nearby binding and catalytic sites. The DPD protein consists of 5 domains [41]: domain I (residues 27–173), and V (residues 1–26 and 848–1025) contain Fe-S clusters, domain II (residues 173–286, 442–524) binds flavin adenine dinucleotide (FAD), domain III (residues 287–441) binds nicotinamide adenine dinucleotide phosphate (NADPH), and domain IV (residues 525–847) binds flavin mononucleotide (FMN). Structural mapping of the missense variants identified in the "toxicity" and "no-toxicity" groups with a MAF< 1 %, including the novel ones, is shown in Fig. 2.

To provide some protein structural parameters to support the functional prediction calculated by the AFP *in silico* tool, the amino acid category, the number of clashes and contacts with the nearest residues at less than 5 Å by the mutant compared to the wild-type residue, and the energy stability of the mutant proteins compared to the wild-type were evaluated. The results are summarized in Table 4.

Six (p.E161K, p.T267I, p.T471P, p.I370M, p.G320R, p.V373I) of 11 missense variants resulted in a change in the total number of potential contacts with nearby amino acids compared to the wild-type protein (Table 4, Supplementary Figures 1, 2, and 3). Of those, the presence of the mutant residue in p.E161K, located in the Fe-S cluster coordination domain, and in p.G320R, located upstream of the loop containing the protease-labile site in the NADPH domain, appear to potentially impair the catalytic activity of DPD, consistent with their predicted deleterious effect on protein function. p.I370M is also located in the NADPH-binding domain but has a stabilizing effect on the hydrophobic pocket and thus does not potentially affect the enzymatic activity od DPD. Accordingly, p.I370M was predicted by AFP to have a tolerated effect on protein function. p.T471P, p.T267I, and p.V373I do not appear to affect any relevant catalytic protein domain.

Regarding the difference in free energy calculated for the wild-type and mutant DPD structure, almost all mutant proteins were found to be largely unstable compared with the wild-type proteins, with the exception of p.I370M (predicted to be tolerated also by AFP tool) and p.E161K (Table 4).

3.6. DPD phenotyping

To validate the predicted functional consequences of the five novel missense variants identified in the "toxicity group", we determined U and UH2 metabolites using archival plasma samples (Supplementary Table 3). For one variant (chr1:98144701 G>A, p.T267I), the plasma sample of the patient was not available and analysis was not performed. For the remaining 4 variants, 3 of which were predicted to be deleterious and 1 to be tolerated, the analysis was successfully performed. The U metabolite resulted greater than 10 ng/ml in all samples tested (range 10.6 – 17.1 ng/ml) with a value higher than the deleteriousness threshold of 16 ng/ml in only one case (chr1:98015229 T > G, p.T471P variant, U=17.1 ng/ml). The missense variant, predicted to be tolerated

Table 3
Mutational germline burden of rare, very rare and novel genetic variants and the risk of developing severe fluoropyrimidine-related toxicity. The risk of acute toxicity (cycle ≤ 1) was also considered. Associations with *P*-value < 0.05 are in bold.

Type of genetic variants	Mean number of variants for 100 patients**			Patients with variants					OR (95 % CI)[a]	p-value	Cycle≤ 1 (n = 49)	
	"no-toxicity" group	"toxicity" group	M-W test (p-value)	"no-toxicity" group		"toxicity" group		Fisher's exact test (p-value)			OR (95 % CI)[a]	p-value
				n	(%)	n	(%)					
(A) All												
MAF< 1 %	11.5	20.2	0.115	12	11.5 %	21	19.3 %	0.133	1.72 (0.78–3.79)	0.176	2.42 (0.96–6.10)	0.062
MAF< 1 % (Del)	4.8	9.2	0.305	5	4.8 %	9	8.3 %	0.410	1.94 (0.61–6.21)	0.264	2.48 (0.65–9.50)	0.186
MAF≤ 0.1 %	5.8	13.8	0.051	6	5.8 %	15	13.8 %	0.065	2.53 (0.92–6.94)	0.071	**4.01 (1.28–12.57)**	**0.017**
MAF≤ 0.1 % (Del)	2.9	7.3	0.143	3	2.9 %	8	7.3 %	0.216	2.64 (0.67–10.48)	0.167	3.99 (0.87–18.36)	0.076
Novel	1.9	6.4	0.104	2	1.9 %	7	6.4 %	0.171	3.07 (0.60–15.69)	0.178	**6.13 (1.13–33.35)**	**0.036**
Novel (Del)	1.9	4.6	0.277	2	1.9 %	5	4.6 %	0.446	2.10 (0.38–11.60)	0.393	3.96 (0.65–24.17)	0.136
(B) Missense												
MAF< 1 %	**1.0**	**10.1**	**0.007**	**1**	**1.0 %**	**10**	**9.2 %**	**0.010**	**11.06 (1.35–90.53)**	**0.025**	**16.20 (1.80–145.75)**	**0.013**
MAF< 1 % (Del)	**1.0**	**8.3**	**0.021**	**1**	**1.0 %**	**8**	**7.3 %**	**0.036**	**8.75 (1.05–73.11)**	**0.045**	**10.59 (1.10–102.29)**	**0.041**
MAF≤ 0.1 %	**1.0**	**8.3**	**0.012**	**1**	**1.0 %**	**9**	**8.3 %**	**0.019**	**9.48 (1.15–78.33)**	**0.037**	**15.79 (1.75–142.71)**	**0.014**
MAF≤ 0.1 % (Del)	**1.0**	**6.4**	**0.037**	1	1.0 %	7	6.4 %	0.066	7.25 (0.86–61.52)	0.069	**10.27 (1.06–99.65)**	**0.044**
Novel	**0.0**	**4.6**	**0.027**	0	0.0 %	5	4.6 %	0.060	–	–	–	–
Novel (Del)	**0.0**	**3.7**	**0.049**	0	0.0 %	4	3.7 %	0.122	–	–	–	–

Abbreviations: del, deleterious; M-W test; Mann-Whitney test.
[a] Odds ratio (OR) and corresponding 95 % confidence intervals (CI) were estimated from unconditional regression model, adjusting for sex, age, cancer site, treatment setting, and Fluoropyrimidines.
** (Number of genetic variants/total number of patients) * 100

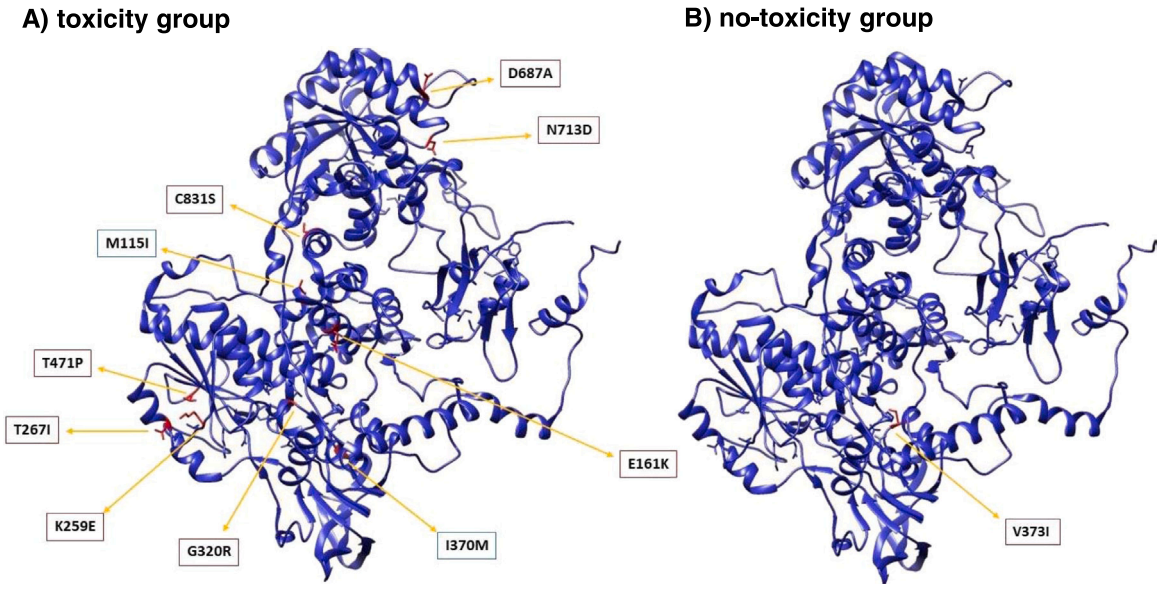

Fig. 2. Structural model of *DPYD* missense variants in A) "toxicity" and B) "no-toxicity" group. In a red rectangle the variants predicted to be deleterious, in a blue rectangle those predicted to be tolerated.

by the *in silico* annotation, has the lowest value of the U metabolite (10.6 ng/ml). The UH2/U ratio ranged from 6.71 to 10.60. The sample (referred to the variant chr1:98015229 T > G, p.T471P) with the highest U value also had the lowest UH2/U ratio, which was below the deleteriousness threshold of 6 (5.91).

4. Discussion

Currently, only a few genetic markers are available as pre-treatment *DPYD* genotyping tests (*DPYD**2A, *DPYD**13, c.2846 A>T, c.1236 G>A-HapB3) to identify patients at risk of developing severe fluoropyrimidine-related toxicity, and a substantial number of severe toxicity events are missed by current targeted genotyping strategies [1,5,6]. Preliminary data have highlighted the great potential of more

Table 4
Summary of changes in protein structural parameters comparing the polymorphic respect to the wild-type residue.

Rs ID	Classification	Amino acid change	Variant localization	Mutant residue	Amino acid category	New contacts /clashes*	Intra-DPD residues involved in the contacts	I-Mutant ΔΔG (kcal/mol)	I-Mutant SVM3 Prediction Effect of Stability
A) TOXICITY GROUP									
na	Novel	p.C831S	DPD domain IV	Ser831	Hydrophobic -> Polar/Small	0	None	-1.01	Large Decrease (RI:5)
rs773407491	Very Rare	p.N713D	DPD domain IV	Asp713	Polar -> Polar/Negative/ Small	0	None	-0.33	Large Decrease (RI:3)
rs755692084	Very Rare	p.D687A	DPD domain IV	Ala687	Polar/Negative- > Hydrophobic/Small	0	None	-0.21	Large Decrease (RI:2)
na	Novel	p.T471P	DPD domain II	Pro471	Hydrophobic/Polar -> Small	8	Val476 and Phe477 in DPD domain II	-1.11	Large Decrease (RI:6)
na	Novel	p.I370M	DPD domain III	Met370	Hydrophobic/Aliphatic -> Hydrophobic	3	Phe345 and Phe363 in DPD domain III	-0.65	Neutral (RI:2)
rs1205376538	Very Rare	p.G320R	DPD domain III	Arg320	Hydrophobic -> Polar/ Positive	12	Asp239 in DPD domain II and Pro906 in DPD domain V	-0.58	Large Decrease (RI: 1)
na	Novel	p.T267I	DPD domain II	Ile267	Hydrophobic/Polar -> Hydrophobic/Aliphatic	5	Thr270 and Asn269 in DPD domain II	-0.01	Large Decrease (RI:3)
rs45589337	Rare	p.K259E	DPD domain II	Glu259	Hydrophobic/Polar/Positive -> Polar/Negative	0	None	-0.46	Large Decrease (RI:1)
na	Novel	p.E161K	DPD domain I	Lys161	Polar/Negative -> Hydrophobic/Polar/ Positive	13	Met128 and Gln157 in DPD domain I	-0.64	Neutral (RI:1)
rs377169736	Very Rare	p.M115I	DPD domain I	Ile115	Hydrophobic- > Hydrophobic/Aliphatic	0	None	-0.88	Large Decrease (RI:1)
B) NO-TOXICITY GROUP									
rs772906420	Very Rare	p.V373I	DPD domain III	Ile373	Hydrophobic&Aliphatic -> Hydrophobic/Aliphatic	1	Pro131 in DPD domain I	-0.56	Large Decrease (RI: 5)

Abbreviation: RI: Reliability Index.
* observed in presence of the mutant residue and after minimization.

comprehensive sequencing approaches to identify rare and novel *DPYD* variants that could be causative of severe fluoropyrimidine-related toxicity [12–15].

This study is the first to report that the burden of rare variants (MAF <1 %) in *DPYD,* including novel variants, identified by NGS, is associated with the risk of developing serious adverse events related to fluoropyrimidines. This effect is even more pronounced when only missense variants with deleterious effects on the DPD phenotype (determined using specifically developed *in silico* tools [17]) are considered. Furthermore, we show here that carriers of rare missense variants in *DPYD* have a 16-fold increased risk in the first cycle of treatment and an 11-fold increased risk during the entire course of chemotherapy of developing severe to fatal toxicities related to fluoropyrimidine treatment.

Previous reports have shown that most of the genetic variants detected in pharmacogenes are rare and often associated with a functional effect. It is estimated that about 30–40 % of the genetically encoded interindividual variation in drug ADME is attributable to rare genetic variants [42]. Indeed, a post-treatment analysis of patients who show unexpected responses to fluoropyrimidines revealed several previously unknown variants in *DPYD* that could be further incorporated into the prospective diagnostic genotyping panel [12–15, 43]. These data are consistent with recent work demonstrating that the high genetic complexity of *DPYD* deficiency benefits from sequencing-based *DPYD* profiling to ensure comprehensive identification of variants that impact fluoropyrimidine-related toxicity [11].

The present study supports the idea that considering the burden of rare variants in a strong candidate gene (such as *DPYD* for fluoropyrimidines) could represent a marker of toxicity that could be prospectively considered in treatment planning. A recent paper proposed a similar approach for cardiological adverse events (i.e., acquired long QT syndrome, aLQTS) associated with drug treatments. The authors demonstrated that the burden of rare non-synonymous variants in genes related to drug metabolism (i.e., cytochrome genes) correlates with the risk of aLQTS and should be considered as a potential pharmacogenomic marker [16]. Similarly, the burden of rare genetic variants in *ABCC1* (MRP1) was identified as a strong predictor of survival in breast cancer patients undergoing therapy with the MRP1 substrates cyclophosphamide and doxorubicin [44].

The results of the present work demonstrate that the burden of rare variants is significantly increased in patients with severe toxicity. This effect becomes even more evident when focusing only on missense variants predicted to have a deleterious effect on the encoded protein. There is therefore an urgent need to improve bioinformatic tools that can provide reliable functional prediction of the deleteriousness of such variants. Advances in this direction have been recently made for missense variants, with the development of high-performance computational methods [17,45,46]. For example, the optimized APF algorithm, which was also used in this study to infer the deleteriousness of missense variants, was demonstrated to achieve an accuracy of prediction (i.e., 91.4 %) very close to that of *in vitro* assays and significantly higher than that of other commonly used *in silico* tools such as SIFT, Polyphen-2, PROVEAN, and CADD. Although further research efforts will certainly be needed to further improve the predictive power of existing computational methods (e.g., population-scale genomic biobanks testing with correlated electronic medical records; increasing amount of systematically collected large-scale functional data), current models may already be sufficiently accurate to identify patients with putatively damaging missense variations who are candidates for dose adjustments or increased clinical monitoring.

To further investigate the functional significance of the novel missense variants identified by NGS and classified as deleterious by *in*

silico bioinformatic analysis, the analysis of U and UH2 metabolites was also performed. This approach was partially consistent with the functional annotation, as two of the four samples analyzed had U levels and UH2/U ratios concordant with the variant effect predicted by *in silico* analysis, according to thresholds currently proposed in the literature [1, 37]. Recently, DPD phenotyping by UH2/U was shown to be poorly related to DPD activity measured in peripheral blood mononuclear cells, which is considered the gold standard for DPD phenotyping, and could be strongly influenced by blood collection time and sample processing [47]; therefore the results should be evaluated with caution.

Molecular modeling of the rare, very rare and novel missense variants identified in the coding region of *DPYD* was performed to visualize the mutant protein structures and attempt to explain the functional effects predicted by *in silico* bioinformatics tools, and was largely in agreement. For ten of the eleven variants, either the evaluation of new contacts established within the protein's catalytic domains or the large variation in the protein's stability supported the functional AFP annotation as "deleterious" or "tolerated". Only for one variant (p.M115I) the evaluation of free energy was discordant with the AFP-predicted functional annotation, and additional underlying mechanisms should be further studied.

When evaluating the 5'UTR and 3'UTR variants no significant effects were found in the present study. This may be due to the fact that less annotation data is available for variants falling in the 5' and 3' regions, and current *in silico* tools still underperform in predicting the deleteriousness of variants located in non-coding or spliced regions [48]. Therefore, particularly for their functional interpretation more extensive methodological advancements are required.

Among the common *DPYD* polymorphisms identified in our case study, the variant rs72981743, located in the non-coding 5' upstream region of the gene, is significantly associated with an increased risk of developing severe fluoropyrimidine-related toxicity. To date, no literature data on the phenotypic or clinical impact of this variant have been published in the literature. However, a preliminary bioinformatic prediction was performed suggesting a potential moderate functional impact. Interestingly, the rs72981743 polymorphism is located in the consensus sequence for the transcriptional factor E2F1, that was preliminarily associated with chemosensitivity to 5-FU [49,50], thus providing a putative mechanistic link between regulation of gene expression and toxicity risk.

Some limitations of the present study need to be considered. First, the study included patients treated mainly with fluoropyrimidines in combination with other chemotherapeutic agents (i.e., irinotecan, oxaliplatin). While particular attention was paid to exclude severe toxicities that were clearly not related to fluoropyrimidines, potential interference in the collection of toxicities that are not all attributed to fluoropyrimidines cannot be excluded. Second, because of the lack of biological material, determination of DPD phenotype was attempted only for a subset of patients and could be performed only by analysis of U and UH2 metabolites rather than by measurement of DPD activity in peripheral blood mononuclear cells, which is considered the elective method for reliable phenotyping [47].

This study pointed out for the first time that rare genetic variants detected by comprehensive analysis of the *DPYD* gene using NGS method significantly improve the percentage of patients at higher risk for severe toxicity who can be identified prior to start the fluoropyrimidine-based therapy. However, some challenges should be overcome before fluoropyrimidine dosing based on *DPYD* sequencing can be implemented into clinical practice. First, to date there are still no dose-adjustment guidelines for carriers of uncharacterized rare and novel variants potentially detectable by *DPYD* sequencing. Our paper is a report of preliminary data that should, firstly, be replicated in future prospective studies to figure out potential dose adjustments based on the presence of *DPYD* rare variants in the patient's genotype. Prospective application in clinical studies as phase II non-inferiority studies would certainly be an ideal setting to clinically validate the approach. We have shown that an approach that couples *DPYD* NGS with functional *in silico* analyses can be helpful in identifying rare variant carriers who are at increased risk for developing a severe adverse event and who should probably be treated with special caution. Only prospective implementation studies could provide practical information on the feasibility of using such genetic markers in clinical practice, including a formal assessment of its cost-effectiveness respect to the conventional screening of known toxicity-associated DPYD variants. Moreover, the requirement of specific instruments and technical equipment to perform a *DPYD* analysis by NGS and an appropriate turn around time for the analysis could represent further obstacles. However, considering the increasingly widespread use of sequencing technologies that is paralleled by decreasing costs, the adoption of *DPYD* sequencing in routine clinical practice might soon be feasible.

5. Conclusion

The present work demonstrated for the first time that *DPYD* rare and novel missense variants could collectively identify an appreciable fraction of patients with DPD deficiency who are at increased risk of developing severe fluoropyrimidine-related toxicity. Assessment of these variants by comprehensive sequencing of *DPYD* could improve the preemptive identification of at-risk patients as compared to when only the four validated *DPYD* variants are tested. If the clinical validity of this approach and the feasibility of its implementation in the clinics will be confirmed by further trials, the introduction of *DPYD* sequencing into routine clinical practice could be an additional reliable tool to prevent severe toxicity events in patients scheduled to receive standard dose of fluoropyrimidine, resulting in reduced patient morbidity and mortality.

Funding

This work was supported by the Italian Ministry of Health (Ricerca Corrente), 5×1000 "Ricerca Sanitaria" and European Union's Horizon 2020 Research and Innovation Programme under grant agreement no. 668353.

CRediT authorship contribution statement

Elena De Mattia: Conceptualization, Visualization, Validation, Funding acquisition, Writing - original draft. **Marco Silvestri:** Data curation. **Jerry Polesel:** Methodology, Formal analysis. **Fabrizio Ecca:** Investigation, Validation. **Silvia Mezzalira:** Investigation. **Lucia Scarabel:** Investigation. **Yitian Zhou:** Data curation, Formal analysis. **Rossana Roncato:** Investigation. **Volker M. Lauschke:** Data curation, Formal analysis. **Stefano Calza:** Data curation. **Michele Spina:** Resources. **Fabio Puglisi:** Resources. **Giuseppe Toffoli:** Funding acquisition, Supervision. **Erika Cecchin:** Conceptualization, Writing – review & editing, Project administration.

Conflict of interest statement

The authors declare no conflict of interest.

Data Availability

Data will be made available on request.

Acknowledgments

The authors wish to thank Dr. Hilde Rosing, from the Department of Pharmacy & Pharmacology - The Netherlands Cancer Institute (Amsterdam, The Netherlands), who performed the U and UH2 quantification.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.biopha.2022.113644.

References

[1] J.E. Knikman, H. Gelderblom, J.H. Beijnen, A. Cats, H.J. Guchelaar, L.M. Henricks, Individualized dosing of fluoropyrimidine-based chemotherapy to prevent severe fluoropyrimidine-related toxicity: what are the options? Clin. Pharm. Ther. 109 (2021) 591–604, https://doi.org/10.1002/cpt.2069.

[2] C. Lunenburg, L.M. Henricks, H.J. Guchelaar, J.J. Swen, M.J. Deenen, J.H. M. Schellens, H. Gelderblom, Prospective DPYD genotyping to reduce the risk of fluoropyrimidine-induced severe toxicity: ready for prime time, Eur. J. Cancer 54 (2016) 40–48, https://doi.org/10.1016/j.ejca.2015.11.008.

[3] M. Boisdron-Celle, O. Capitain, R. Faroux, C. Borg, J.P. Metges, M.P. Galais, M. Kaassis, J. Bennouna, K. Bouhier-Leporrier, E. Francois, I. Baumgaertner, V. Guerin-Meyer, O. Cojocarasu, C. Roemer-Becuwe, C. Stampfli, L. Rosenfeld, T. Lecompte, V. Berger, A. Morel, E. Gamelin, Prevention of 5-fluorouracil-induced early severe toxicity by pre-therapeutic dihydropyrimidine dehydrogenase deficiency screening: assessment of a multiparametric approach, Semin Oncol. 44 (2017) 13–23, https://doi.org/10.1053/j.seminoncol.2017.02.008.

[4] E. Gamelin, M. Boisdron-Celle, V. Guerin-Meyer, R. Delva, A. Lortholary, F. Genevieve, F. Larra, N. Ifrah, J. Robert, Correlation between uracil and dihydrouracil plasma ratio, fluorouracil (5-FU) pharmacokinetic parameters, and tolerance in patients with advanced colorectal cancer: a potential interest for predicting 5-FU toxicity and determining optimal 5-FU dosage, J. Clin. Oncol. 17 (1999) 1105, https://doi.org/10.1200/JCO.1999.17.4.1105.

[5] U. Amstutz, L.M. Henricks, S.M. Offer, J. Barbarino, J.H.M. Schellens, J.J. Swen, T. E. Klein, H.L. McLeod, K.E. Caudle, R.B. Diasio, M. Schwab, Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for dihydropyrimidine dehydrogenase genotype and fluoropyrimidine dosing: 2017 Update, Clin. Pharm. Ther. 103 (2018) 210–216, https://doi.org/10.1002/cpt.911.

[6] C. Lunenburg, C.H. van der Wouden, M. Nijenhuis, M.H. Crommentuijn-van Rhenen, N.J. de Boer-Veger, A.M. Buunk, E.J.F. Houwink, H. Mulder, G.A. Rongen, R.H.N. van Schaik, J. van der Weide, B. Wilffert, V.H.M. Deneer, J.J. Swen, H. J. Guchelaar, Dutch pharmacogenetics working group (DPWG) guideline for the gene-drug interaction of DPYD and fluoropyrimidines, Eur. J. Hum. Genet 28 (2020) 508–517, https://doi.org/10.1038/s41431-019-0540-0.

[7] D. Meulendijks, A. Cats, J.H. Beijnen, J.H. Schellens, Improving safety of fluoropyrimidine chemotherapy by individualizing treatment based on dihydropyrimidine dehydrogenase activity - ready for clinical practice? Cancer Treat. Rev. 50 (2016) 23–34, https://doi.org/10.1016/j.ctrv.2016.08.002.

[8] M. Lek, K.J. Karczewski, E.V. Minikel, K.E. Samocha, E. Banks, T. Fennell, A. H. O'Donnell-Luria, J.S. Ware, A.J. Hill, B.B. Cummings, T. Tukiainen, D. P. Birnbaum, J.A. Kosmicki, L.E. Duncan, K. Estrada, F. Zhao, J. Zou, E. Pierce-Hoffman, J. Berghout, D.N. Cooper, N. Deflaux, M. DePristo, R. Do, J. Flannick, M. Fromer, L. Gauthier, J. Goldstein, N. Gupta, D. Howrigan, A. Kiezun, M.I. Kurki, A.L. Moonshine, P. Natarajan, L. Orozco, G.M. Peloso, R. Poplin, M.A. Rivas, V. Ruano-Rubio, S.A. Rose, D.M. Ruderfer, K. Shakir, P.D. Stenson, C. Stevens, B. P. Thomas, G. Tiao, M.T. Tusie-Luna, B. Weisburd, H.H. Won, D. Yu, D. M. Altshuler, D. Ardissino, M. Boehnke, J. Danesh, S. Donnelly, R. Elosua, J. C. Florez, S.B. Gabriel, G. Getz, S.J. Glatt, C.M. Hultman, S. Kathiresan, M. Laakso, S. McCarroll, M.I. McCarthy, D. McGovern, R. McPherson, B.M. Neale, A. Palotie, S. M. Purcell, D. Saleheen, J.M. Scharf, P. Sklar, P.F. Sullivan, J. Tuomilehto, M. T. Tsuang, H.C. Watkins, J.G. Wilson, M.J. Daly, D.G. MacArthur, C. Exome Aggregation, Analysis of protein-coding genetic variation in 60,706 humans, Nature 536 (2016) 285–291, https://doi.org/10.1038/nature19057.

[9] D.B. Goldstein, A. Allen, J. Keebler, E.H. Margulies, S. Petrou, S. Petrovski, S. Sunyaev, Sequencing studies in human genetics: design and interpretation, Nat. Rev. Genet 14 (2013) 460–470, https://doi.org/10.1038/nrg3455.

[10] J.L. Lopes, K. Harris, M.B. Karow, S.E. Peterson, M.L. Kluge, K.E. Kotzer, G. S. Lopes, N.B. Larson, S.J. Bielinski, S.E. Scherer, L. Wang, R.M. Weinshilboum, J. L. Black III, A.M. Moyer, Targeted genotyping in clinical pharmacogenomics: what is missing? J. Mol. Diagn. 24 (2022) 253–261, https://doi.org/10.1016/j.jmoldx.2021.11.008.

[11] Y. Zhou, C. Dagli Hernandez, V.M. Lauschke, Population-scale predictions of DPD and TPMT phenotypes using a quantitative pharmacogene-specific ensemble classifier, Br. J. Cancer 123 (2020) 1782–1789, https://doi.org/10.1038/s41416-020-01084-0.

[12] X. Garcia-Gonzalez, B. Kaczmarczyk, J. Abarca-Zabalia, F. Thomas, P. Garcia-Alfonso, L. Robles, V. Pachon, A. Vaz, S. Salvador-Martin, M. Sanjurjo-Saez, L. A. Lopez-Fernandez, New DPYD variants causing DPD deficiency in patients treated with fluoropyrimidine, Cancer Chemother. Pharm. 86 (2020) 45–54, https://doi.org/10.1007/s00280-020-04093-1.

[13] L.M. Henricks, E.J.M. Siemerink, H. Rosing, J. Meijer, S.M.I. Goorden, A. M. Polstra, L. Zoetekouw, A. Cats, J.H.M. Schellens, A.B.P. van Kuilenburg, Capecitabine-based treatment of a patient with a novel DPYD genotype and complete dihydropyrimidine dehydrogenase deficiency, Int J. Cancer 142 (2018) 424–430, https://doi.org/10.1002/ijc.31065.

[14] R.C. Ly, R.E. Schmidt, P.J. Kiel, V.M. Pratt, B.P. Schneider, M. Radovich, S. M. Offer, R.B. Diasio, T.C. Skaar, Severe capecitabine toxicity associated with a rare DPYD variant identified through whole-genome sequencing, JCO Precis Oncol. 4 (2020), https://doi.org/10.1200/PO.20.00067.

[15] R. Palmirotta, D. Lovero, H. Delacour, A. Le Roy, S. Cremades, F. Silvestris, Rare dihydropyrimidine dehydrogenase variants and toxicity by floropyrimidines: a case report, Front Oncol. 9 (2019) 139, https://doi.org/10.3389/fonc.2019.00139.

[16] B. Gray, A.E. Baruteau, A.A. Antolin, A. Pittman, G. Sarganas, M. Molokhia, M. T. Blom, R. Bastiaenen, A. Bardai, S.G. Priori, C. Napolitano, P.E. Weeke, S. A. Shakir, W. Haverkamp, J. Mestres, B. Winkel, A.A. Witney, I. Chis-Ster, A. Sangaralingam, A.J. Camm, J. Tfelt-Hansen, D.M. Roden, H.L. Tan, E. Garbe, M. Sturkenboom, E.R. Behr, Rare variation in drug metabolism and long QT genes and the genetic susceptibility to acquired long QT syndrome, Circ. Genom. Precis Med 15 (2022), e003391, https://doi.org/10.1161/CIRCGEN.121.003391.

[17] Y. Zhou, S. Mkrtchian, M. Kumondai, M. Hiratsuka, V.M. Lauschke, An optimized prediction framework to assess the functional impact of pharmacogenetic variants, Pharm. J. 19 (2019) 115–126, https://doi.org/10.1038/s41397-018-0044-2.

[18] G. Toffoli, L. Giodini, A. Buonadonna, M. Berretta, A. De Paoli, S. Scalone, G. Miolo, E. Mini, S. Nobili, S. Lonardi, N. Pella, G. Lo Re, M. Montico, R. Roncato, E. Dreussi, S. Gagno, E. Cecchin, Clinical validity of a DPYD-based pharmacogenetic test to predict severe toxicity to fluoropyrimidines, Int J. Cancer 137 (2015) 2971–2980, https://doi.org/10.1002/ijc.29654.

[19] Babraham Bioinformatics. ⟨http://www.bioinformatics.babraham.ac.uk/projects/fastqc⟩. Accessed 1 April 2022.

[20] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, Bioinformatics 25 (2009) 1754–1760, https://doi.org/10.1093/bioinformatics/btp324.

[21] F. Garcia-Alcalde, K. Okonechnikov, J. Carbonell, L.M. Cruz, S. Gotz, S. Tarazona, J. Dopazo, T.F. Meyer, A. Conesa, Qualimap: evaluating next-generation sequencing alignment data, Bioinformatics 28 (2012) 2678–2679, https://doi.org/10.1093/bioinformatics/bts503.

[22] PICARD. ⟨http://broadinstitute.github.io/picard⟩. Accessed 1 April 2022.

[23] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics 26 (2010) 841–842, https://doi.org/10.1093/bioinformatics/btq033.

[24] NCBI-dbSNP. ⟨https://www.ncbi.nlm.nih.gov/snp/⟩. Accessed 1 April 2022.

[25] J. Bendl, M. Musil, J. Stourac, J. Zendulka, J. Damborsky, J. Brezovsky, PredictSNP2: a unified platform for accurately evaluating SNP Effects by exploiting the different characteristics of variants in distinct genomic regions, PLoS Comput. Biol. 12 (2016), e1004962, https://doi.org/10.1371/journal.pcbi.1004962.

[26] K. Jaganathan, S. Kyriazopoulou Panagiotopoulou, J.F. McRae, S.F. Darbandi, D. Knowles, Y.I. Li, J.A. Kosmicki, J. Arbelaez, W. Cui, G.B. Schwartz, E.D. Chow, E. Kanterakis, H. Gao, A. Kia, S. Batzoglou, S.J. Sanders, K.K. Farh, Predicting splicing from primary sequence with deep learning, Cell 176 (2019) 535–548, https://doi.org/10.1016/j.cell.2018.12.015, e24.

[27] M. Barenboim, B.J. Zoltick, Y. Guo, D.R. Weinberger, MicroSNiPer: a web tool for prediction of SNP effects on putative microRNA targets, Hum. Mutat. 31 (2010) 1223–1232, https://doi.org/10.1002/humu.21349.

[28] C. Liu, F. Zhang, T. Li, M. Lu, L. Wang, W. Yue, D. Zhang, MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs, BMC Genom. 13 (2012) 661, https://doi.org/10.1186/1471-2164-13-661.

[29] Primer3Plus. ⟨http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi⟩. Accessed 1 April 2022.

[30] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera–a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612, https://doi.org/10.1002/jcc.20084.

[31] U.C.S.F. CHIMERA. ⟨https://www.cgl.ucsf.edu/chimera/download.html⟩. Accessed 1 April 2022.

[32] Protein Data Bank ⟨https://www.rcsb.org⟩. Accessed 1 April 2022.

[33] UniProt. ⟨https://www.uniprot.org/uniprot/Q28943⟩. Accessed 1 April 2022.

[34] I.-Mutant Suite Predictor ⟨http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi⟩. Accessed 1 April 2022.

[35] E. Capriotti, P. Fariselli, I. Rossi, R. Casadio, A three-state prediction of single point mutations on protein stability changes, BMC Bioinforma. 9 (Suppl 2) (2008) S6, https://doi.org/10.1186/1471-2105-9-S2-S6.

[36] B.A. Jacobs, H. Rosing, N. de Vries, D. Meulendijks, L.M. Henricks, J.H. Schellens, J.H. Beijnen, Development and validation of a rapid and sensitive UPLC-MS/MS method for determination of uracil and dihydrouracil in human plasma, J. Pharm. Biomed. Anal. 126 (2016) 75–82, https://doi.org/10.1016/j.jpba.2016.04.039.

[37] O. Capitain, V. Seegers, J.P. Metges, R. Faroux, C. Stampfli, M. Ferec, T.M. Budnik, H. Senellart, V. Rossi, N. Blouin, J. Dauve, M. Campone, Comparison of 4 screening methods for detecting fluoropyrimidine toxicity risk: identification of the most effective, cost-efficient method to save lives, 1559325820951367, Dose Response 18 (2020), https://doi.org/10.1177/1559325820951367.

[38] HaploReg v4.1 ⟨https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php⟩. Accessed 1 April 2022.

[39] RegulomeDB v2.0 ⟨https://regulomedb.org/regulome-search/⟩ Accessed April 1, 2022.

[40] Ensembl's Variant Effect Predictor ⟨https://www.ensembl.org/info/docs/tools/vep/index.html⟩. Accessed 1 April 2022.

[41] D. Dobritzsch, K. Persson, G. Schneider, Y. Lindqvist, Crystallization and preliminary X-ray study of pig liver dihydropyrimidine dehydrogenase, Acta Crystallogr D. Biol. Crystallogr 57 (2001) 153–155, https://doi.org/10.1107/s0907444900015250.

[42] M. Kozyra, M. Ingelman-Sundberg, V.M. Lauschke, Rare genetic variants in cellular transporters, metabolic enzymes, and nuclear receptors can be important determinants of interindividual differences in drug response, Genet Med. 19 (2017) 20–29, https://doi.org/10.1038/gim.2016.33.

[43] A.B. van Kuilenburg, J. Meijer, D. Maurer, D. Dobritzsch, R. Meinsma, M. Los, L. C. Knegt, L. Zoetekouw, R.L. Jansen, V. Dezentje, L.H. van Huis-Tanja, R.J. van Kampen, J.M. Hertz, R.C. Hennekam, Severe fluoropyrimidine toxicity due to novel and rare DPYD missense mutations, deletion and genomic amplification affecting DPD activity and mRNA splicing, Biochim Biophys. Acta Mol. Basis Dis. 2017 (1863) 721–730, https://doi.org/10.1016/j.bbadis.2016.12.010.

[44] Q. Xiao, Y. Zhou, S. Winter, F. Buttner, E. Schaeffeler, M. Schwab, V.M. Lauschke, Germline variant burden in multidrug resistance transporters is a therapy-specific predictor of survival in breast cancer patients, Int J. Cancer 146 (2020) 2475–2487, https://doi.org/10.1002/ijc.32898.

[45] S. Shrestha, C. Zhang, C.R. Jerde, Q. Nie, H. Li, S.M. Offer, R.B. Diasio, Gene-specific variant classifier (DPYD-Varifier) to identify deleterious alleles of dihydropyrimidine dehydrogenase, Clin. Pharm. Ther. 104 (2018) 709–718, https://doi.org/10.1002/cpt.1020.

[46] Y. Wu, R. Li, S. Sun, J. Weile, F.P. Roth, Improved pathogenicity prediction for rare human missense variants, Am. J. Hum. Genet 108 (2021) 1891–1906, https://doi.org/10.1016/j.ajhg.2021.08.012.

[47] M. de With, J. Knikman, F.M. de Man, C. Lunenburg, L.M. Henricks, A.B.P. van Kuilenburg, J.G. Maring, M.C. van Staveren, N. de Vries, H. Rosing, J.H. Beijnen, D. Pluim, A. Modak, A.L.T. Imholz, R.H.N. van Schaik, J.H.M. Schellens, H. Gelderblom, A. Cats, H.J. Guchelaar, R.H.J. Mathijssen, J.J. Swen, D. Meulendijks, Dihydropyrimidine dehydrogenase phenotyping using pretreatment uracil: a note of caution based on a large prospective clinical study, Clin. Pharm. Ther. (2022), https://doi.org/10.1002/cpt.2608.

[48] Y. Zhou, V.M. Lauschke, Computational tools to assess the functional consequences of rare and noncoding pharmacogenetic variability, Clin. Pharm. Ther. 110 (2021) 626–636, https://doi.org/10.1002/cpt.2289.

[49] D. Banerjee, P. Mayer-Kuckuk, G. Capiaux, T. Budak-Alpdogan, R. Gorlick, J. R. Bertino, Novel aspects of resistance to drugs targeted to dihydrofolate reductase and thymidylate synthase, Biochim Biophys. Acta 1587 (2002) 164–173, https://doi.org/10.1016/s0925-4439(02)00079-0.

[50] K. Yoshinare, T. Kubota, M. Watanabe, N. Wada, H. Nishibori, H. Hasegawa, M. Kitajima, T. Takechi, M. Fukushima, Gene expression in colorectal cancer and in vitro chemosensitivity to 5-fluorouracil: a study of 88 surgical specimens, Cancer Sci. 94 (2003) 633–638, https://doi.org/10.1111/j.1349-7006.2003.tb01495.x.