UNIVERSITY OF UDINE

DEPARTMENT OF "AREA MEDICA"

XXXII Cycle of

BIOMEDICAL SCIENCES and BIOTECHNOLOGY

PhD Thesis

PHARMACOGENETICS OF FLUOROPYRIMIDINES: FROM THE IMPLEMENTATION OF GUIDELINES TO IDENTIFICATION OF NOVEL GERMLINE VARIANTS RELATED TO TOXICITY

PhD student:
Ecca Fabrizio

Supervisor:
Dr. Cecchin Erika

ACADEMIC YEAR 2018-19

Corso di dottorato di ricerca in:

"Scienze Biomediche e Biotecnologiche"

Ciclo 32°

Titolo della tesi

"PHARMACOGENETICS OF FLUOROPYRIMIDINES: FROM THE IMPLEMENTATION OF GUIDELINES TO IDENTIFICATION OF NOVEL GERMLINE VARIANTS RELATED TO TOXICITY"

Dottorando

Ecca Fabrizio

Supervisore

Dr. Cecchin Erika

Anno 2020

This thesis work
was carried out at the doctoral school
in Biomedical Sciences and Biotechnology,
Department of "Area Medica" of the University of Udine,
directed by Prof. Claudio Brancolini
in collaboration with the Complex Operational Structure
of Experimental and Clinical Pharmacology
of Oncological Reference Center of Aviano (I.R.C.C.S),
directed by Dr. Giuseppe Toffoli

Abstract

Precision medicine finds a valuable ally in pharmacogenomics, which is aimed at identifying genetic variants involved in therapeutic response. In tailored treatments, a specific attention is required for those drugs showing a low therapeutic index, such as fluoropyrimidines (FL) that represent the backbone in the management of many types of solid tumours. FL lead to severe toxicity (G≥3) in a considerable percentage of patients, occasionally resulting in patients death. FL are mainly metabolized by dihydropyrimidine dehydrogenase (DPD; *DPYD*) into inactive metabolite. To date, only four *DPYD* genetic variants (*DPYD* *2A, *DPYD* *13, *DPYD* c.2846A>T and *DPYD*-HapB3) are known to decrease DPD activity and consequently to be associated to FL-related toxicity. The four SNP genotyping test is recommended before a FL-based treatment by international shared pharmacogenetic guidelines, Despite the high specificity, this test demonstrated a low sensitivity, leaving the majority of the toxic events unexplained. Recently, the previously neglected role of novel and rare genetic variants in the intra-individual variability in the response to drugs has emerged.

In this PhD thesis, we aimed at improving the knowledge on the clinical validity and clinical utility of the *DPYD* clinical diagnostic test proposed by the current pharmacogenomics guidelines, by investigating the impact of the four *DPYD* variants on fluoropyrimidines-related toxicity and costs for toxicity management in large cohorts of patients. In addition, we analysed the contribution of novel and rare genetic variants in a set of pharmaco-genes in the development of fluoropyrimidines-related toxicity.

Here, we show how the four-SNPs test for *DPYD* allow to identify a large number of patients who will develop severe adverse reactions to FL-based treatment in a group of 763 oncological patients. This is especially true if the four DPYD SNPs are arranged in the so-called gene activity score (GAS) weighing each SNP according to its specific impact on the DPD protein phenotype. We also demonstrated, in a subset of 550 colorectal cancer patients, how the DPYD genotype for the four-SNPs set or the GAS are related to the management cost for severe toxicities related to FL.

This finding has notable implication in the utility of the test warranting its application in the routine clinical practice. We subsequently analyzed, by means of a next generation sequencing approach, the potential role of novel or rare variants in genes in the detoxification pathway in 108 patients that are wild type for the well-know 4 *DPYD* risk variants and developed severe FL-related toxicity. We identified 9 rare *DPYD* variants, 3 were very rare missense (c.G345C, p.M115I; c.A2060C p.D687A and c.A2137G, p.N713D), and 6 were singleton missense (c.A110G, p.D37G; c.G481A, p.E161K; c.C800T, p.T267I; c.G958A, p.G320R; c.A1110G, p.I370M;c.C1579T, p.P527S). In

comparison, a population of 106 patients treated with a FL-based regimen, without any adverse reaction, displayed a depletion of novel variants (p value =0.001, by Fisher's Exact Test). Also, as expected, we found a statistically significant enrichment of rare, very rare and novel exonic *DPYD* variants in the cohort of patients with severe toxicity as compared to those without (P =0.0187, P =0.0187, P =0.0291, respectively, by Fisher's Exact Test).

Despite the clinical need of tailoring treatment on each patient, to date translating the genetic information of each patient into a therapeutic indication remains a challenging task. The study of the pharmacogenomics in FL-based treatment has already provided useful and valid markers for the clinical practice. Therefore, we believe that the study of rare and novel genetic variants could in the future further improve available tools for a precision medicine in cancer treatment.

Contents

Abstract i
Contents iii
1. Introduction 1
1.1. Fluoropyrimidine 1
1.2. Adverse Drug Reactions (ADRs) 3
1.3. Pharmacogenetics and pharmacogenomics 4
1.4. Pharmacoeconomy 8
1.5. *DPYD*, *MTHFR* and *TYMS* 8
1.6. Rare Variants 11
2. Rationale 13
3. Aims 15
4. Results 16
4.1 Association between *DPYD* genotype and FL-related DLTs 16
4.2. Toxicity and costs associated with its management 19
4.3. Rare variants, cases cohort 23
4.4. Rare variants distribution in the cases cohort 25
4.5. *DPYD* rare, very rare and novel variants 29
4.6. Rare, very rare and novel variants in FL-pathway 32
4.7. Prediction tools analysis 33
4.8. Molecular visualization 35
4.9. Quantification of serum uracil and dihydrouracil concentrations 37
4.10. Rare variants, control cohort 37
4.11. Rare variants distribution in the control cohort 38
5. Discussion 44
6. Material and Methods 50
6.1. Cohorts 50
6.2. Data collection 51
6.3. Toxicity management cost 51
6.4. *DPYD* Genotyping 51
6.4.1. Pyrosequencing 52
6.4.2. Allelic discrimination in real-time PCR by fluorescent-labeled probes 53
6.4.3. Sanger Sequencing for *DPYD* *2A; *13; c.2846A>T, and HapB3 variants validation 54
6.5. Selection of candidate genes for Next Generation Sequencing 56
6.6. Targeted Next Generation Sequencing 58
6.7. Bioinformatic analysis 60
6.8. In silico functional prediction tools 60
6.9. Sanger Sequencing to validate the *DPYD* variants detected by NGS 61
6.10. Molecular visualization 62
6.11. Quantification of serum uracil and dihydrouracil concentrations 62
6.12. Statistical analysis 63
7. References 64
8. Publications 78

1. Introduction

1.1. Fluoropyrimidine

Since the synthesis in 1957 by Heidelberger et al. of 5-Fluorouracil (5-FU) until now fluoropyrimidines (FL) have become the backbone of several anti-tumour therapies.[1] Fl are widely used for the treatment of multiple solid tumours including breast, gastrointestinal, head-neck, pancreatic, ovarian, skin and liver cancer.[2-9] FL are a class of pyrimidine analogue, specifically antineoplastic antimetabolite, that include 5-FU and two inactive prodrugs capecitabine and tegafur, that are metabolized to 5-FU (***Figure 1*).**

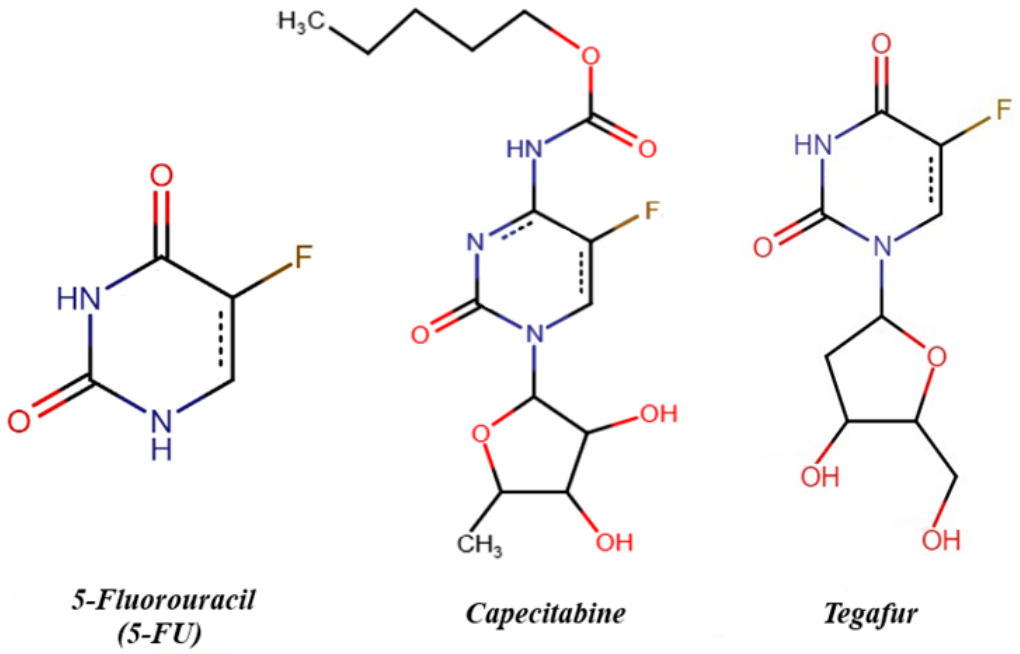


Figure 1: *2D Structure of fluoropyrimidines*

Capecitabine and tegafur are available in oral formulation, thus they are absorbed through the gut wall and are converted into the active metabolite in the liver. On the contrary, 5-FU is commonly administered intravenously and more than 80% is metabolized in the liver by the enzyme dihydropyrimidine dehydrogenase (DPD; *DPYD*) into the inactive metabolite dihydrofluorouracil (DHFU), excreted in urine by kidneys catabolism.[10–14]

Capecitabine (N^4-pentyloxycarbonyl-5'-deoxy-5-fluorocytidine) is converted to 5-FU through a three enzyme cascade and the final step requires the enzyme thymidine phosphorylase (TP;

TYMP).[10] Capecitabine was designed to be highly specific for tumors cells since TP enzyme is significantly more active in tumor than in normal cells.[10]

Tegafur (1-(2-Tetrahydrofuryl)-5-fluorouracil), a prodrug of 5-FU was developed about 50 years ago, is mainly activated in the liver by cytochrome P450 2A6 (CYP2A6) which hydroxylates the drug to 5'-hydroxitegafur, which is spontaneously dehydroxylated to 5-FU.[15,16]

5-FU exhibits its anticancer activity through several mechanism. First of all, it exerts a cytotoxic function by a direct incorporation into the nascent DNA filament during the new DNA synthesis phase (S phase) of the cell cycle resulting in several DNA strand breaks.[17,18] Moreover, 5-FU can be metabolized into fluorouridine diphosphate (FUDP) which is phosphorylated into fluorouridine triphosphate (FUTP) (***Figure 2***). As 5-FU, FUTP can be directly incorporated into RNA chain. The mis-incorporation of FUTP into RNA lead to: inhibition of the pre-rRNA conversion into mature RNA;[19,20] the inhibition of polyadenylation process;[21] the disruption of post-transcriptional modification of tRNAs;[22,23] alteration of the assembly and activity of snRNA/protein complexes with consequences on splicing; and the post-transcriptional conversion of uridine to pseudouridine (Ψ).[24–26] FUDP could be also converted into fluorodeoxyuridine diphosphate (FdUDP) by ribonucleotide reductase enzyme (RR; *RRM1/2*) (***Figure 2***).

FdUDP can be phoshorylated or dephosphorylated into FdUTP or FdUMP, two active metabolites. FdUTP can be incorporated into DNA, thus inhibiting further DNA synthesis.[17] An alternative activation pathway lead to production of FdUMP through thymidine kinase (TK; *TK1*) that inhibits thymidylate synthase (TS; *TYMS*) by forming a stable ternary complex.[27–29] TS is an enzyme belonging to the folate pathway, that converts UMP to dTMP with provision of a carbon donated by 5,10-methylene tetrahydrofolate (5,10-MTHF), the only *de novo* source of thymidylate, which is necessary for DNA replication and repair.[30] Inhibition of DNA and RNA synthesis associated with RNA/DNA misincorporation lead to cell death.

FL are used in different clinical setting, including adjuvant and metastatic treatment, and in combination with several antineoplastic drugs such as Irinotecan and Oxaliplatin or with monoclonal antibodies such as Cetuximab, Bevacizumab and others.[31] FL showing a good efficacy profile are generally well tolerated, thus representing a valuable therapeutic strategy in the management of solid tumours. However, about 10-26% of patients treated with FL-based regimens develops early-onset severe or life-threatening toxicity.[32,33] It is well known that pharmacokinetic parameters could influence the drug toxicity profile generally due to the presence of increase aging, chronic inflammation, oxidative stress and cellular senescence leading to an increased risk to develop severe toxicity, mainly in elderly patients.[34]

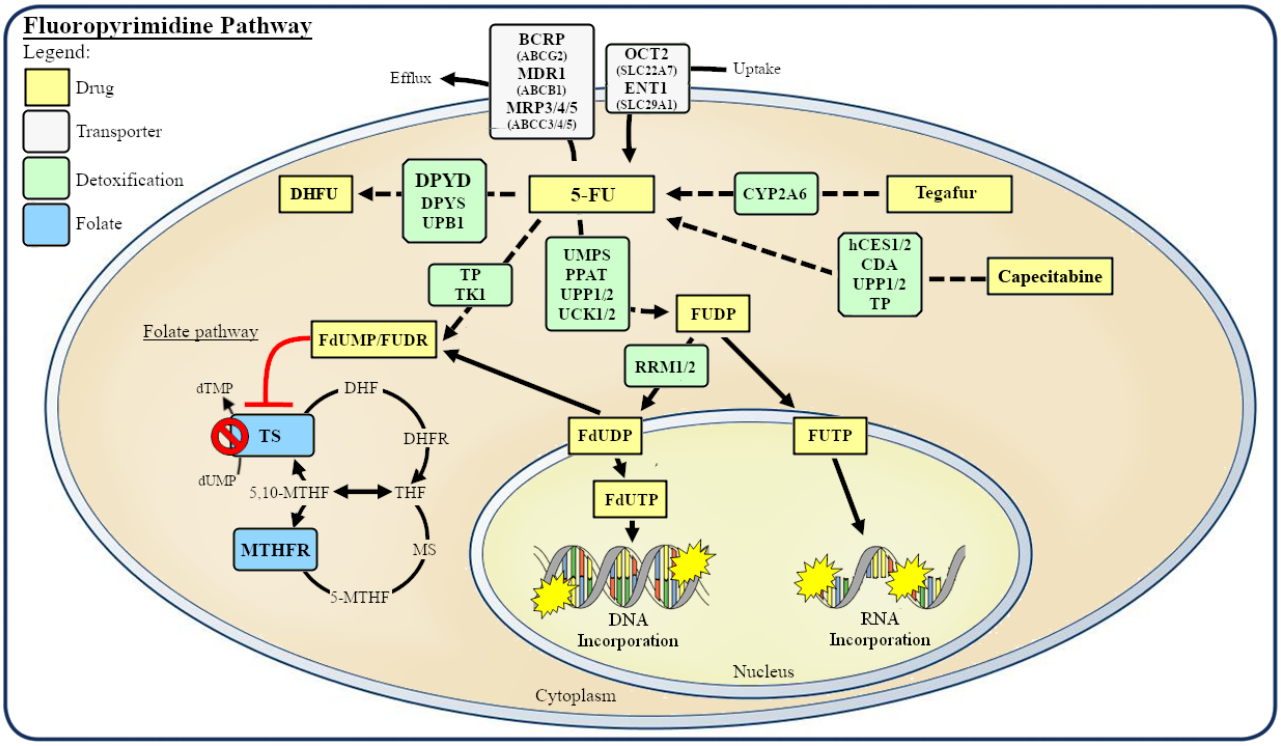


Figure 2: *Fluoropyrimidines Pathway*

1.2. Adverse Drug Reactions (ADRs)

Edwards and Aronson defined adverse drug reaction as "an appreciably harmful or unpleasant reaction, resulting from an intervention related to the use of a medicinal product, which predicts hazard from future administration and warrants prevention or specific treatment, or alteration of the dosage regimen, or withdrawal of the product".[35]

A careful balance between efficacy and toxicity is of primary importance in oncology due to low therapeutic index (TI) of the anticancer drugs. According to Common Terminology Criteria for Adverse Events (CTAEs) the ADRs range from mild and moderate (respectively grade 1 and 2) to severe, life-threatening and death (respectively grade 3, 4 and 5). The combination with other chemotherapeutic drugs exacerbates the appearance and grade of FL-toxicity.

FL-related toxicity can be divided into two main groups: hematological and non-hematological toxicity. Hematological toxicity includes anemia, leucopenia, thrombocytopenia and neutropenia with a consequent increased risk of systemic infections. Non-hematological toxicity includes cardiotoxicity, hepatic toxicity, mucositis and stomatitis, asthenia, hand-foot syndrome, alopecia, nausea/vomiting and diarrhea.[36]

With the high number of patients receiving FL-based therapies, and the significant effects of toxicities on their quality of life, it is of major clinical interest to reduce the incidence of FL-related adverse events. Furthermore, the adverse effects lead to an interruption of the therapy which affects

not only the efficacy of it, but also involves an additional management cost for the national healthcare system.[37] Indeed, adverse reactions might require the use of specific drugs, treatments and hospitalization with the consequence increase of management cost.

Although fluoropyrimidines are highly used in the treatment of solid tumors, they show a high inter-individual heterogeneity in the response to treatment, both in terms of efficacy and in the development of adverse events (AEs). The high heterogeneity may depend on several factors including environmental, physiological and pathological factors. In a similar way to a multifactorial disease, environmental factors affecting the safety profile of FL include life style, smoking, diet and the concomitant presence of other therapies.[38] Nevertheless, physiological factors such as age, sex, renal and hepatic function can also affect treatment response as well as pathological factors such as the stage and type of the disease and the concomitant presence of other diseases.[38] Eventually, differences in human genome mainly contribute to the high heterogeneity in treatment response.[39]

1.3. Pharmacogenetics and pharmacogenomics

Pharmacogenetics, term coined by Motulsky in the 1957, describes the influence of genetic factors on drug therapy.[40]

However, we prefer using the term pharmacogenomics (PGx), which considers the impact of genomics and epigenetics to the drug response, whereas pharmacogenetics does not, although the two terms are often used interchangeably. Pharmacogenomics, as well as personalized medicine, aimed at identifying genetic biomarkers. The identification of genetic biomarkers could allow the prevention of toxicity events and improve the effectiveness of therapy based on the genetic characteristics of each patient.

Genetic biomarkers include single nucleotide variants (SNPs), accounting for the 90% of all human genetics variants, copy number variants (CNVs) and small or large insertion and deletion (indels).[41] Moreover, epigenetic variants and large rearrangement like deletion and inversion, can represent a suitable biomarker.

Pharmacogenomics represents the present and the future of drug therapy due to the rapidly growing field of precision medicine.[38] Pharmacogenomics screening for several drugs remains a very powerful, cost-effective diagnostic tool that allows clinicians to tailoring the therapy for each patient.

Among the limitations of pharmacogenomics to create a clear prescription for all patients, remain the difficult task for clinicians in translating genetic information obtained from the pre-treatment laboratory tests, that is crucial to the right drug's prescription. Clinical Pharmacogenetics

Implementation Consortium (CPIC) and the Royal Dutch Association for the Advancement of Pharmacy (DPWG) aimed to overcome this barrier moving toward the clinical implementation of pharmacogenetics tests by creating clinical practice guidelines.[42,43] Founding in the 2009 with the aim to create a curate, peer-reviewed guidelines, CPIC is composed by members of Pharmacogenomics Research Network, Pharmacogenomics Knowledge Base (PharmGKB) and several experts of pharmacogenomics.

In the case of fluoropyrimidines, the CPIC guidelines consider the four *DPYD* variants well-known for the impact they have on enzymatic function and on toxicity risk: *DPYD* c.1905+1G>A (rs3918290, also known as *2A, IVS14), *DPYD* c.1679T>G (rs55886062, *13, p.I560S), *DPYD* c.2846A>T (rs67376798, p.D949V), and *DPYD* c.1129–5923C>G (rs56038477, HapB3)[42]. *DPYD* gene is located in the long arm of chromosome 1 (Chr1p21.3) and encodes for the initial and rate-limiting enzyme (DPD) involved in the catabolism of the pyrimidine bases: uracil and thymine.[44–46] *DPYD* contains 23 exons that encode for 1,025 amino acids.[47] DPD dimer activity has shown to be highly variable in the population, with an estimated 3% to 5% of the population being partially DPD deficient.[48]

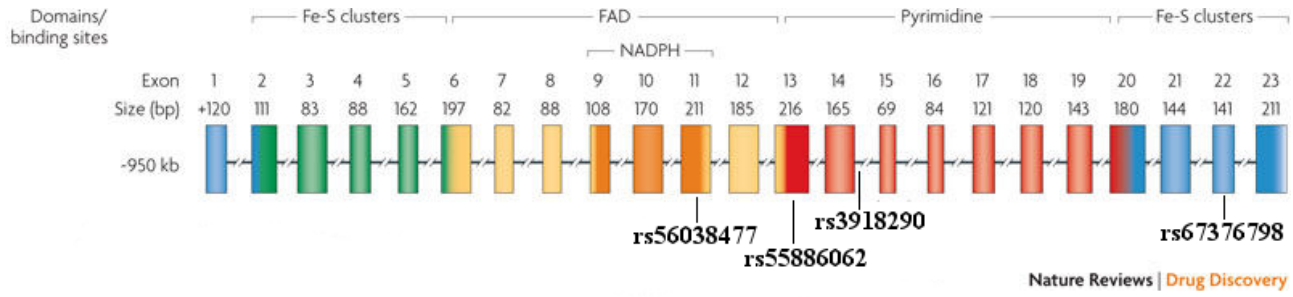


Figure 3: *DPYD gene and risk variants location. Adapted from Goodsaid F. M. et al., 2010*

DPYD rs3918290 and rs55886062 have the most deleterious impact on DPD protein activity while rs67376798 and rs56038477 have a moderate impact in reducing protein activity.

DPYD rs3918290is an intronic splicing variant which leads to the skipping of the whole exon 14, whereas *DPYD* rs55886062 and rs67376798 are two missense mutation located in exon 13 and 11, respectively (***Figure 3***).

Rs56038477 is composed by 3 different variants in linkage disequilibrium and introduces a cryptic splice site in intron 10 which results in a partial production of a nonfunctional transcript (***Figure 4***).

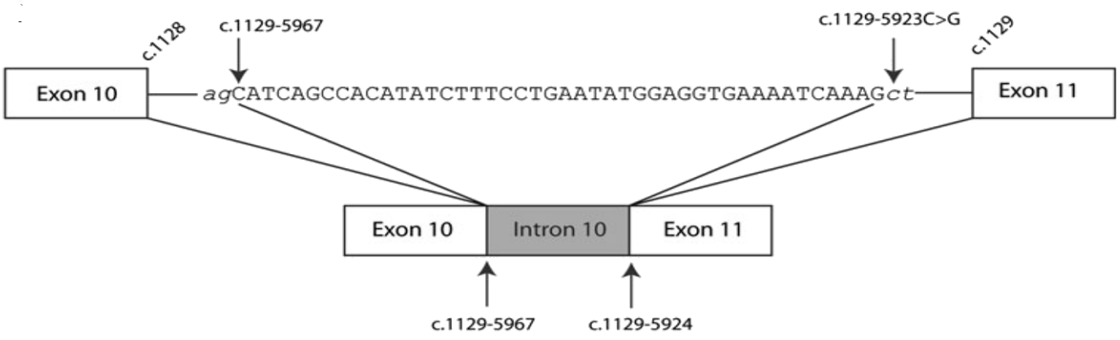


Figure 4: *DPYD Haplotype B3, Van Kuilenburg, A. B. et al. 2010*

Rs56038477 is the most common *DPYD* functional variant in Europe with an allele frequency of 4.7%. On the other hand, *DPYD* rs3918290, rs55886062 and rs67376798 variants are considered rare and common (0.2-1.6%, respectively) in Caucasian population.[37] Taken together, about ~7% of Europeans carries at least one *DPYD* decreased function variant.

To simplify and standardize the effect of these variants, *DPYD* alleles were scored a "gene activity score" (GAS), as firstly described by Steimer et al. for *CPY2D6* gene allele.[49,50]

DPYD alleles are divided in three categories according to the protein activity: normal metabolizer (value of 2), intermediate metabolizer (value of 1.0 or 1.5) and poor metabolizer (value of 0). Based on genotype, gene activity score represents the enzymatic phenotype of the patient. These results represent an easy method to describe the phenotype and can be used for adjusting the dose of FL. CPIC guidelines provide recommendations of different strength about better FL starting dose to use for each patient based on a standard method that grades levels of evidence linking *DPYD* alleles to these phenotypic categories in the literature.

For poor metabolizers patients (GAS = 0.5 or 0), it is strongly recommended to avoid the use of 5-fluorouracil-based regimens. If it is not possible to remove FL, a starting dose with a 75% of reduction is recommended (***Table 1***). Patients who are intermediate metabolizer (GAS = 1.0 or 1.5) should receive a reduced starting dose of 50% compared to the standard dose (***Table 1***).

This method can be used to standardize FL dose adjustments for each patient resulting in optimal safety and effectiveness of FL-based regimens. Moreover, it provides the possibility to include novel SNPs and rare variants which may be identified by future whole exome (WES) and whole genome sequencing (WGS) project.[49]

LIKELY PHENOTYPE	ACTIVITY SCORE	GENOTYPES	IMPLICATIONS	DOSING RECOMMENDATIONS	CLASSIFICATION OF RECOMMENDATIONS
DPYD Normal Metabolizer	2	An individual carrying two normal alleles	Normal DPD activity and "normal" risk for fluoropyrimidine toxicity	Based on genotype, there is no indication to change dose or therapy. Use label-recommended dosage and administration	Strong
DPYD Intermediate Metabolizer	1 or 1.5	An individual carrying one normal function allele plus one no function allele or one decreased function allele, or an individual carrying two decreased function alleles	Decreased DPD activity (leukocyte DPD activity at 30% to 70% that of the normal population) and increased risk for severe or even fatal drug toxicity when treated with fluoropyrimidine drugs	Reduce starting dose by 50% followed by titration of dose based on toxicity or therapeutic drug monitoring (if available). Patients with the c.[2846A>T];[2846A>T] genotype may require >50% reduction in starting dose	Activity score 1: Strong Activity score 1.5: Moderate
DPYD Poor Metabolizer	0 or 0.5	An individual carrying two no function alleles or an individual carrying one no function plus one decreased function allele	Complete DPD deficiency and increased risk for severe or even fatal drug toxicity when treated with fluoropyrimidine drugs	Activity score 0.5: Avoid use of 5- fluorouracil or 5-fluorouracil prodrug-based regimens. In the event, based on clinical advice, alternative agents are not considered a suitable therapeutic option, 5-fluorouracil should be administered at a strongly reduced Activity score 0: Avoid use of 5-fluorouracil or 5-fluorouracil prodrug-based regimens.dose[k] with early therapeutic drug monitoring	Strong

Table 1: *CPIC recommendations*

Current pre-treatment *DPYD* genetic test including the 4 risk variants allows the prediction of the major number of adverse events although several toxic reactions cannot be predicted due to the low sensitivity of the test.[51] It is important to note that patients carrying one of the four variants have a double risk to develop severe toxicity.[51]

Although available CPIC guidelines, the application of pretreatment *DPYD* genotyping for clinical purposes is not mandatory.[52,53] One reason may be the lack of information about the usefulness and cost-effectiveness of pre-treatment genotyping test.[53] Another limiting factor is that toxicity events do not always occur in the intermediate and poor metabolizer. Moreover, only the 50% of patients carrying at least one risk variants develop severe toxicity when treated with the standard doses. This issue may depend on the regimens of treatment administered or on the number of treatment's cycles.[54,55] On the other hand, severe toxicity may occur in wild-type *DPYD* patients due to environmental, physiological, pathological and genetics factors. Indeed, recently common polymorphisms in *DPYD*-regulatory genes and genes involved FL-detoxification pathway have been associated with FL-related toxicity.

1.4. Pharmacoeconomy

The limited diffusion of preemptive *DPYD* tests in routine clinical practice is not only a consequence of the low-test sensitivity, which allow the prediction of a few number of adverse events (AEs), but is also caused by a limited interest from clinicians. Furthermore, the tricky translation of genetic information from laboratory tests to a possible therapy remains one of the hardest tasks for clinicians and limits these test use. Another reason may be the lack of information about the useful and cost-effectiveness of pretreatment test. For these reasons, an economic evaluation of toxicity management costs can be a good tool to point out the utility of preclinical tests.

Our earlier studies in Italian Healthcare setting indicates that patients without any *DPYD* risk variants had more quality adjusted life years (QALYs) and better clinical outcome, when compared with patients carrying at least one *DPYD* risk variants.[37,56] Moreover, patients carrying *DPYD* risk variant experience higher grade toxicity events and higher management cost. Our results demonstrated that toxicity management cost depends on patients genotype and hospitalization is the most expensive intervention.[57] Other studies reported similar results which remark the utility of *DPYD* pre-treatment genotyping strategies to improve patients' therapy, quality of life and to reduce FL-treatment related cost.[58,59] These results support the role of pharmacogenomics *DPYD* test as a powerful and cost-effective tool to improve FL-based treatment.

1.5. *DPYD*, *MTHFR* and *TYMS*

More than 160 SNPs have been reported in *DPYD* gene but only four of them presented a sufficient evidence for clinical impact to be included in the current guidelines.[60]

Defects affecting other regulatory mechanisms in the expression of DPD have been extensively investigated and are deemed to be responsible for FL toxicity events. Indeed, *DPYD* locus can be regulated at several different levels.

At transcriptional level, *DPYD* gene can be regulates by transcriptional factors AP-3 and SP-1 binding and by additional tissue-specific transcriptional factors. *DPYD* promoter lacks the typical TATA or CCAAT boxes but have several GC-rich regions containing potential cis-regulatory elements.[61] On these basis genetic variants in *DPYD* promoter region and in the transcriptional factors can affect the protein expression.

At post-transcriptional level, DPD mRNA can interact with miRNA-27a and miRNA-27b. These two miRNAs repress DPD expression through the interaction with two conserved recognition sites

in the 3'UTR.[62] A common polymorphism (rs895819) in miRNA-27a gene (*MIR27A*) is associated with a decreased expression of DPD and with an early onset of toxicity.[63]

Eventually, the methylation status of *DPYD* promoter seems to correlate with DPD expression in a tissue specific manner. It has been shown that its expression is linked to methylation status of lysine 27 in histone H3 (H3K27) which is in turn regulated by demethylase UTX (*KDM6A*) and methyltransferase EZH2.[64–67]

Several genes, in addition to *DPYD*, have emerged from different studies to be related to the FL-toxicity. Among these genes the most debated are *MTHFR* and *TYMS*.

MTHFR is a gene located in the chromosome 1 (Chr1p36.22) (***Figure 5***), which encodes for methylenetetrahydrofolate reductase (MTHFR). MTHFR protein catalyzes, as homodimer, the conversion of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate, a co-substrate for homocysteine re-methylation to methionine.[68]

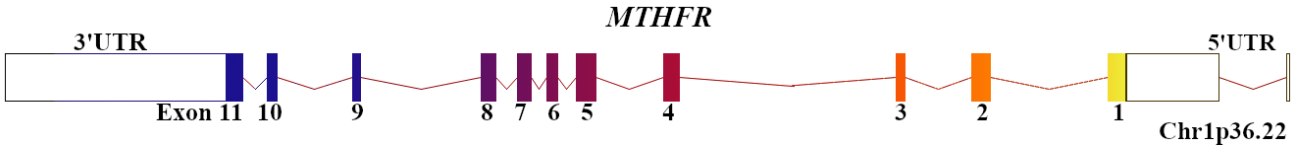


Figure 5: *MTHFR gene.*

MTHFR known variants are associate with several diseases and clinical response to FL treatment. The two most common SNPs in *MTHFR* (rs1801133 and rs1801131) have been largely investigated in tumor and healthy tissues to assess if they may interfere with FL-treatment. Despite the efforts, no clear-cut conclusions for the association with these two polymorphism and FL-related toxicity emerged.[69–72]

Another gene that might play a limited role in FL-related toxicity is *TYMS* which is located in chromosome 18 (Chr18p11.32) and encodes for thymidylate synthase (TS), enzyme responsible for the maintaining of intracellular dTMPs pool, a crucial factor for DNA replication and repair.[73]

TS homodimer is the primary target of FL and in tumor tissue the different mRNA/protein expression of TS may influence FL-based treatments. Indeed, preclinical and *in vitro* studies have demonstrated that TS overexpression leads the development of FL resistance resulting in a poor response to treatment and in a lower frequency of toxicity events.[74–76] On the contrary, low level of TS correlate with an increase sensitivity toward FL, resulting in an increased risk to developing toxicity and in a better response to treatment.[76]

TS protein act as RNA-binding protein that negatively regulate its own mRNA by binding a stem-loop structure in 5'UTR.[77–79] Moreover, TS expression is regulated by the presence of 28bp short tandem enhancer repeats (TSER) at the end of 5'UTR.[80]

In world population the 28bp sequence is frequently present in double (2R) or triple (3R) repetition.[81] *In vitro* studies with reporter gene highlights that 3R/3R genotype have an activity of 2.6 fold higher than that with 2R/2R genotype.[77] The 3R genotype leads to an overexpression of TS mRNA/protein and this protects cells of both normal and tumor tissue against damages induced by FL therapy due to the low efficiency of TS inhibition.[82,83] Patients carrying 3R/3R genotype show a significant lower risk of toxicity and a lower response rate when compared with 2R/2R genotype.[84] However, the role of this region is still misunderstood. A few studies have shown that at least one copy of the 28-bp repeat is require for stem-loop formation and transcription.[85,86] Moreover, this mechanism might need an inverted repeat located upstream from the TSER to enhance expression.[73,87]

Eventually, another mechanism elucidated by Dolnick et al. may influence TS expression and the FL-treatment outcome.[88] The *TYMS* antisense RNA gene, *ENOSF1* encodes for a mitochondrial protein that can generate different isoform through the alternative splicing mechanism. One of these has an overlapped 3'UTR with *TYMS* 3'UTR which allows it to bind and inactivate TS mRNA through the RNA cleaving (***Figure 6***).[88–90]

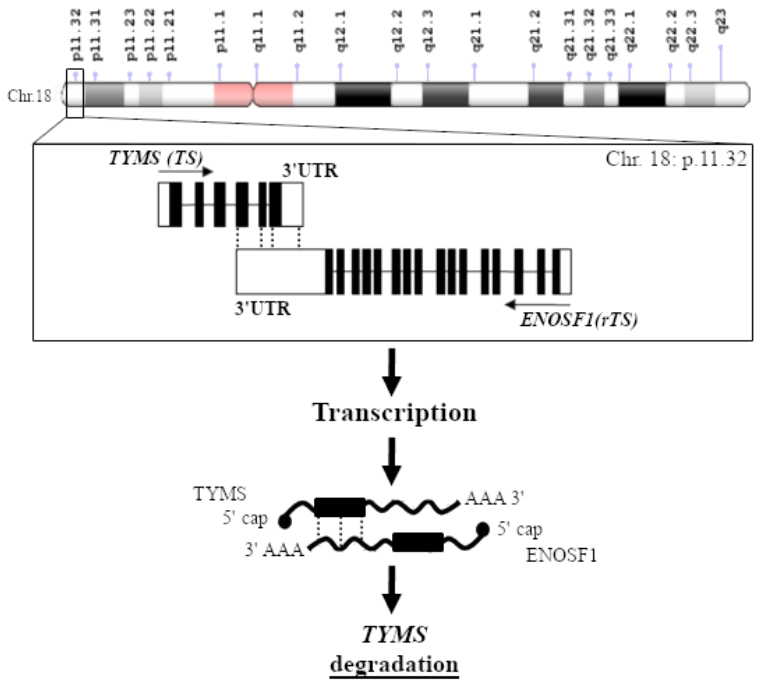


Figure 6: *Post-trascriptional regulation of TYMS mRNA.*

1.6. Rare Variants

In the last 20 years we have been witnessed to the rapid development of DNA sequencing techniques. From the publication of first human genome draft in 2001 to date sequencing technologies have undergone a remarkable evolution in both technique and costs.[91–93] The constant decreasing of genome sequencing costs and time with a consequent increase in the number of sequencing projects has brought to light a large amount of gene variants and has revolutionized the identification of genetic disorders.

It is widely accepted that genetic variations impact on human diseases. However, the increase number of projects and data provided by these platforms has shown how the presence of rare (minor allele frequency (MAF) ≤1) or novel variants greatly contributes to development of human diseases and therefore are a typical characteristic of the patient. In fact, in several databases more than 85% of present variants have been reported as rare.[39,94–96]

Genome-wide association studies (GWAS), whole-genome and whole-exome sequencing (WGS and WES, respectively) revealed that rare gene variants exert larger effect than common SNPs and may account for a part of heritability of some human phenotype.[95–97]

The advent of next generation sequencing (NGS) and the mapping of the human genome transformed the field of pharmacogenetics into pharmacogenomics which genetic variants, rare or common, in association with environmental and life style factors can affect drug adsorption, detoxification, metabolism, elimination (ADME) and toxicity risk.[98]

Several studies reported that more than 90% of the variants in the genes associated with pharmacodynamics (PDs) and pharmacokinetics (PKs) phenotype is characterized by a high frequency of rare variants.[39,98] It has been estimated that rare variants represent the 30-40% of functional variability in pharmacogenes, emphasizing the importance of including these variants in pre-treatment genetic screening in order to understand how genetic variations can be clinically relevant.[99]

The impact of rare variants on human disease and pharmacogenomics it is also indirectly confirmed by the traditional approaches that consider only the common variants (SNPs) as biomarker which fail to fully explain the high interindividual heterogeneity.[100] The rise of the NGS has allowed to find novel biomarkers for each patients leading to the development of precision medicine.[101]

Precision medicine is aimed at identifying genetic variants, which are specific for each patient, in order to select the best pharmacological approach to achieve the therapeutic success, considering not only the efficacy of drugs administered but also their safety profile.

Despite the inclusion of new genetic biomarkers in pre-treatment genetic test by means of high throughput technologies, which may play a pivotal role in the management of therapy, only a few rare variants are reported in the CPIC guidelines to date (e.g. *DPYD**2A).

Among the hurdles in clinical implementations of genetic tests, there is the fact that new discovered genetic variants with unknown functional significance require experimental control before their translational in clinical setting. Notably, due to the huge number of rare variants, their functional confirmation is not feasible.[102] Moreover, association studies aimed at detecting rare variants with clinical impact require high statistical power with consequences prohibitively large sample sizes to be translated into clinical practice.[99,102]

2. Rationale

In the era of precision medicine, clinicians struggle to select the right drug at the right moment, with the right dosage for the right patient. Precision medicine finds a valuable ally in pharmacogenomics which is aimed at identifying genetic variants involved in therapeutic response and that could be exploited as prognostic and predictive marker of drug efficacy and toxicity. This strategy is crucial especially in assessing the therapeutic treatment efficacy.

In tailored treatments, a specific attention is required for those drugs showing a low therapeutic index, such as chemotherapeutic agents. Among these drugs, a canonical example is provided by fluoropyrimidines, structural analogous of pyrimidines, composed by 5-fluorouracil (5-FU) and its two prodrugs capecitabine and tegafur. Fluoropyrimidines represent the backbone in the management of many types of solid tumours and are used in several combination regimens.

Despite fluoropyrimidines are used for the treatment of different malignancies as colorectal, gastrointestinal, breast and head-neck cancers, they display a huge heterogeneity in term of efficacy and toxicity. Indeed, they frequently lead to severe toxicity resulting in treatment interruption and delay in therapy administration. Unfortunately, these outcomes lead not only to treatment failure but also to discomfort for patients. Moreover, the Adverse Drug Reactions (ADRs) lead to a sharp rise in the cost for toxicities management for the National Healthcare Systems (NHSs), as reported by our group.

Fluoropyrimidines display a low therapeutic index which results in a narrow range of plasma concentration between the minimum effective dose (MED) and the maximum tolerated dose (MTD). Given this characteristic, it is reported that about 26% of patients experience severe drug-related hematologic toxicity, non-hematologic toxicity or a combination of them in course of FL-based treatment. Ultimately, FL-related toxicities lead to death in about 0.1% of cases. This phenomenon can be caused by a low clearance of the drug, thus leading patients to be exposed to drug concentration in plasma above the therapeutic window.

The 85% of 5-FU is metabolized in the liver by dihydropyrimidine dehydrogenase (DPD) enzyme into the inactive metabolite dihydrofluorouracil (DHFU), excreted in urine by kidneys catabolism. The Clinical Pharmacogenetics Implementation Consortium (CPIC) suggests a genetic test for the assessment of DPD catabolic activity before fluoropyrimidine-based treatment.

The currently validated genetic screening is based on four genetic variants associated to a reduced activity of DPD enzyme that are known to be related to fluoropyrimidines toxicity. DPD enzyme is encoded by *DPYD* gene, located in p21.3, the small arm near the centromeric region of

human chromosome 1. *DPYD**2A (c.1905+1G>C or rs3918290) and *DPYD**13 (c.1679T>G or rs55886062) are the most deleterious variants described in *DPYD* gene, whereas *DPYD* c.2846A>T (rs67376798) is associated to a less deleterious phenotype. Recently, a novel haplotype (Hap-B3) of *DPYD* gene has been identified and associated to an altered functionality of the enzyme resulting in a higher risk to develop severe toxicity (G≥3). PGx guidelines are not mandatory in clinical practice yet and the genetic test is still scarcely widespread among clinicians. Moreover, the low frequency of *DPYD* variants in the population negatively affect the adoption of pre-treatment genotyping.

The four SNPs-based genotyping tests showed a high specificity in predicting FL-related toxicities but a low sensibility. Consequently, these analyses do not allow the prediction of all toxicity events related to fluoropyrimidines treatment. Several recent studies suggest that novel and rare germline variants in pharmacogenes might contribute to the reported interindividual variability.

3. Aims

In this PhD thesis, we aimed at improving the knowledge on the clinical validity and clinical utility of the *DPYD* clinical diagnostic test proposed by the current pharmacogenomic guidelines, by investigating the impact of the four *DPYD* variants on fluoropyrimidines-related toxicity and cost. In addition, we analysed the contribution of novel and rare genetic variants in the development of fluoropyrimidines-related toxicity.

The project was designed and structured to analyse in depth the impact of patient genotype in the exacerbation of FL-related toxic effects. Therefore, we can summarize our work in three main parts:

- In the first part, we aimed at validating the clinical value of the pre-emptive *DPYD* test on a retrospective population of patients with CRC treated with FL-based regimens. We selected 763 patients with CRC for whom detailed toxicity events reported during FL-based chemotherapy were recorded. We genotyped them for the four *DPYD* risk variants (rs3918290, rs55886062, rs67376798 and rs56038477) by means of different sequencing technologies (Pyrosequencing, Real-Time PCR and Sanger Sequencing). Functional impact of these variants has been assessed by using a "gene activity score" (GAS), according to CPIC guidelines.
- In the second part, we hypothesized that the *DPYD* risk genotype could be not only associated to an increased risk of developing severe toxic events, but also to an increased cost for National Health Care System due to their clinical management. We analysed for the first time a subpopulation of 550 patients affected by CRC and assessed a relationship between *DPYD* genetic variants, toxicity and cost derived from the management of them.
- In the third part, we aimed at investigating the role of rare and novel germline variants in a panel of candidate genes involved in the FL-pathway to assess the relationship between rare variants and severe toxicity onset. To achieve this goal, we sequenced, by means of NGS in Illumina platform (MiSeq), a selected subpopulation of 108 *DPYD* wild type patients who experienced an extreme toxicity phenotype (grade 3 to 5) after or during FL-based treatment. As control group, we sequenced a cohort of 106 *DPYD* wild type patients treated with a fluoropyrimidine-based regimens who did not experienced severe toxicities (grade 0 to 2).

4. Results

4.1 Association between *DPYD* genotype and FL-related DLTs

A final set of 763 patients with CRC treated with FL-based chemotherapy was selected to assess the relationship between the 4 *DPYD* risk variants in clinical genetic test and the risk to develop adverse reactions and eventually stratifying patients according to the *DPYD* genetic activity score (GAS) model. Patients detail concerning genotyping results, age, gender, chemotherapy regimens and radiotherapy exposure are listed in ***Table 2***.

Characteristic	N	%
Sex		
Male	479	62.8
Female	284	37.2
Age (years)		
Median (range)	61 (20–85)	
FL- association therapy		
Monotherapy	229	30
Oxaliplatin	265	34.7
Irinotecan	269	35.3
Radiotherapy		
Yes	197	25.8
No	566	74.2
***DPYD* genotype**		
rs3918290	9	1.2
rs55886062	0	0.0
rs67376798	5	0.7
rs56038477	31	4.1

Table 2: *Cohort of 763 patients and their characteristics*

Concerning the primary tumor site, in this cohort 423/763 (55.4%) patients had colon cancer while 340/763 (44.6%) patients had rectal cancer. As far as the treatment is concerned, 196/763 (25.7%) patients were treated in neo-adjuvant setting, 282/763 (37%) were treated in adjuvant setting, 279/763 (36.6%) were treated with the first-line treatment for metastatic disease while the remaining 6/763 (0.8%) were treated with the second-line.

All 763 patients were successfully genotype for the 4 *DPYD* risk variants. All polymorphisms were detected in heterozygosity and not compound heterozygosity was detected. Only 45 (5.9%) patients carried at least one *DPYD* variants, while the remaining 718 (94.1%) resulted wild type. Specifically, 9 (1.2%) patients carried the *DPYD* rs3918290 variant, 5 (0.7%) and 31 (4%) patients

carried rs67376798 and rs56038477 *DPYD* variants, respectively. No patients carrying *DPYD* rs55886062 allele were detected.

Patients were then classified according to the GAS model. 9 (1.2%) patients were associated to *DPYD* activity score of 1.0, as all of them harbored rs3918290 variant, while the remaining 36 (4.71%) *DPYD* mutated patients were associated to a 1.5 score. The 718 (94.1%) wild-type patients were considered as full-metabolizer (GAS = 2.0).

In the whole population, 178 (23.3%) patients developed at least one toxicity during the entire course of chemotherapy that was classified as a dose-limiting toxicitie (DLT), whereas 583 (76.7%) did not. Among patients who experienced DLT, 38/178 (21.3%) developed grade 4 hematological toxicity as the most severe toxicity reported, 154 (86.5%) developed grade 3 or 4 non-hematological toxicity and 14 (7.9%) experienced both of them. Looking at the acute toxicities, 108 (14.2%) patients develop DLTs within the first three cycles of treatment, and 655 (85.8%) patients did not. Among patients who developed DLTs, 22/108 (20.4%) developed hematological toxicities, 94 (87%) non-hematological and 8 (7.4%) both (***Table 3***). Of the 178 (23.3%) patients experiencing DLTs during the entire course of chemotherapy, the most common adverse events were hand-foot syndrome, neurological disorders, neutropenia, leukopenia and gastrointestinal toxicities, mainly represented by diarrhea and vomiting.

	Acute		Total	
DLT	**N**	**%**	**N**	**%**
All				
No	655	85.8	858	76.7
Yes	108	14,2	178	23,3
Haematological				
No	741	97.1	725	95.0
Yes	22	2.9	38	5.0
Non-haematological				
No	669	87.7	609	79.8
Yes	94	12.3	154	20.2

Table 3: *Distribution of acute and total dose limiting toxicity (DLT)*

As expected, the onset of DLT is significant associated (P=0.0009) with the presence of at least one *DPYD* variant within the 4-SNP panel analyzed. Indeed, carriers of at least one variant had about a 2.7-fold increased risk to develop at least one DLT when considering both acute and total toxicity (***Table 4***). Specifically, patients carrying at least one *DPYD* variant who developed a DLT within the first three cycles (acute toxicity) and within entire course of treatment (overall toxicity) account for the 28.9% and 44.4%, respectively. On the contrary, only the 13.2% and the 22% wild-type *DPYD* patients experienced DLTs during the first three cycles and during the whole course of treatment, respectively.

However, it was shown that the GAS better defined the risk to develop DLTs than the single *DPYD* allele approach. Indeed, the risk (OR) to develop DLT was 2.08 (95% CI = 1.02-4.27) and 7.09 (95% CI = 1.69-29.65) in patients with GAS = 1.5 and 1.0, respectively when compared with patients bearing a GAS = 2 (***Table 4***).

			Acute		Total	
	N	**%**	**OR**	**95% CI**	**OR**	**95% CI**
SNP panel	45	5.9	2.69	1.33-5.41	2.67	1.42-5.04
***DPYD* activity score**						
2	718	94.1	1		1	
1.5	36	4.7	1.80	0.78-4.15	2.08	1.02-4.27
1	9	1.2	10.12	2.55-40.2	7.09	1.69-29.65
χ^2 for trend			*P*=0.0007		*P*=0.0009	

Table 4: *Association of DPYD variants with dose-limiting acute and total toxicity*

For acute toxicity, the risk (OR) was 1.80 (95% CI = 0.78-4.15) and 10.12 (95% CI = 2.55-40.20) in patients with GAS = 1.5 and 1.0, respectively (***Table 4***). Among the patients with a GAS of 1.0, 6/9 (66.7%) showed the onset of DLT during the whole course of treatment and, 5/9 (55.6%) developed DLT within the first three cycles. Among patients with a GAS of 1.0, 14/36 (38.9%) showed DLT during the entire course of chemotherapy and 8/36 (22.2%) reported DLT within the first three cycles of treatment. On the other hand, wild-type patients (GAS = 2.0) presented DLT in 158/718 cases (22.0%) when considering the whole treatment and in 95/718 cases (13.2%) when considering the first three cycles of therapy. The toxicity grade developed in relation to the GAS is reported in ***Figure 7***: within the first three cycles of chemotherapy (A) and during the entire treatment (B). Toxicities were classified as mild (grade from 0 to 2), severe (grade 3) and very severe (grade 4). Patients with a *DPYD* GAS of 1.0 were more likely to develop grade 4 toxicity rather than other patients while patients with a *DPYD* GAS of 1.5 showed an increased risk to develop grade 3 toxicity rather than wild-type *DPYD* patients (***Figure7***).

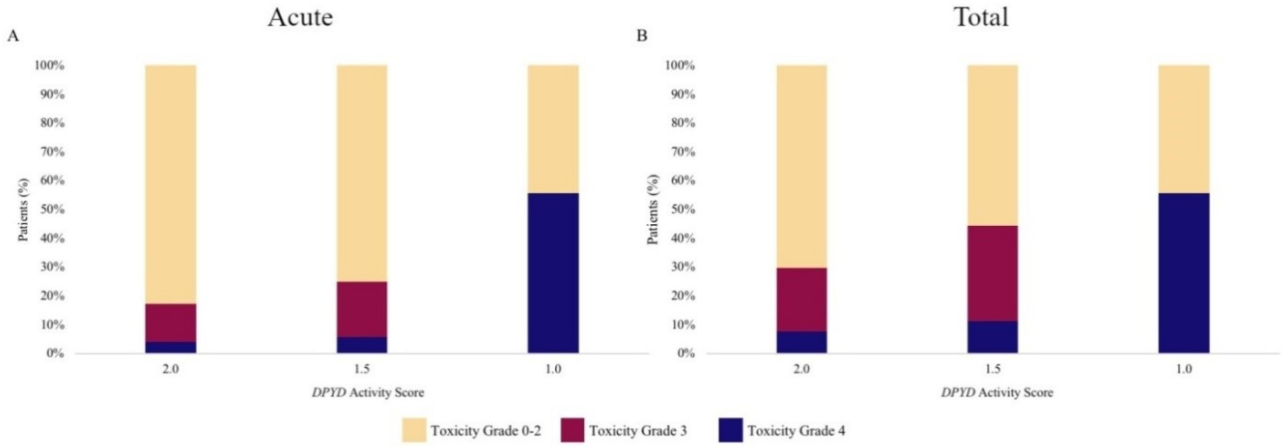


Figure 7: *Toxicity grade (by NCI CTC scale) distribution according to the DPYD activity score model within the first three cycles of chemotherapy (A) and during the whole treatment (B).*

4.2. Toxicity and costs associated with its management

To estimate the relationship between *DPYD* risk genotype, toxicity and costs associated with its clinical management, we selected, based on inclusion criteria, 550 patients affected by CRC and treated with a chemotherapy based on the use of fluoropyrimidines.

Patients' characteristics are listed in ***Table 5***.

		Chemotherapyr egimen		
Characteristics	**Patients**	**Fluoropyrimidine monotherapy**	**Fluoropyrimidine Irinotecan**	**Fluoropyrimidine Oxaliplatin**
All	550	84	265	201
Sex				
Male	330 (60%)	54 (64.3%)	171 (64.5%)	105 (52.2%)
Female	220 (40%)	30 (35.7%)	94 (35.5%)	96 (47.8%)
Age (years)				
<60	208 (37.8%)	17 (20.2%)	107 (40.4%)	84 (41.8%)
60-69	211 (38.4%)	18 (21.4%)	120 (45.3%)	73 (36.3%)
≥70	131 (23.8%)	49 (58.3%)	38 (14.3%)	44 (21.9%)
TNM stage at diagnosis				
I-II	83 (15.1%)	30 (35.7%)	26 (9.8%)	26 (12.4%)
III	275 (50%)	42 (50%)	70 (26.4%)	163 (81.1%)
IV	192 (34.9%)	12 (14.3%)	169 (63.8%)	11 (5.5%)
Setting of treatment				
Adjuvant	270 (49.1%)	71 (84.5%)	5 (1.9%)	194 (96.5%)
First/second line	280 (50.9%)	13 (15.5%)	260 (98.1%)	7 (3.5%)
Number of cycles received				
1–6	196 (35.6%)	49 (58.3%)	115 (43.4%)	32 (15.9%)
7–12	331 (60.2%)	35 (41.7%)	127 (47.9%)	169 (84.1%)
≥13	23 (4.2)	0 (0%)	23 (8.7%)	0 (0%)
Type of fluoropyrimidine				
5-Fluorouracil	485 (88.2%)	58 (69.1%)	265 (100%)	162 (80.6%)
Capecitabine	65 (11.8%)	26 (30.9%)	0 (0%)	39 (19.4%)

Table 5: *Characteristics and demographics of 550 patients*

In this cohort 485 (88.2%) patients were treated with 5-fluorouracil, and 65 (11.8%) with capecitabine. Fluoropyrimidine monotherapy was used in 15.3% of patients and all patients were treated with a full starting dose of FL.

DPYD genotyping was performed retrospectively after treatment conclusion.

To investigate a possible association between DPYD genotype and treatment interruption or discontinuation, we systematically collected patients´ clinical record concerning both treatment interruption and the FL dose reduction during chemotherapy as a consequence of intolerance toward the drug. According to this survey, during the treatment course a FL dose reduction was reported in 123/550 (22.4%) patients. Intriguingly, in 27 of them the treatment discontinuation was required as

a consequence of toxicity reported. In addition, other 86/550 (15.6%) patients discontinued therapy prematurely without undergoing FL dose reduction. A treatment interruption occurred in 47/550 (41.6%) patients as a consequence of chemotherapy-related toxicity, in 21 (18.6%) patients as a consequences of early disease progression, in 14 (12.4%) patients due to patient refusal and in 13 (11.5%) patients due to other causes. Treatment interruption occurred in other 11 (9.7%) patients due to disease progression, in 4 (3.5%) patients due to death and eventually in 3 (2.7%) patients due to lost to follow-up.

In the whole population, during the first 12 cycles of treatment, 4,052 toxic adverse reactions of any grade were reported, with a mean of 8.2 events for patient (range 0-33). The most common toxicities were hematological (1,581; 39%) and gastrointestinal (1,144; 28.2%), followed by neurological (649; 16%) and cardiovascular (18; 0.4%). However, no toxic death was reported.

Among the adverse reactions, 1,214 (30%) events required additional costs for their management and 84 (2.1%) led to patient hospitalization. In this cohort, the average cost for the management of adverse reactions was €930 per patient (range €0-€31,454). All grade 3 and 4 toxicity reactions were associated with a cost for their management, whereas only 19.7% and 38% of grade 1 and 2 were associated with a toxicity management cost, respectively. Grade ≥3 cardiovascular toxicities and all type of grade 4 required hospitalization, whereas the higher cost was attributed to febrile neutropenia (€6,102 per episode). Moreover, the cost of grade 3 thrombocytopenia or anemia was €653 per episode and €75 for nausea or vomiting.

Among this cohort of 550 patients, only 37 (6.7%) patients were *DPYD* carriers and all of them were heterozygous for one of the four risk variants. In these patients carrying at least one *DPYD* variant, no homozygous or compound heterozygous patient was detected. Specifically, 9 (1.6%) patients were heterozygous for rs3918290, 5 (0.9%) for rs67376798, 23 (4.2%) for rs56038477 and finally no one carried rs55886062.

Concerning GAS, 513 (93.3%) patients had a score of 2.0, 28 (5.1%) of 1.5, 9 (1.6%) of 1.0 and no patients had a score of 0. Carriers of one *DPYD* risk variants had a mean toxicity management cost of €2,972 (95% CI = €2,456-€3,505), which was significantly higher than noncarriers (€825; 95% CI = €785-€864; $P<0.0001$). The 37 (6.7%) *DPYD* carriers accounted for €109,964, which is 20.7% of the overall cost (€529,085). The mean cost per patient was inversely related to the GAS ($P<0.0001$) (***Table 6***).

		Cost (€)		
***DPYD* variant status**	**N**	**Mean**	**95% CI**	**ANOVA**
Noncarriers	513	825	785-864	
Carriers	37	2,972	2,456-3,505	*P*< 0.0001
Gene activity score (GAS)				
2	513	825	785-864	
1,5	28	2,188	1,683-2,693	
1	9	5,414	2,268-8,561	
0	0	0	0	*P*< 0.0001

Table 6: *Association between DPYD variants and the toxicity management costs*

Carriers were significantly at increased risk of developing an adverse reaction requiring hospitalization than noncarriers (OR, 4.14; 94% CI = 1.87-9.14; *P* = 0.0004, ANOVA). A display of the incidences of these two types of toxic events by cycle and carrier status is reported in ***Figure 8.***

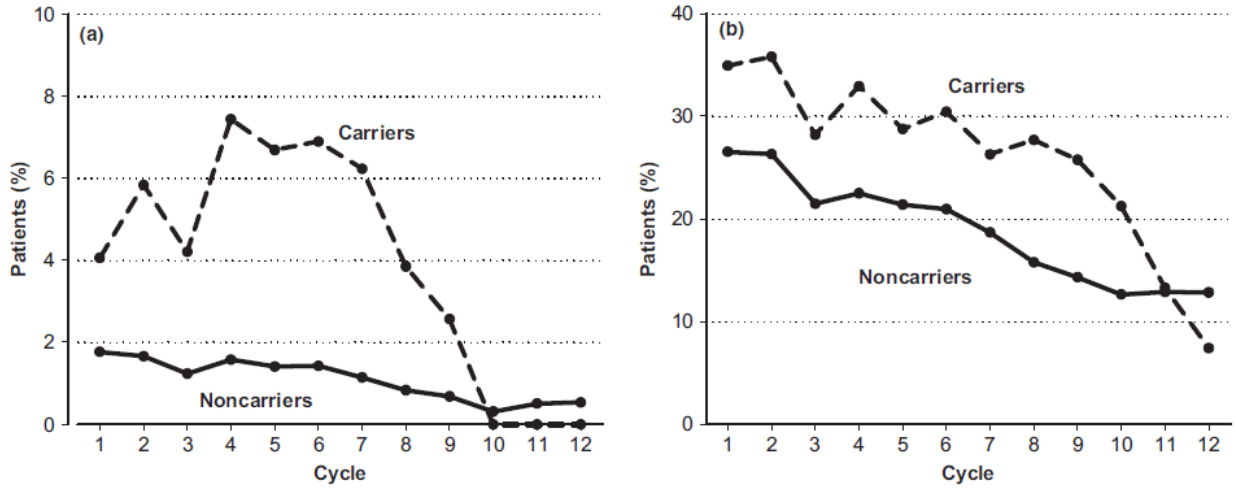


Figure 8: *Percentage of patients experiencing toxicities requiring hospitalization (a) or toxicities associated with any management cost (b)*

No significant association was found with the risk of developing toxic events associated with any management costs (***Table 7***).

	Noncarriers	Carriers	
	N (%)	**N (%)**	**OR (95% CI)**
Patients with toxicities associated with any management cost			
No	147 (28.7)	8 (21.6)	Reference
Yes	366 (71.3)	29 (78.4)	1.52 (0.67-3.47)
			P = 0.3172
Patients with toxicities requiring hospitalization			
No	463 (90.3)	26 (70.3)	Reference
yes	50 (9.7)	11 (29.7)	4.14 (1.87-9.14)
			P = 0.0004

Table 7: *Association between DPYD variants and toxicities associated with any management cost and toxicities requiring hospitalization*

Patients treated with FL monotherapy had a mean toxicity management cost of €352 (95% CI = €269-€409), while patients treated with irinotecan in combination with FL resulted in a mean management cost of €798 (95% CI = €709-€887) and, eventually, patients treated with fluoropyrimidines/oxaliplatin resulted in a mean cost of €1,459 (95% CI = €1,260-€1,659; *P*< 0.0001).

Figure 9 reports the toxicity management cost according to the chemotherapy regimen and treatment cycle, without taking into account the cost of the drugs. It is possible to note that the cost for the three regimens were very similar during the first cycles but that they tend to decrease in the following cycles (cycles 7-9). According to the previous results, hospitalization occurred in 31/201 (15.4%) of patients treated with fluoropyrimidine/oxaliplatin, in 20/265 (7.4%) patients receiving fluoropyrimidine/irinotecan, and in 5/84 (5.9%) patients treated with fluoropyrimidines monotherapy (P = 0.007; χ^2test). ***Figure 9*** showed the trend which is consistent with a difference in hospitalization rates.

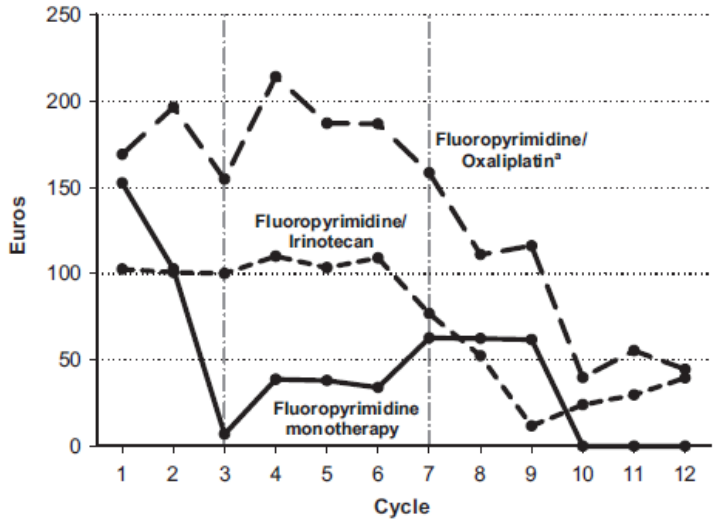


Figure 9: *Toxicity management cost per patient according to the treatment cycle and chemotherapy regimen*

The effects of the *DPYD* status and GAS on toxicity management costs were also evaluated in relation to the different chemotherapy regimens and remained significant higher for *DPYD* risk variants carriers in all three regimens (***Table 8***). Within each chemotherapy regimen, the cost related to *DPYD* carriers was about 3.5-fold higher than for noncarriers. Moreover, we performed a second economic analysis to estimate the cost of *DPYD*-guided toxicity management and the clinical benefit expressed as quality adjusted life years (QALYs)[56].

These studies suggest that the incremental costs were mainly driven by higher percentage of severe ADRs in poor/intermediate metabolizer than in full metabolizer patients.

		Cost (€)		
***DPYD* variant status**	**N (%)**	**Mean**	**(95% CI)**	**ANOVA**
Fluoropyrimidine monotherapy				
Noncarriers	79 (94.1)	320	298-342	
Carriers	5 (6.0)	1,208	0-2,658	*P*< 0.0001
Gene activity score				
2	79 (94.1)	320	298-341	
1.5	3 (3.6)	471	0-1,529	
1	2 (2.4)	2,315	0-11,353	
0	0 (0)	0	0	
Fluoropyrimidine/irinotecan				
Noncarriers	249 (94.0)	662	630-695	
Carriers	16 (6.0)	2,385	918-3,851	*P*< 0.0001
Gene activity score				
2	249 (94.0)	662	630-695	
1.5	12 (4.5)	1,031	618-1,443	
1	4 (1.5)	6,447	2,241-10,654	
0	0 (0)	0	0	
Fluoropyrimidine/oxaliplatin				
Noncarriers	185 (92.0)	1,269	1,212-1,326	
Carriers	16 (8.0)	4,808	1,529-8,087	*P*< 0.0001
Gene activity score				
2	185 (92.0)	1,269	1,212-1,236	
1.5	13 (6.5)	2,68	910-2,754	
1	3 (1.5)	16,183	516-31,850	
0	0 (0)	0	0	

Table 8: *Association between DPYD variants and the toxicity management costs in patients undergoing different chemotherapy regimens*

4.3. Rare variants, cases cohort

To assess the role of novel and rare variants in the onset of fluoropyrimidine-related toxicity, we sequenced a selected cohort of 120 patients through a Miseq (Illumina) platform. Of these, 12 patients were discarded due to the low quality reads. Patients characteristics are listed in ***Table 9***.

Patients were selected according to specific inclusion criteria listed in Material and Methods (section 6.1). Briefly, only patients who experienced at least one event of severe (grade >= 3) toxicity were included. For them, the DPYD genotype for the four variants (rs3918290, rs55886062, rs67376798 and rs56038477) was previously assessed to be sure that only DPYD wild-type patients had been included in this analysis. Thus, only patients with DPYD GAS = 2.0 were considered to rule out the impact of any DPYD variant on patients phenotype (toxicity)

As reported in ***Table 9***, the gender of the population was equally distributed, with a slight majority of female 59 (54.6%) over male (44.4%). Average age of enrollment was 60 years ranging between

30-82. Considering the primary tumors site, 87 (80.6%) patients had colorectal cancer, 6 (5.6%) had breast tumor, 5 (4.6%) had gastric cancer, 2 (1.8%) had head-neck tumor, 1 (0.9%) patient had cervical tumor and 1 (0.9%) had pancreatics cancer, whereas the remaining 6 (5.6%) had not available information about primary tumor site.

Concerning the therapy, fluoropyrimidine monotherapy was used in 8 (7.4%) patients while combination treatment was used in 100 (92.6%) patients. Among these 100 patients, 37 (34.3%) were treated with irinotecan, 37 (34.3%) and 7 (6.5%) patients with cisplatin and oxaliplatin respectively. 6 (5.5%) patients were treated with a combination of irinotecan and cisplatin and 13 (12%) were treated with other combination regimens. Moreover, 87 (80.6%) patients were treated with a base of 5-FU while 21 (19.4%) were capecitabine-based treatment.

Concerning ADR reported in this cohort, 61 (56.5%) patients developed grade 3 toxicity, 46 (42.6%) developed G4 toxicity and 1 patient died for toxicity (grade 5). Moreover, in 98 (90.7%) patients the onset of severe toxicity occurred during the first 3 cycles of treatment (acute toxicity), while in 6 (5.6%) patients it occurred after 3 cycles (total or delayed toxicity). Unfortunately, for 4 (3.7%) patients detailed toxicity data were not available.

	Patients	Treatment	
Characteristic	**N° (%)**	**5-FU N° (%)**	**Capecitabine N° (%)**
All	108	87 (80.6%)	21 (19.4%)
Age			
Mean (range)	60 (30-82)	60 (30-82)	60 (39-79)
Sex			
Male	49 (44.4%)	40 (45.9%)	9 (42.9%)
Female	59 (54.6%)	47 (54.1%)	12 (57.1%)
Type of cancer			
ColoRectal	87 (80.6%)	75 (86.2%)	12 (57.1%)
Breast	6 (5.6%)	4 (4.6%)	2 (9.5%)
Gastric	5 (4.6%)	2 (2.3%)	3 (14.3%)
Head-Neck	2 (1.8%)	2 (2.3%)	/
Cervical	1 (0.9%)	1 (1.2%)	/
Pancreas	1 (0.9%)	/	1 (4.8%)
N.A.	6 (5.6%)	3 (3.4%)	3 (14.3%)
Treatment			
Monotherapy	8 (7.4%)	2 (2.3%)	6 (28.6%)
Combination:	100 (92.6%)	85 (97.7%)	15 (71.4%)
Platin	37 (34.3%)	36 (41.4%)	1 (4.8%)
Irinotecan	37 (34.3%)	37 (42.5%)	/
Oxaliplatin	7 (6.5%)	/	7 (33.3%)
Platin + Irinotecan	6 (5.5%)	4 (4.6%)	2 (9.5%)
Others	13 (12%)	8 (9.2%)	5 (23.8%)

(Continued on the next page)

Continued	Patients	Treatment	
Characteristic	**N° (%)**	**5-FU N° (%)**	**Capecitabine N° (%)**
Max Grade Toxicity (all)			
G5	1 (0.9%)	1 (1.1%)	/
G4	46 (42.6%)	36 (41.4%)	10 (47.6%)
G3	61 (56.5%)	50 (57.5%)	11 (52.4%)
Severe Toxicity (≥G3)			
Acute (During the firsts 3 cycles)	98 (90.7%)	79 (90.8%)	19 (90.5%)
Delayed or Total (After cycle 3)	6 (5.6%)	6 (6.9%)	/
N.A.	4 (3.7%)	2 (2.3%)	2 (9.5%)

Table 9: *Cases cohort characteristics*

4.4. Rare variants distribution in the cases cohort

Sequencing data of the 108 patients from the cases cohort revealed a total of 19,960 variants distributed in 54 candidate genes. Notably, 961 unique variants were detected, which included common polymorphisms, rare and novel variants. In agreement with the fact that the untranslated regions (UTRs) are not high conserved, most variants fall on UTRs (602; 62.64%). Among these, 526 (54.73%) fall on 3'UTRs and 76 (7.91%) fall on 5'UTRs. Exonic variants were divided into missense (168; 17.48%), synonymous (166; 17.27%), stop gain (5; 0.52%), frameshift (4; 0.41%) and others (10; 1.04%) that include in-frame insertion and deletion (***Figure 9***). Then, 6 (0.62%) variants were deemed to influence the splicing sites.

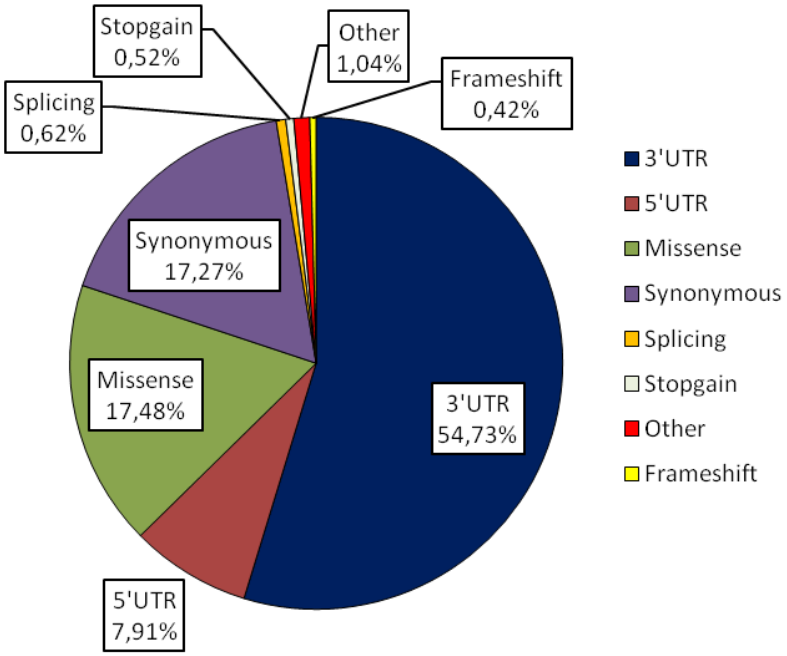


Figure 9: *Variants types distribution across 54 target gene in cases cohorts*

To understand if any gene could be enriched in a specific type of variant, we have plotted the class of variants identified on each gene **(*Figure 10a, b*)**.

Except for a few genes, most genes are enriched for variants in the UTR, as expected. By contrast, the *DPYD* gene, which encodes for the main metabolizer of the drug, is enriched in missense variants (***Figure 10a, b***).

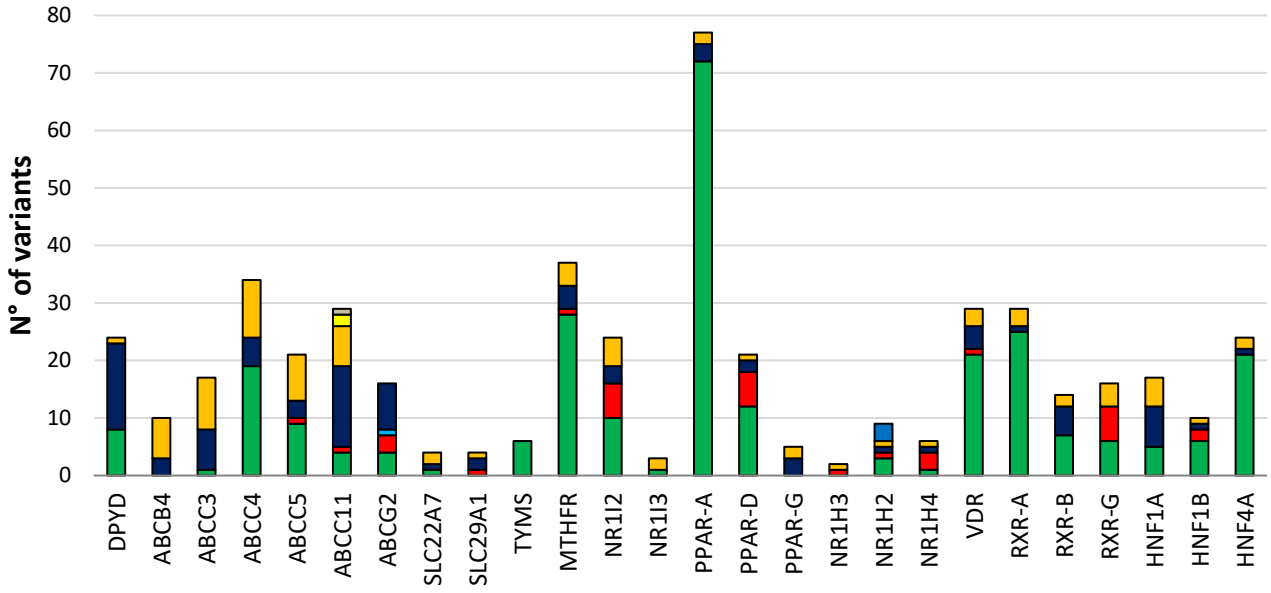


Figure 10a: *Variants distribution per gene*

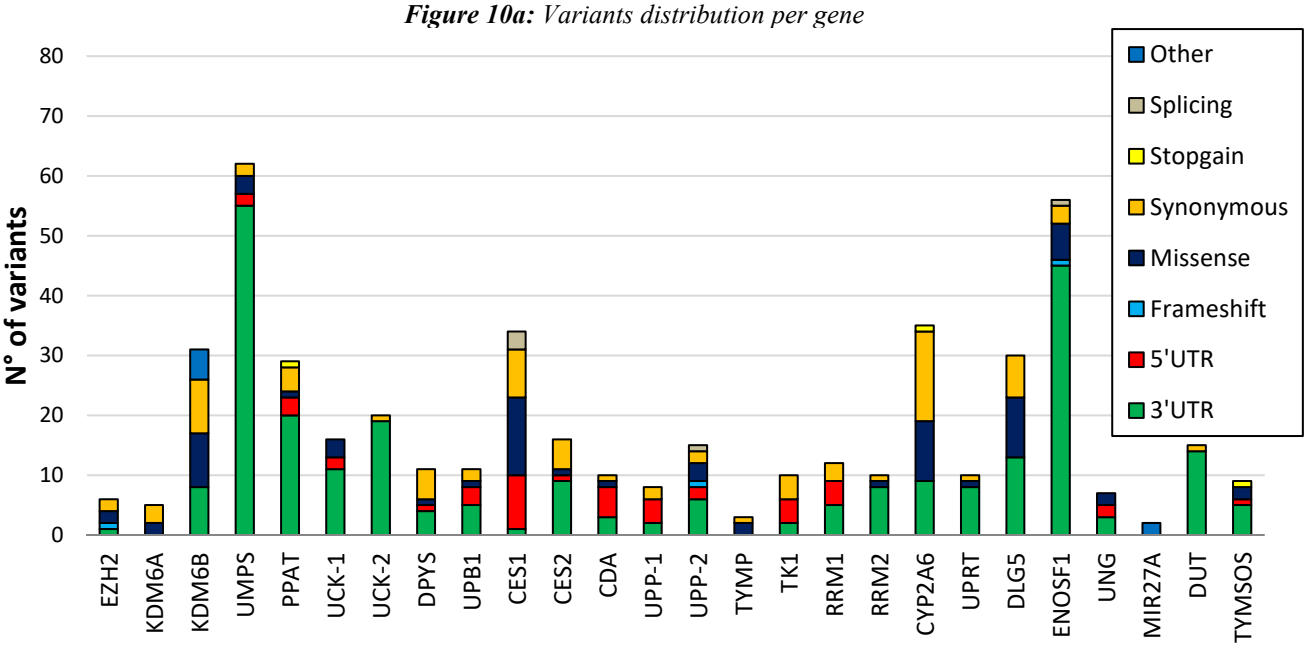


Figure 10b: *Variants distribution per gene*

However, the number of variants per gene may depend on the length of coding sequence (CDS), on the length of UTRs and on other characteristic of gene and DNA sequence. To assess whether the distribution of variants is influenced by the length of the gene, we normalized the number of variants for the length of each gene. With this in mind, we reported in dark blue the total number of variants normalized for the gene length and in red the number of exonic variants normalized for the exon length, respectively (***Figure 11a, b***). After normalization, it appears evident that some genes like *CES1* and *CYP2A6* are hypervariable compared to others due to the greater number of variants

per kbs detected. Moreover, the length of UTRs or genes does not seem to influence the number of variants per genes in all genes.

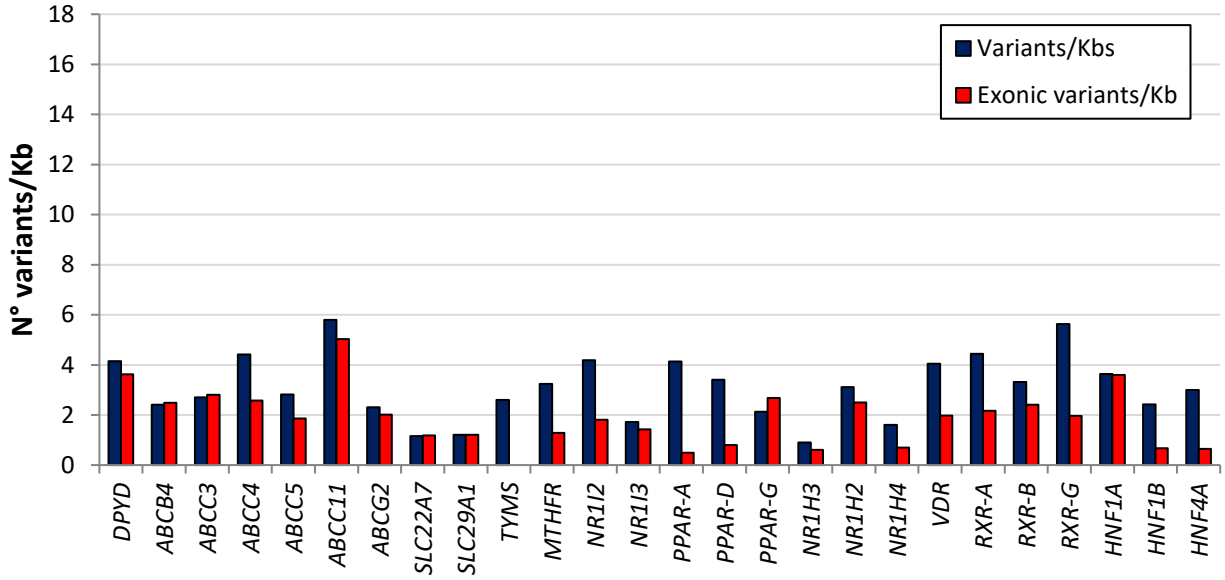


Figure 11a: *Variants distribution per gene normalized on the gene length*

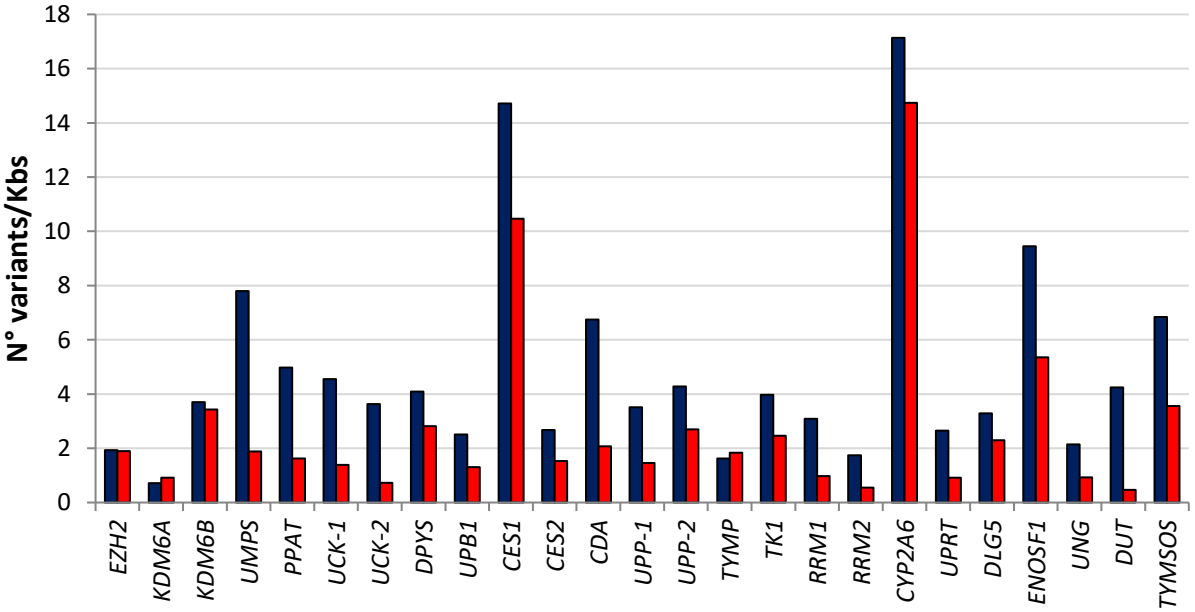


Figure 11b: *Variants distribution per gene normalized on the gene length*

According to the idea that a rare variant is supposed to have a stronger impact on phenotype we investigated the distribution of variants based on their minor allele frequency (MAF). Among these unique 961 variants, 434 (45.2%) are rare with a MAF<0.01, 273 (28.4%) are very rare variant (MAF<0.001) and 176 (18.3%) are novel variants ***(Figure 12)***.

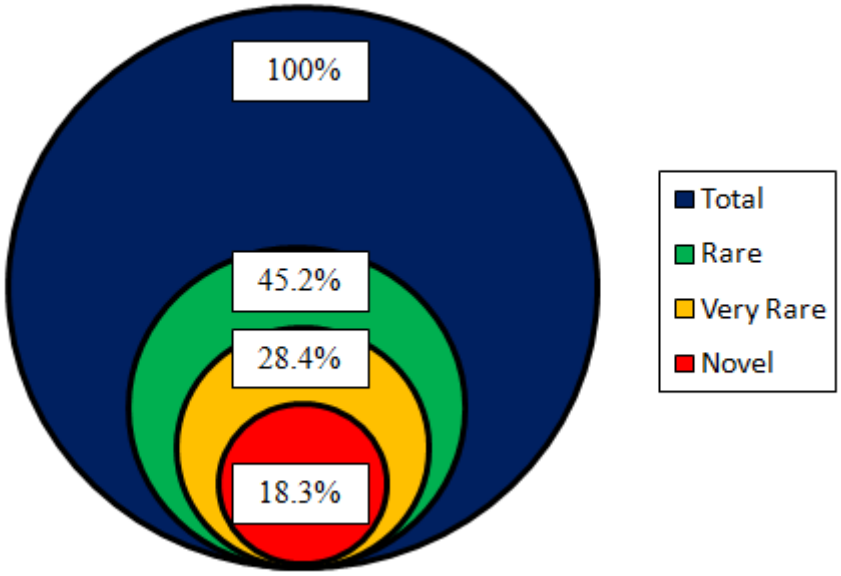


Figure 12: *Variants distribution according to Minor Allele Frequency (MAF)*

In addition, we speculated that the number of variants in a single gene may reflect a difference in the role in the onset of toxicity events. To verify our hypothesis, we clustered the genes into 4 main classes, such as *DPYD*, transporters, genes belonging to folate pathway, nuclear receptors and others.

Bearing in mind that a rare variant has a high probability of being dangerous for the normal function of protein, and we observed that *DPYD* and genes involved in folate pathway show an high number of novel, very rare and rare variants, perhaps suggesting a functional role of these variants ***(Figure 13)***.

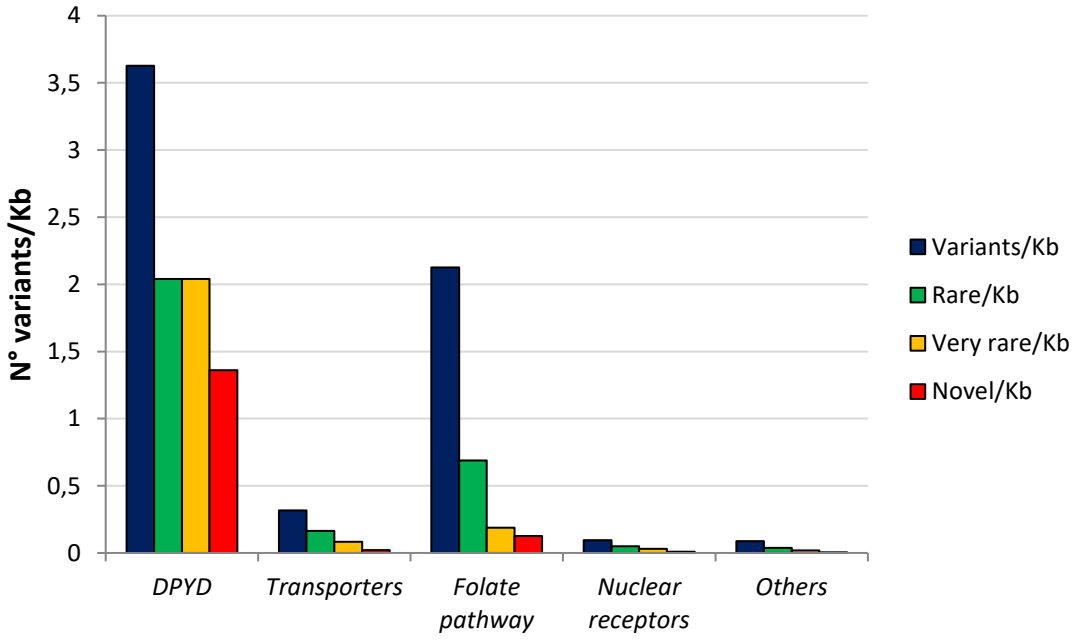


Figure 13**:** *Exonic variants distribution per genes classes normalized on length*

4.5. *DPYD* rare, very rare and novel variants

In the cases cohort, we highlighted 9 different very rare variants (MAF <0.001) on *DPYD*. For them, no compound heterozygous or homozygous patients were found. ***Table 10*** shows the *DPYD* rare variants. All variants were confirmed by Sanger sequencing.

***DPYD*: NM_000110**					
Sample	**Variant**	**Rs**	**Location**	**Functional Consequences**	**Freq ExAC_NFE**
P1	c.G345C; p.M115I	rs377169736	Exon5	Missense	0.0001
P2	c.G481A; p.E161K	/	Exon5	Missense / Splice	/
P3	c.C800T; p.T267I	/	Exon 8	Missense / Splice	/
P4	c.G958A; p.G320R	/	Exon9	Missense / Splice	/
P5	c.A1110G; p.I370M	/	Exon 10	Missense / Splice	/
P6	c.A1411C; p.T471P	/	Exon12	Missense / Splice	/
P7	c.A2060C; p.D687A	rs755692084	Exon 17	Missense / Splice	0.00002*
P8	c.A2137G; p.N713D	rs773407491	Exon 17	Missense / Splice	0.000008247
P9	c.T2491A; p.C831S	/	Exon 20	Missense / Splice	/

Table 10: *DPYD rare variants (MAF <0.01)*

In the P1 proband, a 60 year-old female, we highlighted a very rare missense variant (c.345C>G; p.M115I; rs377169736) which falls on exon 5 of *DPYD* (***Figure 14a***).

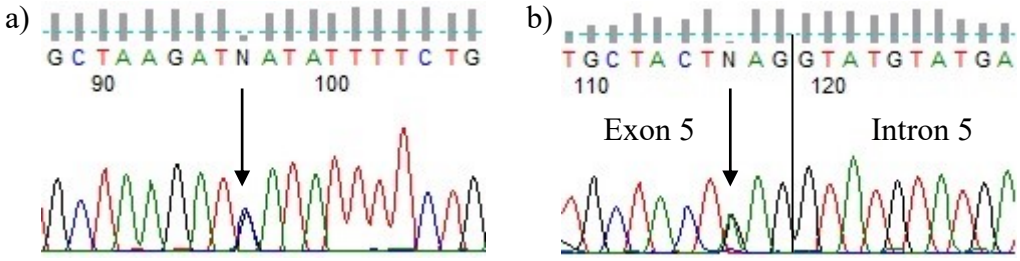


Figure 14: *Electropherogram related to DPYD variants in patient P1(a) and in patient P2 (b)*

The missense variant was reported as a very rare variant from both the databases dbSNP and ExAc. Interestingly, in this patient a common variant (c.T2C; p.M1T; rs2228570) on *vitamin D receptor* (*VDR*) that may lead to severe defect in the normal mRNA translation chemistry was present. P1 was affected by colorectal cancer and was treated in adjuvant setting with a combination of capecitabine and oxaliplatin (CAPOX). The patient developed a severe gastrointestinal toxicity (diarrhea G4) after the first cycle of treatment which led to a reduction in the dosage of the FL drug. The proband P2 was a 42 year-old female treated with cyclophosphamide, methotrexate, and fluorouracil (CMF) combination for the treatment of breast cancer. After only one cycle of therapy she experienced a hematological toxicity (leukopenia G4) that led to FL dose reduction. P2 presented a novel missense variant on *DPYD* (c.481G>A; p.E161K) (***Figure 14b***). Moreover, in this

patient a very rare synonymous (c.4185A>G; p.P1395P; rs778920577) variants on *DPYD*'s regulator *lysine (K)-specific demethylase 6A* (*KDM6A*) was found.

P3 proband was a 47 year-old female affected by CRC and treated with a FOLOFOX-4 regimen that includes 5-FU and oxaliplatin. P3 developed acute non hematological G3 toxicity after the first cycle of therapy. NGS sequencing revealed the presence in heterozygosis of a novel variant into exon 8 (c.800C>T; p.T267I) of *DPYD* gene (***Figure 15a***).

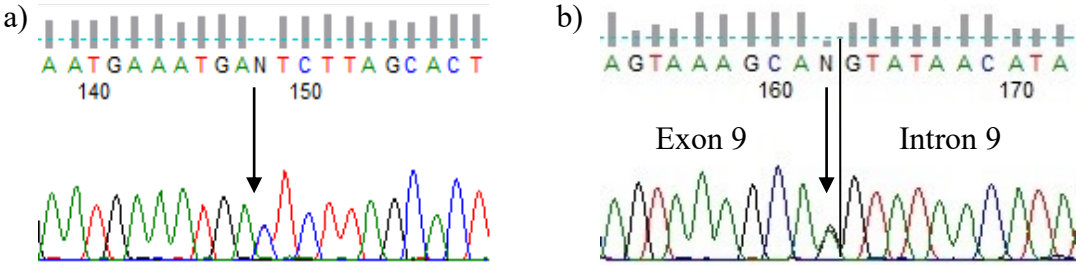


Figure 15: *Electropherogram related to DPYD variants in patient P3(a) and in patient P4 (b)*

Another novel missense (c.958G>A; p.G320R) variant which fall on exon 9 was found in proband P4 (***Figure 15b***). This patient was a 64 year-old male with a CRC treated with FOLFOX-4 regimen. He developed a severe G4 neutropenia which led to dose reduction after the second cycle of therapy.

Proband P5, a 54 year-old male, experienced not hematological G4 toxicity after one cycle of FOLFOXIRI (fluorouracil, oxaliplatinum, irinotecan) regimen plus monoclonal antibody Cetuximab. P5 was characterized by the presence of a novel missense variant (c.1110A>G; p.I370M) in exon 10 of *DPYD* (***Figure 16a***). Interestingly, this patients showed the presence of other variants in *DPYD*'s regulatory genes: one very rare (c.380G>A; p.R127Q; rs371659081) variant on *nuclear receptor subfamily 1, group I, member 2* (*NR1I2*) and one rare missense (c.803C>T; p.A268V; rs1042311) variants in *peroxisome proliferator-activated receptor alpha* (*PPARA*). Furthermore, this patient carried a common polymorphism (n.40A>T; rs895819) on *MIR27A* which is a direct regulator of DPD expression. This polymorphism is already known to lead to an increased risk of developing FL-related toxicity in patients carrying at least one *DPYD* risk variant [103].

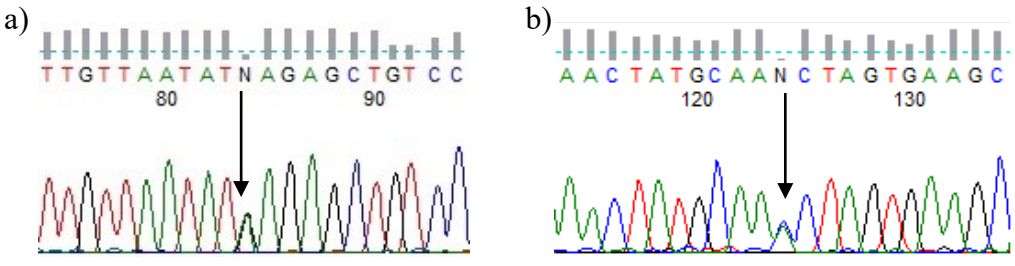


Figure 16: *Electropherogram related to DPYD variants in patient P5(a) and in patient P6 (b)*

P6 proband, was a male 64 year-old patient affected by CRC treated with FOLFIRI regimen who developed a G3 neutropenia after the first cycle of therapy. After target sequencing a novel missense (c.1411A>C; p.T471P) variant on exon 12 of *DPYD* has emerged (***Figure 16b***). As P5, P6 carried a *MIR27A* risk variant (n.40A>T; rs895819) in addition to a rare missense (c.1582G>A; p.A528T; rs45605536) variant on *ATP-binding cassette, sub-family G (WHITE), member 2* (*ABCG2*), a transporter that regulates the intracellular efflux of 5-FU.

In the P7 proband, 57 year-old female, we found in exon 17 of *DPYD* a very rare missense (c.2060A>C; p.D687A; rs755692084) (***Figure 17a***). Moreover, we identified the *MIR27A* risk variant (n.40A>T; rs895819). The proband developed acute toxicity after three cycle of therapy with FOLFIRI regimen, which include a combination of 5-FU and irinotecan. Specifically, she developed neutropenia G3 and stomatitis G3.

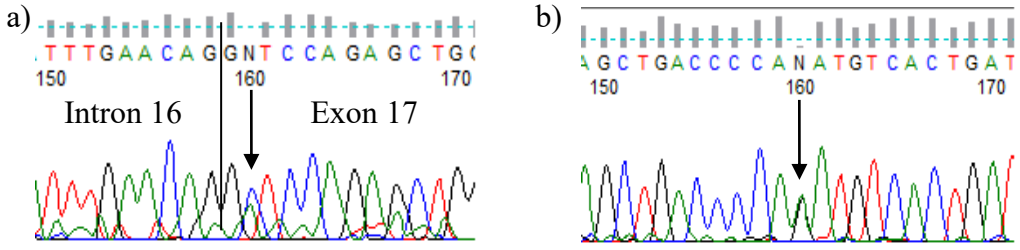


Figure 17: *Electropherogram related to DPYD variants in patient P7(a) and in patient P8 (b)*

Proband P8 carried a very rare missense variants (c.2137A>G; p.N713D; rs773407491) on exon 17 of *DPYD* (***Figure 17b***). Furthermore, P8 proband showed the presence of one rare missense (c.92G>A; p.G31D; rs137853247) variants in *Hepatocyte Nuclear Factor 1-Alpha* (*HNF1A)* and one (c.1126A>C; p.T376P; rs771937539) on *Hepatocyte Nuclear Factor 1-Beta* (*HNF1B)*, two transcriptional factors express in liver involved in the *DPYD*'s expression. P8 was a 77 year-old male affected by CRC who developed non hematological G4 toxicity after the first cycles of capecitabine treatment.

P9 proband is a male patient treated with capecitabine for metastatic disease that developed a severe (G4) hematological and non-hematological toxicity. However, for this patient, information about age, first tumor site and cycle of toxicity were not available. From NGS a novel missense (c.2491T>A; p.C831S) variant that falls in exon 20 has emerged (***Figure 18***). Furthermore, NGS highlighted the *MIR27A* risk variant (n.40A>T; rs895819). Among the others, we found very interesting the novel 43bp long deletion (c.*941_*983del) variant found on overlapped 3'UTR of the *TYMS* and its antisense overlapping gene *ENOSF1*.

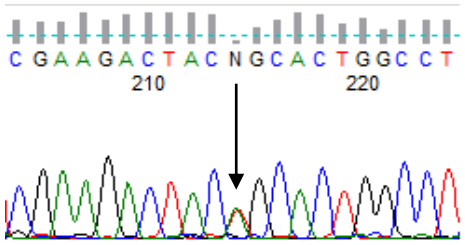


Figure 18*: *Electropherogramrelated to DPYD variants in patient P9

4.6. Rare, very rare and novel variants in FL-pathway

NGS allowed us to highlight 9 rare (MAF <0.01) variants on *DPYD* in a cohort of 108 patients. However, 99 patients did not present any variant on the candidate genes and, even assuming that all the previously nine variants would be pathological, we cannot define a straightforward explanation for all the episodes of severe toxicity. Nevertheless, we found out other five rare variants that may explain the onset of toxicity in other six patients.

The proband P10 was a male 73 year-old affected by CRC and was treated with FOLFOX regimen that led to an acute neutropenia G3 at the third cycle. In P10 we highlighted a novel frameshift deletion (c.1774delT; p.C592fs) on *ABCG2* gene.

Patient P11, a 52 year-old male with CRC, was treated with FOLFIRI regimen. He developed a severe G4 neutropenia at the second cycle of treatment. Sequencing data shows the presence of a rare variant, a stop gain (c.297G>A; p.W99X; rs145048685) mutation, on *ATP-binding cassette, sub-family C (CFTR/MRP), member 11* (*ABCC11*).

Proband P12 carried a very rare stop gain (c.G2490A; p.W830X; rs767682650) mutation on *ABCC11* and the common *MIR27A* risk variant (n.40A>T; rs895819). P12 was a 68 year old female affected by CRC, she developed a non-hematological G3 toxicity during the first cycle of FOLFOX treatment.

P13 shows a novel frameshift insertion (c.1213_1214insC; p.K405fs) on *enhancerof zeste homolog 2(Drosophila)* (*EZH2*). P13, a 66 years-old female affected by CRC, experienced a G3 neutropenia after one cycle of FOLFIRI therapy.

Finally, a novel synonymous variant (c.783C>G; p.G261G) on *MTHFR* genes which seem affected the splicing was found un two unrelated patients (P14 and P15). These patients were treated with FOLFIRI regimen for CRC and both developed a non-hematological G3 toxicity. P14 and P15 were a 53-year old and 64-years old female, respectively.

4.7. Prediction tools analysis

To predict the impact of novel and rare variants on DPD seven prediction tools were leveraged: SIFT[104], Polyphen-2[105], PROVEAN[106], CADD[107], LRT[108], MutationTaster[109] and FATHMM[110]. Each prediction results and score are reported in ***Table 11***.

Variants	SIFT	Polyphen2	PROVEAN	CADD	LRT	MutationTaster	FATHMM
c.G345C; p.M115I	T	B	N	16,01	N	D	T
c.G481A; p.E161K	D	D	D	33	D	D	T
c.C800T; p.T267I	D	P	D	28.5	D	D	D
c.G958A; p.G320R	D	D	D	34	D	D	D
c.A1110G; p.I370M	T	B	N	9.495	D	D	D
c.A1411C; p.T471P	D	D	D	26.9	D	D	D
c.A2060C; p.D687A	T	P	D	21.5	D	D	T
c.A2137G; p.N713D	D	D	D	25.8	D	D	D
c.T2491A; p.C831S	T	P	D	17.12	D	D	D

Table 11: *Results from prediction tools for DPYD variants. D= Deleterious, T=Tolerated, B= Benign, P=Probably Deleterious, N=Neutral, CADD >16 = Deleterious.*

From this analysis, only two *DPYD* variants (p.M115I and I370M) are deemed to have no impact on protein function. However, these computational tools have been optimized on variants associated with disease and main rely on evolutionary conservation. This is a hurdle we trying to assess the phenotypic consequences of variants that fall on a poorly conserved pharmcogenes[111]. Nevertheless, at the best of our knowledge, to date there are no open tools optimized on pharmacogenes.

Since several exonic variants may affect splicing mechanisms, *DPYD* variants were evaluated for a possible impact on splicing through HSF web tool[112]. In ***Table 12*** are reported the results of prediction. These data suggest that *DPYD* variants could have a greater impact on the splicing mechanisms than in amino acid point changes. Furthermore, p.M115I is the only variants that seem not related to aberrant splicing mechanism and is one of two variants predicted from the previous analysis as benign variants. In contrast the other 8 *DPYD* variants, appear to modify the physiological splicing mechanism through modification of non-canonical splicing sites.

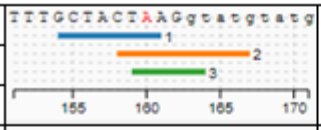
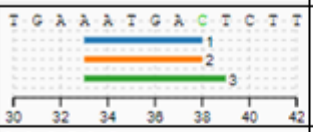
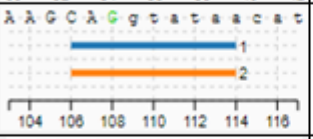
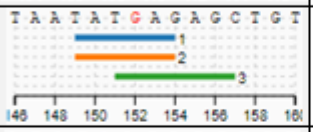
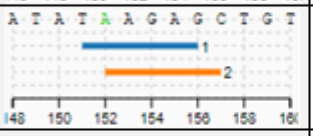
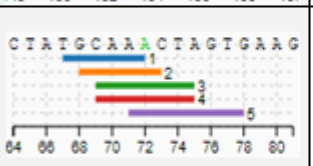
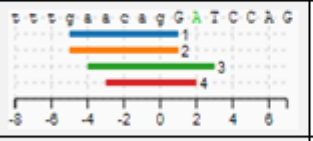
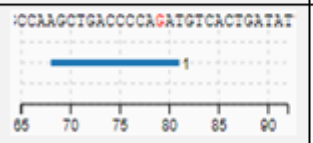
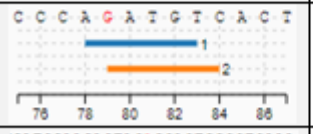
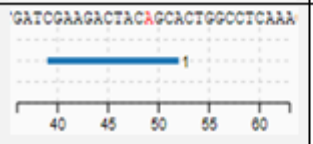
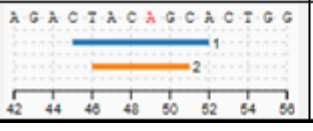
Variant	Predicted signal	Prediction algorithm	cDNA Position	Interpretation
c.G345C p.M115I	No significant splicing motif alteration detected. This mutation has provably no impact on splicing			
c.G481A p.E161K	New ESS Site	1 - Sironi et al. - Motif 3 2 - PESS Octamers from Zhang & Chasin 3 - HSF Matrices - hnRNP A1		Creation of an exonic ESS site. Potential alteration of splicing.
c.C800T;p.T267I	ESE Site Broken	1 - RESCUE ESE Hexamers 2 - ESR Sequences from Goren et al. 3 - EIEs from Zhang et al.		Alteration of an exonic ESE site. Potential alteration of splicing.
c.G958A p.G320R	Broken WT Donor Site	1 - HSF Matrices 2 - MaxEnt		Alteration of the WT donor site, most probably affecting splicing.
c.A1110G p.I370M	New ESS Site	1 - HSF Matrices - hnRNP A1 2 - ESR Sequences from Goren et al. 3 - Sironi et al. - Motif 2		Creation of an exonic ESS site. Potential alteration of splicing.
	ESE Site Broken	1 - HSF Matrices - 9G8 2 - HSF Matrices - Tra2-ß		Alteration of an exonic ESE site. Potential alteration of splicing.
c.A1411C p.T471P	ESE Site Broken	1 - ESR Sequences from Goren et al. 2 - HSF Matrices - 9G8 3 - ESE-Finder - SF2/ASF 4 - ESE-Finder - SF2/ASF(lg) 5 - ESE-Finder - SC35		Alteration of an exonic ESE site. Potential alteration of splicing.
c.A2060C p.D687A	ESE Site Broken	1 - ESE-Finder - SF2/ASF 2 - ESE-Finder - SF2/ASF(lg) 3 - PESE Octamers from Zhang & Chasin 4 - HSF Matrices - 9G8		Alteration of an exonic ESE site. Potential alteration of splicing.
c.A2137G p.N713D	New Acceptor Site	1 - HSF Matrices		Activation of an exonic cryptic acceptor site, with presence of one or more cryptic branch point(s). Potential alteration of splicing.
	New ESS Site	1 - IIEs from Zhang et al. 2 - ESR Sequences from Goren et al.		Creation of an exonic ESS site. Potential alteration of splicing.
c.T2491A p.C831S	New Acceptor Site	1 - HSF Matrices		Activation of an exonic cryptic acceptor site, with presence of one or more cryptic branch point(s). Potential alteration of splicing.
	New ESS Site	1 - Sironi et al. - Motif 1 2 - HSF Matrices - hnRNP A1		Creation of an exonic ESS site. Potential alteration of splicing.

Table 12: *Results of HSF prediction tools for DPYD variants*

4.8. Molecular visualization

To understand how *DPYD* variants may impact the protein function we analyzed 3D protein structure with Chimera software[113] (***Figure 19a, b*** and ***c***). A human reference structure for this protein is missing so we choose the entry 1h7w from the Protein Data Bank (PDB; https://www.rcsb.org/) as this sequence shares more than 90% similarity with the Homo sapiens DPD. The selected reference (1h7w) is DPYD structure from Sus scrofa (Pig) and its resolution is 1.9Å. As shown by ***Figure 19***, amino acid substitutions occurs in surface residues of the protein, mainly in α-helix or loops motifs far from substrate binding pocket and catalytic site.

Moreover, understanding impact of these variants on protein activity is a hard task given that each variant is distant more than 6Å from cofactors or substrate excluding a possible interaction. However, this is only an in silico visualization of variants and thus we cannot rule out a possible aberrant mechanism during co-translation folding that may have a strong impact in the protein structure and function. Moreover, since many of the missense variants fall on exonic splicing sites, the functional impact may be cause by an aberrant splicing mechanism rather than by a single amino acid change.

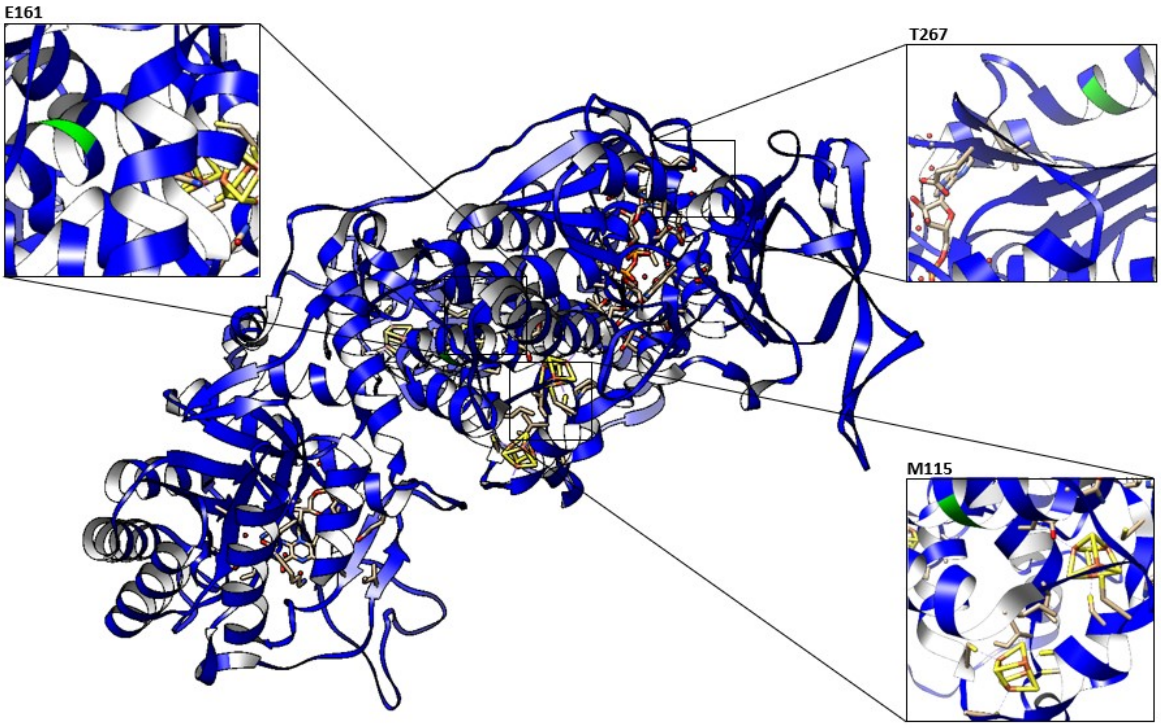


Figure 19a: *3D DPYD monomer protein visualization, green residues represent 3 variants (M115I; E161K and T267I) founded in cases cohort*

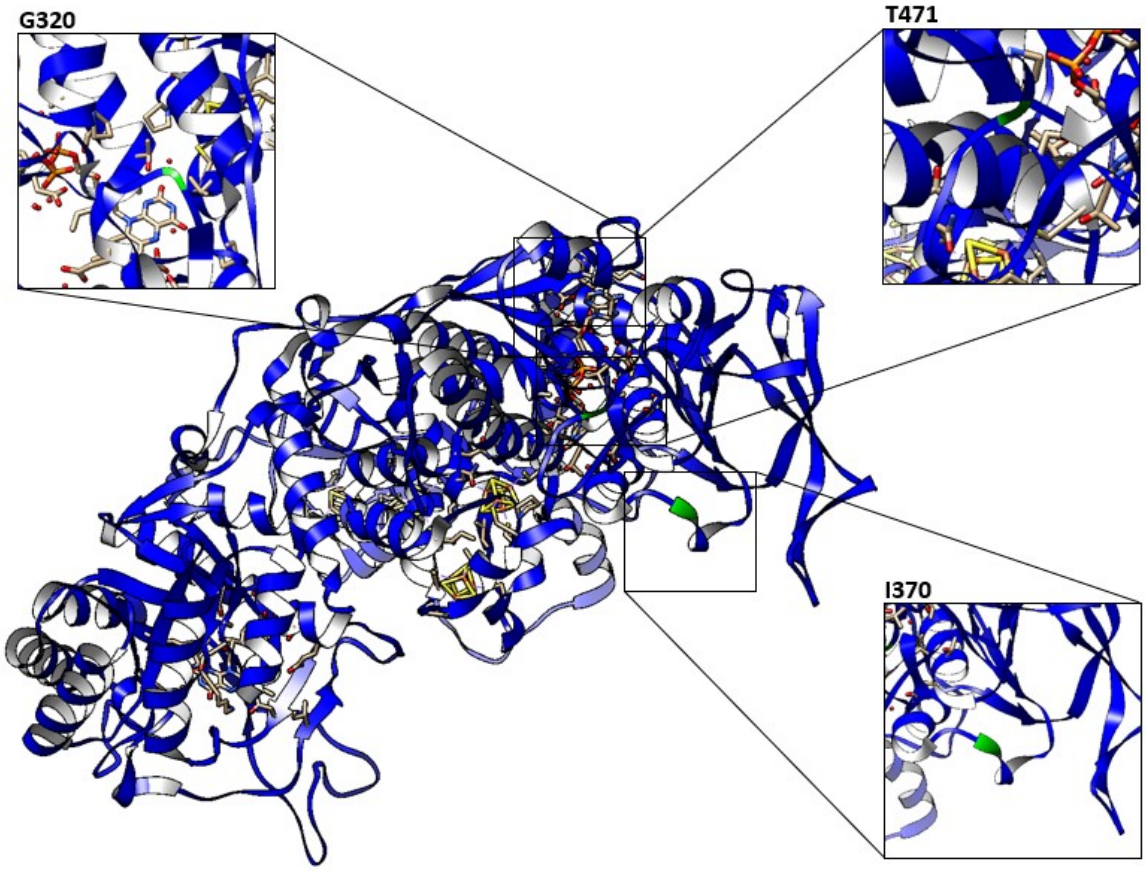


Figure 19b: *3D DPYD monomer protein visualization, green residues represent 3 variants (G320R; I370M and T471P) founded in cases cohort*

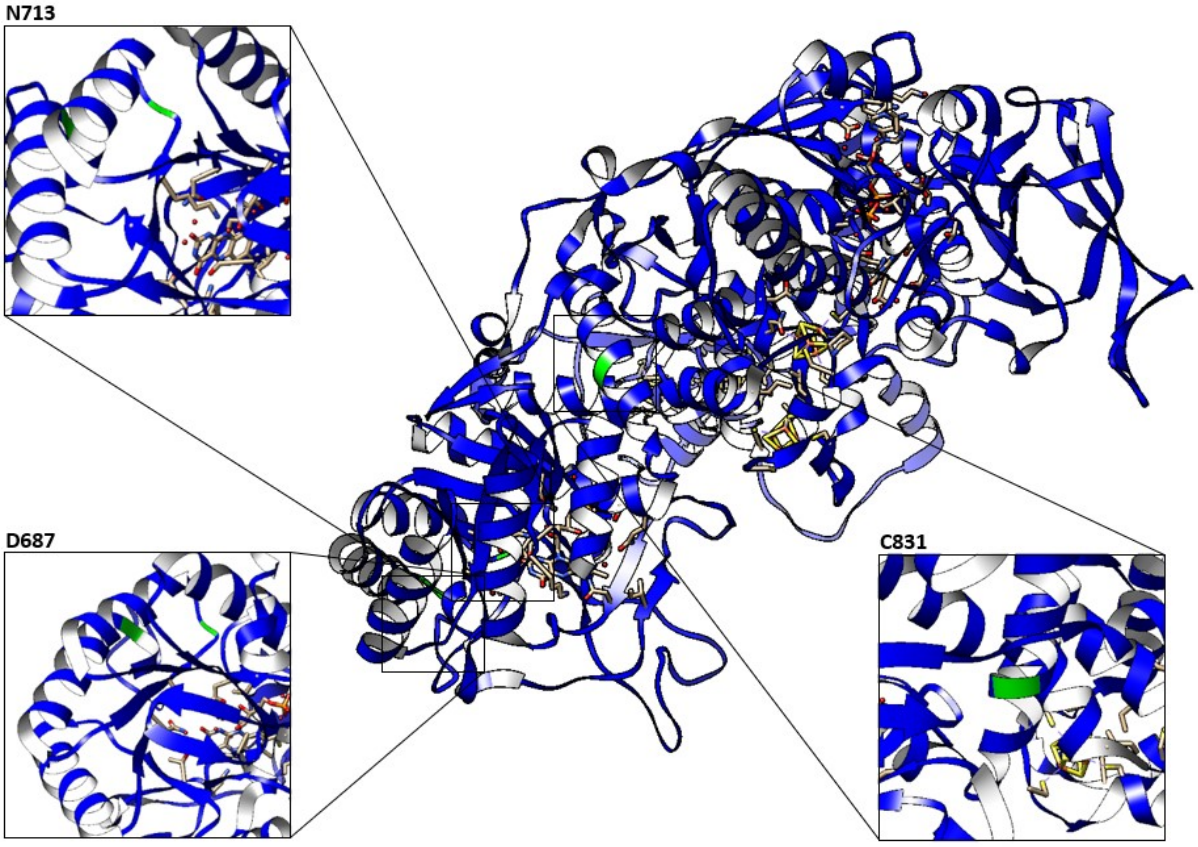


Figure 19c: *3D DPYD monomer protein visualization, green residues represent 3 variants (D687A; N713D and C831S) founded in cases cohort*

4.9. Quantification of serum uracil and dihydrouracil concentrations

Due to the retrospectively nature of this project, only four plasma sample among 9 patients carrying a very rare/novel genetic variant on *DPYD* were available for the measurement of DPD catabolic activity. The concentration of uracil (U), dihydrouracil (UH_2) and their ratio at baseline make possible to estimate the functionality of the enzyme. Metabolites quantification is reported in ***Table 13***.

***DPYD*: NM_000110**					
Sample	**Variant**	**Location**	**Uracil (ng/mL)**	**Dihydrouracil (ng/mL)**	**UH_2:U**
P2	c.G481A p.E161K	Exon5	14	94	6,71
P5	c.A1110G p.I370M	Exon 10	10,6	86,4	8,15
P6	c.A1411C p.T471P	Exon12	17,1	101	5,91
P9	c.T2491A p.C831S	Exon 20	11,8	125	10,59

Table 13: *Quantifications of main DPYD's metabolites, uracil, dihydrouracil and their ratio*

According to literature and laboratory's practice we considered defective *DPYD* patients to have an uracil concentration higher than 16ng/mL. Among these four patients, only proband P6 (c.1411A>C p.T471P) showed an altered uracil concentration, whereas P2 had a borderline DPD activity. However, from literature emerged that many factors beyond the catabolic activity of DPD may influence the uracil, dihydrouracil concentration and their ratio. Thus, we can't completely rule out that the presence of *DPYD* variants does not hamper the enzyme catabolic activity because as many other pharmacokinetic, genetic and related-life style factors might contribute to modify the uracil and dihydrouracil plasmatic levels.

4.10. Rare variants, control cohort

To assess whether the rare variants were differentially distributed in patients experiencing severe (grade ≥3) toxicity and in patients without, we sequenced a cohort of 106 patients who did not experienced severe toxicity in course of FL treatment. Patients were matched with the cohort of patients with toxicity by sex, age, primary disease location, treatment regimen and setting.
Patients characteristics are listed in ***Table 14.***
In this cohort, 53 (50%) patients are male, 50% are female and they ranged between 26 and 96 years old with a average age of 61 years.
Concerning the primary tumor site 92 (86.8%) patients were affected by CRC, 6 (5.7%) by breast cancer, 4 (3.8%) by gastric cancer, 3 (2.8%) by pancreatic and 1 (0.9%) by head-neck tumor.

Regarding the treatment, only 8 (7.5%) patients were treated in monotherapy with capecitabine, while 98 (92.5%) patients were treated with 5-FU based combination regimens. About these 98 patients, 43 (43.9%) were treated in combination with oxaliplatin, 38 (38.8%) with irinotecan, only 1 (1%) with platin, 4 (4.1%) with a combination of platin and irinotecan and 12 (12.2%) were treated with other combination. Concerning the adverse reactions, 37 (35%) patients did not experienced any toxic events, 52 (49%) patients experienced G1 and 17 (16) experienced G2 toxicities. Sequencing was performed on a MiSeq (Illumina) platform by means of the same gene target panel used for the cases cohort.

	Patients	Treatment	
Characteristic	**N° (%)**	**5-FU N° (%)**	**Capecitabine N° (%)**
All	106	87 (82.1%)	19 (17.9%)
Age			
Mean (range)	61 (26-98)	60 (26-98)	66 (38-83)
Sex			
Male	53 (50%)	43(49.4%)	9 (47.4%)
Female	53 (50%)	44 (50.6%)	10 (52.6%)
Type of cancer			
ColoRectal	92 (86.8%)	77 (88.5%)	15 (78.9%)
Breast	6 (5.7%)	6 (6.9%)	/
Gastric	4 (3.8%)	1 (1.15%)	3 (15.8%)
Head-Neck	1 (0,9%)	1 (1.15%)	/
Pancreas	3 (2.8%)	2 (2.3%)	1 (5.3%)
Treatment			
Monotherapy	8 (7.5%)	/	8 (42.1%)
Combination:	98 (92.5%)	87 (100%)	11 (57.9%)
Platin	1 (1%)	/	1 (9.1%)
Irinotecan	38 (38.8%)	38 (43.7%)	/
Oxaliplatin	43 (43.9%)	36 (41.4%)	7 (63.6%)
Platin + Irinotecan	4 (4.1%)	4 (4.6%)	/
Others	12 (12.2%)	9 (10.3%)	3 (27.3%)
Max Grade Toxicity (all)			
G0	37 (35%)	31 (83.8%)	6 (16.2%)
G1	52 (49%)	43 (82.7%)	9 (17.3%)
G2	17 (16%)	13 (76.5%)	4 (23.5%)

Table 14: *Control cohort characteristics*

4.11. Rare variants distribution in the control cohort

From sequencing data of the 106 patients from the control cohort a total of 17,368 variants has emerged. This cohort was composed by patients treated with FL-based regimens, wild-type for the four *DPYD* risk variants and at last with absence of severe adverse reactions. In this cohort 755

unique variants were present and included common polymorphisms, rare and novel variants. Moreover, in this cohort most variants fall on UTRs (386; 51.13%). About these, 330 (43.71%) fall on 3'UTRs and 56 (7.42%) fall on 5'UTRs. Instead, exonic variants were divided into missense (170; 22.52%), synonymous (166; 21.99%), stop gain (4; 0.53%), frameshift (4; 0.53%) and others (12; 1.59%) that include in-frame insertion and deletion (***Figure 20***). Finally, 13 (1.72%) variants were detected to influence the splicing sites. However, variant distribution in the control cohort was very similar to that reported for the case cohort (see ***Figure 12***).

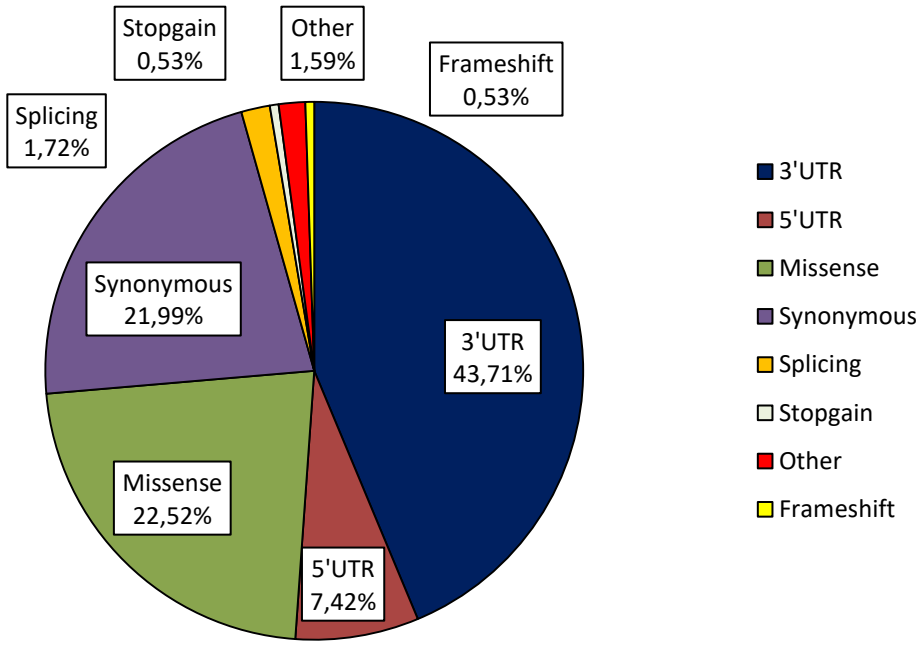


Figure 20: *Variants types distribution across 54 target gene in control cohorts*

As in the cases cohort, we plotted the type of variants identified on each gene to understand if any gene could be enriched with a specific type of variant **(*Figure 21a, b*)**.

Except for a few genes, most genes are enriched for variants in the UTR, according to our expectations. In contrast to the data emerged from cases set, *DPYD* gene seemed to be enriched in UTRs variants (***Figure 21a***).

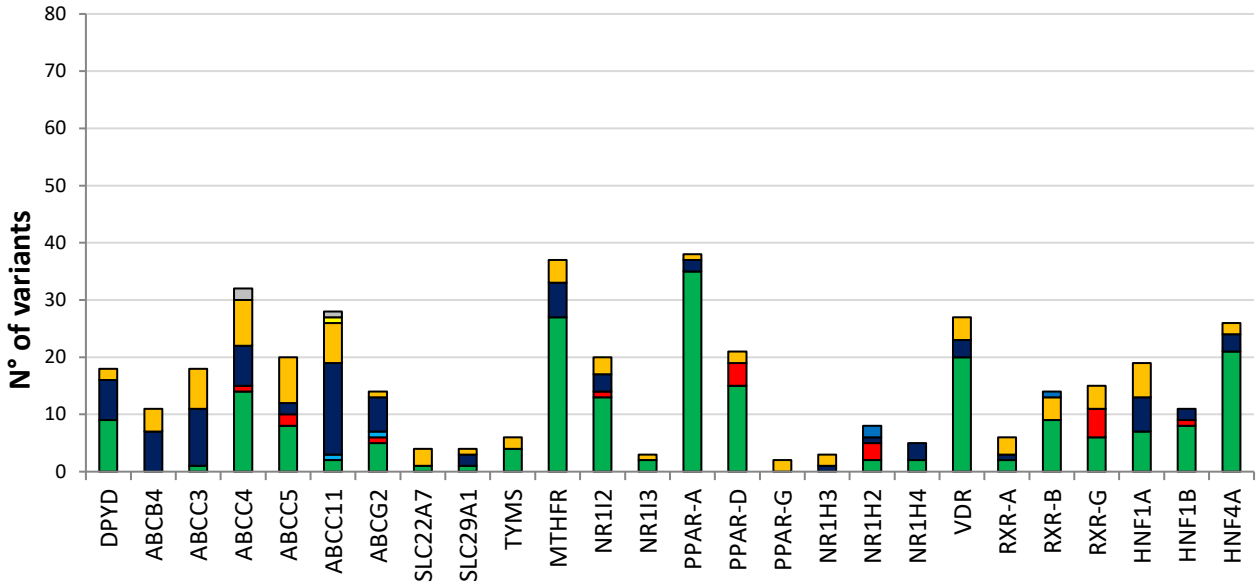


Figure 21a: *Variants distribution per gene in control cohort*

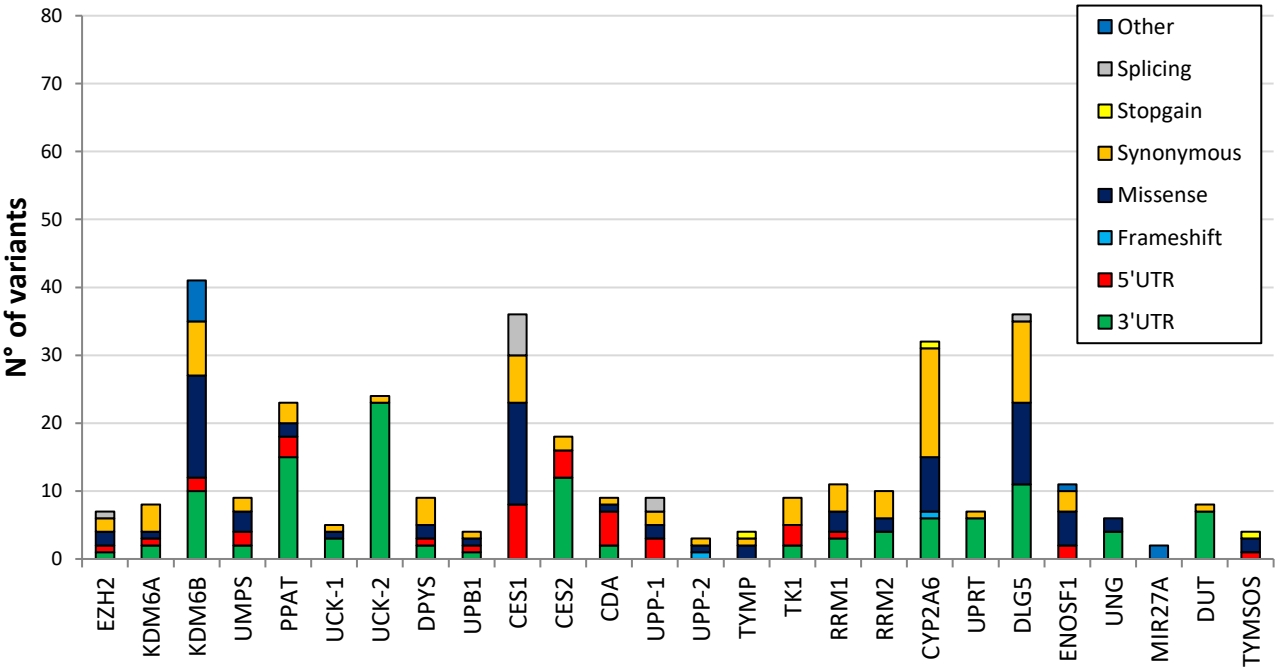


Figure 21b: *Variants distribution per gene in control cohort*

The number of variants per gene may depend on characteristic of gene and sequence. To assess if the distribution of variants is influenced by the length of the gene, we normalize the number of variants on the length of each gene. To investigate this, in ***Figure 22 (a, b)*** we reported in dark blue the total number of variants normalized for the gene length and in red the number of exonic variants normalized for the coding sequences length, respectively. After normalizing the number of variants for the gene's length, it appears evident that some genes, like *CES1* and *CYP2A6,* are hypervariable in both cohorts of patients when compared to other genes, as they have a greater number of variants per kbs. These genes seem to have the same trend in the cases and in the control cohort. Furthermore, *MIR27A* risk variant (rs895819) seem to be enriched in the control cohort.

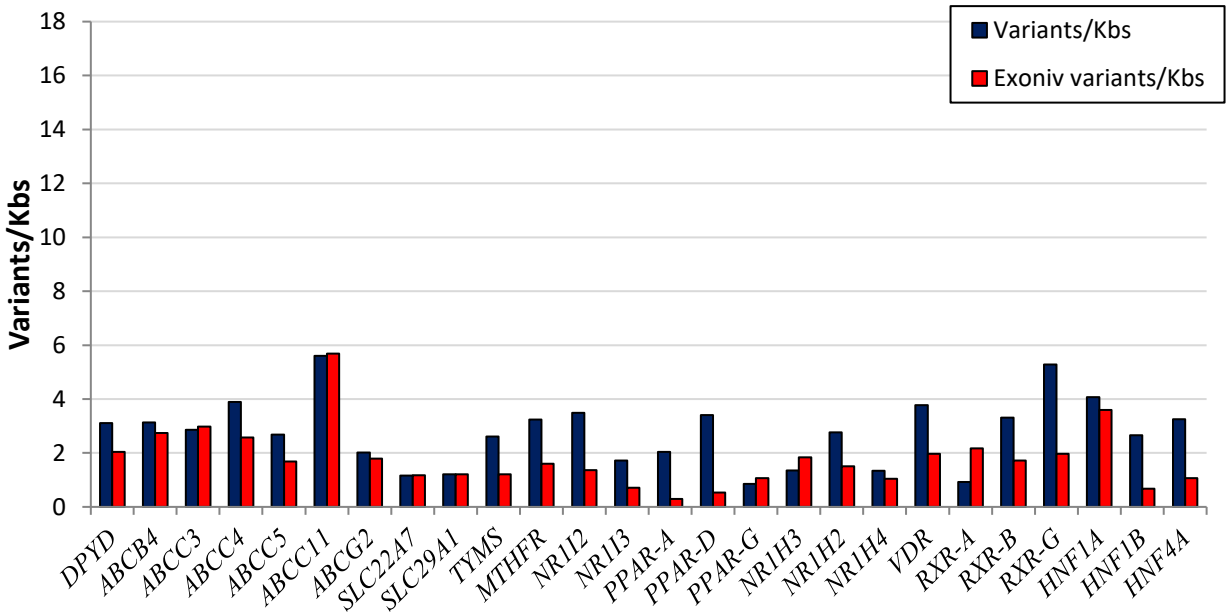


Figure 22a: *Variants distribution per gene normalized on the gene length in the control cohort*

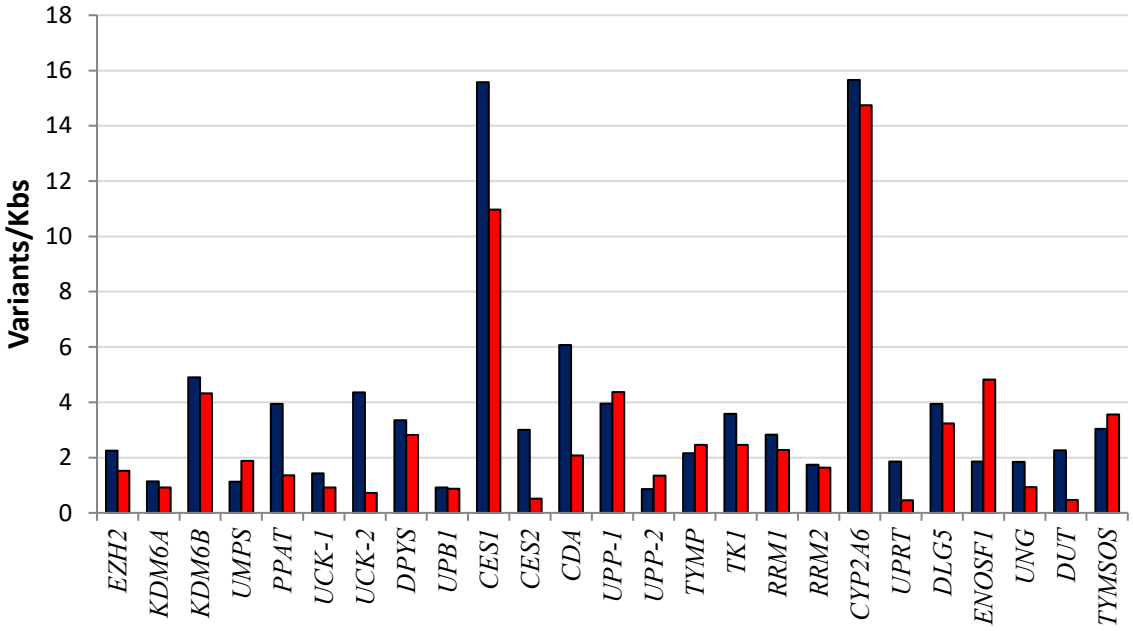


Figure 22b: *Variants distribution per gene normalized on the gene length in the control cohort*

Bearing in mind that a rare variant may have a stronger impact on phenotype than common polymorphisms, we investigated the distribution of variants detected according to their minor allele frequency (MAF). Among these unique 755 variants, 379 (50.2%) were rare (MAF<0.01), 256 (33.9%) were very rare variant (MAF<0.001) and finally only 94 (12.4%) were novel variants (MAF=0.0) ***(Figure 23)***. In the control cohort we observed a depletion of novel variants but no strong difference in very rare and rare variants abundances (Fisher's exact test, p value =0.001).

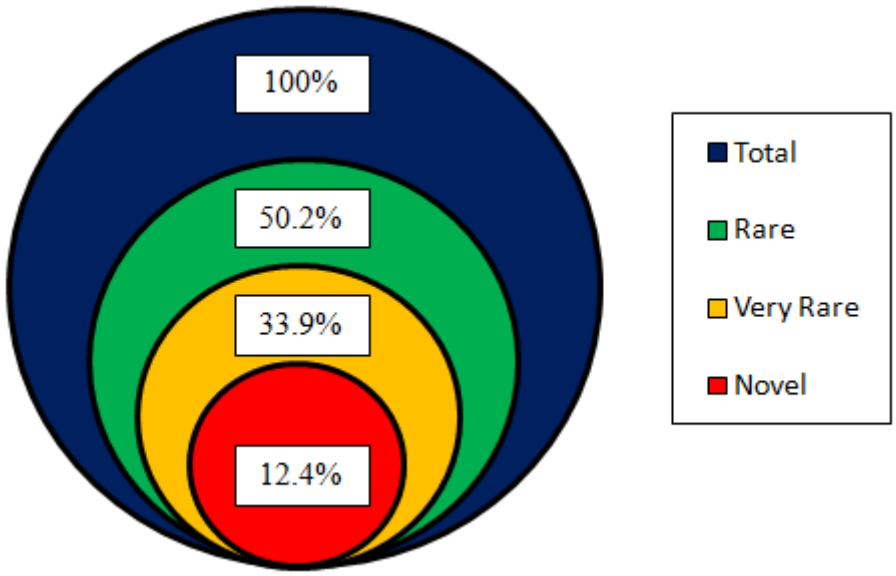


Figure 23: *Variants distribution according to Minor Allele Frequency (MAF) of control cohort*

In addition, we clustered the genes by class like as *DPYD*, transporters, genes belonging to folate pathway, nuclear receptors and others to normalize them on gene's length as shown previously. Genes involved in folate pathway showed a low number of novels, very rare and rare variants per kbs when compared with cases cohort group ***(Figure 24)***.

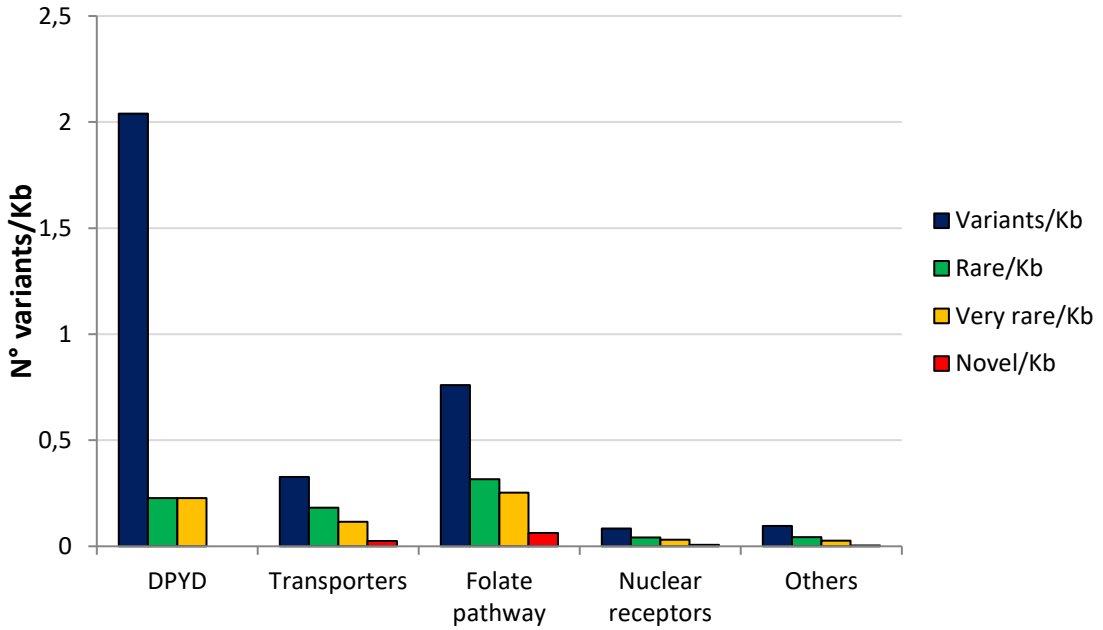


Figure 24: *Exonic variants distribution per genes classes normalized on the class length in the control cohort*

Moreover, in contrast with data emerged from the cases cohort, *DPYD* does not present novel variants and shows a low number of rare and very rare variants per kbs ***(Figure 24)***. To assess whether this difference may be related to the onset of severe toxicity, we analyzed the distribution of rare, very rare and novel *DPYD* variants in these two cohort through Fisher's exact test ***Table 15***. As show in ***Figure 25***, we found an unexpected statistically significant relation with the number of

very rare variants (P =0.0335) but not with the all novel and rare variants. The absence of statistically significant values with all three group may be due to the characteristic of genes where exonic sequences results more conserved than UTRs which may result high heterogeneous region. Indeed, when we consider only exonic variants, we found a significant enrichment of rare, very rare and novel variant of *DPYD* gene in the cases cohort rather than in control cohort (***Table 15***, ***Figure 25***). These data suggest that variants with probably impacting for the onset of toxicity are uncommon (MAF<0.01) variants located in the exons of *DPYD* genes.

ALL (Rare MAF <0.01)						Exonic (Rare MAF <0.01)			
Cohort	Total	Yes (%)	No (%)	OR (95% CI)	P	Yes (%)	No (%)	OR (95% CI)	P
Control	106	5 (4.7%)	101 (95.3%)	1		1 (0.9%)	105 (99.1%)	1	
Cases	108	10 (9.3%)	98 (90.7%)	**0.48 (0.16-1.47)**	**P = 0.284**	9 (8.3%)	99 (91.7%)	**0.1 (0.01-0.84)**	**P = 0.018**
ALL (Very Rare MAF <0.001)						**Exonic (Very Rare MAF <0.001)**			
Cohort	Total	Yes (%)	No (%)	OR (95% CI)	P	Yes (%)	No (%)	OR (95% CI)	P
Control	106	2 (1.9%)	104 (98.1%)	1		1 (0.9%)	105 (99.1%)	1	
Cases	108	10 (9.3%)	98 (90.7%)	**0.19 (0.04-0.88)**	**P = 0.033**	9 (8.3%)	99 (91.7%)	**0.1 (0.01-0.84)**	**P = 0.018**
ALL (Novel MAF =0)						**Exonic (Novel MAF =0)**			
Cohort	Total	Yes (%)	No (%)	OR (95% CI)	P	Yes (%)	No (%)	OR (95% CI)	P
Control	106	1 (0.9%)	105 (99.1%)	1		0 (0%)	106 (100%)	1	
Cases	108	7 (6.5%)	101 (93.5%)	**0.14 (0.02-1.14)**	**P = 0.065**	6 (5.6%)	102 (94.4%)	**/**	**P = 0.029**

Table 15: *Variants distribution according MAF in the two cohort.*

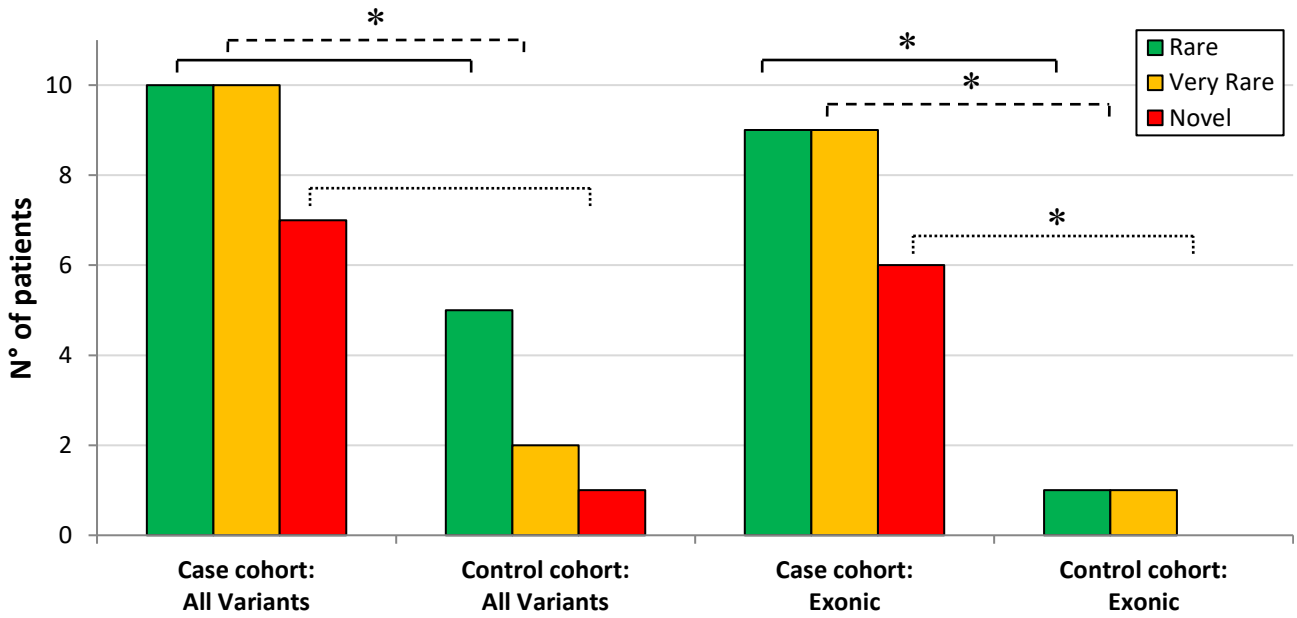


Figure 25: *Distribution of patients carrying DPYD variants in the two dataset. Each patient has only one variant, and these variants have been divided into rare, very rare and novel according to MAF. Fisher's exact test, * p value <0.05*

5. Discussion

The utility of pre-emptive genotyping test translation as a clinical recommendation remains debated, although the pharmacogenetics guidelines for the management of FL-based chemotherapy toxicity agree upon the clinical validity of *DPYD* risk variants.[114,115]

We retrospectively evaluated, in a large cohort of patients affected by colorectal cancer, the effect of dysfunctional *DPYD* variants on the risk to develop DLTs related to a FL-based treatment. DLTs are defined as grade ≥3 for non-hematological and grade ≥4 for hematological toxicities and lead to drug dose reduction or treatment interruption, with potential consequences on the anti-cancer efficacy.

In our study, patients carrying at least one of the four *DPYD* risk variants (rs3918290; rs55886062; rs67376798; rs56038477), were significantly associated to a 2.7-fold increased risk of developing DLTs, both acute and total.

A further step towards a better characterization was achieved using a *DPYD* GAS model where patients are stratified based on their catalytic enzyme activity according to the most updated CPIC guidelines. The general trend is maintained after patients' stratification according to the *DPYD* activity score. Indeed, patients with a GAS of 1.0 (intermediate metabolizer) showed a 10.1-fold increased risk to develop DLTs when compared to full metabolizer (GAS = 2.0). On the other hand, a weaker association was found between 1.5 GAS carriers and DLTs onset. This result is consistent with a milder effect of the variants on the enzyme activity. However, intermediate metabolizer (GAS = 1.0) have a stronger association when considering the first three cycles of treatment (OR = 10.12, 95% CI = 2.55-40.2) rather than entire course of chemotherapy (OR = 7.09 95% CI = 1.69-29.65). In opposite way, only when considering the entire course of chemotherapy, the 1.5 *DPYD* carriers display an increased risk to develop DLTs (OR = 2.08, 95% CI =1.02-4.27).

Moreover, to better understand the toxicity profile in relation to the GAS model, we evaluated the distribution of toxicity events among the three GAS group. We considered separately grade 3 and grade 4 toxicities, without distinguishing between hematological and non-hematological adverse reactions. Patients with a 1.0 GAS was enriched for grade 4 toxicities, which was not maintained when considering patients with a 1.5 GAS which showed an increased onset of grade 3 toxicities.

These results suggested that a lower *DPYD* activity score is strictly related to a more severe and acute toxicity onset while an intermediate activity is strictly related to a more mild and delayed toxicity.

One of the major causes of poorly widespread of pharmacogenomics diagnostic test in clinical practice lies in the lack of information on the usefulness and on economic consequences. Indeed, pharmacogenomics test can help to set the best therapy and at same time decrease the cost of treatments. Our study reports, for the first time, that patients' genotypes for the four *DPYD* risk variants allow to predict those patients with significant higher chemotherapy-related toxicity management costs, among subjects treated with fluoropyrimidines-based treatment. Moreover, patients carrying at least one of the four *DPYD* risk variants have an increased risk to develop toxicity which are associated with a higher toxicity management costs compared with noncarriers, especially when requiring hospitalization. Indeed, hospitalization result to be the most expensive intervention for management of adverse reactions.

By stratifying patients according to the *DPYD* gene activity score, toxicity management cost is inversely proportional to the GAS. The highest management cost was associated with rs3918290 (*DPYD**2A) carriers, and the lowest was associated with carriers of rs67376798 (*DPYD* c.2846) and rs56038477 (*DPYD* Hap-B3). Once again in our cohort, the poor metabolizer has been associated with early severe toxicity and a high toxicity management cost, while patients with a mild impact variant exhibit accumulation toxicity and management cost of lower toxicity.

These finding support the utility of pharmacogenomics classification according to *DPYD* genotype especially if considering that the costs for genotyping will decrease soon.

In the present study, fluoropyrimidines were administered with different combination regimens because monotherapy is not the standard therapy for patients with CRC. However, our analysis allowed a direct correlation of Fl monotherapy management costs with the costs associated with the combination with other chemotherapy drugs. Indeed, the mean cost per patient was highest for patients treated with a combination of fluoropyrimidine with oxaliplatin. This is due not only to the occurrence of severe neuropathy, but also to a general increase in the occurrence of hematological and gastrointestinal toxicities.

DPYD wild-type patients treated with a combination of fluoropyrimidine and oxaliplatin had a mean toxicity management cost that was 4-fold higher than wild-type patients treated with fluoropyrimidine monotherapy. This analysis demonstrates that *DPYD* variants may affect the cost of toxicity management in monotherapy treatment with FL as well as may affect the cost of toxicities in combination regimens.

Despite the reported results a great number of toxic events are still not explained. Indeed, the 4 *DPYD* risk genetic variants allow to predict a limited number of all adverse events.

Pre-treatment phenotyping of DPD activity through measurement of DH2/U metabolite in plasma, by monitoring DPD activity in peripheral blood mononuclear cells (PBMCs), that remain the gold

standard, and other methods can be helpful tools to predict toxicity and efficacy of FL.[49,116–119] However, this phenotypic approach is not easily applicable in the routine clinical practice.[49] One of the major limitations is the lack of fully validated or standardized methods. Furthermore, genotyping methods are easier, faster and very less expensive than phenotyping methods.[49]

In contrast to SNPs genotyping methods, NGS approach on *DPYD* were used and allowed the identification of new deleterious variants as well as common SNPs.[120] Furthermore, recent analysis from different public human genetic variants data sets highlight that deleterious variants have lower allele frequencies than neutral variants due to negative selection.[39,98] However, this analysis was performed on a very large and heterogeneous patients database, indeed patients in these databases result affected by difference human disease or disorders as well as cancer, neurogenerative disease, obesity and others.

Here, we reported for the first time rare and common variants in 54 pharmacogenes involved in 5-FU/Capecitabine pathway detoxification in patients affected by severe adverse reaction (G≥3).

In the first place we analyzed a group of patients with extreme toxicity to highlight possible causative variants. From sequencing data of 108 patients affected by extreme toxicity phenotype emerged a total of 19,960 variants with an average of 184 variants per patient, 961 were unique variants and included common polymorphisms, rare and novel variants.

Keeping in mind that we aimed to identify rare variants which may have a stronger impact on protein function, we clustered these unique 961 variants according Minor Allele Frequency (MAF) of European population.[39,98] From this analysis, 434 (45.2%) are rare with a MAF<0.01, 273 (28.4%) are very rare variants (MAF<0.001) and 176 (18.3%) are novel variants.

We supposed that the number of variants per gene could reflect a difference in the role of these genes in the onset of toxicity. However, many other factors can influence the number of variants including gene length, evolutionary gene conservation and the length of untranslated regions (UTRs).

To exclude these, we normalized the number of variants in each gene on the gene length. Due the fact that an exonic variants may have a bigger impact on the protein function than variants in UTRs, the same analysis was performed considering only exonic variants and normalizing for the coding sequences (CDS) length. Genes like *CES1* and *CYP2A6* appear hypervariable compared to others due to the greater number of variants per kbs in both analyses. Furthermore, the number of variants per gene does not seem to be influenced by the length of UTRs or genes. Moreover, to study the distribution of rare, novel variants we clustered genes by class like as *DPYD*, transporters, genes belonging to folate pathway, nuclear receptors and others. *DPYD* and genes involved in folate pathway showed a relatively higher number of novels, very rare and rare variants ***(Figure 13)***.

We subsequently focused our attention on the major candidate gene in the development of FL-related toxicity, i.e. *DPYD.* Thanks to NGS, we highlighted in the 108 patients with severe toxicity, three very rare missense (c.G345C, p.M115I; c.A2060C p.D687A and c.A2137G, p.N713D) variants (MAF < 0.001) and six singleton missense (c.A110G, p.D37G; c.G481A, p.E161K; c.C800T, p.T267I; c.G958A, p.G320R; c.A1110G, p.I370M; c.C1579T, p.P527S) variants on *DPYD* gene which may be causative of FL toxicity (***Table 10***). All *DPYD* variants were confirmed by Sanger sequencing. About these variants, we evaluated the possible impact on the protein. To do this, we used seven different prediction tools and, based on the provided scores, 7 out of 9 variants (p.D687A, p.N713D, p.D37G, p.E161K, p.T267I, p.G320R, p.P527S) seem to be deleterious for the protein function. However, these prediction tools have been not optimized on variants which fall on pharmcogenes, as *DPYD.*[111,121] Indeed, the prediction tools commonly used to assess the role of the big amount of variants emerged from NGS studies are highly influenced by the conservation score of the gene or better of the nucleotide. The conservation score measures the evolutionary conservation of the gene or of the single nucleotide/amino acid through the alignment of multiple sequences with paralogs and orthologs genes. Highly conserved positions are often indicative of their structural and/or functional importance. It is easy to understand that this score can be very helpful in genes highly conserved but usually found mutated in human diseases like cancers. However, to the best of our knowledge, to date only two works developed specific prediction tools for pharmacogenomics variants that unfortunately are not yet available.[111,121]

Moreover, since these missense variants fall on exonic splicing sites we evaluated a possible impact on splicing. HSF web tool suggested that *DPYD* variants could have a greater impact on the splicing mechanisms rather than in amino acid point changes.[112]

To better understand how this 9 variants can be deleterious, we analyzed 3D protein structure with UCSF Chimera.[113] Amino acid substitutions occur in surface residues of the protein, mainly in α-helix or loops motifs far from substrate binding pocket and catalytic site. Moreover, understanding impact of these variants on protein activity is a hard task given that this is only an in silico visualization of variants and thus we cannot rule out a possible aberrant mechanism during co-translation folding that may have a strong impact in the protein structure and function. Moreover, since many of the missense variants fall on exonic splicing sites, the functional impact may be caused by an aberrant splicing mechanism rather than by a single amino acid change.

Thanks to the collaboration with Dr Hilde Rosing of Department of Pharmacy and Pharmacology in The Netherlands Cancer Institute, we evaluated the DPD activity by measuring plasma pretreatment concentration of uracil (U) and dihydrouracil (UH_2) through a validated LC-MS/MS method. According to literature and laboratory's practice we considered defective *DPYD* patients when

uracil concentration is higher than 16ng/mL.[122] However, due to the retrospective nature of this project, only four plasma samples among 9 patients are available for measurement of DPD activity. About these four patients, patient carrying p.T471P variants shows an altered uracil concentration whereas patient carrying p.E161K variant shows a borderline activity. The same trend was observed when we considered UH2:U ratio. Thanks to phenotyping test, we are confident that the p.T471P *DPYD* mutation is the cause of neutropenia G3 that affect this patient during the first treatment cycle. However, gender and other genetic factors may influence the uracil, dihydrouracil concentration and consequently their ratio (UH2:U).[116]

More than 90% of the variants present in the pharmacogenes are rare variants, which represent the 30-40% of functional variability in pharmacogenes.[39,98-100] Due to the high human interindividual heterogeneity we cannot exclude that these data are related to the investigated phenotype rather than human heterogeneity. In order to exclude that, we compared the data of *DPYD* wild type patients that experienced adverse events (the case cohort) with those of a control cohort. Control cohort was composed by 106 patients who did not experienced severe toxicity (G<3).

From NGS data, from control cohort emerged a total of 17,368 variants with an average of 163 variants per patients. About these 17,368 variants,755 are unique.

In control cohort we observe a depletion of novel variants (p value =0.001) which may indicate a functional impact of novel variants in the onset of adverse reactions.

Clustering the genes by class and normalizing for the CDS length as previously done for the case cohort, we showed a significant (low number of rare, very rare and novel exonic variants per kbs (***Figure 25***) (P =0.0187, P =0.0187, P =0.0291, respectively). *DPYD* did not present novel variants in 106 patients of control cohort, however, emerged only a very rare missense (c.G1117A; p.V373I; rs772906420) variant.

Despite some limitations such as the heterogeneity of the patients, of the treatment regimens, of the sites of the primary tumor in this study, we have highlighted as common and new variants in the 5FU/Capecitabine detoxification pathway may affect the development of severe toxicity. However, given the heterogeneity of treatment regimens and co-treatments it is impossible to discriminate the effect of individual drugs in the development of severe adverse events.

Moreover, the retrospective nature of the study prevented the access to suitable biological material to allow the assessment of the functional impact of all these variants. Furthermore, this work of thesis focused only on a preliminary analysis of the data almost exclusively limited to the results obtained for the *DPYD* gene. However, we are already working on the analysis of the bulk of data deriving from the entire panel of 54 genes through focused bioinformatics analysis. Finally, due to the huge number of rare variants and the difficulty to evaluate their functional impact, we will need

large prospective studies to validate the role of these rare variants before they can be used in clinical practice.

Concluding, we showed how pre-emptive *DPYD* genotype test could be helpful in reducing the number of adverse reaction and at the same time allow to reduce the management cost for severe toxicities in monotherapy treatment but also in combination regimens. Furthermore, in this work, we showed how novel variants in 5-FU/capecitabine pathway detoxification can explain the development of severe toxicity phenotype in all those patients that are wild type for the well-know 4 *DPYD* risk variants.

Despite the recent efforts of precision medicine in optimizing treatments on each patient, to date translating the genetic information of each patient into a therapeutic indication remains a hard and time-consuming task. Indeed, further efforts are needed to identify new markers of severe toxicity development that in the future can be used in clinical practice.

6. Material and Methods

6.1. Cohorts

To validate the clinical value of the pre-emptive *DPYD* test, from an already existing database and biobank of patients with cancer we selected a retrospective multicenter population of 763 patients. Eligibility criteria we applied were: (i) histologically confirmed diagnosis of colorectal cancer (CRC), (ii) fluoropyrimidine-based treatment, (iii) availability of a peripheral blood sample, and (iv) signed informed consent approved by the local Ethical Committee.

Secondly, to assess the relationship between *DPYD* genotype and management-related cost we selected a population of 550 patients affected by CRC with FL-based therapy (***Figure 25***).

Eventually, to investigate the role of rare and novel germline variants in the toxicity onset, we selected a subpopulation of 120 *DPYD* wild-type patients (***Figure 25***). This subpopulation is characterized by: (i) fluoropyrimidine-based treatment, (ii) experience of extreme toxicity (grade 3 to 5), (iii) absence of 4 risk variants, (iv) availability of a peripheral blood sample, and (v) signed informed consent approved by the local Ethical Committee.

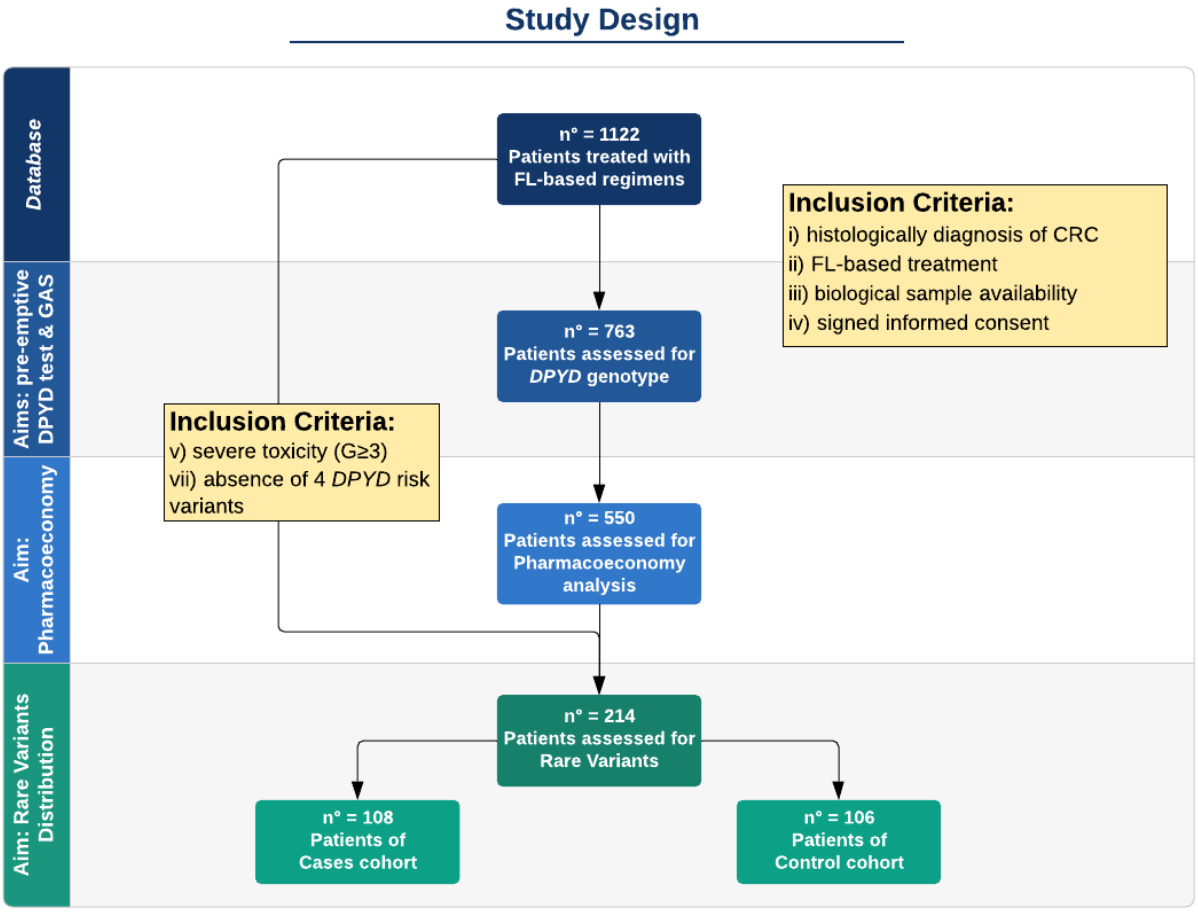


Figure 25: *Study design*

6.2. Data collection

The baseline clinical and demographic characteristics were retrieved for each patient, which include: chemotherapy schedule, dosage, deviation from the treatment schedule (i.e., drug dose reduction, delays, or early treatment interruption), and the complete list of treatment-related toxic events at each chemotherapy cycle, until drug discontinuation for any reason. Clinical and toxicity data for these patients were collected by the Experimental and Clinical Pharmacology Unit of CRO-Aviano. The causality of the toxic events was assessed by the physician at the time of occurrence, and only the chemotherapy-related events were recorded. All the toxic events were graded according to NCI-CTC for Adverse Events version 5.0.

6.3. Toxicity management cost

The cost of each toxicity event develop during treatment was listed based on three expert clinical oncologists from CRO-Aviano. Only the chemotherapy-related toxicity events were recorded by clinicians at the time of occurrence at each treatment cycle. The total cost was calculated as the sum of whole cost management occurring during each toxicity events, limited to the first 12 cycles of treatment. However, some patients were enrolled in different parts of Italy, most of them were recruited in Friuli Venezia Giulia Region. The Italian Public Heat Care System is regulated at regional level, so for this reason the analysis cost derived from the following: (i) the Health Agency of the Friuli Venezia Giulia Region (http://www.egas.sanita.fvg.it/) and (ii) the current version of the Friuli Venezia Giulia Regional Health System website.
The analysis covers whole cost related to clinical management including drugs, laboratory and instrumental examinations, nurses, and physicians for the needed time of hospitalization. The remain indirect costs such as productivity loss, transport, accommodation and others were not considered in this study.

6.4. *DPYD* Genotyping

DNAwas extracted with BioRobot EZ1and EZ1 DNA Blood Kit (Qiagen, Hilden, Germany) according to manufactures' instruction. Genotype was assessed for the four *DPYD* risk variants (c.1905+1G>C, *2A; c.1679T>G, *13; c.2846A>T, D949V and c.1236G>A, HapB3) by Pyrosequencing or Real-Time PCR and confirmed with Sanger sequencing.

In ***Table 16*** are listed the primer sequence used for the amplification and sequencing of the four risk variants.

***DPYD*: NM_000110**		
Primer		***Sequence: 5'→ 3'***
HapB3	**F**	*TGGTGAAAGAAAAAGCTGCAT*
	R	*TTGCATCACACATTTCAGCTC*
HapB3 (Pyro)	**F**	*[Btn]AACCAAAGGCACTGATGACC*
	R	*AATTTCTGCCATTCCTGTCC*
	S	*CAGTTTGTTCGGACAG*
**2A*	**F**	*ATGTATGGCCCTGGACAAAG*
	R	*ATGCATCAGCAAAGCAACTG*
**2A (Pyro)*	**F**	*CGGCTGCATATTGGTGTCAA*
	R	*[Btn]CACCAACTTATGCCAATTCTCTTGT*
	S	*GGCTGACTTTCCAGACA*
**13*	**F**	*CGGATGCTGTGTTGAAGTGATTT*
	R	*GTGTAATGATAGGTCTTGTCAAATAGT*
**13 (Pyro)*	**F**	*[Btn]CCTTTTGGTCTTGCTAGCGC*
	R	*AGTTTTGGTGAGGGCAAAACC*
	S	*CTTCAAAAGCTCTTCG*
2846	**F**	*GGTCCAAAAATGAGAAAAAGTTAGCC*
	R	*TCTCTCTAATGTTGTGGCTGATGA*
2846 (Pyro)	**F**	*[Btn]GCAGTACCTTGGAACATTTGGT*
	R	*AGGTCATGTAGCATTTACCACAGT*
	S	*CACAGTTGATACACATTTTCT*

***Table 16*:** *Primer sequence used to genotype analysis*

6.4.1. Pyrosequencing

Pyrosequencing is a sequencing-by-synthesis technique based on the detection of light. The first step consists in the amplification of DNA template with a biotinylated primer. Then, using streptavidin-coated magnetic beads the template is immobilize on the bead's surface. With a second primer the complementary strand is synthesize one base pair at a time. Each base incorporated releases a pyrophosphate (PPi) molecule because of nucleotide chain elongation. The enzyme ATP sulfurylase converts PPi to ATP in the presence of adenosine 5' phosphosulfate (APS). The ATP generated acts like as substrate for the enzyme luciferase which converts luceferin to oxyluciferin, thus generating light at a wavelength of 560 nm. The intensity of light emitted is proportional to the amount of ATP. At the end of each cycle, the unincorporated dNTPs and ATP are degraded by the apyrase.

For the amplification of DNA target regions, the "AmpliTaq Gold™ DNA Polymerase with Gold Buffer and $MgCl_2$kit" (Applied Biosystem Inc., Carlsbad, CA, USA) has been used, according to the mix reported in ***Table 17***.

Reagent	Volume (µl)	FinalConcentrazion
DNA	*1,5*	
Buffer 10X	*5*	1X
dNTPs[25mM]	*0,25*	*0,125mM*
$MgCl_2$ [25mM]	*5*	*2,5mM*
Primer F [100µM]	*0,1*	*0,2µM*
Primer R [100µM]	*0,1*	*0,2µM*
AmpliTaq Gold® DNA Polymerase (5U/µl)	*0,2*	1U /rnx
H_2O	*37,85*	
Total	*50*	

Table 17: *Pyrosequencing mix composition*

The thermal cycle used for DNA amplification was:

- Polymerase activation at 95 °C per 10'
- 35 cycles of:
    - Denaturation at 95 °C for 30"
    - Annealing at 56 °C for 30"
    - Extension at 72 °C for 1'
- Final elongation at 72 °C for 5'

Reagents used for pyrosequencing was magnetic beads (3µl) are load on the disk and mixed with 10µl of sample on PyroMark Q48 Autoprep (Qiagen, Hilden, Germany). 2 µl of sequencing primers (4µM) are mix with an 8µl of annealing buffer for a final concentration of 800nM and are loaded in injector catrigdes. The four dNTPs, binding buffer (5µl for each sample), substrate, enzyme and apyrase are load in automatic injectors catrigdes according to manufactures' instruction.

6.4.2. Allelic discrimination in real-time PCR by fluorescent-labeled probes

Real-Time PCR (RT-PCR or qPCR) is a technique that allows the amplification and the simultaneous quantification of target DNA. Through a pair of sequence-specific fluorescent probe that hybridize with the target sequence, this technique allows to discriminate patient's genotype. In these analyses we used a dual label TaqMan probes (FAM and VIC) that emit respectively at 517nm and 551nm. The detection is allowed by using the thermocycler 7500Real-Time PCR System (Applied Biosystem Inc., Carlsbad, CA, USA).

Each sample was analysed according to mix reported in ***Table 18*** with the follow thermal cycle:

- Polymerase activation at 95 °C per 10'
- 55 cycles of:
    - Denaturation at 92 °C for 15''
    - Annealing/Extension at 60 °C for 1'

Reagent	Volume (µl)	FinalConcentrazion
PCR product	*1,5*	
MasterMix 2X	*10*	*1X*
AssayMix 20X	*1*	*1X*
H_2O	*7,5*	
Total	*20*	

Table 18: *Real-Time mix composition*

For the *DPYD* c.2846 variants assay, primers and probes are reported in ***Table 19***.
HapB3 was screening according to the commercial TaqMan® SNP Genotyping Assay (Assay ID: C__25596099_30) (Thermo Fisher Scientific, Wilmington, DE, USA). No information about sequences of primers and probes have been released by the company.

DPYD: NM_000110		
Primer		***Sequence: 5'→ 3'***
2846	**F**	TGAATTGAGCAACGTAGAGCA
	R	TGTAGCATTTACCACAGTTGATACACA
Probe		***Sequence: 5'→ 3'***
2846		VIC-TGGCTATGATTGATGAAGAA-MGB
		6-FAM-TGGCTATGATTGTTGAAGAA-MGB

Table 19: *Primers and probes for RT-PCR*

6.4.3. Sanger Sequencing for *DPYD* *2A; *13; c.2846A>T, and HapB3 variants validation

Sanger sequencing was adopted to confirm the presence of four risk variants previously highlight by Pyrosequencing and Real-Time PCR.PCR was used to amplified DNA following the reported thermal cycle.
c.1905+1G>C and c.1679T>G:

- Polymerase activation at 95 °C per 10'
- 35 cycles of:
    - Denaturation at 95 °C for 30"
    - Annealing at 56 °C for 30"
    - Extension at 72 °C for 1'
- Final elongation at 72 °C for 5'

c.2846A>T:

- Polymerase activation at 95 °C per 10'
- 35 cycles of:
    - Denaturation at 95 °C for 30"
    - Annealing at 63 °C for 30"
    - Extension at 72 °C for 30''
- Final elongation at 72 °C for 7'

c.1236G>A:

- Polymerase activation at 95 °C per 10'
- 35 cycles of:
    - Denaturation at 95 °C for 30"
    - Annealing at 58 °C for 30"
    - Extension at 72 °C for 30''
- Final elongation at 72 °C for 10'

The PCR product have been subjected to electrophoresis on agarose gel at 1% in 1X TBE buffer. For the Sanger sequencing, the amplified were firstly purify with Diffinity RapidTip®2 (Sigma-Aldrich, Inc., Milwaukee, WI, USA), then were prepared through a three-step method. The first step consists in labelling amplified DNA whit the fluorescent nucleotide. A mix with polymerase, primer, dNTPs and label ddNTPs from "BigDye® Terminator v3.1 Cycle Sequencing Kit" (Applied Biosystem Inc., Carlsbad, CA, USA) are used like reporter in ***Table 20***. Label ddNTPs lack the 3'-OH group of dNTPs that is essential for polymerase-mediated strand elongation in a PCR. Thermal cycle:

- Polymerase activation at 96 °C per 1'
- 30 cycles of:
    - Denaturation at 96 °C for 1'
    - Annealing at 50 °C for 30"
    - Extension at 60 °C for 2'

Reagent	*Volume (µl)*	*Final Concentrazion*
PCR product	*2*	
Buffer	*1.2*	
BigDye Terminator3.1	*1*	
Primer [3,3µM]	*1*	*0,3µM*
H_2O	*4,8*	
Total	*10*	

Table 20: *Label mix for Sanger sequencing*

The labeled fragments are purified through precipitation. 10µL of label DNA was mixed with 75µL of magnesium ethanolate and incubated for 10' at RT. Then it was centrifuged at 13000G for 15' at 4 °C to allow the precipitation of DNA. Eventually, the supernatant was promptly discarded, and the DNA pellet was resuspended in 15µL of HiDi™ Formamide (Applied Biosystem Inc., Carlsbad, CA, USA). Isolated DNA was then denatured at 96 °C for 2' and loaded in the automatic DNA sequencing.

Finally, the automatic DNA sequencing (AB3130xl Applied Biosystem Inc., Carlsbad, CA, USA) was used to separate electrophoretically the DNA fragment. The emitted fluorescent light is revealed, and sequencing data are transferred tothe instrument computer that, by an algorithm, converts them into electropherograms. To display the electropherograms the "FinchTV" (Geispiza) program was used.

6.5. Selection of candidate genes for Next Generation Sequencing

Based on PubMed database (https://www.ncbi.nlm.nih.gov/pubmed/) and PharmGKB (https://www.pharmgkb.org/), we selected 54 genes including *DPYD* and its regulators regions, genes involved in tegafur and capecitabine pathways, transporters and nuclear receptors related to fluoropyrimidines pharmacokinetics and pharmacodynamics (***Table 21***).

Gene	Transcript ID	RefSeq	RefSeqNP	Gene	Transcript ID	RefSeq	RefSeqNP
DPYD	ENST00000370192.3	NM_000110	NP_000101	Others			
Folate Pathway				EZH2	ENST00000320356.2	NM_004456	NP_004447
TYMS	ENST00000323274.10	NM_001071	NP_001062	KDM6A	ENST00000377967.4	NM_021140	NP_066963
MTHFR	ENST00000376590.3	NM_005957	NP_005948	KDM6B	ENST00000254846.5	NM_001080424	NP_001073893
Transporters				MIR27A	ENST00000385073.1	NR_029501	/
ABCB4	ENST00000265723.4	NM_000443	NP_000434	MIR27B	ENST00000385129.1	NR_029665	/
ABCC3	ENST00000285238.8	NM_003786	NP_003777	UMPS	ENST00000232607.2	NM_000373	NP_000364
ABCC4	ENST00000376887.4	NM_005845	NP_005836	MIR23A	ENST00000385245.1	NR_029495	/
ABCC5	ENST00000334444.6	NM_005688	NP_005679	PPAT	ENST00000264220.2	NM_002703	NP_002694
ABCC11	ENST00000394748.1	NM_032583	NP_115972	UCK1	ENST00000372215.4	NM_031432	NP_113620
ABCG2	ENST00000237612.3	NM_004827	NP_004818	UCK2	ENST00000367879.4	NM_012474	NP_036606
SLC22A7	ENST00000372589.3	NM_006672	NP_006663	DPYS	ENST00000351513.2	NM_001385	NP_001376
SLC29A1	ENST00000393841.1	NM_001078177	NP_001071645	UPB1	ENST00000326010.5	NM_016327	NP_057411
Nuclear Receptors				CES1	ENST00000360526.2	NM_001025195	NP_001020366
NR1I2	ENST00000393716.2	NM_003889	NP_003880	CES2	ENST00000317091.4	NM_003869	NP_003860
NR1I3	ENST00000367980.2	NM_001077482	NP_001070950	CDA	ENST00000375071.3	NM_001785	NP_001776
PPARA	ENST00000262735.5	NM_001001928	NP_001001928	UPP1	ENST00000395564.4	NM_003364	NP_003355
PPARD	ENST00000360694.3	NM_006238	NP_006229	UPP2	ENST00000005756.4	NM_173355	NP_775491
PPARG	ENST00000309576.6	NM_138712	NP_619726	TYMP	ENST00000252029.3	NM_001113755	NP_001107227
NR1H2	ENST00000253727.5	NM_007121	NP_009052	TK1	ENST00000301634.7	NM_003258	NP_003249
NR1H3	ENST00000441012.2	NM_005693	NP_005684	RRM1	ENST00000300738.5	NM_001033	NP_001024
NR1H4	ENST00000548884.1	NM_001206979	NP_001193908	RRM2	ENST00000360566.2	NM_001165931	NP_001159403
VDR	ENST00000549336.1	NM_000376	NP_000367	CYP2A6	ENST00000301141.5	NM_000762	NP_000753
RXRA	ENST00000481739.1	NM_002957	NP_002948	UPRT	ENST00000373383.4	NM_145052	NP_659489
RXRB	ENST00000374680.3	NM_001270401	NP_001257330	DLG5	ENST00000372391.2	NM_004747	NP_004738
RXRG	ENST00000359842.5	NM_006917	NP_008848	ENOSF1	ENST00000251101.7	NM_017512	NP_059982
HNF1A	ENST00000541395.1	NM_000545	NP_000536	UNG	ENST00000336865.2	NM_003362	NP_003353
HNF1B	ENST00000225893.4	NM_000458	NP_000449	DUT	ENST00000331200.3	NM_001025248	NP_001020419
HNF4A	ENST00000316099.4	NM_178849	NP_849180	TYMSOS	ENST00000323813.3	NM_001012716	NP_001012734

Table 21: *54 genes selected for NGS screening*

6.6. Targeted Next Generation Sequencing

To perform NGS analysis, we designed a hybridization based custom NimbleGenSeqCap EZ Choice Library (Roche, Inc., Madison, WI, USA) to target the 5' and 3' UTR and coding sequence of genes. 3 kb of genomic sequence flanking at the 5′UTR ends of *DPYD*, *MTHFR* and *TYMS* gene was included, accounting for a total of 186.535 bp.

Libraries were constructed according to the NimbleGen SeqCap EZ Library SR User's Guide v3.0 (Roche, Inc. Madison, WI, USA) (***Figure 26***). Pooled libraries were loaded according to the DNA Truseq protocol (Illumina, Inc., San Diego, CA, USA).

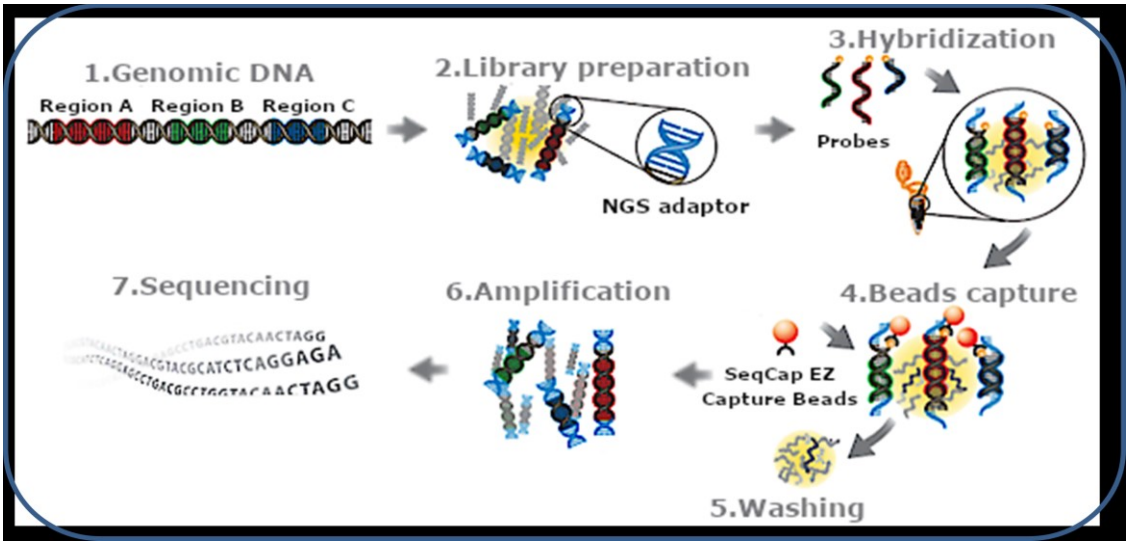


Figure 26: *DNA library preparation workflow*

Briefly, genomic DNA from blood was firstly purified using Agencourt AMPure XP Beads (Beckman Coulter, Brea, California, USA), then the quality was assessed through NanoDrop (Thermo Fisher Scientific, Wilmington, DE, USA) and Quantus Fluorometer (Promega, Madison, Wisconsin, USA).

100 ng of each samples were diluted into 35µL and fragmented with eight different restriction enzymes for 15' at 37 °C to obtain fragments of approximately 200bp following ***Table 22***.

Reagents	Volume(µL)
gDNA	*35*
KAPA Frag Buffer (10X)	*5*
KAPA FragEnzyme	*10*
Total	*50*

Table 22: *Fragmentation mix*

Then, DNAs were subjected to three enzymatic steps: end repair, A-tailing and ligation to Illumina paired-end indexed adapters (***Table 23*** and ***Table 24***).

Reagents	Volume(µL)
KAPA End Repair & A-tailing Buffer	7
KAPA End Repair & A-tailing Enzyme Mix	3
Total	10

Table 23: *End repair and A-tailing mix*

Reagents	Volume(µL)
PCR-grade water	5
KAPA Ligation Buffer	30
KAPA DNA Ligase	10
Total	45

Table 24: *Index ligation mix*

Once the DNA libraries were indexed, a double-sided selection was performed to obtain fragments of approximately 300-350bp, then they were PCR-amplified according to the follow thermal cycles:

- Activation of polymerase at 98 °C for 45''
- 12 amplification cycles split in:
  - Denaturisation at 98 °C for 15''
  - Annealing at 60 °C per 30''
  - Extension at 72 °C per 30''
- Final elongation at 72 °C per 1'

The amplified fragments were quantified by NanoDrop (Thermo Fisher Scientific) and Quantus (Promega). To assess the fragments' length distribution, each library was analyzed through 2200 High Sensitivity TapeStation (Agilent Technologies, Santa Clara, CA, USA) according to manufacturer's instruction (***Figure 27***).

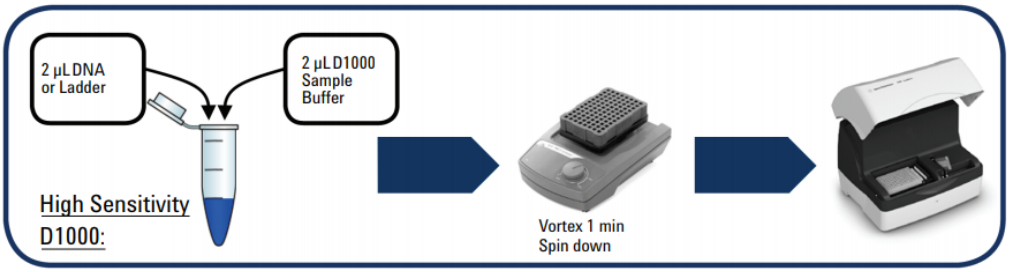


Figure 27: *TapeStation workflow*

45ng of amplified DNA from each sample were pooled, then were in-solution hybridizated to a custom NimbleGenSeqCap EZ Choice Library (Roche, Inc.) of complementary oligonucleotide DNA baits. After washing, the captured fragments were amplified following this thermal cycle:

- Activation of polymerase at 98 °C for 45''
- 7 amplification cycles split in:
  - Denaturation at 98 °C for 15''

- Annealing at 60 °C per 30''
- Extension at 72 °C per 30''

- Final elongation at 72 °C per 1'

Then the pooled-libraries were sent for sequencing to generate 2 × 150 bp paired-end reads using MiSeq platform (Illumina, Inc.). The resulting fastQ files were analyzed using an in-house-developed pipeline described below.

6.7. Bioinformatic analysis

FASTQ sequencing reads were initially trimmed using trimmomatic-0.36 and assessed in terms of sequencing quality, those falling below the required quality threshold were excluded from subsequent analyses.[123] Then, the FASTQ sequencing trimmed reads were mapped to human reference genome (UCSC hg19 NCBI build 37.1) assembled using Burrows-Wheeler Aligner (BWA).[124] The Bam files obtained were sorted, indexed and visualized by Integrative Genomics Viewer (IGV).[125]

Local realignment around indels and PCR duplicate removal were performed by Genome Analysis Toolkit (GATK) (Broad Institute, Cambridge, MA)mand picard MarkDuplicates (http://broadinstitute.github.io/picard).[126,127] Single nucleotide variations and indels were called through VarScan2.[128] Variant effect prediction was done using Annovar and annotated using population based annotations such as dbSNP, 1000 Genomes Project, NHLBI Exome Project (ESP6500) (Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (http://evs.gs.washington.edu/EVS/)) and Exome Aggregation Consortium (ExAC, http://exac.broadinstitute.org/).[95,129,130]

6.8. In silico functional prediction tools

To evaluate the functional impact of exonic variants we used five different prediction algorithms that are SIFT, PolyPhen, PROVEAN, CADD**,** LRT**,** MutationTaster and FATHMM.[104-110] In addition, for variants that fall near the exon junctions we integrated the results from HSF.[112] HSF combine different tools and algorithms that allow to study canonical as well as non-canonical splicing regions including: matrices for SR protein from ESEfinder, exonic and intronic sequence region from RESCUE-ESE hexamers, algorithms to find PESE and PESS octamers motif identified by Zhang and Chasin, exon and intron-identity element (EIE and IIE) defined by Zhang and

colleagues, ESR sequence identify by Goren and a new algorithms for silencer sequences defined by Sironi and co-workers and ESS decamers.[131-137]

6.9. Sanger Sequencing to validate the *DPYD* variants detected by NGS

Sanger sequencing was adopted to confirm the presence of variants previously highlight by NGS.
All primers were designed through Primer3Plus and evaluated for the specificity by NCBI BLASTn.[138,139] In ***Table 25*** are present the primer sequence used for the amplification and sequencing.
For the amplification has been used the "AmpliTaq Gold™ DNA Polymerase with Gold Buffer and $MgCl_2$kit" (Applied Biosystem Inc., Carlsbad, CA, USA Inc., Carlsbad, CA, USA) following the same procedure previously reported in ***Table 17***.
Each PCR reaction was submitted to the follow thermal cycle:

- Activation of polymerase at 95 °C for 10'
- 35 amplification cycles of:
    - Denaturisation at 95 °C per 30''
    - Annealing ±4 °C at 60 °C per 30''
    - Extension at 72 °C per 30"
- Final elongation at 72 °C for 5'

DPYD: NM_000110		
Variants		**Primers Sequence: 5'→ 3'**
c.G345C; p.M115I	**F**	TGTTTGTCGTAATTTGGCTGTT
	R	GTGCATGGTGATGGTAGTGG
c.G481A; p.E161K	**F**	TGGTCTGACTTGTGGAATGGT
	R	GGGTATCAACAGAGCACCAAG
c.C800T; p.T267I	**F**	TTTGCTTACAGATGTTTTCCTCT
	R	TGCTGCTGAGCTTGATTTTG
c.G958A; p.G320R	**F**	AAAAGCCCCTCCTCCTGCTAA
	R	TGTGCTGCTGAGCTTGATTTTGA
c.A1110G; p.I370M	**F**	TGCCTTTGACTGTGCAACAT
	R	GAAACAATTATGTGTAGCCCTGAA
c.A1411C; p.T471P	**F**	TGATTGTCTATTGTTTTTCACTTGTTT
	R	GATGGCAAATGCCTACCTGT
c.A2060C; p.D687A	**F**	GGTGTAATACTGAGGAGGAATTTGA
	R	TTCTTACCTTCCTTTGCAGCTC
c.A2137G; p.N713D	**F**	CAGCTTTGCTGTTGTTCCAG
	R	TGTTTGTAGGAAAAGACATACTTGA
c.T2491A; p.C831S	**F**	CTCCAGACGGCTACTGATCC
	R	AATCACATCCAGGAGGCACT
c.*431A>G	**F**	ATTGCCCTATGCTGTGCTCC
	R	AGCAATTTCAGCGAAGGGGA

Table 25: *Primers for variants control through Sanger sequencing*

6.10. Molecular visualization

To visualize DPYD 3D protein we use Chimera software.[113] Protein structure (1h7w) from *Sus scrofa (Pig)* was downloaded from the Protein Data Bank (PDB). Protein structure was at 1.9Å resolution.

6.11. Quantification of serum uracil and dihydrouracil concentrations

Thanks to the collaboration with Dr Hilde Rosing of Department of Pharmacy and Pharmacology in The Netherlands Cancer Institute, we were able to estimate DPD activity by measuring plasma pretreatment concentration of uracil (U) and dihydrouracil (UH_2) through a validated LC-MS/MS analysis[122]. Briefly, after protein precipitation, clear supernatants were collected and separated with a short chromatographic run with Acquity UPLC system (Waters, Milford, MA, USA) coupled to a QTrap 5500 triple quadrupole spectrometer (Sciex, Framingham, MA, USA).

According to laboratory practice the threshold to identify deficient DPD enzyme was put at 16ng/ml.[140]

6.12. Statistical analysis

The association between the *DPYD* genotype and the risk to develop DLT was evaluated grouping patients according to the presence of at least one risk variant in the 4-SNP panel and accordingly to GAS model. SNPs frequency were compared with those reported for Caucasian population in dbSNP (https://www.ncbi.nlm.nih.gov/SNP/) and then tested for deviation from Hardy-Weinberg equilibrium. The association between patients carrying DPYD risk allele and toxicity occurrence was estimated through an unconditional logistic regression model, adjusted for gender, age, chemotherapy scheme and radiotherapy exposure. Odds ratio (OR) and 95% confidence interval (CI) were calculated and the statistical significance was set at $P < 0.05$ (two-sided).

For the pharmacoeconomics analysis, patients were stratified into non-carrier group (patients without *DPYD* variants) and carrier group (patients with at least one *DPYD* risk variant, rs3918290, rs55886062, rs67376798 and rs56038477). Moreover, patients were categorized according to the *DPYD* gene activity score (GAS).

For the statistical analysis we applied general linear models with γ distribution and log-link function. Management cost was predicted on a model equation that included physiological factors such as sex, age, and treatment factors like as setting (adjuvant/metastatic), number of cycles, and chemotherapy regimen. The statistical difference between mean toxicity management costs in each genotype group or chemotherapy regimen was evaluated by ANOVA.

Assessment of association between patients with *DPYD* variants and hospitalization or toxicity occurrence was estimated through an unconditional logistic regression model, adjusted for sex, age, setting (adjuvant/metastatic), number of cycles, and chemotherapy scheme.

The association between *DPYD* genotype and hospitalization was tested by Fisher's exact test. OR and 95% CI were computed, and statistical significance was set at $P < 0.05$. Using the $\chi 2$ test we assessed the statistically significant differences in hospitalization rates within patients undergoing different chemotherapy regimens.

Difference in variants distribution of rare, very rare and novel variants on *DPYD* genes between the cases and control cohort were analyzed using Fisher's exact test, statistical significance was set at $P < 0.05$.

7. References

1. Heidelberger, C. *et al.* Fluorinated pyrimidines, a new class of tumour-inhibitory compounds. *Nature* **179**, 663–666 (1957).
2. Ivy, H. K. Treatment of Breast Cancer with 5-Fluorouracil. *Ann. Intern. Med.* **57**, 598 (1962).
3. Moertel, CharlesG., Reitemeier, RichardJ., Childs, DonaldS., Colby, MalcolmY. & Holbrook, MargaretA. COMBINED 5-FLUOROURACIL AND SUPERVOLTAGE RADIATION THERAPY OF LOCALLY UNRESECTABLE GASTROINTESTINAL CANCER. *The Lancet* **294**, 865–867 (1969).
4. Posner, M. R. *et al.* Cisplatin and fluorouracil alone or with docetaxel in head and neck cancer. *N. Engl. J. Med.* **357**, 1705–1715 (2007).
5. Therapy of locally unresectable pancreatic carcinoma: A randomized comparison of high dose (6000 rads) radiation alone, moderate dose radiation (4000 rads + 5-fluorouracil), and high dose radiation + 5-fluorouracil. The gastrointestinal tumor study group - Moertel - 1981 - Cancer - Wiley Online Library. https://onlinelibrary.wiley.com/doi/abs/10.1002/1097-0142(19811015)48:8%3C1705::AID-CNCR2820480803%3E3.0.CO;2-4.
6. Look, K. Y., Muss, H. B., Blessing, J. A. & Morris, M. A phase II trial of 5-fluorouracil and high-dose leucovorin in recurrent epithelial ovarian carcinoma. A Gynecologic Oncology Group Study. *Am. J. Clin. Oncol.* **18**, 19–22 (1995).
7. Préfontaine, M., Donovan, J. T., Powell, J. L. & Buley, L. Treatment of refractory ovarian cancer with 5-fluorouracil and leucovorin. *Gynecol. Oncol.* **61**, 249–252 (1996).
8. Sadek, H. *et al.* Treatment of advanced squamous cell carcinoma of the skin with cisplatin, 5-fluorouracil, and bleomycin. *Cancer* **66**, 1692–1696 (1990).
9. Zhou, J. *et al.* Capecitabine inhibits postoperative recurrence and metastasis after liver cancer resection in nude mice with relation to the expression of platelet-derived endothelial cell growth factor. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **9**, 6030–6037 (2003).

10. Miwa, M. *et al.* Design of a novel oral fluoropyrimidine carbamate, capecitabine, which generates 5-fluorouracil selectively in tumours by enzymes concentrated in human liver and cancer tissue. *Eur. J. Cancer Oxf. Engl. 1990* **34**, 1274–1281 (1998).
11. Toide, H., Akiyoshi, H., Minato, Y., Okuda, H. & Fujii, S. Comparative studies on the metabolism of 2-(tetrahydrofuryl)-5-fluorouracil and 5-fluorouracil. *Gan* **68**, 553–560 (1977).
12. Fujii, S., Ikenaka, K., Fukushima, M. & Shirasaka, T. Effect of uracil and its derivatives on antitumor activity of 5-fluorouracil and 1-(2-tetrahydrofuryl)-5-fluorouracil. *Gan* **69**, 763–772 (1978).
13. Saif, M. W., Syrigos, K. N. & Katirtzoglou, N. A. S-1: a promising new oral fluoropyrimidine derivative. *Expert Opin. Investig. Drugs* **18**, 335–348 (2009).
14. Miura, K. *et al.* 5-FU Metabolism in Cancer and Orally-Administrable 5-FU Drugs. *Cancers* **2**, 1717–1730 (2010).
15. Giller, S. A., Zhuk, R. A. & Lidak, M. I. [Analogs of pyrimidine nucleosides. I. N1-(alpha-furanidyl) derivatives of natural pyrimidine bases and their antimetabolities]. *Dokl. Akad. Nauk SSSR* **176**, 332–335 (1967).
16. Ikeda, K. *et al.* Bioactivation of tegafur to 5-fluorouracil is catalyzed by cytochrome P-450 2A6 in human liver microsomes in vitro. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **6**, 4409–4415 (2000).
17. Schuetz, J. D., Wallace, H. J. & Diasio, R. B. 5-Fluorouracil incorporation into DNA of CF-1 mouse bone marrow cells as a possible mechanism of toxicity. *Cancer Res.* **44**, 1358–1363 (1984).
18. Glazer, R. I. & Peale, A. L. The effect of 5-fluorouracil on the synthesis of nuclear RNA in L1210 cells in vitro. *Mol. Pharmacol.* **16**, 270–277 (1979).
19. Kanamaru, R., Kakuta, H., Sato, T., Ishioka, C. & Wakui, A. The inhibitory effects of 5-fluorouracil on the metabolism of preribosomal and ribosomal RNA in L-1210 cells in vitro. *Cancer Chemother. Pharmacol.* **17**, 43–46 (1986).

20. Ghoshal, K. & Jacob, S. T. Specific inhibition of pre-ribosomal RNA processing in extracts from the lymphosarcoma cells treated with 5-fluorouracil. *Cancer Res.* **54**, 632–636 (1994).

21. Carrico, C. K. & Glazer, R. I. Effect of 5-fluorouracil on the synthesis and translation of polyadenylic acid-containing RNA from regenerating rat liver. *Cancer Res.* **39**, 3694–3701 (1979).

22. Randerath, K., Tseng, W. C., Harris, J. S. & Lu, L. J. Specific effects of 5-fluoropyrimidines and 5-azapyrimidines on modification of the 5 position of pyrimidines, in particular the synthesis of 5-methyluracil and 5-methylcytosine in nucleic acids. *Recent Results Cancer Res. Fortschritte Krebsforsch. Progres Dans Rech. Sur Cancer* **84**, 283–297 (1983).

23. Santi, D. V. & Hardy, L. W. Catalytic mechanism and inhibition of tRNA (uracil-5-)methyltransferase: evidence for covalent catalysis. *Biochemistry* **26**, 8599–8606 (1987).

24. Patton, J. R. Ribonucleoprotein particle assembly and modification of U2 small nuclear RNA containing 5-fluorouridine. *Biochemistry* **32**, 8939–8944 (1993).

25. Doong, S. L. & Dolnick, B. J. 5-Fluorouracil substitution alters pre-mRNA splicing in vitro. *J. Biol. Chem.* **263**, 4467–4473 (1988).

26. Samuelsson, T. Interactions of transfer RNA pseudouridine synthases with RNAs substituted with fluorouracil. *Nucleic Acids Res.* **19**, 6139–6144 (1991).

27. Santi, D. V., McHenry, C. S. & Sommer, H. Mechanism of interaction of thymidylate synthetase with 5-fluorodeoxyuridylate. *Biochemistry* **13**, 471–481 (1974).

28. Danenberg, P. V. & Danenberg, K. D. Effect of 5, 10-methylenetetrahydrofolate on the dissociation of 5-fluoro-2'-deoxyuridylate from thymidylate synthetase: evidence for an ordered mechanism. *Biochemistry* **17**, 4018–4024 (1978).

29. Danenberg, P. V., Langenbach, R. J. & Heidelberger, C. Structures of reversible and irreversible complexes of thymidylate synthetase and fluorinated pyrimidine nucleotides. *Biochemistry* **13**, 926–933 (1974).

30. Carreras, C. W. & Santi, D. V. The catalytic mechanism and structure of thymidylate synthase. *Annu. Rev. Biochem.* **64**, 721–762 (1995).

31. Longley, D. B., Harkin, D. P. & Johnston, P. G. 5-fluorouracil: mechanisms of action and clinical strategies. *Nat. Rev. Cancer* **3**, 330–338 (2003).

32. Methotrexate and Fluorouracil Toxicities: A Collaborative Practice Approach to Prevention and Treatment - The ASCO Post. http://www.ascopost.com/issues/may-1-2014-supplement/methotrexate-and-fluorouracil-toxicities-a-collaborative-practice-approach-to-prevention-and-treatment/.

33. Meta-Analysis Group In Cancer *et al.* Toxicity of fluorouracil in patients with advanced colorectal cancer: effect of administration schedule and prognostic factors. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **16**, 3537–3541 (1998).

34. Tecza, K., Pamula-Pilat, J., Lanuszewska, J., Butkiewicz, D. & Grzybowska, E. Pharmacogenetics of toxicity of 5-fluorouracil, doxorubicin and cyclophosphamide chemotherapy in breast cancer patients. *Oncotarget* **9**, 9114–9136 (2018).

35. Edwards, I. R. & Aronson, J. K. Adverse drug reactions: definitions, diagnosis, and management. *The Lancet* **356**, 1255–1259 (2000).

36. Palmirotta, R. *et al.* Rare Dihydropyrimidine Dehydrogenase Variants and Toxicity by Floropyrimidines: A Case Report. *Front. Oncol.* **9**, (2019).

37. Toffoli, G. *et al.* The Genotype for DPYD Risk Variants in Patients With Colorectal Cancer and the Related Toxicity Management Costs in Clinical Practice. *Clin. Pharmacol. Ther.* (2018).

38. Ventola, C. L. Role of Pharmacogenomic Biomarkers In Predicting and Improving Drug Response. *Pharm. Ther.* **38**, 545–560 (2013).

39. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).

40. Motulsky, A. G. Drug reactions enzymes, and biochemical genetics. *J. Am. Med. Assoc.* **165**, 835–837 (1957).

41. Crews, K. R., Hicks, J. K., Pui, C.-H., Relling, M. V. & Evans, W. E. Pharmacogenomics and Individualized Medicine: Translating Science Into Practice. *Clin. Pharmacol. Ther.* **92**, 467–475 (2012).

42. Amstutz, U. *et al.* Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Dihydropyrimidine Dehydrogenase Genotype and Fluoropyrimidine Dosing: 2017 Update. *Clin. Pharmacol. Ther.* **103**, 210–216 (2018).

43. Swen, J. J. *et al.* Pharmacogenetics: from bench to byte--an update of guidelines. *Clin. Pharmacol. Ther.* **89**, 662–673 (2011).

44. Van Kuilenburg, A. B. *et al.* Genotype and phenotype in patients with dihydropyrimidine dehydrogenase deficiency. *Hum. Genet.* **104**, 1–9 (1999).

45. Yokota, H. *et al.* cDNA cloning and chromosome mapping of human dihydropyrimidine dehydrogenase, an enzyme associated with 5-fluorouracil toxicity and congenital thymine uraciluria. *J. Biol. Chem.* **269**, 23192–23196 (1994).

46. Takai, S., Fernandez-Salguero, P., Kimura, S., Gonzalez, F. J. & Yamada, K. Assignment of the human dihydropyrimidine dehydrogenase gene (DPYD) to chromosome region 1p22 by fluorescence in situ hybridization. *Genomics* **24**, 613–614 (1994).

47. Wei, X. *et al.* Characterization of the human dihydropyrimidine dehydrogenase gene. *Genomics* **51**, 391–400 (1998).

48. Capecitabine-based treatment of a patient with a novel DPYD genotype and complete dihydropyrimidine dehydrogenase deficiency - Henricks - 2018 - International Journal of Cancer - Wiley Online Library. https://onlinelibrary.wiley.com/doi/full/10.1002/ijc.31065.

49. Henricks, L. M. *et al.* Translating DPYD genotype into DPD phenotype: using the DPYD gene activity score. *Pharmacogenomics* **16**, 1277–1286 (2015).

50. Steimer, W. *et al.* Allele-specific change of concentration and functional gene dose for the prediction of steady-state serum concentrations of amitriptyline and nortriptyline in CYP2C19 and CYP2D6 extensive and intermediate metabolizers. *Clin. Chem.* **50**, 1623–1633 (2004).

51. Dalle Fratte Chiara, C. E., Polesel Jerry, Roncato Rossana, De Mattia Elena, Ecca Fabrizio, Bignucolo Alessia, Garziera Marica, Dreussi Eva, Palazzari Elisa, Buonadonna Angela, Guardascione Michela, Berretta Massimiliano, Foltran Luisa, Sartor Franca, D'Andrea Mario, Favaretto Adolfo, Mini Enrico, Nobili Stefania, De Paoli Antonino, Toffoli Giuseppe. *DPYD* Gene Activity Score Predicts Dose-Limiting Toxicity in Fluoropyrimidine-Treated Colorectal Cancer Patients. *J. Mol. Clin. Med.* **1**, 143 (2018).

52. Schirripa, M., Procaccio, L., Lonardi, S. & Loupakis, F. The role of pharmacogenetics in the new ESMO colorectal cancer guidelines. *Pharmacogenomics* **18**, 197–200 (2017).

53. Gillis, N. K. & Innocenti, F. Evidence required to demonstrate clinical utility of pharmacogenetic testing: the debate continues. *Clin. Pharmacol. Ther.* **96**, 655–657 (2014).

54. Morel, A. *et al.* Clinical relevance of different dihydropyrimidine dehydrogenase gene single nucleotide polymorphisms on 5-fluorouracil tolerance. *Mol. Cancer Ther.* **5**, 2895–2904 (2006).

55. Schwab, M. *et al.* Role of genetic and nongenetic factors for fluorouracil treatment-related severe toxicity: a prospective clinical trial by the German 5-FU Toxicity Study Group. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **26**, 2131–2138 (2008).

56. Fragoulakis, V. *et al.* Estimating the Effectiveness of DPYD Genotyping in Italian Individuals Suffering from Cancer Based on the Cost of Chemotherapy-Induced Toxicity. *Am. J. Hum. Genet.* **104**, 1158–1168 (2019).

57. Roncato, R. *et al.* Cost Evaluation of Irinotecan-Related Toxicities Associated With the UGT1A1*28 Patient Genotype. *Clin. Pharmacol. Ther.* **102**, 123–130 (2017).

58. Henricks, L. M. *et al.* DPYD genotype-guided dose individualisation of fluoropyrimidine therapy in patients with cancer: a prospective safety analysis. *Lancet Oncol.* **19**, 1459–1467 (2018).

59. Henricks, L. M. *et al.* A cost analysis of upfront DPYD genotype–guided dose individualisation in fluoropyrimidine-based anticancer therapy. *Eur. J. Cancer* **107**, 60–67 (2019).

60. The use of pharmacogenetics to increase the safety of colorectal cancer patients treatedwith fluoropyrimidines. https://cdrjournal.com/article/view/2994.

61. Molecular cloning and characterization of the human dihydropyrimidine dehydrogenase promoter. - PubMed - NCBI. https://www.ncbi.nlm.nih.gov/pubmed/11072080.

62. microRNAs miR-27a and miR-27b directly regulate liver dihydropyrimidine dehydrogenase expression through two conserved binding sites. - PubMed - NCBI. https://www.ncbi.nlm.nih.gov/pubmed/24401318.

63. Polymorphisms in MIR27A Associated with Early-Onset Toxicity in Fluoropyrimidine-Based Chemotherapy. - PubMed - NCBI. https://www.ncbi.nlm.nih.gov/pubmed/25655103.

64. Methylation of the DPYD promoter and dihydropyrimidine dehydrogenase deficiency. - PubMed - NCBI. https://www.ncbi.nlm.nih.gov/pubmed/16778115.

65. Methylation of the DPYD Promoter: An Alternative Mechanism for Dihydropyrimidine Dehydrogenase Deficiency in Cancer Patients | Clinical Cancer Research. http://clincancerres.aacrjournals.org/content/11/24/8699?ijkey=4d70366735e35a5d89cef9e064bf5b26378bd75d&keytype2=tf_ipsecsha.

66. Suppression of DPYD expression in RKO cells via DNA methylation in the regulatory region of the DPYD promoter: a potentially important epigenetic m... - PubMed - NCBI. https://www.ncbi.nlm.nih.gov/pubmed/17612628.

67. Histone H3K27 trimethylation modulates 5-fluorouracil resistance by inhibiting PU.1 binding to the DPYD promoter. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5093042/.

68. Guenther, B. D. *et al.* The structure and properties of methylenetetrahydrofolate reductase from Escherichia coli suggest how folate ameliorates human hyperhomocysteinemia. *Nat. Struct. Biol.* **6**, 359–365 (1999).

69. Veronese, M. L. *et al.* Phase I trial of UFT/leucovorin and irinotecan in patients with advanced cancer. *Eur. J. Cancer Oxf. Engl. 1990* **40**, 508–514 (2004).

70. Lu, J.-W. *et al.* [Relationship of methylenetetrahydrofolate reductase C677T polymorphism and chemosensitivity to 5-fluorouracil in gastric carcinoma]. *Ai Zheng Aizheng Chin. J. Cancer* **23**, 958–962 (2004).

71. Ruzzo, A. *et al.* Pharmacogenetic profiling in patients with advanced colorectal cancer treated with first-line FOLFIRI chemotherapy. *Pharmacogenomics J.* **8**, 278–288 (2008).

72. Cohen, V. *et al.* Methylenetetrahydrofolate reductase polymorphism in advanced colorectal cancer: a novel genomic predictor of clinical response to fluoropyrimidine-based chemotherapy. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **9**, 1611–1615 (2003).

73. Kaneda, S. *et al.* Structural and functional analysis of the human thymidylate synthase gene. *J. Biol. Chem.* **265**, 20277–20284 (1990).

74. Copur, S., Aiba, K., Drake, J. C., Allegra, C. J. & Chu, E. Thymidylate synthase gene amplification in human colon cancer cell lines resistant to 5-fluorouracil. *Biochem. Pharmacol.* **49**, 1419–1426 (1995).

75. Joerger, M. *et al.* Germline TYMS genotype is highly predictive in patients with metastatic gastrointestinal malignancies receiving capecitabine-based chemotherapy. *Cancer Chemother. Pharmacol.* **75**, 763–772 (2015).

76. van Kuilenburg, A. B. *et al.* Clinical implications of dihydropyrimidine dehydrogenase (DPD) deficiency in patients with severe 5-fluorouracil-associated toxicity: identification of new mutations in the DPD gene. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **6**, 4705–4712 (2000).

77. Kawakami, K. *et al.* Different Lengths of a Polymorphic Repeat Sequence in the Thymidylate Synthase Gene Affect Translational Efficiency but Not Its Gene Expression. *Clin. Cancer Res.* **7**, 4096–4101 (2001).

78. Pullarkat, S. T. *et al.* Thymidylate synthase gene polymorphism determines response and toxicity of 5-FU chemotherapy. *Pharmacogenomics J.* **1**, 65–70 (2001).

79. Lecomte, T. *et al.* Thymidylate synthase gene polymorphism predicts toxicity in colorectal cancer patients receiving 5-fluorouracil-based chemotherapy. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **10**, 5880–5888 (2004).

80. Horie, N., Aiba, H., Oguro, K., Hojo, H. & Takeishi, K. Functional analysis and DNA polymorphism of the tandemly repeated sequences in the 5'-terminal regulatory region of the human gene for thymidylate synthase. *Cell Struct. Funct.* **20**, 191–197 (1995).

81. Marsh, S. & McLeod, H. L. Thymidylate synthase pharmacogenetics in colorectal cancer. *Clin. Colorectal Cancer* **1**, 175–178; discussion 179-181 (2001).

82. Kawakami, K., Omura, K., Kanehira, E. & Watanabe, Y. Polymorphic tandem repeats in the thymidylate synthase gene is associated with its protein expression in human gastrointestinal cancers. *Anticancer Res.* **19**, 3249–3252 (1999).

83. Yawata, A. *et al.* Polymorphic tandem repeat sequences of the thymidylate synthase gene correlates with cellular-based sensitivity to fluoropyrimidine antitumor agents. *Cancer Chemother. Pharmacol.* **56**, 465–472 (2005).

84. Shahrokni, A., Rajebi, M. R. & Saif, M. W. Toxicity and Efficacy of 5-Fluorouracil and Capecitabine in a Patient With TYMS Gene Polymorphism: A Challenge or a Dilemma? *Clin. Colorectal Cancer* **8**, 231–234 (2009).

85. Kaneda, S. *et al.* Regulatory sequences clustered at the 5' end of the first intron of the human thymidylate synthase gene function in cooperation with the promoter region. *Somat. Cell Mol. Genet.* **18**, 409–415 (1992).

86. Kaneda, S. *et al.* Role in translation of a triple tandemly repeated sequence in the 5'-untranslated region of human thymidylate synthase mRNA. *Nucleic Acids Res.* **15**, 1259–1270 (1987).

87. Takeishi, K. *et al.* Human thymidylate synthase gene: isolation of phage clones which cover a functionally active gene and structural analysis of the region upstream from the translation initiation codon. *J. Biochem. (Tokyo)* **106**, 575–583 (1989).

88. Dolnick, B. J. Cloning and characterization of a naturally occurring antisense RNA to human thymidylate synthase mRNA. *Nucleic Acids Res.* **21**, 1747–1752 (1993).

89. Dolnick, B. J., Black, A. R., Winkler, P. M., Schindler, K. & Hsueh, C. T. rTS gene expression is associated with altered cell sensitivity to thymidylate synthase inhibitors. *Adv. Enzyme Regul.* **36**, 165–180 (1996).

90. Chu, J. & Dolnick, B. J. Natural antisense (rTSalpha) RNA induces site-specific cleavage of thymidylate synthase mRNA. *Biochim. Biophys. Acta* **1587**, 183–193 (2002).

91. Initial sequencing and analysis of the human genome. *Nature* **409**, 860 (2001).

92. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).

93. The Cost of Sequencing a Human Genome | NHGRI. https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost.

94. Kobayashi, Y. *et al.* Pathogenic variant burden in the ExAC database: an empirical approach to evaluating population data for clinical variant interpretation. *Genome Med.* **9**, (2017).

95. 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).

96. Tennessen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–69 (2012).

97. Kimura, H. *et al.* M108 - A NOVEL RARE VARIANT R292H IN RTN4R AFFECTS GROWTH CONE FORMATION AND POSSIBLY CONTRIBUTES TO SCHIZOPHRENIA SUSCEPTIBILITY. *Eur. Neuropsychopharmacol.* **29**, S1014 (2019).

98. Kozyra, M., Ingelman-Sundberg, M. & Lauschke, V. M. Rare genetic variants in cellular transporters, metabolic enzymes, and nuclear receptors can be important determinants of interindividual differences in drug response. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **19**, 20–29 (2017).

99. Lauschke, V. M. & Ingelman-Sundberg, M. Requirements for comprehensive pharmacogenetic genotyping platforms. *Pharmacogenomics* **17**, 917–924 (2016).

100. Targeted Next-Generation Sequencing for Comprehensive Genetic Profiling of Pharmacogenes. - PubMed - NCBI. https://www.ncbi.nlm.nih.gov/pubmed/27727443.

101. Precision Medicine from a Public Health Perspective | Annual Review of Public Health. https://www.annualreviews.org/doi/10.1146/annurev-publhealth-040617-014158.

102. Ioannidis, N. M. *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.* **99**, 877–885 (2016).

103. Meulendijks, D. *et al.* Rs895819 in MIR27A improves the predictive value of DPYD variants to identify patients at risk of severe fluoropyrimidine-associated toxicity. *Int. J. Cancer* **138**, 2752–2761 (2016).

104. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).

105. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).

106. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the functional effect of amino acid substitutions and indels. *PloS One* **7**, e46688 (2012).

107. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).

108. Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes. *Genome Res.* **19**, 1553–1561 (2009).

109. Schwarz, J. M., Cooper, D. N., Schuelke, M. & Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat. Methods* **11**, 361–362 (2014).

110. Shihab, H. A. *et al.* Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum. Mutat.* **34**, 57–65 (2013).

111. Zhou, Y., Mkrtchian, S., Kumondai, M., Hiratsuka, M. & Lauschke, V. M. An optimized prediction framework to assess the functional impact of pharmacogenetic variants. *Pharmacogenomics J.* **19**, 115–126 (2019).

112. Desmet, F.-O. *et al.* Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* **37**, e67 (2009).

113. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

114. Lunenburg, C. A., van Staveren, M. C., Gelderblom, H., Guchelaar, H.-J. & Swen, J. J. Evaluation of clinical implementation of prospective DPYD genotyping in 5-fluorouracil- or capecitabine-treated patients. *Pharmacogenomics* **17**, 721–729 (2016).

115. Meulendijks, D., Cats, A., Beijnen, J. H. & Schellens, J. H. M. Improving safety of fluoropyrimidine chemotherapy by individualizing treatment based on dihydropyrimidine dehydrogenase activity - Ready for clinical practice? *Cancer Treat. Rev.* **50**, 23–34 (2016).

116. Launay, M. *et al.* UPFRONT DPD DEFICIENCY DETECTION TO SECURE 5-FU ADMINISTRATION: PART 2- APPLICATION TO HEAD-AND-NECK CANCER PATIENTS. *Clin. Cancer Drugs* **4**, 122–128 (2017).

117. Van Kuilenburg, A. B. P., Meinsma, R., Zoetekouw, L. & Van Gennip, A. H. Increased risk of grade IV neutropenia after administration of 5-fluorouracil due to a dihydropyrimidine dehydrogenase deficiency: high prevalence of the IVS14+1g>a mutation. *Int. J. Cancer* **101**, 253–258 (2002).

118. Boisdron-Celle, M. *et al.* Prevention of 5-fluorouracil-induced early severe toxicity by pre-therapeutic dihydropyrimidine dehydrogenase deficiency screening: Assessment of a multiparametric approach. *Semin. Oncol.* **44**, 13–23 (2017).

119. van Staveren, M. C. *et al.* Evaluation of an oral uracil loading test to identify DPD-deficient patients using a limited sampling strategy. *Br. J. Clin. Pharmacol.* **81**, 553–561 (2016).

120. Rosmarin, D. *et al.* A candidate gene study of capecitabine-related toxicity in colorectal cancer identifies new toxicity variants at DPYD and a putative role for ENOSF1 rather than TYMS. *Gut* **64**, 111–120 (2015).

121. Shrestha, S. *et al.* Gene-Specific Variant Classifier (DPYD-Varifier) to Identify Deleterious Alleles of Dihydropyrimidine Dehydrogenase. *Clin. Pharmacol. Ther.* **104**, 709–718 (2018).

122. Jacobs, B. A. W. *et al.* 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**, 75–82 (2016).

123. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma. Oxf. Engl.* **30**, 2114–2120 (2014).

124. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **26**, 589–595 (2010).

125. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).

126. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

127. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).

128. Koboldt, D. C., Larson, D. E. & Wilson, R. K. Using VarScan 2 for Germline Variant Calling and Somatic Mutation Detection. *Curr. Protoc. Bioinforma.* **44**, 15.4.1-17 (2013).

129. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).

130. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).

131. Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q. & Krainer, A. R. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* **31**, 3568–3571 (2003).

132. Fairbrother, W. G., Yeh, R.-F., Sharp, P. A. & Burge, C. B. Predictive identification of exonic splicing enhancers in human genes. *Science* **297**, 1007–1013 (2002).

133. Zhang, X. H.-F. & Chasin, L. A. Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev.* **18**, 1241–1250 (2004).

134. Zhang, C., Li, W.-H., Krainer, A. R. & Zhang, M. Q. RNA landscape of evolution for optimal exon and intron discrimination. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 5797–5802 (2008).

135. Goren, A. *et al.* Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers. *Mol. Cell* **22**, 769–781 (2006).

136. Sironi, M. *et al.* Silencer elements as possible inhibitors of pseudoexon splicing. *Nucleic Acids Res.* **32**, 1783–1791 (2004).

137. Wang, Z. *et al.* Systematic identification and analysis of exonic splicing silencers. *Cell* **119**, 831–845 (2004).

138. Untergasser, A. *et al.* Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* **35**, W71-74 (2007).

139. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

140. Meulendijks, D. *et al.* Pretreatment serum uracil concentration as a predictor of severe and fatal fluoropyrimidine-associated toxicity. *Br. J. Cancer* **116**, 1415–1424 (2017).

8. Publications

Related

1. Estimating the Effectiveness of DPYD Genotyping in Italian Individuals Suffering from Cancer Based on the Cost of Chemotherapy-Induced Toxicity.

   Fragoulakis V, Roncato R, Fratte CD, Ecca F, Bartsakoulia M, Innocenti F, Toffoli G, Cecchin E, Patrinos GP, Mitropoulou C.
   Am J Hum Genet. 2019 Jun 6;104(6):1158-1168. doi: 10.1016/j.ajhg.2019.04.017. Epub 2019 May 30.

2. The Genotype for DPYD Risk Variants in Patients With Colorectal Cancer and the Related Toxicity Management Costs in Clinical Practice.

   Toffoli G, Innocenti F, Polesel J, De Mattia E, Sartor F, Dalle Fratte C, Ecca F, Dreussi E, Palazzari E, Guardascione M, Buonadonna A, Foltran L, Garziera M, Bignucolo A, Nobili S, Mini E, Favaretto A, Berretta M, D'Andrea M, De Paoli A, Roncato R, Cecchin E.
   Clin Pharmacol Ther. 2019 Apr;105(4):994-1002. doi: 10.1002/cpt.1257. Epub 2018 Nov 22.

3. Host genetic profiling to increase drug safety in colorectal cancer from cases to implementation.

   Cecchin E, De Mattia E, Ecca F, Toffoli G.
   Drug Resist Updat. 2018 Jul;39:18-40. doi: 10.1016/j.drup.2018.07.001. Epub 2018 Jul 10. Review.

4. DPYD Gene Activity Score Predicts Dose-Limiting Toxicity in Fluoropyrimidine-Treated Colorectal Cancer.

   Dalle Fratte C, Polesel J, Roncato R, De Mattia E, Ecca F, Bignucolo A, Garziera M, Dreussi E, Palazzari E, Buonadonna A, Guardascione M, Berretta M, Foltran L, Sartor F, D'Andrea M, Favaretto A, Mini E, Nobili S, De Paoli A, Toffoli G., Cecchin E.
   J. Mol. Clin. Med. 1, 143 (2018).DOI: 10.31083/j.jmcm.2018.03.003

5. The use of pharmacogenetics to increase the safety of colorectal cancer patients treatedwith fluoropyrimidines.

   E. De Mattia, R. Roncato, C. Dalle Fratte, F. Ecca, G. Toffoli, E. Cecchin
   Cancer Drug Resist. (2019). doi:10.20517/cdr.2019.04

Unrelated

6. Clonal Evolution of *TP53* c.375+1G>A Mutation in Pre- and Post- Neo-Adjuvant Chemotherapy (NACT) Tumor Samples in High-Grade Serous Ovarian Cancer (HGSOC).

   Garziera M, Cecchin E, Giorda G, Sorio R, Scalone S, Mattia E, Roncato R, Gagno S, Poletto E, Romanato L, Ecca F, Canzonieri V, Toffoli G.
   Cells. 2019 Oct 1;8(10). pii: E1186. doi: 10.3390/cells8101186.

7. Virtual screening identifies a PIN1 inhibitor with possible antiovarian cancer effects.

   Russo Spena C, De Stefano L, Poli G, Granchi C, El Boustani M, Ecca F, Grassi G, Grassi M, Canzonieri V, Giordano A, Tuccinardi T, Caligiuri I, Rizzolio F.
   J Cell Physiol. 2019 Jan 29. doi: 10.1002/jcp.28224. [Epub ahead of print]

8. Immunogenetics of prostate cancer: a still unexplored field of study.

   Dreussi E, Ecca F, Scarabel L, Gagno S, Toffoli G.
   Pharmacogenomics. 2018 Feb;19(3):263-283. doi: 10.2217/pgs-2017-0163. Epub 2018 Jan 12. Review.