Doctoral Program in

Biomedical Sciences and Biotechnology

Cycle XXXIV

PhD Thesis

Sex-specific DPYD pharmacogenomic markers of the fluoropyrimidines treatment safety and efficacy in colorectal cancer patients

PhD student: Silvia Mezzalira

Supervisor: Dr. Erika Cecchin

Co-supervisor: Dr. Elena De Mattia

External reviewer: Dr. Barbara Gatto

2022

This doctoral thesis has been carried out at the

Experimental and Clinical Pharmacology Unit

of the National Cancer Institute (IRCCS - Centro di Riferimento Onclogico di Aviano)

directed by Dr. Giuseppe Toffoli

To my family, to my love

Table of Contents

List of Figures

List of Tables

Abbreviations

Abstract

1 INTRODUCTION 1

1.1 FLUOROPYRIMIDINES, FP, AND PATHWAY 1

1.1.1 Fluoropyrimidine Toxicity 5

1.2 INTERINDIVIDUAL VARIABILITY 5

1.3 PERSONALIZED MEDICINE 6

1.3.1 Pharmacogenetics and Its Implementation 7

1.4 FLUOROPYRIMIDINE AND GENETICS 9

1.4.1 DPYD Variants with PGx Guidelines 11

1.4.2 New DPYD Variants 15

1.4.3 Haplotype Analysis 16

1.5 SEX DIFFERENCES AND SEX MEDICINE 17

1.5.1 Fluoropyrimidine clinical outcome and the role of patients sex 19

1.6 TUMOR BIOMARKER EXPRESSION 20

1.7 COLORECTAL CANCER, CRC 22

1.8 RECTAL CANCER 23

1.8.1 Rectal Cancer Staging 26

1.9 LOCALLY ADVANCED RECTAL CANCER, LARC 27

1.9.1 Treatment Response Assessment 30

2 RATIONAL 33

3 AIMS 35

4 MATERIALS AND METHODS 37

4.1 PATIENTS' SELECTION AND STUDY DESIGN 37

4.2 INCLUSION CRITERIA FOR THE CASE STUDY POPULATION 39

4.2.1 Fluoropyrimidine-based treatment patients' selection for DPYD selected genotype analysis (Case Study 1) 39

4.2.2 LARC patient's selection for NGS analysis (Case Study 2) 39

4.2.3 LARC patient's selection for IHC analysis (Case Study 3) 40

4.3 PATIENTS DATA COLLECTION AND DATABASE PRODUCTION 40
4.4 ASSESSMENT OF THE TOXICITY GRADE TO TREATMENT 41
4.5 ASSESSMENT OF THE PATHOLOGICAL RESPONSE TO TREATMENT 41
4.6 BLOOD PROCESSING 42
4.7 GENOMIC DNA EXTRACTION AND QUANTIFICATION 44
4.7.1 Automated extraction method 45
4.7.2 Manual extraction method 47
4.8 DNA SAMPLE QUANTIZATION 47
4.9 DNA PURIFICATION 49
4.10 GENETIC ANALYSIS 51
4.10.1 Polymerase Chain Reaction (PCR) 51
4.10.1.1 PCR evaluation 53
4.10.2 Pyrosequencing 54
4.10.3 Real Time PCR (RT PCR) for 3 new variants of DPYD gene 58
4.10.4 NGS Analysis 62
4.10.4.1 Genes Selection and Customized Panel Design 65
4.10.4.2 Coverage setting and instrument/reagents selection 65
4.10.4.3 Library preparation 66
4.10.4.4 Library Normalization and library sequencing in Miseq instrument 71
4.10.4.5 Experimental set-up 74
4.10.4.6 Data analysis 74
4.10.5 Haplotype analysis 75
4.11 IMMUNOHISTOCHEMISTRY (IHC) ANALYSIS 76
4.11.1 Biomarkers selection 78
4.11.2 Protein expression evaluation 79
4.12 STATISTICAL METHODS 80

5 RESULTS 82

5.1 CASE STUDY 1- COLORECTAL CANCER PATIENTS 82
5.1.1 Patients' Characteristics and toxicity assessment 82
5.1.2 Genotyping analysis 84
*5.1.3 Impact Of Clinically Validated With DPYD rs3918290, *2A; rs55886062, *13; rs67373798, c.2846A>T, rs56038477 c. 1236 HapB3 variants on FP- associated toxicity 85*
5.1.4 Impact of new un-investigated DPYD rs59353118 and rs4294451 variants on FP-associated toxicity 86

5.1.4.1 Patients' Characteristics and toxicity assessment 86
5.1.5 Sex disaggregated data on the impact of DPYD variants on FP toxicity 91
5.1.5.1 Impact of Of Clinically Validated With DPYD rs3918290, *2A; rs55886062, *13; rs67373798, c.2846A>T, rs56038477 HapB3 variants on toxicity : sex disaggregated analysis 93
5.1.5.2 Impact of new un-investigated DPYD variants (rs59353118, rs4294451) : sex disaggregated analysis 95
5.2 CASE STUDY 2: LOCALLY ADVANCED RECTAL CANCER PATIENTS 98
5.2.1 Patients' Characteristics 98
5.2.2 NGS analysis in Locally Advanced rectal Cancer patients 98
5.2.2.1 NGS analysis set-up 98
5.2.2.2 Quality analysis of the NGS experiments 102
5.2.3 Molecular Characterization of LARC patients DNA: Haplotype generation 104
5.2.4 LARC patients' selection for FP-based therapy efficacy evaluation 106
5.2.5 Assesment of the DPYD rs59353118 and rs4294451 variants on therapy efficacy: overall population and sex-disaggregated data 108
5.2.6 Assessment of Response to nCRT in LARC patients: Analysis of DPYD haplotype and sex impact 110
5.2.7 Assessment of Safety profile of nCRT in LARC patients: Analysis of DPYD haplotype and sex impact 114
5.3 CASE STUDY 3: EFFECT OF TUMOR BIOMARKERS IN PRE-TREATMENT TUMOR TISSUE EVALUATING IN IMMUNOHISTOCHEMISTRYOF LARC ON TREATMENT EFFICACY 122
5.3.1 LARC patients' selection for FP-based therapy response evaluation of tumor biomarkers 122
5.3.2 Assessemnt of nCRT response in case study 3 LARC patients 124
5.3.3 Impact of the IHC biomarkers on nCRT response in case study 3 LARC patients 125
5.3.3.1 Analysis of IHC biomarkers between TRG1 vs TRG2-4 patients 126
5.3.3.2 Sex-disaggregated analysis of IHC biomarkers between TRG1 vs TRG2-4 patients 127
5.3.3.3 Analysis of IHC biomarkers between TRG1-2 vs TRG3-4 patients 131
5.3.3.4 Sex-disaggregated analysis of IHC biomarkers between TRG1-2 vs TRG3-4 patients 131
6 DISCUSSION AND CONCLUSION 133

6.1 DPYD New Un-Investigated Variants Analysis In CRC Patients To Predict nCRT Toxicity Risk 136
6.2 *DPYD* un-investigated variants and Haplotype Analysis In LARC Patients To Predict nCRT Outcome And Toxicity Risk 139
6.3 Tumor Biomarkers IHC Evaluation in LARC Patients to Predict nCRT Outcome 143
6.4 Conclusions and Outlooks 148
7 REFERENCES 151
8 PUBBLICATIONS 178
Acknowledgment 182

List of Figures

Figure 1.1 The chemical structure of capecitabine (left) and fluorouracil (right).
Figure 1.2 Capecitabine intestine assorbtion, liver metabolism and convertion into active metabolite FU.
Figure 1.3 Phases of the cell cycle.
Figure 1.4 Fluoropyrimidine Pathway, Pharmacodynamics (www.pharmgkb.org)
Figure 1.5 Fluoropyrimidine Pathway, Pharmacokinetics (www.pharmgkb.org)
Figure 1.6 LocusZoom plots of cis-eQTLs (expression quantitative trait loci) in *DPYD* rs59353118 (located in *DPYD* intron 14).
Figure 1.7 Rectal cancer diagnosis. DRE, digital rectal examination.
Figure 1.8 visualization of the surgery site of Totam Mesorectal Excision, TNM.
Figure 1.9 Temporal evolution of therapeutic approaches in locally advanced rectal cancer
Figure 2.1 PhD project: study design.
Figure 4.2 EDTA blood collection tube.Figure 4.3 Blood sollection tube (without EDTA) and STRECK tube.
Figure 4.4 BioRobot® EZ1 system workstation (Qiagen N.V., Germany) and preparation of DNA extraxtion using EZ1 DNA Blood 200 µL Kit (48).
Figure 4.5 Phases of DNA automated extraction with magnetic beads.
Figure 4.6 NanoDrop 2000 Spectrophotometer (Thermo Fischer Scientific, USA).
Figure 4.7 Quantus Fluorometer instrument (Promega).
Figure 4.8 Pyrophosphate (PPi) release reaction after each dNTP incorporation.
Figure 4.9 Chemical reaction for light signal production.
Figure 4.10 Example of a Pyrogram.
Figure 4.11 Example of conditions with TaqMan probe and primer paired to DNA.
Figure 4.12 Detachment of the reporter from TaqMan probe during PCR.
Figure 4.13 Genotyping cluster plot using Real-Time technology.
Figure 4.14 Example of an amplified HyperPlus sample library analyzed using the Agilent 4200 TapeStation and High Sensitivity DNA assay.
Figure 4.15 Miseq Reagent Kit, Illumina, with cartridge, HT1 buffer, Incorporation Buffer and Flow Cell.
Figure 4.16 IHC reaction steps: Indirect staining methodology.
Figure 4.17 IHC reaction steps: Direct staining methodology.
Figure 5.1 Cumulative incidence of toxicity according to rs4294451 A>T polymorphism.

Figure 5.2 Odds ratio (OR) and corresponding 95% confidence interval (CI) for toxicity according to mutation in women (red) and men (blue).
Figure 5.3 Fragmentation profile at Tape-station of the sample used for the time Set-up.
Figure 5.4 Fragmentation profile at Tape-station of one of the training run sample.
Figure 5.5 Tape-Station profile of the final loading pool for the trsining run.
Figure 5.6 Improvement of the coverage per sample and per gene between the training run and the first run.
Figure 5.7 Distrubution of 95 LARC patients according to sex.
Figure 5.8 Scatter plot of CXCR4, COX2, RAD51 and CXCL12 tumor biomarkers in the totalstudy population and by sex.

List of Tables

Table 1.1 Gene Activity Score, GAS, level, description and example of *DPYD* allele
Table 1.2 Activity Score of the *DPYD* rs3918290, *2A; rs55886062, *13; rs67373798, c.2846A>T and rs56038477, c.1236G>A/HapB3 variants.
Table 1.3 Clinical Pharmacogenetics Implementation Consortium (CPIC) Dosing recommendation of Fluoropyrimidine therapy by genotype/phenotype.
Table 1.4 Royal Dutch Association for the Advancement of Pharmacy - Pharmacogenetics Working Group (DPWG) dosing recommendation of Fluoropyrimidine therapy by genotype/phenotype.
Table 1.5 Rectal cancer staging based on Tumor, Node and Metastasi, TNM, value. Definition of the TNM value.
Table 1.6 Definition of tumor regression grading, TRG, systems
Table 4.1 PCR condition
Table 4.2 Sequences of forward and reverse primer for *DPYD* *2A, *13, c.2846A>T and c.1236C>T for PCR..
Table 4 1 Reagents conditions for *DPYD* *2A, *13, c.2846A>T and c.1236C>T PCR.
Table 4.2 Sequences of forward and reverse primer for rs59353118, rs114170368 and rs4294451 for Real-Time analysis.
Table 4.3 Reagents conditions for rs59353118, rs114170368 and rs4294451 for Real-Time analysis.
Table 4.4 Fragmentation reaction condition.
Table 4.5 End Repair & A-Tailing reaction condition.
Table 4.6 Ligation reaction condition.
Table 4.7 Adapter ligated libraries amplification conditions.
Table 4.8 Hybridization reaction condition.
Table 4.9 Enriched multiplex DNA amplification reaction condition.
Table 4.10 List of the antibodies used for IHC of selected biomarkers in tumor biopsy.
Table 5.1 Socio-demographic and clinical characteristic of the entire case study 1 population of 689 CRC patients.
Table 5.2 Distribution of acute and cumulative toxicity among total patients of case study 1 (689 patients) treated with FP-based chemotherapy.
Table 5.3 Odd ratio (OR) and corresponding confidence intervals (CI) for Acute G3-4 toxicity according to the presence of at least 1 *DPYD* polymorphisms.
Table 5.4 Odd ratio (OR) and corresponding confidence intervals (CI) for Cumulative G3-4 toxicity according to the presence of at least 1 *DPYD* polymorphisms.

Table 5.5 Socio-demographic and clinical characteristic of selected 645 CRC patients.
Table 5.6 Distribution of late toxicity among 645 CRC patients treated with FP-based chemotherapy.
Table 5.7 Odd ratio (OR) and corresponding confidence intervals (CI) for G4 toxicity according to *DPYD* rs59353118 polymorphisms.
Table 5.8 Odd ratio (OR) and corresponding confidence intervals (CI) for G4 toxicity according to *DPYD* rs4294451 polymorphisms.
Table 5.9 Hazard ratio (HR) and corresponding confidence intervals (CI) for G4 toxicity according to *DPYD* rs593531118 and rs4294451 polymorphisms.
Table 5.10 Distribution of Acute toxicity among 645 patients treated with FP-based chemotherapy by sex.
Table 5.11 Distribution of Cumulative toxicity among 645 patients treated with FP-based chemotherapy by sex.
Table 5.12 Odd ratio (OR) and corresponding confidence intervals (CI) for G3-4 acute toxicity according to the presence of at least 1 *DPYD* polymorphisms in male patients.
Table 5.13 Odd ratio (OR) and corresponding confidence intervals (CI) for G3-4 cumulative toxicity according to the presence of at least 1 *DPYD* polymorphisms in male patients.
Table 5.14 Distribution of *DPYD* rs59353118, rs4294451among 645 patients treated with FP-based chemotherapy by sex.
Table 5.15 Hazard ratio (HR) and corresponding confidence intervals (CI) for G4 toxicity according to *DPYD* rs563531118 A>T and rs4294451 A>T polymorphisms, by sex.
Table 5.16 Odd ratio (OR) and corresponding confidence intervals (CI) for G4 acute and late toxicity according to *DPYD* polymorphisms and sex.
Table 5.17 Fragmentation reaction condition for fragmentation time Set-up.
Table 5.18 Loading condition performance of the 8 run.
Table 5.19 Characteristics of *DPYD* variants identified by NGS in the case study 2 population of 229 LARC patients.
Table 5.20 Linkage Disequilibrium (LD) between rs2297595, rs1801160 and rs1801265.
Table 5.21 List of Haplotype generated and respective allele of rs2297595, rs1801160 and rs1801265.
Table 5.22 Socio-demographic and clinical characteristic of selected 212 patients.
Table 5.23 Odd ratio (OR) and corresponding confidence intervals (CI) for TRG distribution according to sex.

Table 5.24 Distribution of rs59353118 and rs4294451 in 212 LARC patients according to sex.
Table 5.25 Odd ratio (OR) and corresponding confidence intervals (CI) for TRG according to *DPYD* polymorphisms by sex.
Table 5.26 Distribution of *DPYD* haplotypes among 212 LARC patients according to sex.
Table 5.27 Haplotype correlation with TRG in 212 LARC patients.
Table 5.28 Haplotype correlation with TRG in 212 LARC patients by sex.
Table 5.29 Odd ratio (OR) and corresponding confidence intervals (CI) for TRG according to *DPYD* haplotypes by sex.
Table 5.30Socio-demographic and clinical characteristic of selected 189 LARC patients.
Table 5.31 Distribution of late toxicity among 189 LARC patients treated with FP-based chemotherapy.
Table 5.32 Distribution of toxicity among 189 LARC patients treated with FP-based chemotherapy.
Table 5.33 Distribution of *DPYD* haplotypes among 189 LARC patients treated with FP-based chemotherapy.
Table 5.34 Distribution of toxicity among 189 patients treated with FP-based chemotherapy according to *DPYD* haplotypes.
Table 5.35 Distribution of toxicity among 189 patients treated with FP-based chemotherapy according to *DPYD* haplotypes by sex.
Table 5.36 Odd ratio (OR) and corresponding confidence intervals (CI) for G3-4 toxicity according to *DPYD* haplotypes in the total 189 LARC study population and by sex.
Table 5.37 Socio-demographic and clinical characteristic of selected 95 LARC patients.
Table 5.38 Treatment characteristic of 95 LARC patients.
Table 5.39 Treatment outcome of 95 LARC patients.
Table 5.40 Distribution of treatment outcome of 95 LARC patients by sex.
Table 5.41 Evaluable biopsies for selected biomarkers in 95 LARC patients and by sex.
Table 5.42 Median value and interquartile range (Q1-Q3) of H-score and neoplastic cellularity for selected parameters according to TRG status comparing TRG1 vs TRG2-5.
Table 5.43 Median value and interquartile range (Q1-Q3) of H-score and neoplastic cellularity for selected parameters according to TRG status in male patients comparing TRG1 vs TRG2-5.
Table 5.44 Median value and interquartile range (Q1-Q3) of H-score and neoplastic cellularity for selected parameters according to TRG status in female patients comparing TRG1 vs TRG2-5.
Table 5.45 Median value and interquartile range (Q1-Q3) of H-score and neoplastic cellularity for selected parameters according to TRG status comparing TRG1-2 vs TRG3-5.

Table 5.46 Median value and interquartile range (Q1-Q3) of H-score and neoplastic cellularity for selected parameters according to TRG status comparing TRG1-2 vs TRG3-5 by sex.

Abbreviations

5'-dFCR: 5'-deoxy-5-fluorocytidine

5-FU: 5-Fluorouracil

5FUH2: 5-dihydrofluorouracil

ADE: Adverse Drug Event

ADME: Absorption, Distribution, Metabolism, And Excretion

AE: Adverse Event

AIFA: Agenzia Italiana del Farmaco

AIOM-SIF: Associazione Italiana Di Oncologia Medica-Società Italiana Di Farmacologia

AJCC: American Joint Committee on Cancer

APS: adenosine 5' phosphosulfate

ATP: Adenosine Triphosphate

bp: base pair

BV: Bevacizumab

CR: Clinical Response

cCR: Complete Clinical Response

CEA: Carcinoembryonic Antigen

CES: Carboxylesterase

CI: Confidence Interval

CNV: Copy Number Variations

COX-2: Cyclooxygenase-2

CPIC: Clinical Pharmacogenetic Implementation Consortium

CPNDS: Canadian Pharmacogenomics Network for Drug Safety

CR: Complete Response

CRC: Colorectal Cancer

CRM: Circumferential Resection Margin

CRO: Centro di Riferimento Oncologico

CRT: Chemoradiotherapy

CT: Computed Tomography

CTC-AE: Common Terminology Criteria For Adverse Events

CTX: Cetuximab

CXCR4: C-X-C Chemokine Receptor type 4

CyD: Cytidine Deaminase

CYP: Cytochrome

DFS: Disease Free Survival

dNTP: Deoxynucleotide Triphosphate

DPWG: Royal Dutch Association for the Advancement of Pharmacy - Pharmacogenetics Working Group

DPYD: Dihydropyridine Dehydrogenase

DPYS: Dihydropyrimidinase

dsDNA: Double Strand DNA

dThdPase: Thymidine Phosphorylase

EDTA: Ethylenediaminetetraacetic Acid

EM: Extensive Metabolizer

EMA: European Medicine Agency

eQTL: Expression Quantitative Trait locus

ESMO: European Society for Medical Oncology

FBAL: Fluoro-beta-alanine

FDA: Food and Drug Administration

FdUMP: Fluorodeoxyuridine Monophosphate

FdUTP: Fluorodeoxyuridine Triphosphate

FFPE: Formalin Fixed Paraffin Embedded

FP: Fluoropyrimidine

FUH2: 5-fluoro-5,6-dihydro-fluorouracil

FUPA: fluoro-beta-ureidopropionate

FUTP: Fluorouridine Triphosphate

GAS: Gene Activity Score

H&F syndrome: Hand And Foot Syndrome

HR: Hazard Ratio

H-score: histo-score

IHC: Immunohistochemistry

IM: Intermediate Metabolizer

KRAS: Kirsten Rat Sarcoma Viral Oncogene Homolog

LARC: Locally Advanced Rectal Cancer

LD: Linkage Disequilibrium

MMR: Mismatch Repair

MRI: Magnetic Resonance Imaging

NGS: Next-Generation Sequencing

NIH: National Institutes of Health

NM: Normal Metabolizer

NTC: No Template Control

OR: Odds Ratio

OS: Overall Survival

P: p-value

pCR: Pathological Complete Response

PCR: Polymerase Chain Reaction

PET: Positron Emission Tomography

PFS: Progression Free Survival

PGx: Pharmacogenetics

PM: Poor Metabolizer

PolyPhen: Polymorphism Phenotyping

PPi: Pyrophosphate

RCF: Relative Centrifugal Force

RFS: Relapse Free Survival

RT: Room Temperature

SAM: Sequence Alignment/Map

SBS: Sequencing by Synthesis

SCPRT: Preoperative Short-Course Radiotherapy

SCRT: Short-Course Radiotherapy

SIFT: Sorting Intolerant From Tolerant

SNP: Single Nucleotide Polymorphism

SV: Structural Variants

TDM: Therapeutic Drug Monitoring

TME: Total Mesorectal Excision

TNM; Tumor, Node, and Metastasis

TNT: Total Neoadjuvant Therapy

TP: Thymidine Phosphorylase

TRG: Tumor Regression Grade

Tris-HCl: Trisaminomethane Hydrochloride

TYMS: Thymidylate Synthetase

UDI: Unique Dual-Indexed

UICC: Union for International Cancer Control

UM: Ultrarapid Metabolizer

UMI: Unique Molecular Indexes

UPB1: Beta-Ureidopropionase

UTR: Untranslated Region

VEGFR: Vascular Endothelial Growth Factor

Abstract

Background: Fluoropyrimidines (FP), including 5-FU and capecitabine, are antiblastic drugs widely used for over 30 years in the treatment of many solid tumors, including colorectal cancer (CRC). Despite a good tolerability profile, severe and, in rare cases, fatal FP-related toxicities occur in 10-40% of patients, with large interindividual differences. Dihydropyrimidine dehydrogenase (DPD), the rate-limiting enzyme in the metabolism of FP, is encoded by *DPYD*, a highly polymorphic gene with four known variants (rs3918290 (*DPYD*2A*), rs55886062 (*DPYD*13*), rs67376798 (*DPYDc.2846A >T*), rs75017182 (*DPYD HapB3*)) associated with impaired enzyme function and a higher incidence of severe and life-threatening toxicity. Evidence-based dosing guidelines exist for adjusting therapy dosage in patients carrying these mutations. Nevertheless, a high percentage of severe adverse drug reactions (ADRs) may occur regardless the presence of those variants. The identification of additional *DPYD* single nucleotide polymorphisms, SNP, causative of toxicity could improve the safety of treatment with FP. Recently a paper was published highlighting three new expression quantitative traits loci (rs59353118, rs114170368, and rs4294451) in *DPYD* by an *in silico* analysis, but no clinical validation of their role has been reported up-to-date. Furthermore, 5- FU is a prominent example of a drug with significant sex-specific interindividual variation in both drug clearance and exposure, with a 26% higher exposure in women possibly related to a different outcome.

Beside FP-related toxicity, the efficacy of FP could also be affected by both genetic and sex variables. In the context of FP-based neo-adjuvant treatment of Locally Advanced Rectal Cancer (LARC) patients, where the tumor response to FP is strongly associated with prognosis and is used to adjust the therapeutic strategy, predictive biomarkers could be very helpful.

The aim of this study was to investigate the effect of new patients' genetic variants in *DPYD*, as well as the influence of patients' sex on clinical outcome of a FP-based therapy in selected CRC patients. Toxicity was evaluated in a group of CRC patients at different tumor stages, whereas tumor response was evaluated in a subgroup of clinically homogeneous LARC patients treated with FP -based chemotherapy.

Materials and Methods: A retrospective case study of 689 CRC patients with available biological material and clinical data was selected in the Experimental and Clinical Pharmacology Unit of Centro di Riferimento Oncologico (CRO) IRCCS in Aviano (Italy). The germline DNA was analyzed for the presence of genetic variants in the *DPYD* gene: four mutations were analyzed by pyrosequencing (rs3918290 (*DPYD*2A*), rs55886062 (*DPYD*13*), rs67376798 (*DPYDc.2846A >T*), rs75017182 (*DPYD HapB3*)), while three newly discovered SNPs (rs59353118, rs114170368, and rs4294451) were analyzed by fluorescence-based allelic discrimination technique. In a subgroup of 229 LARC patients treated with nCRT (monotherapy with capecitabine plus concomitant radiotherapy) a custom NGS panel of 106 genes associated with both FP metabolism and LARC pathologic features was applied. NGS data from *DPYD* sequencing were used to develop *DPYD* haplotypes. In a subset of those LARC patients with available pre-treatment tissue biopsy, immunohistochemistry, IHC, analyzes were performed to evaluate expression level of protein biomarkers potentially related to tumor progression. Patient demographic together with, clinical, therapeutic, toxicity, and response data were collected where not available. The Common Terminology Criteria of Adverse Events (CTC-AE) version 5.0 was used to assess the degree of toxicity in the overall study population. In

addition, limited to LARC cohort, response to nCRT was assessed using the Mandard's tumor regression grade (TRG) scale.

Results: In a final retrospective case study of 689 patients the association between *DPYD*2A*, *DPYD*13*, *DPYD c.2846A>T and DPYD HapB3* and risk of severe toxicity was investigated, confirming that the presence of at least one of these variants significantly increased the risk for both haematologic and non-haematologic grade 3-4 toxicity. New *DPYD* variants (rs59353118, rs114170368, and rs4294451) were analysed excluding patients (n=44) that were carriers of one of the four previously mentioned validated variants. In the final case study population encompassing 645 CRC patients, the *DPYD* rs593531118 variant exposed patients to a higher risk of developing grade 4 haematologic and non-haematologic toxicity (OR=1.92, 95% CI:1.07-3.42), whereas the *DPYD* rs4294451 variant allele was associated with a reduced risk of grade 4 haematologic/non-haematologic and all toxicities (OR=0.50, 95% CI:0.26-0.96; OR=0.49, 95% CI:0.25-0.94, respectively). In addition, when considering the time to toxicity occurrence, a protective effect of rs4294451 variant was highlighted according to an additive model for the development of risk for haematologic, haematologic & non-haematologic, and all toxicities (HR=0.28, 95% CI:0.09-0.83; HR=0.40 95%, CI:0.17-0.93; HR=0.40, 95% CI:0.18-0.90, Fine-Grey test, p-value=0.0444, respectively), suggesting an increased likelihood of developing early toxicity in patients harbouring the variant allele compared to wild-type patients.

Women resulted more exposed than men to grade 4 non haematological and all toxicities.

When looking at the interaction between *DPYD* new variants and sex, no significant difference in toxicity was observed for the two variants (rs59353118 and rs4294451), although a trend toward a higher protective effect of variant rs4294451 in female than in male patients and a slightly increased risk of developing toxicity for the male population carrying variant rs59353118 was observed.

In 189 LARC patient the effect of *DPYD* haplotypes on the risk of toxicity was investigated. Six haplotypes were identified comprising 3 variants (2194 G > A (*6), c.85 T > C (*9A), and c.496 A > G) that resulted in high linkage disequilibrium, LD. A correlation of the *DPYD* haplotype HAP2 with the FP-related toxicity was observed with a significant protective effect against G3-4 non-haematologic toxicity risk (p-value=0.04). The presence of 1 minor allele of HAP2 was significantly correlated with 8.7-fold higher risk (95% CI:1.45-51.92, p-value=0.008) and 4.08-fold higher risk (95% CI:0.96-17.26, p-value=0.028) of developing G3-4 "other" and "all" toxicity, respectively. Regarding the sex effect, the same correlation was maintained only in the female case population.

The clinically homogeneous group of 212 LARC patients was also used as a study group for *DPYD* markers of treatment response (efficacy). No statistical difference in therapy response in term of TRG emerged across the two sexes while, differently, a slight trend toward a better clinical therapy response in female population have been reported considering pathological complete response, pCR, (TRG1) within TRG2. A significant association was observed for rs59353118 and rs4294451: patients carrying rs4294451 variant tended to present a worse response than wild-type subjects (OR= 21.084, 95% CI:1.0929- 4.0677, p-value=0.0261) while rs59353118 carriers showed a better pathological response (p-value = 0.04, Chi Square test). These data are consistent with what observed in toxicity analysis. No differences emerged on the impact of this variants in the separate assessment by sex. Regarding *DPYD* haplotype analysis, HAP1was associated with a higher risk of TRG 2-5, but this effect was limited to the male population (OR: 2.40, 95 CI:1.109-5.2 p-value=0.013). In addition,

for 95 LARC patients for whom a pre-treatment biopsy sample was available, 3 markers of treatment efficacy were found in the IHC study: Ki67, CXCR4, COX-2 (p-value=0.059; p-value=0.010 and p-value=0.030). Separate analysis by sex showed that the association remained significant only among females.

Conclusions: Newly identified exonic and intronic *DPYD* variants and the corresponding genetic haplotypes may improve the identification of patients at high risk of developing toxicity or treatment failure in CRC patients treated with FP-based treatment. In addition, the evaluation of specific protein markers in pre-treatment tumor tissue biopsy provides further important information on LARC patients clinical outcome. The contribution of sex as a key variable in the safety and toxicity profile of FP proved to be very important and to provide new insights in the predictive effect of the pharmacogenetic markers. Further investigation on the field is warranted and that sex together with *DPYD* variants must be considered in the clinical process of therapeutic strategy selection in CRC.

1 INTRODUCTION

1.1 FLUOROPYRIMIDINES, FP, AND PATHWAY

Fluoropyrimidines (FP) are anticancer agents that belong to the class of drugs known as antimetabolites [1]. These are substances that exert a toxic effect at the cellular level by stopping DNA synthesis and thus cell proliferation [2]. Their toxic effect interferes with the metabolism of both normal and neoplastic cells. Expeccially in cancer therapy, FP action slows down the growth of tumor tissue. Due to their mechanism of action, the class of antimetabolites is widely used in antineoplastic and antibiotic therapy and in diseases where antivitamin, glucose and amino acid antagonists are administered [3].

FP are a group of antiblastic drugs primarily used to treat many solid tumors, including cancers of gastrointestinal tumors, breast and cancers of the head and neck [4–6]. FP has been widely used for more than 30 years both in early and advanced tumor stages and can be administered as monotherapy or in combination with other antiblastic agents such as platinum derivatives and/or irinotecan or docetaxel [7]. In colorectal cancer (CRC) therapy, FP are used almost exclusively in combination with oxaliplatin and/or irinotecan or radiotherapy, whereas for head and neck cancer they are used in combination with cisplatin and radiotherapy [8–11]. Recently, the combination of FP with biologica agents (monoclonal antibodies) such as cetuximab (CTX) and bevacizumab (BV) has been intoduced in the first-line treatment of metastatic colon and rectal cancer [12,13]. In certain clinical setting, FP-based chemotherapy is combined with concomitant radiotherapy [14].

The most used FPs in oncology are capecitabine and 5-fluorouracil (5- FU) (Figure 1.1) while tegafur is still today removed from clinical protocols.

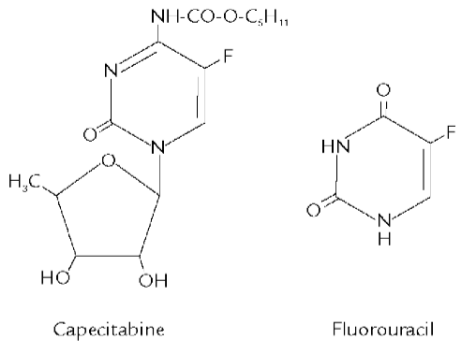


Figure 1.11 The chemical structure of capecitabine (left) and 5-fluorouracil (right).

Capecitabine is a tumor-activated antineoplastic agent developed as a novel oral FP that selectively generates 5-FU in tumor cells [15]. Capecitabine is administered orally in tablet form as a prodrug, absorbed through the intestine, and selectively activated in tumor tissue by thymidine phosphorylase (dThdPase) to form the cytotoxic moiety, 5- FU.

It is absorbed unchanged in the gastrointestinal tract and metabolised mainly in the liver. The duration of absorption could likely be altered by food intake, reducing both the rate and the extent of absorption although the clinical significance is still unclear. Based on efficacy and safety data, it is recommended that capecitabine be taken within 30 minutes of eating.

Since dThdPase is present in a greater amount in tumor cells than in normal tissues, capecitabine was developed to overcome problems associated with off-target activity and to improve the safety profile of these chemotherapeutic agents. The active drug derives from three enzymatic steps with two intermediate metabolites [16]. The first step is mediated by hepatic carboxylesterase (CES), which produces 5'-deoxy-5-fluorocytidine (5'-dFCR). The 5'-dFCR is then converted to 5'-dFUR by the enzyme cytidine deaminase (CyD), and finally dThdPase mediates the final transition of 5'-dFUR to 5-FU (figure 1.2). Plasma data indicate extensive and rapid conversion of the first two steps to 5'-dFCR and 5'-dFUR, with peak plasma concentrations reached approximately 1.5 to 2.0 hours after capecitabine administration. Binding of capecitabine and its metabolites to plasma proteins is greater than 60%, primarly to human albumin (35%). Concentrations decrease exponentially with half-lives of 0.85 h, 11 h, and 0.66 h for 5-FU, 5'-dFCR, and 5'-dFUR, respectively. Over 70% of the administered capecitabine dose is excreted primarily in the urine, with an average of 84% excreted within the first 24 h and 96% excreted over 7 days as unaltered drug or metabolites.

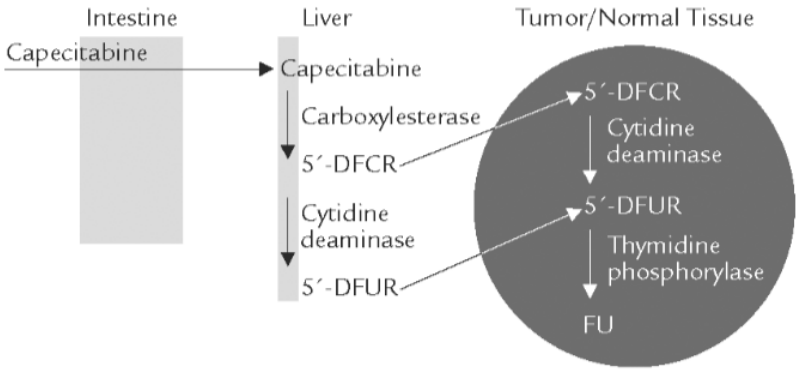


Figure 1.12 Capecitabine intestine assorbtion, liver metabolism and convertion into active metabolite FU.

The active product 5-FU, which can be administered directly into the bloodstream as a bolus, elastomeric or continous infusion, is a structural fluorinated pyrimidine analogue that differs from

uracil in the presence of a fluorine atom in the C-5 position, that stabilises the molecule and is responsible for its activity. The anticancer activity may vary depending on the route of administration and is exerted via:

- The inhibition of the enzyme thymidylate synthetase (TYMS),
- The integration into RNA after being ribosylated and phosphorylated,
- the inhibition of the synthesis of new RNA by blocking one of the enzymes involved in this process, the uracil phosphatise [4,17,18]..

Particularly the bolus treatment promotes RNA damage, while continuous infusion favours DNA damage.

Specifically, 5-FU, by blocking TYMS activity, completely inhibits the synthesis of thymidine, a nucleoside necessary for DNA replication. Since DNA and RNA are essential for cell division and replication, 5-FU, by inhibiting their synthesis, leads to unbalanced cell growth, resulting in the death of the cell itself. Its cytotoxic activity causes cellular damage by stopping the cell cycle at S phase, where synthesis and duplication of genetic material normally occurs (Figure 1.3). It has been reported that more than 80% of the substance is rapidly metabolised in the liver and 5-FU has a half-life of about 10-20 minutes [19].

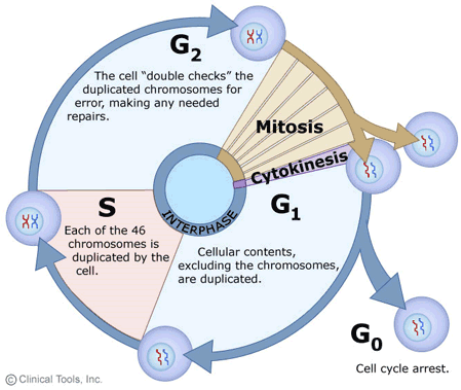


Figure 1.13 Phases of the cell cycle.

Fluorodeoxyuridine monophosphate (FdUMP), fluorodeoxyuridine triphosphate (FdUTP), and fluorouridine triphosphate (FUTP) are the three main metabolites that have pharmacodynamic effects. FdUMP forms a covalent complex with TYMS and prevents the binding and conversion of dUMP to dTMP, which is necessary for pyrimidine and DNA synthesis [20]. Simultaneously, it blocks the conversion of 5,10-methylenetetrahydrofolate to dihydrofolate, a key component of the folate pathway that recycles methyl groups and synthesizes methionine. Inhibition of TYMS leads to results in misincorporation of dUTP into DNA [21]. The FdUMP–TYMS complex is stabilized by the

simultaneous administration of folate analogs. FUTP and FdUTP are incorporated into RNA and DNA, respectively, contributing to the pharmacodynamic effects of FPs (Figure 1.4) [22].

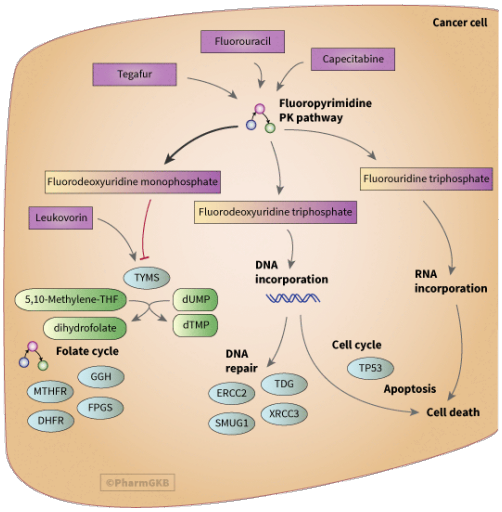


Figure 1.14 Fluoropyrimidine Pathway, Pharmacodynamics (www.pharmgkb.org)

As for the catabolic pathway, the dihydropyrimidine dehydrogenase (DPD) enzyme is responsible for the degradation of about 80-90% of 5-FU to 5-fluoro-5,6-dihydro-fluorouracil (FUH2), the excretion product which is much less toxic and inactive [23]. Subsequently, DHFU is converted by dihydropyrimidinase (DPYS) or β-ureidopropionase (UPB1) to fluoro-β-ureidopropionate (FUPA) and then to fluoro-β-alanine (FBAL), which is excreted via the lungs. Only a small portion (about 20%) is excreted via the kidneys (Figure 1.5).

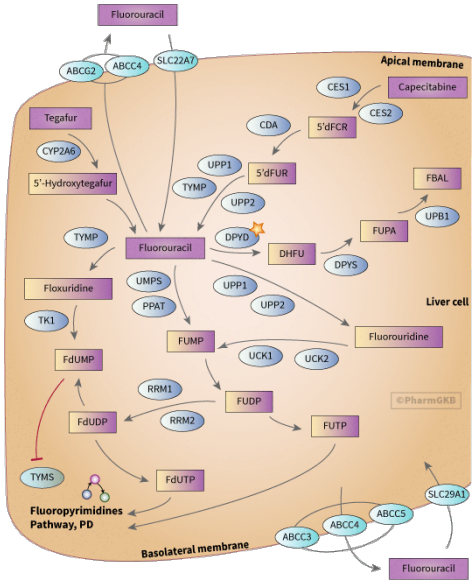


Figure 1.15 Fluoropyrimidine Pathway, Pharmacokinetics (www.pharmgkb.org).

Like capecitabine, tegafur is a prodrug that is administered orally and metabolized by the cytochrome, CYP, 1A2, CYP2A6, and CYP2C8 enzyme comp in 5-FU. As with capecitabine, the final conversion to an active metabolite is mediated by thymidine phosphorylase, TP [24].

1.1.1 Fluoropyrimidine Toxicity

Despite their good antiblastic efficacy in various treatment settings, the occurrence of severe FP-related toxicities, leading to severe morbidity or discontinuation of treatment, represents an important drawback of these drugs. Depending both on the treatment regimen and the drugs used in combination with, 10–40% of patients experience severe and, in rare cases (0.2–0.5%), even fatal FP-related toxicity in early chemotherapeutic cycles [25,26]. Due to their narrow therapeutic window, clinical 5-FU toxicity may include fever, mucositis, stomatitis, nausea, vomiting, and diarrhea as non-haematologic adverse events (AEs), and thrombocytopenia, leukopenia, and neutropenia as haematologic side effects [27]. The qualitative spectrum of toxicity of FP varies widely depending on drug dosage, schedule, and patient characteristics. Bolus infusion of 5-FU is usually associated with a higher risk of myelosuppression as the main toxicity, and to a lesser extent mucositis and diarrhea [28]. High-dose infusion has been associated with mucositis, while low-dose continuous infusion of 5-FU is associated with minor myelosuppression and palmar-plantar dysesthesia (hand-foot syndrome) [29]. 5-FU/calcium folinate therapies may cause the combination of mucositis, diarrhea, and myelosuppression. Older patients (> 70 years) are more likely to experience mucositis and neurotoxicity (myelosuppression), while female patients are statistically more likely to be affected by all 5-FU toxicities, as sex is another risk factor to consider [30]. FP is contraindicated in patients hypersensitive to capecitabine or 5-FU, in patients with severe renal insufficiency (creatinine clearance less than 30 mL/min or 0.5 mL/s) and in patients known to lack DPD activity.

1.2 INTERINDIVIDUAL VARIABILITY

It is well known that there is considerable interindividual variability in drug response and toxicity in patients treated with the same therapy. The efficacy, safety and tolerability of drugs depend on numerous factors that influence their body disposition. These factors may be extrinsic or intrinsic and may be related to the drug itself, to patient characteristics, and to drug-drug interactions resulting in subtherapeutic or toxic plasma drug concentrations [31–33].

Drug-related interindividual variability factors include dosage, dosage form, and therapeutic regimen, which are responsible for drug-drug interactions. Patient-specific factors include age, sex, diet, natural physiological cycles, pregnancy, acute illness, hepatic and renal dysfunction, and other

chronic disease processes [34]. One of the major causes of variability in drug effects is drug own metabolism. Approximately 20-30% of interindividual variability in drug response is correlated with genetic single nucleotide polymorphisms (SNPs) in genes encoding phase 1 and phase 2 enzymes, besides that diet and concomitant therapies may induce or inhibit their activity. The expression and activity of enzymes that metabolize the drug might also depend on the intrinsic factors abovementioned [35].

This interindividual variability has not been studied in the long and expensive process of drug development. The entire process, which can take up to 15 years for an anticancer drug with an average cost of about $2.6 billion, involves phase 1 studies in a large group of healthy volunteers to evaluate the drug's safety, dosage, and potential side effects, phase 2 trials to evaluate the drug's efficacy; and phase 3 trials involving larger numbers of subjects [36,37].

To adequately evaluate established clinical endpoints of efficacy and toxicity, appropriate clinical trial design is required. Homogeneity of the study sample is ensured by defining the inclusion/exclusion criteria, and by randomization, a technique commonly used in clinical trials to exclude potential bias and confounding factors [38]. Besides the aforementioned strategy to ensure proper evaluation of new therapies and avoid imbalances between groups, the well-known interindividual variables are still not considered during enrolment in clinical trials.

In particular, the lack of consideration of genetic makeup of enrolled patients, which can lead to the inclusion of patients in the clinical trial who are carriers of highly impacting mutations, emerges as a critical issue.

In addition, analysis of clinical trial quality conducted by several authors have shown that women are significantly underrepresented in pivotal Phase 1, 2 and 3 clinical trials. Indeed, the lack of sex diversity in clinical trials has long been controversial. Traditionally, trials have mostly included adult white Caucasian adult males, particularly in Phase I clinical trials investigating tolerability, clinical pharmacology, dose-related adverse effects, and early evidence of efficacy [39].

1.3 PERSONALIZED MEDICINE

Recently, the therapeutic strategy is considerably changed with a srowing attention towards interindividual differences. In this context, personalized medicine or precision medicine is a healthcare model in which medical decisions, practices, interventions, and therapies are tailored to each individual based on their predicted response or disease risk factors [40]. Patients are divided into

different groups and treated separately to achieve the best individual benefit. This new approach put an end to the traditional "trial and error" practice drug prescription and aims to provide safer, more effective, and cost-efficient drug therapy [41]. In this framework, personalized medicine is described as a therapeutic strategy that has a high potential to achieve the best response and the highest pharmacological safety margin. According to Mathur and Sutton *et al.*, it aims to provide the best care for the patient by diversifying the therapeutic approach based on the individual characteristics of both the disease and the patient [40].

Personalized medicine has become increasingly common in recent years given the growing diagnostic and informatics approaches, particularly in genetic field. Understanding the molecular basis of disease and metabolism has enabled researchers to stratify patients to select both the appropriate diagnostic tests and therapies based on clear evidence. In particular, the use of genetic information has played such a crucial role in personalized medicine that the term pharmacogenomics was first coined in the context of genetics and then evolved into a new branch of pharmacology [42].

1.3.1 Pharmacogenetics and Its Implementation

Pharmacogenetics (PGx) aims to personalize therapy based on patients' genetic make-up [43]. Genes that influence the effect of a drug are called pharmacogenes [44]. According to many authors, adjusting dosages and medications taking intoaccount genetic variants is the first true clinical application of genetics in the post-genomic era [45]. Exon DNA point mutations, substitutions, insertions, rearrangements, or deletions can result in structural changes in the encoded protein that affect its functionality while mutations in the untranslated sequences of the gene, as regulatory or splice regions, lead to a quantitative change in protein expression. In this sense, patients' genotype could provide more detailed information to help physicians make the more cost-effective and accurate treatment decisions. Specifically, somatic mutations as well as germline mutations have been considered in the personalization of tumor treatment as well as in the treatment of other common diseases.

It has been shown that the standard dose of a drug may increase the risk of overdose or underdose in carriers of certain genetic variants, to which *ad personam* dosage have been required. In fact, the presence of a variant in a gene encoding metabolic enzymes may increase or decrease its enzymatic activity, with significant implications for efficacy and/or the development of toxicity. Genetic variants could be related to i) lower or higher drug exposure, ii) increased toxic metabolite concentration, iii) altered interaction with the target drug, and iv) idiosyncratic drug toxicity due to immune activation [46].

For proteins involved in the drug Absorption, Distribution, Metabolism, and Excretion (ADME), genotype-to-phenotype translation has linked a specific SNP to the corresponding enzyme function. An estimated of 90% of the population carries at least one variant in a gene related to drug metabolism or mechanism of action, or a variant indicating a higher risk for hypersensitivity reactions [47,48].

Phenotype could be divided into the following groups:

- "Ultrarapid metabolizers" (UM) with increased enzyme activity;
- "Normal metabolizers" (NM) with a normal and weak enzyme activity;
- "Intermediate metabolizers" (IM) / “Extensive metabolizer” (EM) with an intermediate enzyme functionality;
- "Poor metabolizers" (PM), when the enzyme is completely dysfunctional [49].

It is estimated that 7% of the 1200 Food and Drug Administration (FDA) approved drugs are associated with clinically actionable variant, and these drugs account for 18% of the ~4 billion prescriptions in the United States [50–52]. Consequently, the inability to obtain appropriate therapy, even when PGx information is used, translates to the so-called "Therapeutic Odyssey" for the patient. This condition is characterized by the constant search for the most effective therapy, ineffective treatments, frequent visits, alternative treatment regimens based on the use of nutritional supplements, phytotherapeutics or polytherapeutics agents, the deterioration of the patient's condition and the treatment failure. PGx testing prior to treatment could limit this phenomenon and many of the above cases could be identified a priori to avoid side effects or therapeutic inefficacy [53].

The adoption of PGx into clinical practice depends largely on the availability of approved and validated PGx tests, the ability of clinicians to use and understand them, and evidence-based recommendations for altered therapeutic management from standard approaches. Currently, the slow clinical adoption of PGx is primarly due to barriers that delay the implementation of PGx into clinical practice. The lack of standardized PGx guidelines and their difficult interpretation by clinicians, which demonstrated resistance towards PGx information, are just some of these barriers [45,54]. To overcome the main problem of lack of clinical guidelines, several Institutions have established dedicated working groups with the aim of developing pharmacogenomics-based therapeutic dose recommendations. Nowadays, the inclusion of genetic information in the decision-making process has become possible due to the establishment of national and international guidelines. As of today, authoritative consortia worldwide aim to develop guidelines that can help clinicians to translate available genetic test results into clinical decisions. The Clinical Pharmacogenetics Implementation

Consortium (CPIC), the Royal Dutch Association for the Advancement of Pharmacy-Pharmacogenetics Working Group (DPWG), the Canadian Pharmacogenomics Network for Drug Safety (CPNDS) and other professional societies are working to facilitate the incorporation of PGx testing into clinical practice and to translate genetic information into prescribing recommendations [55–59]. PharmGKB website, an NIH (National Institutes of Health)-funded resource, collects international guidelines and provides information on how genetic variations affect drug response making this knowledge available to clinicians and researchers. To date in PharGKB are available 165 clinical guidelines, 780 drug labels, and 151 curated information trials [60].

Available guidelines describe the starting dose optimization for all drug, including anti-cancer treatment (e.g., FPs and irinotecan).

1.4 FLUOROPYRIMIDINE AND GENETICS

Given the extreme heterogeneity of outcomes in patients treated with FP, the identification of biomarkers responsible for different entities of both response and toxicity is becoming increasingly important.

Thanks to the development and improvement of genetic analysis technique, it has been possible to identify biological and genetic biomarkers associated with differential response to therapy in both prospective and retrospective studies. The improvements in the field of PGx have been useful in stratifying patients for selection of the proper therapy and dosing regimens.

Response to therapy and development of toxicity depend on two types of genomes: the germline genetic profile of the patient and the somatic profile of the cancer cell. Specifically, the former may primarily affect the pharmacokinetics (PK) and the pharmacodynamic (PD) drug pathway if germline mutations are present in genes encoding metabolic enzymes, transporters, or drug targets and directly affect plasma levels of drugs and their metabolites.

As aforementioned, DPD is the first rate-limiting enzyme in the degradation pathway of FP [61,62]. It is encoded by *DPYD* gene, located on chromosome 1p22, and includes 23 coding exons with 136,751 bases. The protein contains 519 amino acids with a molecular mass of 56,630 Da. DPD mediates over than 80% of the FP metabolism and large-scale studies have shown that variations in the enzymatic activity of the DPD result in altered and abnormal plasma levels of drugs with a variable disposition rate in treated patients [23].

Literature reports that several thousand mutations have been identified in the *DPYD*, of which more than 50 are associated with protein deficiency. Mutations in the *DPYD* gene specifically interfere with the degradation of uracil and thymine, which directly affects their variable blood and urine concentrations. In addition, altered DPD function impairs the degradation of pyrimidine-analogue drugs. As a result, the half-life of the drug increases with higher accumulation in the body leading to a higher risk of drug-induced toxicity [62–65]. DPD deficiency in childhood occurs in 3-5% of the general population and is a disorder characterized by a wide range of severity and a clinical spectrum of symptoms that varies greatly from person to person [64,66,67]. The disorder develops in childhood when recurrent seizures and mental retardation become more prominent. Neurological problems include epileptic seizures, mental retardation, microcephaly, hypertension, delayed development of motor skills, and autistic behaviours. Other features include growth retardation, unusual facial morphology, ocular malformations, and abnormal brain anatomy [68–70]. In addition, some studies have shown that DPD enzymatic activity appears to be influenced by circadian rhythms through epigenetic regulatory mechanisms as well as by sex-specific variability [71].

Any alteration in DPD activity due to mutation in the encoding gene could lead to toxic accumulation of the FP and increase the risk of AEs. Indeed plasma concentrations of 5-FU are up to 15-fold higher in patients with decreased enzyme activity than in patients with normal DPD enzyme activity [72–74].

If no variant is detected (formerly known as the *1 allele), normal enzyme activity is present. Individuals who have 2 copies of *DPYD* alleles with normal activity are referred to as NM and have fully functional DPD enzyme activity. The *DPYD* alleles c.1601G>A (*4, rs1801158), c.1627G>A (*5, rs1801159), c.2194G>A (*6, rs1801160), and c.85T>C (*9A, rs1801265) are also considered to have normal function, although conflicting results are reported in literature [75].

The Gene Activity Score (GAS) model, originally described by Henricks *et al*. in 2015, aims to translate genetic test results into the most likely phenotypic outcome [76]. The GAS predicts DPD activity and could be used to optimize the patient' starting dose therapy. The GAS ranges from 0 (no DPD activity) to 2 (normal DPD activity) [77].

The total GAS of the gene is calculated by summing the activity score of the 2 alleles, assigning an acticity value of 1 to a fully functional allele and an activity value of 0 to a non-functional allele. A 0.5 value is assigned to an allele with an intermediate functionality. The possible total number GAS comprises five groups, as described in Table 1.1.

Table 7.1 Gene Activity Score, GAS, level, description and example of *DPYD* allele.

GAS	Description	Possible DPYD genotype
0	An individual carrying two no function alleles	*2A/*2A, *13/*13
0.5	An individual carrying one no function plus one decreased function allele	1236A/1236A, 2846T/2846T
1	An individual carrying one normal function allele plus one no function allele or an individual carrying two decreased function alleles	*1/*2A or *1/*13
1.5	An individual carrying one normal function allele plus one decreased function allele	*1/1236A or *1/2846T
2	An individual carrying two normal alleles	*1/*1

Based on the GAS, the patient's phenotypes are divided into *DPYD* NM (GAS=2), *DPYD* IM (GAS=1 or 1.5), and *DPYD* PM (GAS=0.5 or 0).

Prior to treatment, DPD testing must be considered to avoid the risk of developing toxicities. Based on the GAS, the available international PGx guidelines recommend standard FP dose reduction - as detailed in the following sections -, and avoiding the use of FP in the case of near absence of DPD activity.

1.4.1 *DPYD* Variants with PGx Guidelines

Many variants in *DPYD* gene have been described and the association between the presence of genetic polymorphism and alterations in DPD protein functionality is well documented, although 98% of these genetic alterations are found in uncoded regions of the gene.

To date, four *DPYD* SNPs have been found to be associated with significant changes in DPD protein activity and are present in nearly 7% of Europeans: rs3918290, *2A; rs55886062, *13; rs67373798, c.2846A>T and rs56038477 which characterizes haplotype B3 [78].

DPYD rs3918290 mutation (*2A, IVS14+1 G>A, c1905+1 G>A) is the most investigated mutation associated with DPD loss of function [79]. It is a point mutation located in the splice recognition sequence of intron 14 which results in the substitution G>A. The presence of this mutation leads to the skipping of the entire exon and results in a 165-bp deletion in the DPD mRNA [80]. The resulting truncated protein is characterized by a complete loss of enzyme activity, leading to a complete deficiency in patients carrying homozygous mutated *DPYD**2A genotype, and in a partial deficiency of DPD activity in patients with heterozygote genotype [81].

DPYD rs55886062 (*13, c.1679 T>G) is a missense mutation that causes a Ile560Ser amino acid change in the flavine mononucleotide binding domain of DPD. This mutation and the consequent amino acid substitution leads to a non-functional protein [82,83].

The non-synonymous variant *DPYD* rs67373798 (c.2846A>T) results in an amino acid substitution Asp949Val localized near an iron-sulphur motif leading to a decrease in the catalytic activity of the encoded enzyme [82].

Recent studies have focused on the intronic region of *DPYD*, leading to the identification of a new haplotype, or haploblock (rs75017182 *DPYD HapB3*). A haploblock is an allelic combination of different genetic mutations characterized by being inherited simultaneously. These mutations are in linkage disequilibrium (LD) and, for its definition, the determination of one of these mutations is sufficient to determine the presence of the others. The *DPYD* HapB3 consists of three intronic variants (c.483+18 G>A rs56276561, c.680+139 G>A rs6668296, and c.959-51 T>C rs115349832) and one synonymous variant (c.1236 G>A, E412E; rs56038477) [25]. For the *DPYD* HapB3, the so-called tagging SNP, is the intronic mutation c.1236G>A rs56038477, which is responsible for the insertion of 44 bp in intron 10 of the mRNA and promotes the formation of a premature stop codon of exon 11. The resulting protein is characterized by lower functional activity. In Table 1.2 the values for activity assigned to alleles of *DPYD* are reported.

Table 1.8 Activity Score of the *DPYD* rs3918290, *2A; rs55886062, *13; rs67373798, c.2846A>T and rs56038477, c.1236G>A/HapB3 variants.

Activity Score	Alleles
0	*DPYD**2A (rs3918290)
	*DPYD**13 (rs55886062)
0.5	c.2846A>T (rs67376798)
	c.1236G>A/HapB3 (rs56038477)
1	*DPYD**1 (wild-type)

5.1.1 PGx Guidelines for Fluoropyrimidine Treatment

The introduction of PGx into clinical practice depends largely on the availability of approved and validated PGx tests, the ability of clinicians to use and understand them, and the availability of evidence-based recommendations for therapeutic management. Nowadays, there is a plethora of evidence that genetic variants in the DPD gene are predictive of severe FP-related toxicities, and there are international clinical practice recommendations for *DPYD* genotype-guided FP dosing and therapeutic drug monitoring (TDM) [84]. Despite both solid evidence, *DPYD* genotyping which

representa a standard of care in many European countries, has not yet gained worldwide acceptance [26]. Recommendations for FP dose adjustment proposed by the CPIC and DPWG Consortia are listed in Table 1.3 and Table 1.4 respectively [77,78].

Table 1.9 Clinical Pharmacogenetics Implementation Consortium (CPIC) Dosing recommendation of Fluoropyrimidine therapy by genotype/phenotype.

PHENOTYPE	GAS	IMPLICATIONS	DOSING RECOMMENDATIONS	CLASSIFICATION OF RECOMMENDATIONS
DPYD Normal Metabolizer	2	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	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	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 dose[k] with early therapeutic drug monitoring.[l] Activity score 0: Avoid use of 5-fluorouracil or 5-fluorouracil prodrug-based regimens.	Strong

Table 1.10 Royal Dutch Association for the Advancement of Pharmacy - Pharmacogenetics Working Group (DPWG) dosing recommendation of Fluoropyrimidine therapy by genotype/phenotype.

ALLELE/ GENOTYPE/ PHENOTYPE	DESCRIPTION	RECOMMENDATION
Activity Score 0	The gene variation increases the risk of severe, potentially fatal toxicity. A reduced conversion of fluorouracil/capecitabine to inactive metabolites means that the standard dose is a more than 100-fold overdose.	1. Avoid fluorouracil and capecitabine. Tegafur is not an alternative, as this is also metabolised by DPD. 2. If if is not possible to avoid fluorouracil and capecitabine: determine the residual DPD activity in mononuclear cells from peripheral blood and adjust the initial dose accordingly. A patient with 0.5% of the normal DPD activity tolerated 0.8% of the standard dose (150 mg capecitabine every 5 days). A patient with

		undetectable DPD activity tolerated 0.43% of the standard dose (150 mg capecitabine every 5 days with every third dose skipped).
Activity Score 1	The gene variation increases the risk of severe, potentially fatal toxicity. A reduced conversion of fluorouracil/capecitabine to inactive metabolites means that the normal dose is an overdose.	Start with 50% of the standard dose or avoid fluorouracil and capecitabine. Adjustment of the subsequent dose should be guided by toxicity and effectiveness. However, in one study involving 17 patients with gene activity 1, the average dose after titration was 57% of the standard dose. Tegafur is not an alternative, as this is also metabolised by DPD.
Activity Score 1.5	The gene variation increases the risk of severe, potentially fatal toxicity. A reduced conversion of fluorouracil/capecitabine to inactive metabolites means that the normal dose is an overdose.	Start with 50% of the standard dose or avoid fluorouracil and capecitabine. After starting treatment, the dose should be adjusted based on toxicity and effectiveness. In a study involving 17 patients with genotype *1/2846T*, the average dose after titration was 64% of the standard dose. For 51 patients with genotype *1/1236A*, the average dose after titration was 74% of the standard dose. Tegafur is not an alternative, as this is also metabolised by DPD.
FENO	The gene variation increases the risk of severe, potentially fatal, toxicity. A reduced conversion of fluorouracil/capecitabine to inactive metabolites means that the normal dose is an overdose.	It is not possible to recommend a dose adjustment for this patient based on the genotype only. Determine the residual DPD activity in mononuclear cells from peripheral blood and adjust the initial dose based on phenotype and genotype, or avoid fluorouracil and capecitabine. Tegafur is not an alternative, as this is also metabolized by DPD.

At present, the feasibility and safety of FP genotype-guided dosing has been demonstrated by several groups [85]. Based on data from the Deenen *et al*. and Henricks *et al*. studies, the CPIC has recommended upfront dose reductions for carriers of at least one minor allele of the common *DPYD* SNPs [76,86]. In a prospective clinical trial, patients starting FP therapy were screened for *DPYD*2A*. Consistent with non *DYPD*2A* carriers, a comparable severe toxicity incidence emerged among carriers with an upfront dose reduction. Additionally respect to a cohort of **2A* carriers receiving the full dose, the risk of drug-related death was reduced from 10% to 0% in **2A* carriers receiving a reduced dose. Subsequently, the prospective evaluation of *DPYD* genotype guided dosing was extended to four *DPYD* variants (*2A or c.1905+1G>A, c.2846A>T, *13 or c.1679T>G, and HapB3/c.1236G>A). It has been reported that heterozygous variant allele carriers receiving initial chemotherapy dose reductions showed a reduction in the relative risk of severe FP -related toxicity in all groups compared to historical controls.

On April 30, 2020, the European Medicines Agency (EMA) published recommendations on laboratory tests that should be performed before treatment with FP [87]. European Medicine Agency, EMA, has adopted the recommendations of the scientific societies AIOM-SIF, CPIC and DPWG, which emphasise the importance of checking for the presence of *DPYD* mutations that predict toxicity before starting treatment [88].

Although the implementation of PGx guidelines evaluating 4 *DPYD* variants has hugely change the treatment landscape of patients with solid cancer for many others, who are not carriers of these variants, the therapy remains challenging because of the severe toxicities developed.

Identification of additional genetic markers is therefore necessary and the concomitant variables involved in individual variability should be further investigated.

1.4.2 New *DPYD* Variants

Although many studies have been performed on the *DPYD* variants and their contribution to the toxicological profile of FP, only 4 variants with PGx guidelines are clinically implemented. Noqwadays, after 30 years, new potential mutations have finally been identified.

Recently, new studies have focused on assessing the expression Quantitative Trait Locus (eQTL) in the human liver, which has been shown to be crucial in determining how genetic affects disease risk and therapeutic outcomes [89]. An eQTL is a genetic variant that can influence gene expression by altering gene transcription and transcript stability [90].

A study by Etheridge *et al*. identified three novel variants that affect DPD protein expression: rs59353118, rs114170368, and rs4294451 [89].

rs59353118 appears to be the most significant cis-eQTL in a haplotype block associated with *DPYD* expression, with the minor allele associated with reduced expression (Figure 1.6). rs59353118 is a point mutation located in intron 14 leading to substitution A>T, resulting in LD with rs72728438 and rs12022243 [91]. In addition, these two variants have been associated with decreased DPD activity in mononuclear cells and increased risk of capecitabine toxicity, respectively. Notably, the genetic position of rs72728443 in an open chromatin region enriched in histone modifications and in a p53 binding motif could alter the binding of the *DPYD* repressor p53. This is proposed as the mechanism of action by which the variants alter *DPYD* expression in the liver, whereas the T genotype of intron variant rs12022243 has been associated to a greater extent with capecitabine toxicity in humans with colorectal cancer (CRC) than allele C [92].

rs114170368, c.1339+3875 is an intronic SNP of *DPYD* that correlates with lower enzyme expression leading to a substitution of G for a C allele. LD is reported for this SNP with *DPYD* HapB3 variant [93].

Additionally, rs4294451 was correlated to enhanced *DPYD* expression. The variant is located in an intron region near *DPYD* gene. rs4294451 is a highly conserved potential regulatory element that could modulate educational competencies and/or schizophrenia risk [94,95]. Since DPD functionality also plays a role in several neurological and psychiatric disorders, these variants could influence DPD activity itself.

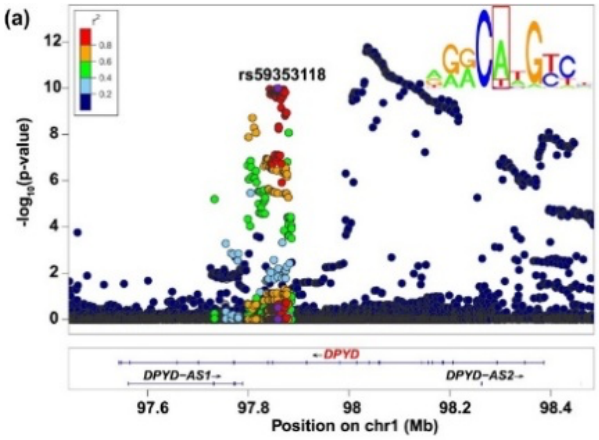


Figure 1.16 LocusZoom plots of cis-eQTLs (expression quantitative trait loci) in *DPYD* rs59353118 (located in *DPYD* intron 14).

These new variants may help update strategies to maximize the efficacy and safety of clinical interventions in the future, but further investigation is needed.

1.4.3 Haplotype Analysis

Analysis of variation in human populations is essential to understand how gene variants affect response to treatment. Instead of the well-known variants that affect the function of enzymes directly involved in the degradation/metabolism of various drugs, new approaches that focus on the influence of haplotypes rather than specific and individual SNPs have gained increasing interest in the clinical field. Indeed, analysis of haplotypes in association studies to identify common variants may be more informative than single allele studies [96].

In fact, often the SNPs approach is not sufficient to predict a phenotype or to explain the totality of toxicities associated with 5-FU. The evaluation of combinations of SNPs, as a haplotype, provides an attractive and informative tool for stratifying the toxicity risk of patients in pre-emptive genotyping.

A haplotype is a set of DNA SNPs inherited together due to their close positions in the DNA sequence [97]. The International HapMap Project, which aims to determine the common patterns of DNA sequence variation in the human genome by characterizing sequence variants, has provided an excellent tool for identifying these regions of haplotypes and obtaining a consistent collection of information needed to study the influence of genes on diseases in genetic studies [98].

Recently two common *DPYD* variants (c.85T>C and c.496A>G) emrge of high interest but controversial data were published on their effects on DPD phenotype and their role in 5-FU toxicity. Recently Hamzic *et al*, based on previous study results suggesting LD between the two variants, discovered a possible effect of c.85T>C and c.496A>G on 5-FU toxicity depending on the haplotype structure [63,99]. The addition of *DPYD* rs1801160 *6c.2194G>A Val732Ile was investigated as a candidate *DPYD* haplotype to improve toxicity prediction. Multi-SNPs analysis by Gentile *et al*. allowed the detection of two haplotypes containing the minor alleles of SNPs rs1801160, rs1801265 and rs2297595, that were significantly associated with decreased 5-FUDR level [100].

The haplotype approach seems promising in identifying novel risk factors for toxicity but despite the reassuring data, it is necessary to proceed with subsequent validation and new studies.

1.5 SEX DIFFERENCES AND SEX MEDICINE

Given the great importance that the personalized therapy approach currently reach in the treatment of diseases, especially cancer, the systematic inclusion of sexual biology in clinical research and in the field of pharmacology is of great importance, as it is the most relevant aspect that characterizes an individual in his identity [101].

Sex medicine is an innovative approach to precision medicine that aims to improve care and treatment for men and women alike, overcoming the administration of drugs on a milligramme/kilogramme-based basis or "one size fits all," drug administration, which often leads to treatment failures [102,103]. Sex differences affecting virtually all body compartment correlate with protection or susceptibility to cancer, cancer progression, and response to therapy [104,105].

Indeed, a careful analysis of the literature showed that there are sex differences not only in disease incidence but also in treatment outcome where patients' sex could affets both PK and PD drug pathway [105-107].

Women are more susceptible to overdose because of lower volume of distribution, higher body fat percentage, larger free fraction of drug, and slower xenobiotic clearance. Body composition parameters that could affect the drug distribution include total body water, extracellular water, intracellular water, total blood volume, plasma volume, and red blood cell volume, which are greater in men than in women, resulting in different responses to the drug. Plasma protein binding has also been shown to vary between the two sexes, due to the influence of oestrogens, which increase the concentration of serum binding globulins. Oral contraceptives, pregnancy, and menopause are cited as sex-specific conditions that affect the pharmacokinetics and pharmacodynamics of various drugs. Another important difference between the two sexes is related to the composition of the gut microbiota [108,109].

In addition, lifestyle factors such as diet, physical inactivity, tobacco use, and excessive alcohol intake, which are known to have a direct impact on drug response, differ greatly between men and women as reported in several studies. In addition, polypharmacy could have an impact on treatment response. The literature reports that women take a greater number of medications than men, leading to drug-drug interactions.

Moreover, PK and PD play a crucial role in this context, with both phase I and phase II metabolism enzymes activity significantly differing by sex. The phase I enzymes CYP3A4 and CYP2D have higher activity in females, wheras CYP1A has higher metabolic activity in males. These differences result in sex-specific changes in exposure to various drugs, including clomipramine, clozapine, olanzapine, acetaminophen, codeine, diazepam, fentanyl, statins, and tamoxifen. Phase 2 enzymes such as UDP-glucuronosyltransferases and methyltransferases have been found to be more active in males than in females, and directly affect the clearance of ibuprofen, acetaminophen, azathioprine, dopamine, oxazepam, and levodopa [110,111].

In parallel, the most striking differences between men and women are sex chromosomes: alteration of the X inactivation process, loss of the entire Y chromosome, gene mutation or deletion, epigenetic deregulation, and miRNA effect, have been associated with altered drug metabolism and, moreover, with cancer predisposition [112–114]. Many genetic SNPs have also been shown to exhibit sex specificity [115]. A recent genome-wide association study (GWAS) of psychiatric disorders identified genetic variants that differ between males and females. Variants in genes related to neuronal development, immune system, and vascular function have been associated with sex-dependent effects on the onset of schizophrenia, major depressive disorder, and bipolar disorder [116].

While there is evidence of sex differences in the efficacy and toxicity of many drugs, including chemotherapeutic agents, sex medicine still remains a mirage [117–119]. As mentioned earlier, both

animal studies and phase 3 clinical trials of chemotherapeutic agents showed a sex imbalance exclusively in preference of men enrollment. The low proportion of women in randomized clinical trials has also been accompanied by the lack of female animal models, which also confirms the lack of cell difference data from *in vitro* studies [120,121]. Consequently, in clinical practise, the same dosage is administered to men and women indifferently with higher incidence rate of toxic effect specifically in women. In 1977 FDA excluded women from clinical research because of the risk of birth defects until 1993 when it revised its position and issued the Guidance for the Study and Evaluation of Sex Differences in the clinical evaluation of drugs. In those years, however, and still today, albeit to a lesser extent, there was a lack of sex-specific data in drugs that are still used today [122]. In addition, there are no accurate studies of the sex-specific effects of PGx on patient treatment, as in other areas of pharmacological research.

1.5.1 Fluoropyrimidine clinical outcome and the role of patients' sex

In the treatment of various cancers, 5-FU has been observed to produce different toxicities depending on sex [123,124]. Clearance of 5-FU is lower in females than in males (Milano *et al*., 1992), resulting in higher toxicity in females and affecting therapeutic efficacy [125,126]. Zalcberg *et al* in 1998 already reported that haematological and non-haematological toxicities after 5-FU and leucovorin administration in patients with advanced CRC are influenced by sex [7]. With 5-FU-based chemotherapy, women experienced more frequently severe toxicities such as stomatitis, leukopenia, alopecia, and diarrhea with respect to men. These data were recently confirmed by Wagner *et al*. in an analysis of women with colon cancer who received adjuvant FP-based chemotherapy [127]. The mechanism for which the PK of 5-FU is known to be sex-dependent when administered as an infusion, with drug clearance being lower in females, is still unclear, but the authors postulated a role for sex differences in DPD activity [128]. DPD activity is decreased in females and the difference in DPD activity between males and females is 15% [129]. Such general sex differences in the 5-FU-toxicity risk could also have implication for the discovery of genetic associations of *DPYD* variants with 5-FU toxicity, as the proportion of toxicity cases explained by *DPYD* variants may differ between men and women [63]. Lower DPD activity associated with toxicity in females could be explained with the DPD deficiency syndrome [130]. In this context, another interesting finding is that the effect of the c.1905+1G>A variant on the toxicity of 5-FU has been reported to besex-specific in a study by Schwab *et al*. [131]. While a significant association was found with this variant in the male patients, no effect was seen in females patients. Lee *et al*. observed a stronger effect of *DPYD*2A* in males

compared with females, suggesting an interaction between sex and gene for different effects of the same genetic variant [132]. In contrast, the same c.1905+1G>A variant was associated with severe 5-FU toxicity in four female patients, questioning the proposed sex-specific role of this variant [133]. Thus, the result suggests that sex might also influence the effects of *DPYD* variants on 5-FU toxicity. Althogh there are conflicting data, further analysis is needed to better understand the role of the patient's sex on the outcome of FP-based therapies.

Besides safety profile, sex differences continue to be observed in the outcomes of CRC treatment. There are few reports of molecular and biological differences between the sexes in CRC, although clinical differences between them have occasionally been reported. In CRC, some previous studies have shown that women may benefit more from adjuvant chemotherapy and that sex and tumor location influence the survival benefit of chemotherapy [134,135]. In addition, a study of 52 men and 49 women with pterygium conducted by Shah *et al* found a statistically insignificant difference between efficacy and sex after 5 FU treatment [136]. These sparse data show that sex difference is still a little discussed topic in the era of precision medicine.

1.6 TUMOR BIOMARKER EXPRESSION

Besides patients' genetics and patients' constitutive characteristics, the biomolecular characterization of tumor biomarkers expression and complex interaction with pathology outcome is the milestone on which precision oncology is based.

The concept of precision oncology primarily involves the search for specific mutations in a particular gene sequence, RNA, or protein that, if present in a patient with a particular tumor type, are potential candidates for individualized treatment. Therefore, based on the molecular characteristics of the candidate's tumor, patients are selected for a specific target treatment that will provide greater and longer-lasting potential clinical benefit compared to standard therapy.

In addition to the presence of specific mutations, the expression level of certain tumor proteins has also been associated with different treatment outcomes after tumor tissue biopsy analysis. Several tumor-associated proteins have already been investigated as relevant prognostic markers in rectal cancer [137]. However, despite advances in the understanding of the rectal cancer pathogenesis, no molecular marker has achieved a general consensus for clinical decision making and has been shown to provide consistent and common prognostic information [138–140]. For this reason, further investigation of features related to the patient's response to therapy is essential to avoid unnecessary treatment. In this context, immunohistochemistry (IHC) analysis is used to identify molecules with

prognostic and/or therapeutic significance, as well as for histogenetic classification of anaplastic tumors and to determine the metastatic tumors' site origin.

Recent studies have highlighted the importance of upregulation and aberrant staining patterns of several markers in patient prognosis. Thus, reduced membranous staining for β-catenin, absence of cytoplasmic staining for β-catenin, reduced membranous staining for E-cadherin, and absence of cytoplasmic staining for E-cadherin correlate with metastatic disease [141].

An example of proteins associated with response to chemoradiotherapy, CRT, in rectal cancer are those involved in the mismatch repair (MMR) pathway. Lack of expression of some of these proteins in tumor tissue has been evaluated as an independent prognostic factor for both overall survival, OS, and local recurrence. This is true for MutS Homolog 6 (MSH6) expression in tissue before treatment or for MSH6 expression after chemoradiation [142]. Expression of proteins involved in the inflammatory process, angiogenesis, cell proliferation, and hypoxia has recently been associated also with disease free survival (DFS) or OS in cancer patients.

IHC is a widely used technique for the identification and histological localization of antigens, cellular and tissue components *in situ*. In recent years, it has become a fundamental tool for the diagnosis of many diseases and has undergone a remarkable methodological evolution [143].

This technique has enabled the identification of new tumor antigens, oncofetal proteins and proteins encoded by oncogenes. It is also useful for the classification of neoplasms, the evaluation of prognostic value of proteins, the identification of pathological deposits such as immunoglobulins and amyloid, and for their use in infectious diseases and the origin and function of different cell types. In addition, IHC is rapid, and inexpensive, and has the advantage of eliminating the error of sampling test tissues by identifying tumor cells [144].

IHC examination, which highlights the presence of a specific molecule (antigen) in the tissue under study by exploiting antibody-specific binding, is performed on histological sections obtained from a biopsy specimen. The binding between the antibody and the antigen is visualized by staining mediated by enzymatic reactions. The protein of interest can be expressed at the cell membrane level as well as in the cytoplasm or nucleus. In addition to detecting the protein, IHC analysis also provides quantitative information on the percentage of cells expressing the protein and the extent of its expression (intensity of staining). Currently, the histo-score (H-score), has been introduced as a clearer interpretation of the comprehensive IHC results of cellular status, since there is no homogeneous scoring system for semi-quantification of biomarker expression. Variuous H-scores were proposed, each of which scored the IHC results differently for both intensity and cell positivity. In the H-score by Hatanaka *et al*, the intensity of nuclear staining in 1000 randomly selected cells from five areas (two different lesions were selected in each area) was assessed under microscopic

observation at 400× magnification, and the intensity was divided into 4 levels: 0 negative; 1+ weakly positive; 2+ moderately positive; and 3+ strongly positive. The calculation of H-Score was based on the percentage of positive nuclei (Pi), varying from 0 to 100%, and their staining intensity value (i) (0, 1, 2, or 3), each calculated according to the formula: H-score = $\sum$ (i x Pi). The final H-Score value was graded from 0 to 300 [144].

Similarly, Huang MY *et al*. proposed another H-score value which considers the intensity and the percentage of positively stained tumor cells [145]. The score for intensity was graded on a scale ranging from 0 to 3:

- 0 = no staining;
- 1 = weak staining in more than 50% of positive cells or moderate staining in less than 50% of cells;
- 2 = moderate positive staining in more than 50% of cells or with strong staining in less than 50% of cells;
- 3 = strong staining in more than 50% of cells.

The score for percentage of positive cells was based on 4 values, with 0 and 1 representing no overexpression, and 2 and 3 representing overexpression of the marker in tumor tissue. The final staging score was calculated multiplying the intensity score by the percentage score, resulting in a scale value from 0 to 9.

In calculating the IHC score of Huang F *et al*, each sample was assigned a score from 0 to 3 depending on the intensity of staining of the membrane or nucleus:

- 0 = no staining or not detected;
- 1 = weak staining/light yellow;
- 2 = moderate staining/yellowish brown;
- 3 = strong staining/brown.

In addition, as the other H-Score, a numerical value was assigned to the percentage of staining as follow: 0% = 0, 1-24% = 1, 25-49% = 2, 50-74% = 3, 75-100% = 4. The final H-score was determined multiplying the intensity value by the extent of stained cells, resulting in a score value ranging between 0, the minimum value, and 12, the maximum value [146].

1.7 COLORECTAL CANCER, CRC

CRC is the third most commonly diagnosed solid cancer and represents the fourth leading cause of cancer-related death worldwide [147]. In general, the incidence of CRC in Italy is estimated at about 13% in both sexes, with diagnosis at the median age of 50 years. In men it is the third most common neoplasm after prostate and lung cancer, while in women it is the second most common tumor after breast cancer. The incidence is estimated at 125,000 new cases per year at European level and 44,180 new cases in the United States in 2019, accounting for approximately 30-35% of all CRC cases [148]. The incidence of CRC has declined in recent decades except in people younger than 50 years age, which may be related to the increasing screening program.

Regarding CRC risk factors, it has been reported that both strong environmental associations and genetic play a crucial role. Epidemiological studies indicate a clear environmental and lifestyle link for CRC. Increased CRC risks are influenced by obesity, diet, tobacco, alcohol, androgen deprivation therapy, and cholecystectomy [149]. In addition to environmental factors, genetic mutations has also been suggested as a cause of tumor development. It is estimated that 5% of patients with CRC have a hereditary disease and 5% of these are due to known genetic mutations [150]. In particular, alteration of the normal colonic epithelium requires accumulation of either somatic and/or germline (inherited) mutations over a period of approximately 10-15 years. Moreover, chromosomal instability (CIN), MMR, and CpG hypermethylation are the major pathways involved in the development of CRC [151].

Diagnostic or screening colonoscopy is required for pathological confirmation. Additional computed tomography (CT) and carcinoembryonic antigen (CEA) determination are useful for CRC staging [152,153]. Indeed, pathologic stage at diagnosis is the most important prognostic factor for survival. As for the treatment of patients, the therapy of CRC strictly depends on the stage of the tumor, the location of the tumor mass, and the general condition of the patient (performance status, PS). However, surgery remains the therapy of choice for the treatment, especially in early-stage CRC. In stage III CRC adjuvant 5-FU chemotherapy is the cornerstone of treatment and increases survival in high-risk patients CRC [154,155]. In patients with local recurrence and with liver and lung metastases, different therapeutic approaches are used, while in patients with inoperable CRC, systemic palliative therapy is given to improve quality of life and prolong life expectancy [156].

1.8 RECTAL CANCER

Rectal cancer belongs to a subcategory of CRC and differs from colon cancer mainly by the distance to the anal margin, which is measured by endoscopic examination and has a cutoff value of 15 cm.

(Figure 1.7). Rectal cancer is therefore divided into low (up to 5 cm), middle (from 5 to 10 cm), and high (10 cm or more) [157].

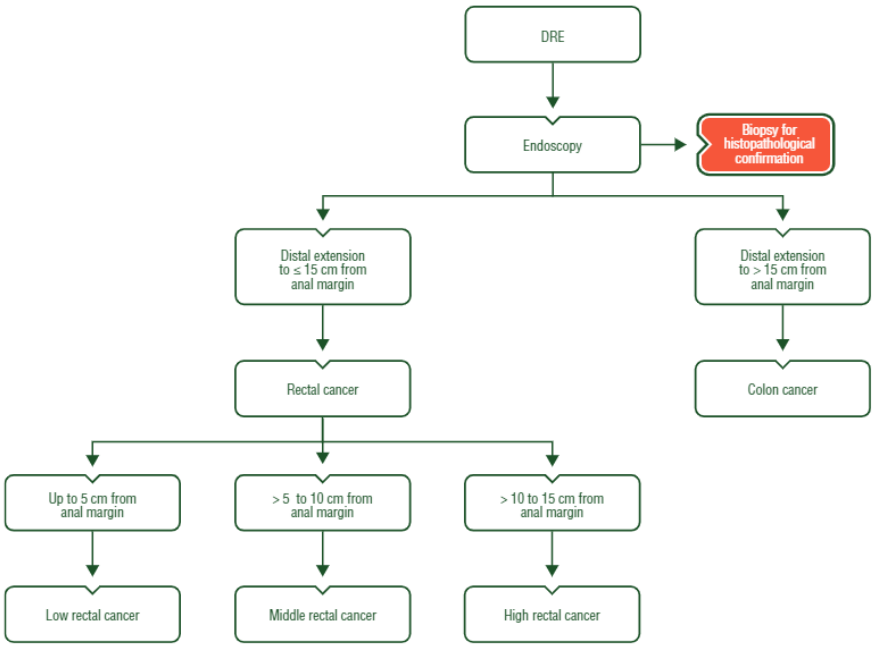


Figure 1.17 Rectal cancer diagnosis. DRE, digital rectal examination. (ESMO)

Rectal cancer is a common malignancy especially in Western countries with 704,376 cases worldwide and 310,394 deaths in 2018 [158]. Approximately 28% of all CRC occur in the rectum, with a difference in frequency between men and women of 32% and 25% [159]. There was also a difference in mortality rates between the sexes, with males having a 30-40% higher rate than females, although it varied by age.

For rectal cancer incidence in both sexes has increased in Italy since the 2000s, partly due to the development of numerous screening programs. In recent years, the mortality rate in Italy has decreased despite the increase in cases [160].

The 5-year rate OS for rectal cancer is 67%, with differences according to the stage at diagnosis:

- Stage I: 88% ;
- Stage IIA: 81%;
- Stage IIB: 50%;
- Stage IIIA: 83%;
- Stage IIIB: 72% for stage IIIB;
- Stage IIIC: 58% for stage IIIC;
- Stage IV: 13% for stage IV.

The 5-year rate OS after local and distant recurrence following pCRT is 6-9% [161–163]. In addition, the survival rate varies depending on the stage: a survival rate of 89% is estimated if the diagnosis is made at a localized stage, 71% if the tumor has spread to surrounding tissue/local lymph nodes and 15% if the tumor has metastasized, i.e., if the cancer has spread to distant parts of the body [164].

Rectal cancer is characterized by several morphologic and molecular features that allow it to be distinguished from CRC, with different associated causes and risk factors [157]. Although the exact cause of rectal cancer is not yet known, environmental and patient-specific risk factors, along with age, sex, race, and family history, contribute to the cancer predisposition [165].

Indeed, age itself is an independent risk factor: rectal cancer is a disease with a higher incidence in the population between 40 and 50 years of age [166]. In addition, comorbidities such as diabetes, chronic inflammatory bowel disease, Cronh's disease and metabolic syndrome have been associated with an increased risk of rectal cancer occurrence [167,168]. Protective factors include a proper diet rich in fruits and vegetables, physical activity, consumption of calcium, garlic, milk, and use of non-steroidal anti-inflammatory drugs (NSAID) [169–171].

From a genetic point of view, molecular alterations common in rectal cancers include CIN, mainly due to a deficiency of the MMR system, and mutation in the *APC*, *TP53*, *KRAS, PIK3CA, BRAF* and *NRAS* genes. Mutations in genes as *MLH1, MSH2, MSH6, PMS2* or *EPCAM* promote MMR deficiencies and higher cancer risk [172]. The DNA MMR system is a highly conserved repair mechanism in cellular evolution whose variation increases susceptibility to cancer. MMR deficiency or high-grade microsatellite instability (MSI-H) are indeed correlated with a high developmental risk of cancers in general, including rectal cancer [173].

Symptoms of rectal cancer may vary from patient to patient. In summary they include rectal bleeding, dark brown or bright red blood in the stool, sudden changes in bowel habits, weakness or fatigue, and abdominal pain with unexplained weight loss [148]

Screening, performed every 2 years after age 50 or more frequently if there is a family history of hereditary syndromes, has allowed early diagnosis of the disease and a reduction in mortality. Screening consists of a noninvasive fecal occult blood test. If the screening result is positive the AIOM guidelines recomend a colonoscopy and clinical evaluation. Common cancer diagnostic procedures include digital rectal examination in symptomatic patients and histopathologic examination of a biopsy. In this setting, physicians evaluate parameters such as the distance to the anal margin, the colon, the presence of bleeding, the volume of the tumor mass, the condition of the anal sphincter, and the degree of fixation [160,174].

Additional blood CEA levels, liver, and kidney function, genetic and imaging tests (e.g., Magnetic Resonance Imaging (MRI), CT, and Positron Emission Tomography (PET)) are used to fully assess the patient's condition and provide additional information about the stage and the presence of disease outside the pelvis [175].

1.8.1 Rectal Cancer Staging

Each case of rectal cancer is discussed by a tumor board involving clinitiants from as radiologists, surgeons, radiation oncologists, oncologists and pathologists. In rectal cancer treatment, correct tumor staging based on diagnostic evaluation (rectal exploration and rectal sigmoidoscopy) is crucial for choosing the best therapeutic option. The 2018 Aiom guidelines established diagnostic-therapeutic algorithms for determining treatment, that also consider infiltration of the submucosa and the presence or absence of other risk factors [160].

Based on the exams results, it is possible to classify the disease according to the Tumor, Node, and Metastasis (TNM) classification (Table 1.5). The TNM classification of the Union for International Cancer Control (UICC) (8th edition) is used to distinguish the different tumor stages, thus helping the physicians to choose the right therapy and allowing patients to participate in clinical trials [176].

The TNM staging system uses alphanumeric codes to distinguish the different cancer stages based on 3 parameters such as tumor size, lymph node involvement and the presence of metastases. The different staging systems derived from the Dukes staging system developed in 1932 [177].

Table 1.11 Rectal cancer staging based on Tumor, Node and Metastasi, TNM, value. Definition of the TNM value.

Stage	Definition
Primary tumor (T)	
TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
Tis	Carcinoma *in situ:* intraepithelial or invasion of lamina propria
T1	Tumor invades submucosa
T2	Tumor invades muscularis propria
T3	Tumor invades through muscularis propria into the subserosa or into nonperitonealized pericolic or perirectal tissues
T4	Tumor perforates visceral peritoneum or directly invades other organs or structures
Regional lymph nodes (N)†	
NX	Regional lymph nodes could not be assessed
N0	No regional lymph node metastases
N1	Metastases in one to three regional lymph nodes
N2	Metastases in four or more regional lymph nodes
Distant metastases (M)	
MX	Distant metastases could not be assessed
M0	No distant metastases
M1	Distant metastases
Extent of resection (R)‡	
RX	Presence of residual tumor cannot be assessed
R0	No residual tumor
R1	Microscopic residual tumor
R2	Macroscopic residual tumor

In accordance with the TNM cluster scores, rectal stages could be grouped into 4 stages as follows:

- Stage 0 which comprehends Tis N0 M0;
- Stage I which comprehends T1-2, N0 M0;
- Stage II which comprehends:
    - IIa: T3, N0, M0;
    - IIb: T4a, N0, M0;
    - IIc: T4b, N0, M0.
- Stage III which comprehends:
    - IIIa: T1-2, N1, M0;
    - IIIb: T3-4, N1, M0;
    - IIIc: T3-4b, N2, M0.
- Stage IV which comprehends any T, any N, M1.

1.9 LOCALLY ADVANCED RECTAL CANCER, LARC

The Locally Advanced Rectal Cancer (LARC) accounts for 60% of all rectal cancers and is defined as stage II/III rectal tumor with or without lymph node involvement and without concurrent metastases.

The standard of care for the clinical management of LARC consists in FP-based neoadjuvant chemoradiotherapy (nCRT) treatment, followed by radical surgery, mainly represented by total mesorectal excision (TME) (Figure1.8) and optionally adjuvant chemotherapy. Indeed, depending on the postoperative evaluation and pathological stage, adjuvant treatment or close follow-up are the main treatment options. It has been that nCRT is crucial for the proper treatment of LARC, as it can improve the rate of curative resection and significantly reduce the rate of local recurrence [161,178].

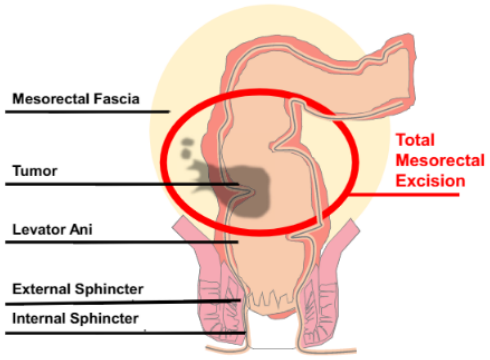


Figure 1.18 Visualization of the surgery site of Totam Mesorectal Excision, TNM.

To date, standard treatment is based on nCRT with a recommended RT dose of 45-50 Gy in 25-28 fractions over five weeks, with a preoperative boost of an additional 5.4 Gy in 3 fractions in case of high risk circumferential resection margin feature, combined with FP-based chemotherapy.

Preoperative nCRT, aimed at reducing tumor mass before surgery, has been associated with increased sphincter preservation rate and high pathologic complete response, CR, rate in approximately 10-30% of patients, due to reduction of the tumor size [179,180].

Although up to 20% of LARC patients respond completely to therapy, patients who do not respond at all or respond minimally have been shown to have a higher likelihood of disease recurrence, worse OS and worse progression free survival, PFS [162,181,182]. Indeed, up to 15-30% of patients undergoing nCRT have been shown to progress or develop distant metastases during therapy, after nCRT, or after surgery [183–186].

Approach has changed over the years in the treatment of LARC patients. The results of randomized phase III trials, as well as diagnostic tools and surgical/non-surgical strategies, have allowed for a significant evolution in LARC treatment (Figure 1.9).

Besides nCRT which is considered the gold-standard, treatment options for LARC patients included CRT, short-course RT (SCRT), or induction of chemotherapy followed by CRT. Starting in 1986, when the only treatment option was the surgical approach with TME, an initial clinical trial allowed the SCRT to be introduced into clinical practice and improved the local control rate. In the following years, the German Rectal Cancer Study CAO/ARO/AIO-94 published the results of 823 LARC patients in whom adjuvant CRT and neoadjuvant long-term CRT were evaluated. This showed a significant improvement in local control, a substantial reduction in grade 3 acute toxicity and grade 3 late toxicity [187]. In addition, nCRT showed superiority in reducing the local recurrence rate in two studies compared with only preoperative RT [188].

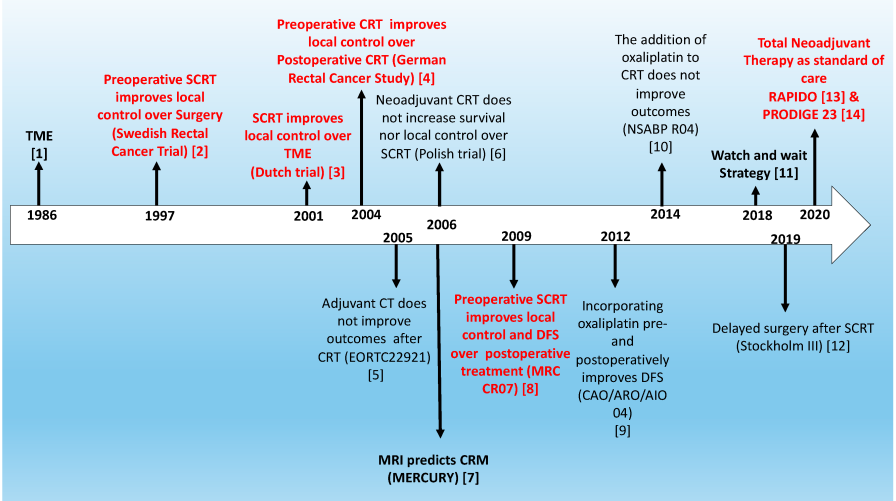


Figure 1.19 Temporal evolution of therapeutic approaches in locally advanced rectal cancer.

Recently, several working groups have focused on novel neoadjuvant approaches combining different chemotherapeutic agents or RT regimens to investigate potential improvement in clinical efficacy. Concurrent administration of irinotecan, oxaliplatin, gefitinib and cetuximab with FP has been tested in several studies with contrasting results [189–193]. Oxaliplatin-based neoadjuvant and adjuvant chemotherapy has been explored, although no clear evidence of improvement in pathologic control has been provided. In addition, due to the increasing interest in immunotherapy, the use of immune checkpoint inhibitors in the clinical treatment of LARC has been investigated in recent years to expand therapeutic options in the future [194,195].

Preoperative short-course radiotherapy (SCPRT), based on a total dose of 25 Gy at 5 Gy/fraction for one week, followed by surgery 10 days after the first radiation fraction, SCPRT with delayed surgery, long-term RT and LCRT are some of the possible treatment options according to the 2017 updated

European Society for Medical Oncology, ESMO, clinical practise guidelines for resectable LARC [157]. Conversely, “high-risk patients”, defined as those with T4 rectal tumors, invasion of the anal sphincter, a radiologically threatened circumferential radial margin, tumors with extramural venous invasion, and extra-mesorectal lymph nodes with clinically malignant features, are scheduled for an intensified treatment programme. These patients are eligible for so-called total neoadjuvant therapy (TNT), which includes systemic chemotherapy before or after nCRT [196].

Several variables related to the efficacy of LARC therapy, have emerged from various studied conducted to determine the appropriate LARC therapy. These variables, which are currently being investigated in several clinical trials, are the optimal chemotherapy dose, fractionation of radiotherapy, the proper sequence of therapeutic approaches, and the proper time interval between chemotherapy and surgery.

In addition to differences in therapeutic efficacy, the toxicity profile of the chemotherapy regimen was of particular importance in the setting of LARC therapy. The toxicity of therapy could directly contribute to its efficacy if patients had to interrupt, discontinue, or prematurely terminate therapy due to the development of side effects.
For this reason, there is an increasing need for predictive biomarkers of response that could be incorporated into the decision algorithm for selecting optimal therapy. Additional and more effective stratification criteria are needed to improve the currently available algorithms for risk stratification in LARC [197,198]. Although many biomarkers have been studied, their clinical availability is still lacking.

1.9.1 Treatment Response Assessment

Clinical Response

After CRT, clinical response is assessed by a combination of physical, instrumental, and imaging examinations that determine the appropriate surgical approach and optional follow-up treatment. Complete clinical response (cCR) is defined as the absence of clinically detectable tumor after nCRT, whereas incomplete response is determined by the presence of a palpable nodule, ulceration, or significant stenosis. Evaluation of cCR has recently been investigated as a substitute for pathologic CR, in which histopathologic examination of a surgical specimen is performed to identify residual cancer cells. Proper determination of cCR could avoid surgical intervention in patients eligible for a “watch and wait” approach [199]. “Watchful waiting” involves close monitoring of patients with cCR

and has been shown to provide good outcomes without oncologic compromise [200]. However, the definition of cCR can vary between different medical centres and is often imprecise [201].

Pathological Response

Assessment of response after surgery is performed by analysis of surgical specimens, which is nowadays the gold standard after neoadjuvant treatment. Pathologic staging after surgery includes TNM staging and, in addition, evaluation of tumor regression grade (TRG) in tumor specimen, which is a prognostic factor. LARC patients with pathologically complete regression and partial tumor regression have been shown to have better DFS and longer OS, as well as a lower risk of local recurrence and development of distant metastases [202]. Different TRG systems have been proposed, corresponding to different classifications of the tumor regression with different prognostic significance [203]. The most commonly used in clinical practice are the TRG system of Mandard, Dworak, American Joint Committee on Cancer (AJCC), and Ryan [204–207].

Ryan and AJCC (2010) TRG systems include a three-point scale for tumor regression, whereas the TRG classifications of Mandard and Dworak include a five-point grades scale and are based on residual tumor and fibrosis. In contrast, the Ryan and AJCC systems are based on volume of residual tumor cells.

The Mandard classification system was published in 1994 and initially used to assess tumor regression in squamous cell carcinoma of the oesophagus after neoadjuvant treatment with cisplatin and radiotherapy. The TRG system was first described in 1997. The definition of the score is based on the analysis of 17 patients have received preoperative radiochemotherapy with 5-FU/50 Gy

Mandard TRG scale is expressed with numerical values from 1 to 5 as follows:

- TRG 1 = no residual tumor cells;
- TRG 2 = occasional residual tumor cells with marked fibrosis;
- TRG 3 = marked fibrosis with tumor cells scattered or in groups;
- TRG 4 = abundant cells tumors with low fibrosis;
- TRG 5 = tumor non-regression. [204]

Dworak TRG scale is expressed with numerical values from 0 to 4 which have clinical significance, unlike those of the Mandard scale:

- TRG 0 = no regression;
- TRG 1 = minor regression;

- TRG 2 = moderate regression;
- TRG 3 = good regression;
- TRG 4 = complete regression.[205].

The TRG systems are showed in Table 1.6.

Table 1.12 Definition of tumor regression grading, TRG, systems

	Dworak	Mandard	Ryan	AJCC
Complete regression	No tumor cells (TRG 4)	No residual cancer cells (TRG 1)	No viable cancer cells, or single cells, or small groups of cancer cells (TRG 1)	No viable cancer cells (TRG 0)
Near complete regression	Very few tumor cells (TRG 3)	Rare residual cancer cells (TRG 2)	-	Single or small groups of tumor cells (TRG 1: moderate response)
Moderate regression	Dominantly fibrotic changes with few tumor cells or groups (TRG 2)	Predominant fibrosis with increased number of residual cancer cells (TRG 3)	Residual cancer outgrown by fibrosis (TRG 2)	Residual cancer outgrown by fibrosis (TRG 2: minimal response)
Minimal regression	Dominant tumor mass with obvious fibrosis (TRG 1)	Residual cancer outgrowing fibrosis (TRG 4)	Significant fibrosis outgrown by cancer, or no fibrosis with extensive residual cancer (TRG 3)	Minimal or no tumor cells killed (TRG 3: poor response)
No regression	No regression (TRG 0)	No regressive change (TRG 5)	-	-

Furthermore, Becker *et al*., in 2003, proposed a different classification system for advanced gastric carcinomas treated with cisplatin-based neoadjuvant chemotherapy, and Rödel *et al*. in 2005, have used a five-level score to assess tumor regression in rectal carcinomas [208,209].

2 RATIONAL

FP-based chemotherapy, which involves the administration of 5-FU and capecitabine, is considered standard therapy for the treatment of solid tumors such as gastrointestinal, breast, or head and neck cancer. Despite advances in new cancer therapies, FP in monotherapy or combination regimens still represents one of the most effective treatments for CRC at any stage of the disease. Despite a good tolerability profile, FP can lead to severe and sometimes life-threatening toxicities, which are exacerbated by co-administration of additional chemotherapy agents and occur in up to one third of treated patients. Better knowledge of individual variability and implementation of influencing factors into treatment personalization, tailoring drug selection and dosing to patients' features, may be a promising strategy to improve patient's clinical benefit. Available data suggest that both genetic and non-genetic factors play a role in 5- FU toxicity.

Among the causes of inter-individual variability, there are some genetic SNPs of the *DPYD* gene, which encodes the enzyme DPD, the rate-limiting enzyme involved in the first step of FP catabolism and responsible for about 85% of FP liver detoxification. Four SNPs in *DPYD* associated with decreased enzyme function and a higher frequency of developing severe and life-threatening toxicity are now recommended by EMA for testing prior to initiating therapy. Besides those validated *DPYD* gene SNPs, which account for approximately 10% of all FP, novel mutations require to be investigated as predictive factors for 5-FU toxicity and their impact on the patient's clinical outcome need to be assessed. There is a need to explore the clinical role of newly identified *DPYD* SNPs as predictive PGx markers of severe FP-related toxicity. Innovative molecular approaches to examine the genome allow the identification of new genetic variants that, after further validation, could be used in clinical practice.

In parallel, patients' sex has recently emerged as a variable of considerable interest that can influence and modify the outcome of therapy. 5-FU is a prominent example of a drug with substantial sex-specific interindividual variability in clearance/exposure, with ∼26% higher exposure in women. Given the increasing interest in the association between sex and toxicity effects, a meta-analysis conducted by Markus Diefenhardt *et al*. found an increased rate of acute toxicity effects in women with CRC treated with FP. Moreover, two recent studies have shown that female patients suffer more frequent and severe 5-FU toxicity than men. Further research is requested to better understand the understudied interaction between patient genetics and sex in defining the clinical outcome phenotype and to integrate sex as a variable in the optimization of pharmacological therapeutic strategy in a scenario where the so-called "gender problem" has not yet been investigated.

Beside toxicity that has been widely reported to be affected by multiple factors, FP efficacy could be impacted by genetics and sex variables.

The clinical context of homogeneous LARC patients treated with a neo-adjuvant chemoradiotherapy (a subset of the whole CRC cases considered in the study), could be a model to study the interplay between sex and genetics not only on the safety profile but also on the evaluation of efficacy. The selected clinical endpoint in this context has been the pCR after neoadjuvant chemoradiotherapy, that is associated with a favourable prognosis. When considering tumor response to treatment, the tumor tissue profile could also play a fundamental role, therefore for this task the tissue expression of specific protein biomarkers was considered, in patients with pre-treatment tissue available.

3 AIMS

The main aim of this doctorate thesis was the assessment of the predictive role of PGx markers, tumor tissue protein expression, patients' sex and their interaction on the outcome of a FP-based treatment in term of treatment safety and efficacy.

Specific PhD aims were:

i) The assessment of the predictive role of new previously un-investigated *DPYD* variants and patients sex affecting the risk of FP- related toxicity in a group of CRC patients,
ii) The evaluation of the interaction of patients' sex and *DPYD* variants with FP safety profile in CRC cases study population;
iii) The assessment in a subgroup of LARC patients, clinically homogeneous by treatment regimen (capecitabine administration in monotherapy and concomitant radiotherapy) of new previously un-investigated *DPYD* variants and haplotypes on FP efficacy profile;
iv) The evaluation of the interaction of patients' sex and new previously un-investigated variants and *DPYD* haplotypes on FP efficacy profile in a subset of LARC patients;
v) The evaluation of the interaction of patients' sex *DPYD* haplotypes on FP safety profile in a subgroup of LARC patients;
vi) The assessment of new IHC biomarkers in tumor biopsy in LARC patients, and their sex-specific effect as prognostic predictive factors of therapy efficacy

To achieve these goals the following laboratory activities were performed:

1. Selection of a study population of patients, treated with FP-based chemotherapy, with biological material and clinical data availability. For patients whose data were missing, clinical data were collected, and a complete database was created.
2. Pyrosequencing and Real Time analysis of 7 *DPYD* germline variants on a total of 689 CRC and 212 LARC patients,
3. Targeted hybrid capture-based custom Roche Exome NGS panel including of 106 FP-, radiotherapy- and rectal cancer-related genes to assessed genetic variability in of all exons and their adjacent splice junctions and, specifically, for *DPYD* haplotype investigation in a subgroup of 229 homogeneously treated LARC patients;
4. IHC, analysis of the expression of 11 biomarkers on a subset of 95 LARC patients with available pre-treatment tumor tissue;
5. Statistical analysis of the distribution of patients' toxicities by sex, the correlation of *DPYD* mutations and IHC markers with the FP efficacy/safety profile and to evaluate the interplay

between patients' sex and the above mentioned markers in predicting therapy toxicity/efficacy.

4 MATERIALS AND METHODS

4.1 PATIENTS' SELECTION AND STUDY DESIGN

Patients' selection for the current PhD project was carried out by the screening of a 1,934 clinical cases database including two different case study populations. The study design of the PhD project is summarized in Figure 1.4.

The first cohort is a historical study population which comprising 1,122 patients with CRC (any stage). Between 1999 and 2015, detailed clinical data and biological samples (blood) from these patients who received FP-based chemotherapy were collected at the Clinical and Experimental Pharmacology Unit of Centro di Riferimento Oncologico (CRO), in Aviano (PN), and DNA was extracted and stored.

The second case study population includes 812 LARC patients enrolled in a prospective clinical study protocol in the Department of Radiation Oncology at CRO. Recruiment, which began in 2003, is currently active and the case study continues to expand. A biobank of biological samples from patients enrolled in this prospective clinical trial protocol was established in the Experimental and Clinical Pharmacology Unit at IRCCS CRO in accordance with ethics committee guidelines. For each patient, a biological blood sample was collected, appropriately processed, and stored at -80°C for later analysis.

All patients who agree to participate in a clinical study were adequately informed by the medical personnel regarding the methods of sampling the biological material and the aims / purposes of the research. All recruited patients provided informed consent for data and blood collection by signing the consent form.

Three subgroups of patients were selected from this existing biobank for the purpose of this PhD work according to different eligibility criteria (Figure 4.1).

1. The first case population included CRC patients treated with FP-based chemotherapy with clinical data available in 2019. The germline genetic material of the included patients was used for the analysis of 7 germline variants of the *DPYD* gene by Pyrosequencing and Real-Time techniques.

2. The second case series includedd selected LARC patients homogeneously treated with capecitabine as monotherapy and concomitant radiotherapy. The germline DNA of the totality of the patients was evaluated by NGS technology using a custom panel of 106 selected genes with a role in the FP

ADME, in radiotherapy effect and with the patient's prognosis. Subsequently, correlation analysis could be performed only for patients with available clinical data.

3. The third case study involved LARC patients for whom it was possible to obtain pre-treated tumor tissue for IHC analysis of cellular expression of 11 biomarkers. This evaluation was possible thanks to the collaboration with Dr. Vincenzo Canconieri of the Pathology Unit of the CRO.

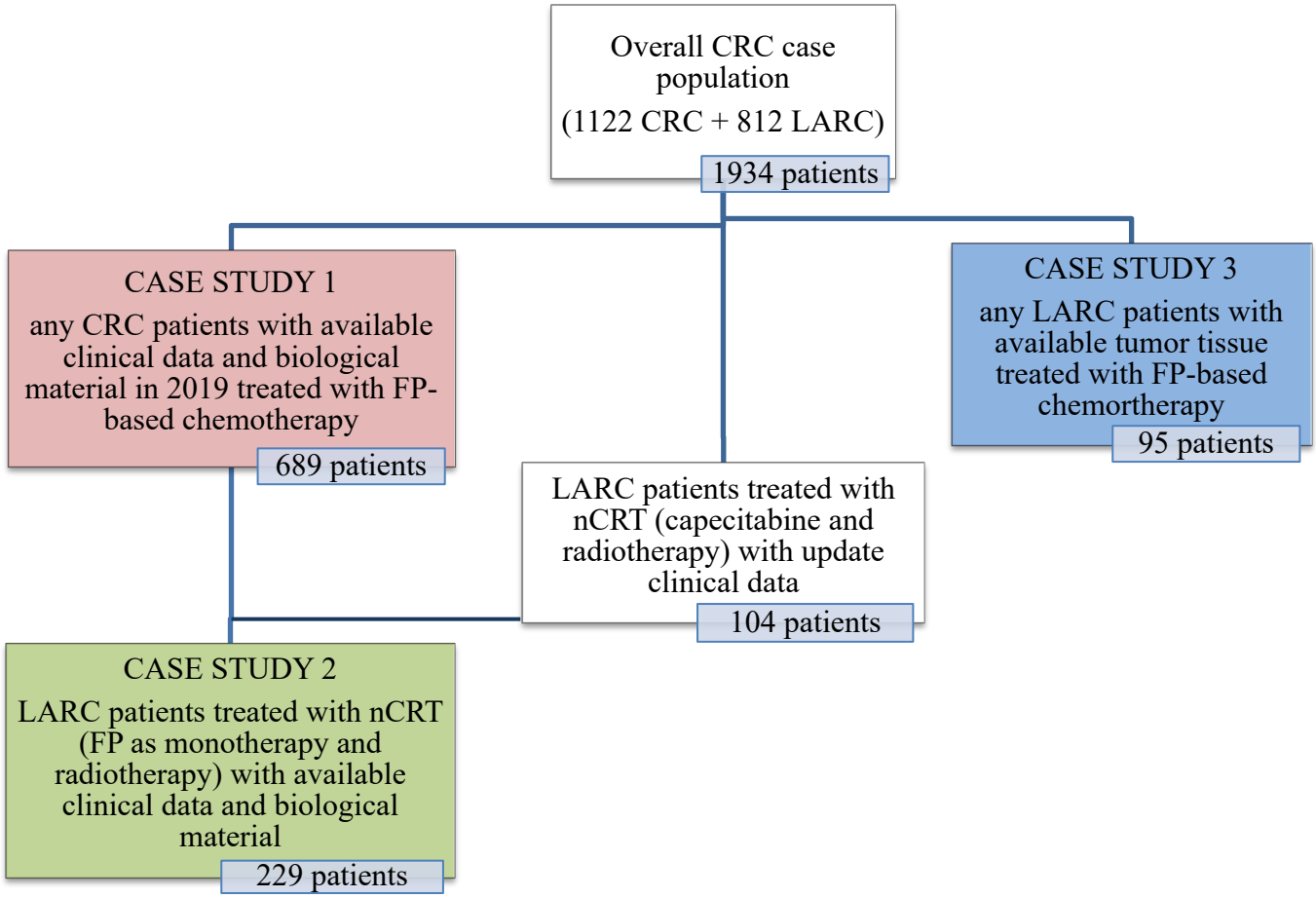


Figure 4.1 PhD project: study design.

Data on the characteristics of the patients included in the case studies are carefully descriebd in the next paragraphs.

4.2 INCLUSION CRITERIA FOR THE CASE STUDY POPULATION

4.2.1 Fluoropyrimidine-based treatment patients' selection for *DPYD* selected genotype analysis (Case Study 1)

From a database of 1,934 clinical cases, we selected patients based on the following inclusion criteria for *DPYD* selected genotype analysis:

- Confirmed diagnosis of colorectal/LARC cancer;
- Peripheral blood/DNA samples availability stored at the Clinical and Experimental Pharmacology Unit Biobank of IRCCS CRO;
- Assumption of FP-based chemotherapy;
- Detailed clinical data availability;
- Age ≥ 18 years.
- Signed written informed consent approved by the local Ethical Committee for clinical and biological data analysis.

4.2.2 LARC patient's selection for NGS analysis (Case Study 2)

A retrospective case population of LARC patients was selected from the entire database of 812 clinical cases admitted to the Radiation Oncology Unit at CRO between 2003 and 2021, according to the following criteria:

- Availability of biological material stored in the Biobank of the Clinical and Experimental Pharmacology Unit Biobank of IRCCS CRO;
- Histologic confirmation of stage II or III rectal adenocarcinoma;
- Clinical confirmation of the absence of visible metastatic sites;
- FP-based nCRT treatment with 5-FU/capecitabine administration as monotherapy regimen and concomitant radiotherapy;
- Availability of detailed clinical data, including TNM stage at diagnosis, CEA levels at diagnosis and after CRT, date of start and end of treatment, date and type of surgery, clinical and pathologic assessment of treatment response, TRG score, date of disease progression; chemotherapy/radiotherapy dose modification-suspension-early discontinuation; days of

suspension; entity of dose reduction and any toxicity with appropriate Grade data that develops during nCRT;

- Age ≥ 18 years;
- Signed written informed consent approved by local Ethical Committee for clinical and biological data analysis.

4.2.3 LARC patient's selection for IHC analysis (Case Study 3)

A retrospective study population of LARC patients was selected from the entire database of 1,378 clinical cases admitted from 2003 to 2021 in the Surgery division at CRO between 2005 and 2014 for IHC analysis. Specific criteria for the selection of the subpopulation included:

- Histopathologically confirmed resectable stage II-III rectal adenocarcinoma;
- Absence of distant metastases;
- Disease stage T3-4 e N0-2;
- Performance status 0-2;
- Normal renal, liver and bone narrow function as determined by biochemical analysis;
- FP-based nCRT treatment;
- Availability of a pre-treatment tumor biopsy;
- Age ≥ 18 years;
- Signed written informed consent approved by local Ethical Committee data analysis.

4.3 PATIENTS DATA COLLECTION AND DATABASE PRODUCTION

The collection of clinical data of patients in the study was complete and usable for analysis for 125 LARC patients and 514 CRC patients at the beginning of my PhD project. For a group of recently enrolled LARC patients, it was necessary to create an updated database and, in collaboration with radiotherapy clinicians, especially with the support of Dr. Elisa Palazzari, clinical data were progressively collected.

Patients medical records were reviewed to collect the following clinical information: (i) baseline patient assessment, (ii) chemotherapy information, (iii) toxicity data, (iv) chemotherapy information including FP-dose, concomitant chemotherapeutic agents, start and end of therapy, therapy discontinuation, (v) toxicity data on each chemotherapy cycle, including severity grade and start date, (vi) patient follow up data for cancer progression and/or death.

4.4 Assessment of the Toxicity Grade to Treatment

ADEs were recorded from the first visit after therapy initiation to the last follow-up visit at the end of CRT. Toxicity was assessed according to the National Cancer Institute CTC-AE (version 5.0).

The NCI CTC-AE is a descriptive terminology which can be used for AE reports. An AE is defined as any unfavourable and unintended abnormal laboratory finding, symptom, or transient condition during a medical treatment or procedure.

A severity scale is given for each AE term. The CTCAE indicates grades 1 through 5 with unique clinical descriptions of severity for each AE based on this general guideline:

- Grade 1: Mild; asymptomatic or mild symptoms; clinical or diagnostic observations only; intervention not indicated;
- Grade 2: Moderate; minimal, local or non-invasive intervention indicated; limitation of age-appropriate instrumental ADL;
- Grade 3: Severe or medically significant but not immediately life-threatening; hospitalization or prolongation of hospitalization indicated; disabling; limiting self-care ADL;
- Grade 4 Life-threatening consequences; urgent intervention indicated;
- Grade 5 Death related to AE.

4.5 Assessment of the Pathological Response to Treatment

TRG of gastrointestinal carcinomas to assess the degree of tumor regression aims to categorize the entity of regressive changes after chemotherapy/radiotherapy treatment. In fact, TRG is the reference system for pathological staging that evaluates the degree of tumor regression which realistically describes the effects of the chemo-radiotherapy. In recent years, objectively determinable

histopathologic findings after patient treatment have emerged as a potential prognostic factor in rectal cancer. The TRG systems according to Mandard, Becker, Dworak, and Rödel are examples of commonly used TRGs [204,205,208,209].

In this work, the TRG scoring system of Mandard was adopted [204]. Residual tumor content in surgical specimens, residual tumor size, degree of differentiation, presence of necrotic and fibrotic content, lymph node status, and development of metastatic lesions over fibrosis are variables included in TRG classification.

Mandard's TRG has been classified into five grades as follow

- TRG 1: CR with absence of residual cancer and fibrosis extending through the tumor margin;
- TRG 2: presence of residual isolated cells scattered through the fibrosis;
- TRG 3: increase in the number of residual cancer cells, but fibrosis still predominant;
- TRG 4: residual cancer outgrowing fibrosis;
- TRG 5: absence of regressive changes.

4.6 BLOOD PROCESSING

Patients, visited by clinicians, belonging to the radiotherapy or oncology departments of CRO, who met the inclusion criteria for the different study populations were enrolled. The clinicians requested the blood collection, and tubes were immediately brought to the Experimental and Clinical pharmacology laboratories. The blood sample was accepted as soon as possible and a unique numeric code for its identification was assigned.

Two 7.5 mL tubes of blood containing ethylenediaminetetraacetic acid, EDTA, as an anticoagulant were collected from each patient (Figure 4.2). Upon arrival at the laboratory, the sample was processed at room temperature (RT) according to the following steps:

1. 200 µL of whole blood were collected and aliquoted in a specific tube for DNA extraction;
2. A centrifugation was performed at RT (~ 21 °C) for 10 minutes at 1,589 RCF without brake to avoid cell lysis (1st centrifugation) for the two 7.5 mL tubes. In this first centrifugation it is possible to separate the plasma at the surface, from the corpuscular part, which settles into a buffy coat in the middle (ring of white blood cells) and a part of red blood cells at the bottom;

3. The supernatant (plasma) was collected, leaving a small aliquot to avoid aspirating the corpuscular portion, and transferred to a falcon tube;
4. The buffy coat was removed after the 1st centrifugation and aliquoted into 2 tubes of 1.5 mL labelled with the patient's code;
5. The plasma was subjected to a 2nd centrifugation at RT (~ 21 °C) for 10 minutes 3,493 RCF (corresponding to the maximum speed of the centrifuge) without brake (2nd centrifugation);
6. The supernatant (plasma) of the 2nd centrifugation was aliquoted (avoiding aspirating the forming pellet) into tubes of 1.5 mL labelled with the patient's code.

Figure 4.2 EDTA blood collection tube.

Specifically, the following tubes were collected for LARC patients' samples:

- 1 STRECK tube (for circulating DNA), hand-signed with the patient code;

- 1 tube for serum, hand-signed with the patient code.

Plasma (also suitable for cfDNA analysis) and buffy coat were obtained from the STRECK tube, whereas serum was obtained from the serum tube containing no anticoagulant (Figure 4.3). To obtain the serum, the whole blood was allowed to clot spontaneously for at least 30 minutes and then centrifuged.

Figure 4.3 Blood sollection tube (without EDTA) and STRECK tube.

The processing procedure for LARC blood samples consisted in:

1. Centrifugation of the STRECK tube at 2,600 RCF (1st centrifugation) for 10 minutes without brake at RT (~ 21 °C);
2. After centrifugation, the supernatant/plasma was transferred into a new falcon with a pipette;
3. The buffy-coat ring was aliquoted into two tubes of 1.5 mL labelled with the patient's code;
4. Centrifugation of the falcon containing plasma (2nd centrifugation) together with the serum tube at 4,000 RCF for 10 minutes without brake at RT (~ 21 °C);
5. The plasma was aliquoted into three tubes of 1.5 mL. Similarly, the serum was aliquoted into three tubes of 1.5 mL labelled with the patient's code;
6. At the end of the processing, 3 tubes of plasma, 3 of serum, and 2 of buffy coat were obtained.

The tubes containing the buffy coat and the plasma were stored at –80 °C according to the storage protocol.

4.7 GENOMIC DNA EXTRACTION AND QUANTIFICATION

DNA extraction is a crucial step for molecular analysis in the diagnostics, forensics or research. The extraction can be performed starting from whole blood, buffy coat or tissue according to specific methods. The procedure is divided into four steps, which are summarized below:

- Cell lysis by enzymatic digestion;
- Inactivation of DNases and RNases that hydrolyze nucleic acids released from the nucleus;

- Isolation of nucleic acids by both organic solvents and adsorption in silica gel matrices in the presence of chaotropic salt;
- Elution of nucleic acids.

Germinal DNA was extracted by two different techniques: with an automated method using the EZ1 DNA Blood 200 µL Kit and BioRobot EZ1, and a manually via specific silica gel columns from the GeneJet Whole Blood DNA Purification Mini Kit No. K0871 50prep from Thermo Scientific™.

4.7.1 Automated extraction method

Genomic DNA extraction can be performed using the automated extractor, BioRobot® EZ1 system workstation (Qiagen N.V., Germany) and EZ1 DNA Blood 200 µL Kit (48) cat. no. 951054 and Biorobot® EZ1 DNA blood card. The BioRobot® EZ1 workstation allows the extraction of highly purified DNA from whole blood, tissue, and buffy coat samples, up to a maximum of 6 samples per run (Figure 4.4). The purity of the extracted DNA enables its use for subsequent laboratory analysis (e.g., PCR, sequencing). The extraction process of the BioRobot® EZ1 workstation exploits an aqueous liquid formulation with neutral pH for the separation of the DNA and for the elution of the nucleic acids.

For the analysis, the extraction was performed starting from 200 µL of buffy coat or whole blood to obtain a final DNA volume of 200 µL. The instrument automatically performs with the extraction phases using the reagents contained in a single cartridge (magnetic beads, lysis buffer, washing buffer, and elution buffer). The cartridge was placed in the appropriate holder in the same number as the DNA to be analysed.

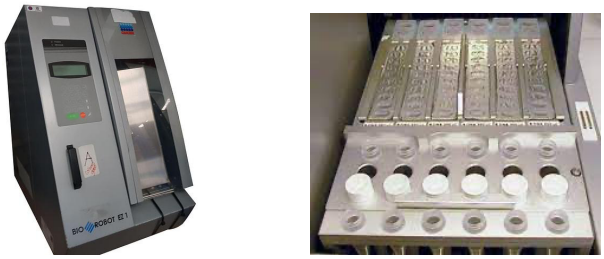


Figure 4.4 BioRobot® EZ1 system workstation (Qiagen N.V., Germany) and preparation of DNA extraxtion using EZ1 DNA Blood 200 µL Kit (48).

The extraction process takes about 20 minutes. The procedure is schematically represented in Figure 4.5 and consists of the following steps:

1. Lysis of white blood cells by adding the buffer containing Proteinase K (extracted from *Tritirachium album*);
2. Addition of the magnetic beads coated with silica;
3. Binding of DNA to the silica surface of the magnetic beads in the presence of a chaotropic salt by mixing the solution;
4. Magnetic separation to remove excess lysate;
5. Washing of the DNA-bound to the magnetic beads with washing Buffer and its subsequent elimination by magnetic separation;
6. Addition of Elution buffer;
7. Elution of 200 µL of DNA solution into the appropriate test tubes.

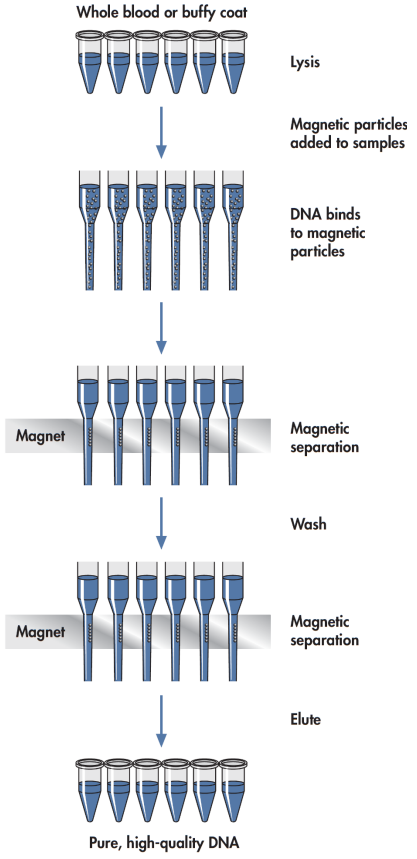


Figure 4.5 Phases of DNA automated extraction with magnetic beads.

4.7.2 Manual extraction method

Manual DNA extraction was performed starting from 200 µL buffy coat or whole blood sample and elution was performed in 200 µL.

The extraction was performed using silica gel columns from the GeneJet Whole Blood DNA Purification Mini Kit No. K0871 50prep from Thermo Scientific™.

The steps for the extraction are as follows:

1. Addition of 20 µL of Proteinase K to 200 µL of whole blood or buffy coat;
2. Addition of 400 µL of the Lysis Solution buffer;
3. Incubation at 56 °C for 10 minutes;
4. Transfer of the solution to a column connected to a 2 mL tube;
5. Centrifugation at 8,000 RCF for 1 minute;
6. Addition of 500 µL of buffer WB1 (washing buffer 1) and centrifugation at 8,000 RCF for 1 minute;
7. Addition 500 µL of buffer WB2 (washing buffer 2) to a new column and centrifugation at 14,000 RCF for 3 minutes;
8. Centrifugation for 1 minute at 14,000 RCF to eliminate the residual WB2;
9. Transfer the column to a new 1.5 mL tube with lid and addition of 210 µL of galenic water;
10. Incubation at RT for 10 minutes, followed by centrifugation at 8,000 RCF for 1 minute.

The extracted DNA can be stored at a temperature of +2-8 °C or stored in the freezer at –80 °C.

4.8 DNA SAMPLE QUANTIZATION

DNA extracted from blood was quantified using NanoDrop 2000c, Thermo Fisher Scientific, Wilmington, DE, U.S.A, a full-spectrum UV-VIS spectrophotometer, to assess sample purity of the sample and calculate DNA concentration, and Quantus Fluorometer (Promega).

The NanoDrop 2000c spectrophotometer measures the absorbance at three different wavelengths (260 nm, 280 nm, 230 nm) of small sample volumes (up to 0.5 µL). The absorbance at 260 nm is

proportional to the concentration of DNA in solution, while the 260 nm/280 nm and 260 nm/230 nm absorbance ratios evaluate sample purity by providing indication of the presence of contaminating proteins in the sample. The optimal values for A260/A280 and A260/A230 ratios are 1.8 and 2.0, respectively. Indeed, phenol contamination could cause anomalous spectra between 220 and 240 nm and spectral shifts at values from 260 to 280 nm. Conversely, guanidine residues or proteins could alter the peak at 230 nm and cause a shift from 230 nm to about 240 nm. Analysis of the concentration (expressed in ng/µL) and quality of DNA based on a spectrophotometric method allows the desired parameters to be read using 1 µL of DNA placed in the appropriate optical reader of the instrument.

Quantification of the samples with the NanoDrop 2000 spectrophotometer is carried out in the following steps (Figure 4.6).

First, a blank measurement was performed by adding 1 µL of deionized water to the lower optical surface and selecting the “blank” option in the software. The result of the blank measurement should be 0 to proceed with the measurements of the samples.

Optical surfaces must be cleaned between each measurement. After the blank mesurement, 1 µL of the nucleic acid sample was added to the lower optical base and the lever arm was closed. Selecting "Measure", the software automatically calculates the nucleic acid concentration and purity ratios. After measuring the sample, it is necessary to examine the spectral image to assess the quality of the sample.

Figure 4.6 NanoDrop 2000 Spectrophotometer (Thermo Fischer Scientific, USA).

In addition, sample concentrations were calculated using a fluorimetric method involving the use of the Quantus Fluorometer Instrument (Promega) and QuantiFluor® ONE dsDNA Dye Reagent (Promega) (Figure 4.7). The Quantus™ Fluorimeter is a compact dual-channel fluorometer designed

for high-sensitivity fluorescence detection in nucleic acid quantification and is recommended for next-generation sequencing (NGS) applications.

Figure 4.7 Quantus Fluorometer instrument (Promega).

The reagent must be equilibrated at RT for 30-40 minutes before use, taking care not to expose it to light as it is photosensitive. After equilibration, sample concentration evaluation is performed at RT.

After selecting the desired protocol from the instrument and setting the volume/units and sample volume for analysis, we sterted evaluationg the DNA concentration from 1 µL of DNA sample.

The sample preparation involves mixing of 1 µL of DNA of unknown concentration and 199 µL of QuantiFluor® ONE dsDNA Dye in a suitable 0.5 mL PCR tube. The obtained sample must be vortexed in the Vortex, STARLAB Smart Instrument. After a light protected incubation of 5 minutes at RT protected from light, the samples can be loaded into the tube holder. When you closed the lid, the instrument automatically measures the fluorescence of the sample, and the calculated nucleic acid concentration is displayed.

4.9 DNA PURIFICATION

To increase DNA purity and improve A260/A280 and A260/A230 ratios, genomic DNA was purified using Agencourt AMPure XP Beads (Beckman Coulter, Brea, CA, USA). Starting with 400 ng of DNA in 50 µL and using 100 µL beads, the purified DNA was reanalysed for the above parameters. This step is useful to improve the performance of the subsequent analysis and to remove any reagents from the previous steps that could affect the following reactions.

DNA from selected patients, which was tested for quality using the Nanodrop and quantified using Quantus, was subsequently purified. The method consists of purifying the DNA samples with magnetic beads to obtain DNA of appropriate quality for NGS analysis, as required by the Roche protocol. Specifically, the DNA sample must be free of EDTA contamination and have optimal absorbance ratios, as measured by the Nanodrop (values for A260/A280 between 1.7-2.0 and for A260/A230 between 1.8-2.2).

The following reagents and materials are required for sample purification:

- Agencourt AMPure XP Beads, brought at RT, vortexed and kept away from light until used;
- Fresh Ethanol solution (80%) using absolute Ethanol and steril water;
- Sterile water S.A.L.F. 100 mL package (order from DSC), present on the DNA counter;
- 96 PCR plate without skirt, Sarstedt;
- Ambion RNA Magnet by Life Technologies-Magnetic Stand-96, DynaMag Magnet - 96 Side Skirted Invitrogen by Thermo Fisher Scientific;
- Centrifuge 5810 from Eppendorf.

Quantus Fluorometer (Promega, Madison, WI, U.S.A.) with the QuantiFluor dsDNA Dye (Promega, Madison, WI, U.S.A.) data were used to calculate the initial amount of DNA to be purified. For DNA purification, we assumed a total initial amount of aproximately 400 ng (in mass) of DNA. DNA volume (µL) was calculated for each sample using the following formula:

$$\mu L\ DNA = 400 / Quantus\ concentration\ (ng/\mu L)$$

DNA volume for each patient was aliquoted into a 96 PCR plate without skirt, Sarstedt, and brought to a final volume of 50 µL with sterile water. A duplicate volume of Agencourt AMPure XP Beads was added to each sample, mixed, and incubated for 5 minutes at RT. This step allows the beads to bind only to the DNA. The 96 PCR plate is moved to the Ambion RNA magnet from Life Technologies-Magnetic Stand-96 to allow the beads to form the pellet. The pellet formation takes approximately 5 minutes at RT. When the solution became clear, the supernatant was discarded, and two successive washes with 80% ethanol were performed on the magnet: 200 µL of 80% ethanol was quickly added to each well and after 30 seconds of incubation – during which the plate was rolled on the magnet – the supernatant was discarded. Immediately thereafter, a second wash step was performed using the same procedure. After the second wash, the supernatant was completely removed, the 96 PCR plate was breafly centrifuged and placed back on the magnet. This rapid centrifugation step is performed to remove any residual 80% ethanol on the bottom of the wells. Leaving the plate on the magnet, allow the beads pellet to dry for 5 minutes at RT until it appears flat and opaque, avoiding excessive dry.

By removing the plate from the magnet, each sample is eluted in 23 µL of sterile water by adding the water directly to the beads pellet and gently mixing until the beads are completely resuspended. The

plate was then placed on the DynaMag - 96 Side Skirted Invitrogen by Thermo Fisher Scientific Magnet for approximately 2-5 minutes at RT. Finally, 20 µL of the supernatant containing the DNA is remuved and aliquoted into a well of the column to obtain 20 µL of purified DNA.

The concentration and quality of the DNA samples after purification were reassessed using Nanodrop and Quantus analysis.

4.10 GENETIC ANALYSIS

4.10.1 Polymerase Chain Reaction (PCR)

The Polymerase Chain Reaction (PCR), is a technique developed by K. Mullis in 1983 and allows to selectively amplify specific DNA sequences. It is the first step of pyrosequencing analysis [210].

The principle of PCR is based on the use of the endonuclease activity of the enzyme DNA polymerase and is used in numerous analytical techniques including sequencing and single mutation analysis. The PCR reaction requires a single-stranded DNA template, specific primers and the enzyme polymerase.

Primers are short sequences of different nucleotides in the forward and reverse directions based on the sense or antisense strand to be synthesized. They are designed to selectively match selectively the sequence to be amplified. These short sequences appear at the 5' and 3' ends of the fragment of interest and serve as starters for the amplification reaction.

The required reagents for a PCR are:

- A buffer solution (Tris-HCl and HCl) which set up the optimal conditions for the reaction;
- $MgCl_2$: the magnesium ion in the PCR reaction allows the polymerase to be in the active spatial conformation. Its concentrations in the reaction mix can vary (1.5-2.5 mM) and are therefore evaluated when setting PCR conditions;
- DNA-polymerase enzyme: the enzyme enables the sequential incorporation of complementary nucleotides into ssDNA. Consequently, a double-stranded DNA will be synthesized from the ssDNA. In subsequent cycles, the newly formed dsDNA serves as a template for 2 ssDNA, which forms the template for the subsequent elongation reaction. Generally, the most commonly used enzyme is Taq polymerase, which is derived from the bacterium *Thermophilus Acquaticus*. The maximum enzymatic activity occurs at 72 °C where it is capable of incorporating 50-60 nucleotides per second. It requires an activation phase as it is supplied in an inactive state;

- dNTPs: dNTPs are triphosphate nucleotides used in the formation of the new DNA chain. Namely during the polymerization reaction, two phosphates are removed from, the triphosphate nucleoside and it is therefore incorporated as a monophosphate nucleotide. The optimal concentration is 0.2 mM in the reaction mix. Higher or lower concentrations may lead to incorporation errors or low polymerization efficiency;
- Primers: primers are short sequences of about 20 bases specifically designed to amplify target regions of the genome. Primer design is usually performed using Primer3plus software. The basic properties of the primers on which the efficiency of the PCR reaction depends, are the individual melting temperature of each primer which must be comparable to avoid the formation of dimers, and the length which must be suitable for a specific coupling. When designing the primers, some requirements have to be considered:
- The annealing temperature (T_a) of the probe in the elongation phase of RT PCR;
- The melting temperature (T_m) of the primers must be at least 5 °C higher than the T_m of the two primers;
- The primers must have a length of about 20-30 bp and a G-C content of 50%;
- The elongation temperature phase must take place at a temperature lower than the usual temperature of 72 °C used in PCR;
- The primers must not form dimers or even pair with themselves.

The preparation of the PCR reaction mix must avoid contamination and must be performed in a laboratory other than the one where the DNA is normally stored.

Therefore, the prepared mix was aliquoted into 96-well plate and the DNA to be analysed was added.

Amplification of the selected DNA sequence includes the following steps:

1. A denaturation phase in which the temperature is brought to 95 °C and generally maintained for 10 minutes during which the two paired DNA filaments are separated into two single filaments by breaking the hydrogen bonds between the nitrogen bases. DNA polymerase is also activated during this phase;
2. An amplification phase that allows to achieve the exponential amplification of the fragment of interest by repeating an n-fold number of thermal cycles, where 3 steps can be distinguished:

a. Primers pairing (annealing): the temperature is brought to 63-65 °C to allow primers to anneal to the complementary ssDNA strand. The annealing temperature is calculated based on the melting temperature of the primers and evaluated during reaction setup;

b. Elongation (extension): increasing the temperature to 72 °C promotes the polymerase activity of Taq-polymerase which, starting from the primer, starts inserting the dNTPs complementary to the mold filament;

c. Termination phase: in this last step the temperature is kept constant at 72° C for 5 minutes to complete the elongation.

At the end of the PCR reaction, 2*n*-2 times of the starting material is obtained, where *n* is the number of amplification cycles.

PCR reaction thermocycler condition for the analysis of *DPYD* variants rs3918290 IVS14+1G>A (*2A); rs55886062 1679T>G (*13); rs67376798 (c2846A>T) and rs75017182 c.1236C>T /HapB3 are listed in Table 4.1.

Table 4.1 PCR condition.

Reaction condition	N° of cycles
10' at 95 °C	X 1
30'' at 95 °C	X 35
30" at 56 °C	
1' at 72 °C	
7' at 72 °C	X 1

4.10.1.1 PCR evaluation

To evaluate the amplification process, electrophoresis of the amplification products is performed on agarose gel. This technique allows to verify the efficiency of the amplification and the fairness of the obtained fragments by comparing their weight with a molecular weight marker loaded in parallel into the gel. Briefly, this technique takes advantage of the electrical charge and size of DNA molecules, which, when exposed to a magnetic field, can migrate at different speeds in the agarose matrix. Agarose is a linear polymer that, when hot dissolved in a specific TAE buffer and than cooled, solidifies to form a semisolid and porous matrix. An intercalating agent, GelRed Nucleic Acid Gel Stain 10,000X in water, is added to this mixture to make the DNA fragments visible as it binds them

and absorbs UV rays by re-emitting in the visible spectrum DNA molecules, which are negatively charged due to the phosphate groups, could migrate from the negative to the positive pole when exposed to a magnetic field. Given the uniform charge of the DNA, the speed of migration hardly depends on the size of the DNA fragments. In our case, the electrophoresis is performed in 3% agarose gel in TAE buffer with a 100 V electric field. After the gel is prepared and solidified in special containers with combs forming wells for the sample distribution, the gel was immersed in the electrophoresis chamber containing TAE buffer. 10 µL of each sample is mixed with loading buffer (Gel Loading Buffer 10X Blue Juice, a compound of glycerol, bromophenol blue, SDS and EDTA that allows the DNA to precipitate inside the well) and loaded into the agarose gel to evaluate the migration front. At the end of the electrophoresis run, the bands are visualized using the transilluminator, Bio Rad gel doc EZ Imager(fig).

Biotinylated reverse or forward primers were used for pyrosequencing as streptavidin-coated Sepharose beads were used to immobilize DNA fragments.

Since streptavidin-coated Sepharose beads were used to immobilize DNA fragments, reverse or forward biotinylated primers were used to take advantage of biotin-streptavidin binding capacity

Thus, the amplification products evaluated with electrophoresis, constitutes the starting material for numerous subsequent (or consecutive) analysis including pyrosequencing and sanger sequencing

4.10.2 Pyrosequencing

Pyrosequencing is a technique that allows the identification of the presence of SNPs in a DNA fragment previously amplified by PCR. The rapid and efficient analysis allows simultaneous testing of many samples and different SNPs with different primers, called sequence primers.

The starting material for pyrosequencing are dsDNA fragments with a biotinylated end (especially the 5' end). This feature is exploited in the pyrosequencing phase to isolate the single DNA strand by biotin-streptavidin interaction with the beads.

After confirming the fairness of the PCR reaction, the pyrosequencing analysis is performed.

The principle of the technique is based on light emission resulting from the release of pyrophosphate groups released by the dNTPs each time they are incorporated into the nascent DNA strand complementary to the template.

The reagents necessary for the analysis are the following:

- PyroMarkQ48 Magnetic Beads (300), QIAGEN, streptavidin-coated sepharose beads;
- Four nucleobases: adenine (A), cytosine (C), guanine (G), thymine (T);
- Annealing buffer;
- Denaturation solution;
- Binding buffer;
- Lyophilized enzyme and substrate rehydrated with 660 µl of annealing buffer and stored at -20 °C;
- PyroMarkQ48 Discs. Each disc contains 48 wells that can be used to perform 4 analyses for 12 samples simultaneously
- PyroMarkQ48 Absorber Strips stored at RT used to absorb the reagents during analysis, where the disc rotates at 6 RCF, or during the cleaning step. Only the DNA bound to the magnetic beads is retained in the wells, the rest is ejected and absorbed by the strip;
- Sequence specific primer: the lyophilized primer is rehydrated with water for injection and brought to concentration of 100 µM, subsequently diluted to 40 µM with annealing buffer and stored at -20 °C. It must be further diluted to 4 µM with annealing buffer before the analysis;
- Pyrophosphatase inorganic BM0361L 100 units/mL, 50 units, stored at -20 °C.

The instrument used is the Pyromark Q48 Autoprep, which automatically sets the correct volume of reagents and performs the analysis. The pyromark Q48 Autoprep uses proven sequence-based real-time Pyrosequencing technology. It can be used for detection and quantification in genetic analysis and epigenetic methylation studies, analysing up to 48 samples simultaneously. It is a user friendly and automated protocol that prepares single-stranded DNA samples without manual intervention by the user. Pyrosequencing reaction includes 5 steps starting from a dsDNA previously amplified by PCR using a biotinylated primer. The biotinylated strand serves as the pyrosequencing template.

1. After denaturation, the biotinylated single-stranded PCR strand is isolated due to the binding of biotin with the streptavidin coated beads;
2. The hybridized primer and single-stranded template are incubated with the DNA polymerase enzyme, ATP sulfurylase, luciferase and apyrase, and the substrates adenosine 5' phosphosulfate (APS) and luciferin;
3. The first deoxyribonucleotide triphosphate (dNTP) is added to the reaction and complementarily incorporated to the DNA template by DNA polymerase;

4. Each incorporation causes the release of pyrophosphate (PPi) in an equimolar amount to that of the incorporated nucleotide (Figure 4.8);

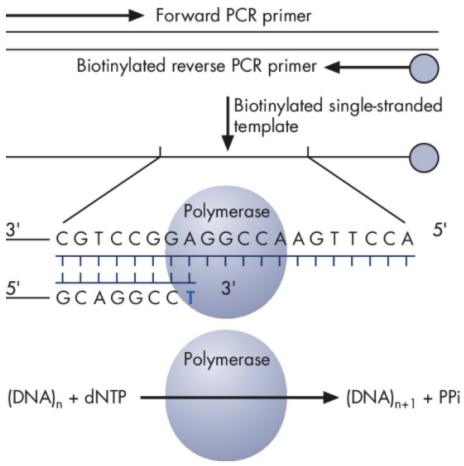


Figure 4.8 Pyrophosphate (PPi) release reaction after each dNTP incorporation.

5. ATP sulfurylase converts PPi to ATP in the presence of adenosine 5'-phosphosulfate (APS). This ATP is used in the luciferase-mediated conversion of luciferin to oxyluciferin. This reaction generates visible light proportional to the amount of ATP (Figure 4.9). The light is detected by CCD sensors and displayed as a peak in the raw data output (Pyrogram). The height of each peak (light signal) is proportional to the number of nucleotides.

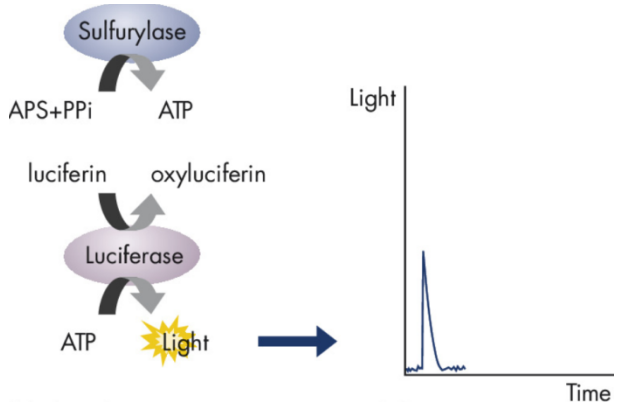


Figure 4.9 Chemical reaction for light signal production.

Unincorporated nucleotides are degraded by apyrase, a nucleotide-degrading enzyme. When degradation is complete, dNTPs are added sequentially.

The sequence of the DNA fragment is determined from the signal peaks in the Pyrogram (Figure 4.10).

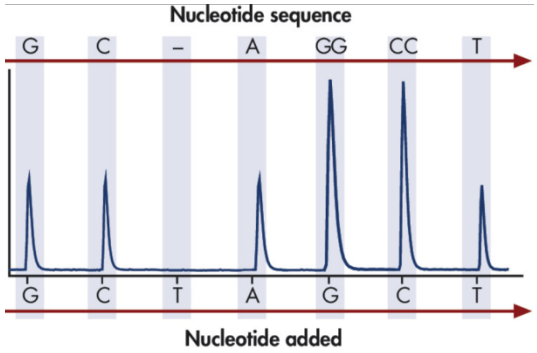


Figure 4.20 Example of a Pyrogram.

Pyrosequencing analysis is performed for the determination of 4 *DPYD* variants: rs3918290 IVS14+1G>A (*2A); rs55886062 1679T>G(*13); rs67376798 (c2846A>T) and rs75017182 c.1236C>T /HapB3.

The sequence primers for each *DPYD* variant are listed in Table 4.2

Table 4.2 Sequences of forward and reverse primer for *DPYD* *2A, *13, c.2846A>T and c.1236C>T for PCR.

SNP	Forward primer sequence	Reverse primer sequence
***2A**	CGGCTGCATATTGGTGTCAA	[Btn]CACCAACTTATGCCAATTCTCTTGT
***13**	[Btn]CCTTTTGGTCTTGCTAGCGC	AGTTTTGGTGAGGGCAAAACC
c.2846A>T	[Btn]GCAGTACCTTGGAACATTTGGT	AGGTCATGTAGCATTTACCACAGT
c.1236C>T	[Btn]AACCAAAGGCACTGATGA	AATTTCTGCCATTCCTGTCC

For the master mix preparation, the Kit Ampli Taq Gold®DNA polymerase Applied Biosystems (by Thermo Fisher Scientific) was used including 10X PCR Gold Buffer, $MgCl_2$ Solution 25mM and Ampli Taq GoldTM 5U/µl. dNTP set, 100mM of A-C-G-T PROMEGA are diluted to a final concentration of 25 mM (Table 4.3).

Table 4.11 Reagents conditions for *DPYD* *2A, *13, c.2846A>T and c.1236C>T PCR.

Reagents condition per 1 sample (volume)	
Buffer	**3,0 µL**
$MgCl_2$	**3,0 µL**
Taq polymerase	**0,3 µL**
dNTPs	**0,2 µL**
Forward primer	**0,2 µL**
Reverse primer	**0,2 µL**
Water	**23,1 µL**

4.10.3 Real Time PCR (RT PCR) for 3 new variants of DPYD gene

Real-time PCR is a quantitative molecular biology laboratory technique that allows allelic discrimination of two variants of a single nucleic acid sequence. It is possible to monitor the amplification of a targeted DNA molecule during PCR, and this technique provides both quantitative, semi-quantitative, and qualitative information about the DNA sample. Real-Time systems allow simultaneous analysis of DNA fragments and time tracking of the amplification process. It combines a thermocycler with a fluorometer and makes it possible to monitor in real time the amount of amplification products formed in the exponential phase of amplification and also to obtain an estimate of the initial DNA concentration (quantitative PCR). The detection of amplified DNA is possible with specific probes for the fragments to be amplified, which generate fluorescence signals.

The ABI7500 Real Time PCR System allows simultaneous genotyping of many samples as it can load a 96-well plate.

The basic principle of this technique is based on the use of (i) non-specific fluorescent dyes that intercalate with each dsDNA during PCR and (ii) sequence-specific primers consisting of oligonucleotides labelled with a fluorescent reporter that allows detection only after hybridization of the probe with its complementary sequence.

Potential applications of real-time PCR analysis include SNP genotyping, quantification of gene expression, quantification of viruses (e.g., SARS-CoV-2, HIV), assay quality control and validation, bacteriological analysis, control of OGMs and mitochondrial DNA and miRNA analysis.

The reaction components include:

- Target DNA;
- Two oligonucleotides;
- Fluorescent probe;
- Universal master mix containing buffer, $MgCl_2$, DNA polymerase and dNTPs.

TaqMan® probes are dual-labelled oligonucleotide fragments complementary to the target sequence within the PCR primers. The probe is labelled with a fluorophore at the 5‘ end and a quencher at the 3’ end. The 5' reporter is a high-energy dye that emits fluorescence, while the 3' quencher is a low-energy dye that turns off fluorescence of the reporter by energy transfer. The distance between the fluorophore and the quencher is sufficent to block the emission of fluorescence (FIG).

The probe is designed to hybridize with the fragment amplified in the PCR reaction. When DNA-polymerase encounters the hybridized probes due to given its 5' → 3' exonuclease activity during the elongation PCR phase, it begins to degrade them (Figure 4.11).

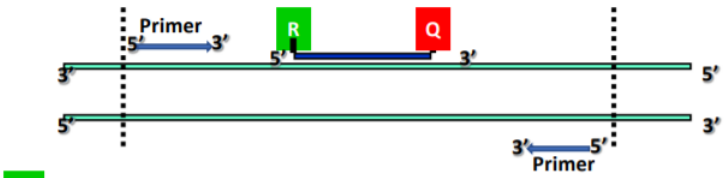


Figure 4.11 Example of conditions with TaqMan probe and primer paired to DNA.

The separationof the reporter and the distancing from the quencher molecules results in the emission of fluorescent signal that is automatically registered by the instrument. Accumulation of the amplified product is indicated by observation of the fluorescence increase of the reporter (Figure 4.12).

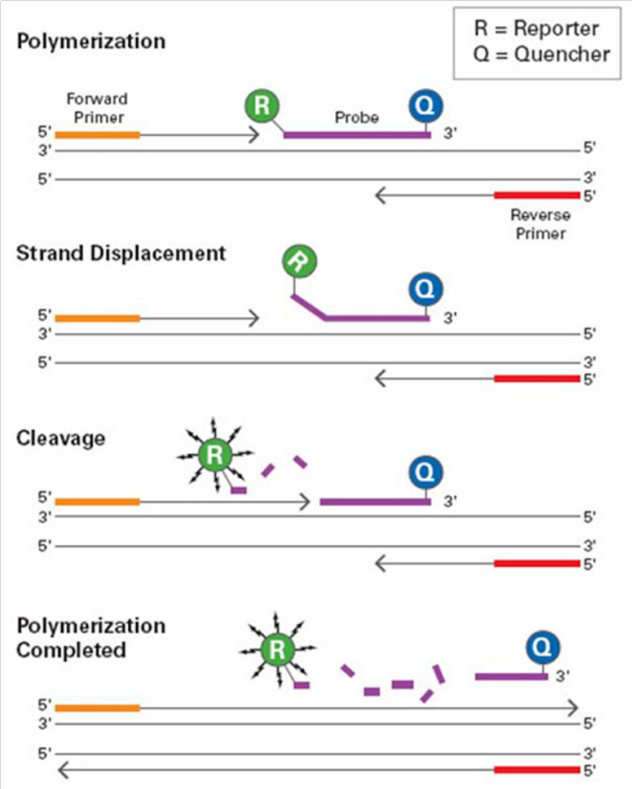


Figure 4.22 Detachment of the reporter from TaqMan probe during PCR.

By recording the fluorescence emission for each cycle, it is possible to monitor the PCR: the emission of the neo-synthesized products is directly proportional to the initial concentration of the DNA template in the sample.

When designing PCR primers, the annealing temperature (Ta), the melting temperature (Tm), the probe length and the dimer formation are key parameters.

Spectrophotometric data of fluorescence emission are acquired at a wavelength between 500 and 600 nm and the software converts the raw data, expressed as fluorescence signal versus λ, into clean colour signals using specific algorithms. The data provided by the Taq Man® assay is diplayed as a two-dimensional graph. Homozygous mutated, heterozygous, and homozygous wild-type samples are distinguished in coloured groups where each dot represents a sample. Generally, a No Template Control (NTC), a homozygous mutated control and a heterozygous control are loaded simultaneously with the samples. Since the NTC does not contain DNA, it should not develop a fluorescent signal, and in the diagram its position should coincide with 0 of the Cartesian axes (Figure 4.13).

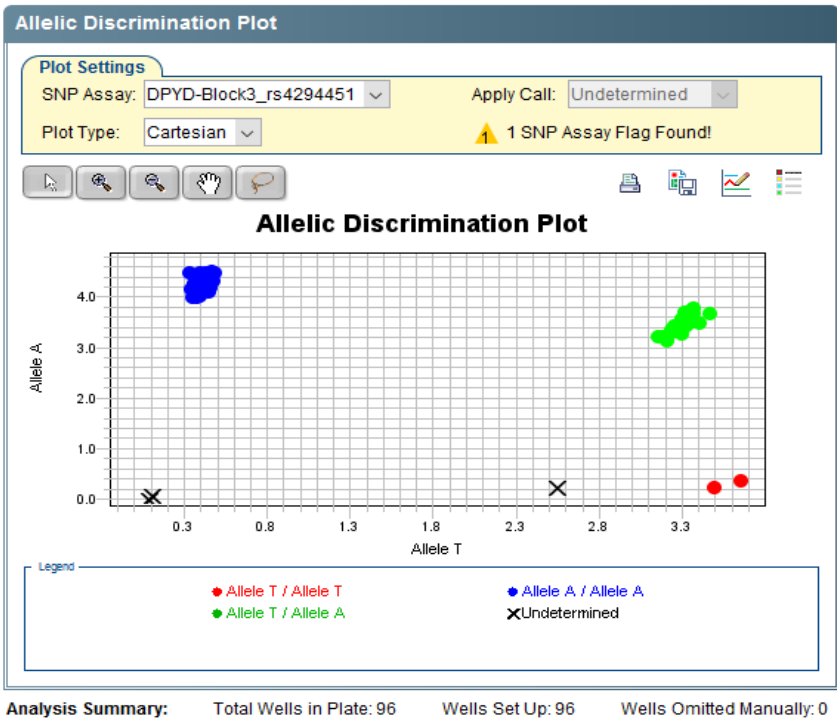


Figure 4.13 Genotyping cluster plot using Real-Time technology.

Real-Time analysis is performed for the assessment of *DPYD* variants rs59353118, rs114170368 and rs4294451 using the ABI7500 Real Time PCR System. The specific assay mix contains primers (forward and reverse) and probes for the analysis of the 3 novel *DPYD* haploblock. The primers and probe sequences are listed in Table 4.4.

Table 4.12 Sequences of forward and reverse primer for rs59353118, rs114170368 and rs4294451 for Real-Time analysis.

SNP	Amplificated sequence	Chromosome variant Position
rs59353118	AAATAGCCTATTTTCTTTTTTTTTT [A/T] AATTATTATACTTTAAGTTCTAGGG	Chr.1: 97393302 on GRCh38
rs114170368	TTCCCACTCCCTGGTACAAT [C/T] CAAAAAGTGATGGAAGAAAC	Chr.1:97569885 on GRCh38
rs4294451	TTTCCTTTTTTTATTTATTTTTCTT [T/A] TCTGGGGGTTGGTGTTGCGGGTGCT	Chr.1: 97930158 on GRCh38

The analysis involves the preparation of a reaction mix by combining the assay-specific mix with the master mix. 15 µL of the master mix is aliquoted into a 96-well plate and finally 1 µL of the genomic

DNA samples was added. The plate was covered with an adhesive seal and then centrifuged for 1 minute to remove air bubbles at the bottom of the wells. The plate was placed in the ABI7500 Real Time PCR System instrument for analysis. The thermal conditions of the RT PCR are as follows (Table 4.5).

Table 4.13 Reagents conditions for rs59353118, rs114170368 and rs4294451 for Real-Time analysis.

Reaction condition	N° of cycles
5' at 50 °C	X 1
10' at 95 °C	X 1
15" at 92 °C	X 50
1' at 60 °C	

4.10.4 NGS Analysis

Rationale

The Next Generation Sequencing (NGS), or so-called second-generation sequencing, indicates a series of time-spending technologies that allow for the sequencing of large genomes. Therefore, it is possible to quickly obtain information regarding the DNA of organisms, animals, and plants, fundamental in the studies of medical, molecular, population, conservation genetics, and genomics. At the same time, it is possible to characterize the genomes of many patients, identify balanced and unbalanced chromosomal rearrangements, identify deletions, and copy number variations (CNV), evaluate the presence of SNPs and characterize them later through a bioinformatics analysis.

By means of the NGS technique, it is possible to analyse the entire exome, the entire genome, targeted genes, amplicons previously produced in a PCR reaction, the transcriptome intended as total RNA, mRNA, or small RNA (<30 nt), epigenomic variants and the methylation profile.

The technique provides the preparation of the library according to the protocol and its sequencing. The NGS, is divided into four main phases:

1. Library preparation;

2. Sequencing;
3. Data analysis.

Before starting the NGS workflow, purification of nucleic acid is highly recommended since some DNA extraction methods can introduce inhibitors, which negatively affect the enzymatic reactions that occur in the NGS workflow.

The first step of library preparation is crucial to the success of the NGS workflow. Generally, libraries are created by randomically fragmenting DNA and adding specific adapters and unique indexes to both ends. In the sequencing workflow, these adapters contain complementary sequences that allow the DNA fragments to bind to the flow cell. To save resources and time, multiple libraries can be pooled together and sequenced in the same run. This process is called multiplexing. During adapter ligation, unique index sequences are added to each library which allow to distinguish single samples within the library during data analysis. Thus, fragments generated can then be amplified and purified and they represent the starting sequencing library.

During the second step of the NGS workflow - the sequencing -, multiplexed libraries at appropriate concentrations are loaded into a flow cell and placed on the sequencer. The sequencing process starts with the cluster's generation process in which each DNA fragment bounded to the flow cell is amplified. This amplification process is known as "bridge amplification" resulting in millions of copies of single-stranded DNA around, equal to the original fragment. Specifically, the strand fragments complementary to the single strand fragment hybridized to the flow cell are synthesized. Due to the presence on the flow cell surface of multiple adapter complementary sequences, the neo-formed double strand fragment folds back to form a bridge-like structure. A denaturation step allows to obtain two single strand filaments adhered to the flow cell. The synthesis-folding and denaturation process continues until the generation of clusters. In the end, linearization is performed to prepare filaments for sequencing.

In the step called sequencing by synthesis (SBS), chemically modified nucleotides bind to the DNA template strand through natural complementarity. Each nucleotide contains a fluorescent tag and a reversible terminator that blocks the incorporation of the next base. Once a dNTPs is incorporated into, a fluorescence signal is generated, and the emission is detected. The software can record the different fluorescence emissions and convert them into the corresponding nucleotide sequence. Sequencing can take place starting from a single end (single-end reads) or from both ends (paired-end reads). Regarding paired-end sequencing, this method consists in the sequencing of the forward DNA strand, the wash of the reads, and the sequencing of the reverse DNA strand.

After sequencing, the core of the entire process is involved the data analysis. It is possible to import sequencing data into a standard analysis tool or set up a specific pipeline. These tools provide sequence alignment, variant calling, data visualization, or interpretation.

The analysis of the data can be divided into five steps:

1. Quality assessment of the raw data;
2. Read alignment to a reference genome;
3. Variant identification;
4. Annotation of the variants;
5. Data visualization.

Following the sequencing, the instrument software identifies nucleotides in a step called base calling, and the predicted accuracy of those base calls. From this process, we obtained FASTQ data, a text-based format for storing biological sequence and its corresponding quality scores. It has become the standard for storing the output of high throughput sequencing instruments.

FASTQ contains quality information as Phred quality scores Q, defined as a property logarithmically related to the base-calling error probabilities and P, defined as the probability that the corresponding base call is incorrect. In the quality assessment step, the quality of NGS reads is evaluated and reads that do not meet specific standard points are removed, corrected, or trimmed. Indeed, during the sequencing process some errors in the incorporation can occur, leading to calling errors and the generation of poor-quality reads. Given to softwares such as FASTQC, in this step the abovementioned errors are identified by a score quality consideration.

Secondly, after the quality assessment, the reads are aligned to the reference genome associating the precise location in the genome from which each base pair (bp) of each sequencing read comes. The Sequence Alignment/Map (SAM) format resulting from this step is a file for the storage of sequence alignments and their mapping coordinates.

Following the quality assessment and the reads alignment, the step of variant identification and annotation is performed where variants are identified using different tools. Based on the starting material analysed and the type of variants in which the experiment is focused on, available tools could be divided in 4 categories: (i) germline callers, (ii) somatic callers, (iii) Copy Number Variants (CNV) identification and (iv) Structural Variants (SV) identification. In the variant calling step, the variants are identified, called, and characterised with their position in the chromosome and the corresponding

rs if present in available databases. The main parameter involved is the coverage, as the identification of a variant is more certain if the number of readings is higher. This process led to a file called VCF.

The following step, the annotation of variants, consists of the association of the variants with a biological significance. ANNOVAR, Sorting intolerant from tolerant (SIFT) and SNP phenotyping (PolyPhen) -2 are 3 of the available tools for the variants' annotation which provide information on the functional impact of the variant, on the pathogenetic prediction, and on its characteristics.

4.10.4.1 Genes Selection and Customized Panel Design

To investigate the effect of germline variants in efficacy and safety profile of FP-based treatment, a list of 105 genes have been selected to analyze. Selected genes include genes associated with FLs pathway, ABC/SLC transporters, nuclear receptors, DNA repair genes, cell cycle checkpoint genes, inflammation-related genes, genes of angiogenesis and immune response, oxidative stress, hypoxia genes and genes closely related with rectal cancer.

Hybrid capture-based custom Roche assay ("KAPA HyperChoice"; Roche Diagnostics) was used to analyse purified DNA extracted from peripheral blood of LARC patients to assessed the genetic variability of all exons and their adjacent splice junctions (35 bases up- and downstream of the exon). The custom design was carried out by HyperDesign software (Genome Build hg38/GRCh38).

4.10.4.2 Coverage setting and instrument/reagents selection

In the context of NGS, coverage indicates the average number of reads that "cover" a specific target region. Therefore, coverage always describes a relationship between the number of reads and a reference region and can be expressed in terms of percentage or average coverage (e.g., 100X means that on average the target regions are covered by 100 reads). To obtain a desired coverage, the cluster density parameter is of considerable importance. Cluster density, which consist on fragment clonal clusters on flow-cell surface, is a critically important metric that influences run quality, reads passing filter and Q30 scores and directilt proportional to pool loading concentration. For our analysis, an expected coverage of 110X and a cluster density of around 1100 K/mm2 were set. Through a calculation by the Illumina software using as variable the expected coverage, it was possible to calculate the maximum number of samples to be loaded in the same analysis. The instrument considered for our analysis is the Miseq Illumina and the reagents for loading the Miseq v2 Reagent

Kit. To maximize their use and to obtain good quality results in terms of coverage, it was decided to load 64 samples simultaneously.

4.10.4.3 Library preparation

Reagents

For the library preparation the following reagents were used:

- KAPA HyperPlus Kit, Roche, containing:
    - Fragmentation reagents: KAPA Frag Buffer and KAPA Frag Enzyme;
    - Reagents for End Repair and A-Tailing: KAPA End Repair & A-Tailing Buffer and KAPA HyperPlus End Repair & A-Tailing Enzyme Mix;
    - Ligation reagents: KAPA Ligation Buffer and KAPA DNA Ligase;
    - KAPA HiFi Hotstart ReadyMix.
- KAPA Universal Adapter cod. 09 063 781 001, Roche;
- KAPA UDI Primer Mixes, 1-96 cod. 09 134 336 001, Roche. Lyophilized UDI Primer Mixes in plates have been individually resuspended with 10 µL of steril water;
- KAPA HyperCapture Reagent Kit cod. 09 075 810 001, Roche containing:
    - Hybridization reagents: COT Human DNA, Universal Enhancing Oligos, Hybridization Buffer and Hybridization Component H;
    - Reagents for Wash and Capture: 10X Stringent Wash Buffer; 10X Wash Buffer I; 10X Wash Buffer II; 10X Wash Buffer III and 2.5X Bead Wash Buffer;
    - Reagents for Post-Capture Amplification: KAPA HiFi HotStart ReadyMix (2X) and Post-Capture PCR Oligos. Post-Capture PCR Oligos are lyophilized and must be resuspended upon first use by adding 480 µL of water for injections;
- KAPA HyperCapture Bead Kit containing the Capture Beads cod. 09 075 780 001, Roche used for the capture step;
- KAPA HyperPure Beads cod. 08 963 843 001, Roche used for washes;
- KAPA Target Enrichment Probe (Kapa HyperChoice MAX 3Mb T1, 12rnx), Roche. These are the specific probes for the custom assay. Dried probed must be resuspended with the

volume indicated in the vial of KAPA Probes Resuspension Buffer cod. 09 075 879 001. Aliquots of 4 µL were prepared and stored until use at -20 °C;

- Sterile water for injections S.A.L.F. 100 mL package;
- Fresh prepared 80% ethanol;
- Tris-HCl 0.1 M pH = 8. The TRIS buffer solution was prepared using Tris base TRIZMA base, Sigma, and HCl 37%, Sigma-Aldrich.

Protocol

Targeted libraries were prepared starting from 100 ng of DNA according to the KAPA HyperCap Workflow v3.0. Purified DNA samples of different concentrations (10-25 ng/µL) were diluted in water to a final volume of 35 µL into a well of a PCR plate with 10 mM Tris-HCl, pH 8.0.

The commercially available KAPA HyperPlus Kit was used for the enzymatic fragmentation, End Repair and A-Tailing Reaction and Adapter Ligation Reaction.

Fragmentation reaction was assembled on ice with 5 µL of KAPA Frag buffer and 10 µL of Enzyme for each sample. Enzymatic fragmentation was performed using the following incubation program as describe in Table 4.6 with the lid temperature to ≤ +50 °C, fragmentation: 25 minutes at +37 °C.

Table 4.14 Fragmentation reaction condition.

Reaction condition	**N° of cycles**
4 °C	HOLD
25' at 37 °C	X 1

After fragmentation, the plate was quickly placed on ice for the End Repair and A-Tailing Reaction. A master mix of KAPA End Repair & A-Tailing Buffer and KAPA HyperPlus End Repair & A-Tailing Enzyme Mix was prepared according to the manufacturer's instructions and the End Repair and A-Tailing step incubation was performed in a thermocycler as described in Table 4.7 with the lid temperature set to +85 °C.

Table 4.15 End Repair & A-Tailing reaction condition.

Reaction condition	N° of cycles
65 °C	HOLD
30' at 65 °C	X 1

After End Repair and A-tailing reaction, the DNA fragments were ligated at their 5' ends KAPA Universal Adapter, added to each well individually prior to the addition of the Ligation Master Mix prepared with KAPA Ligation Buffer and KAPA DNA Ligase. The ligation thermocycler program is described in Table 4.8 with the lid set to +105 °C.

Table 4.16 Ligation reaction condition.

Reaction condition	N° of cycles
20 °C	HOLD
20' at 20 °C	X 1

Post-Ligation Clean-up with two 80% ethanol washing steps was performed using KAPA HyperPure Beads and Adapter ligated samples DNA were eluted in 20 μL of nuclease-free water as described in the "manufacturer's instruction manual".

Amplification and Purification of the Sample Library was performed using the KAPA Unique Dual-Indexed (UDI) Primer Mix and KAPA HyperPure Beads.

The KAPA Unique Dual-Indexed (UDI) Primer Mix was used in conjunction with KAPA Universal Adapters to generate uniquely labelled libraries from individual biological DNA samples. Sample indexing during library amplification allows the libraries' pooling prior to target capture or cluster generation and enables multiplexed sequencing. Each KAPA UDI Primer Mix is a pre-mixed combination of forward and reverse primers. Primers contain a non-redundant (unique) 8-nucleotide index designed to mitigate index misalignment ("index hopping") on Illumina sequencers that employ patterned flow cells and exclusion amplification chemistry. Adapter ligated libraries were amplified in the regions of interest using the following reaction condition with the lid temperature set to +105 °C (Table 4.9).

Table 4.17 Adapter ligated libraries amplification conditions.

Reaction condition	N° of cycles
45'' at 98 °C	X 1
15" at 98 °C	
30" at 60 °C	X 6
30" at 72 °C	
1' at 72 °C	X 1

The number of PCR cycles equal to 6 is based on the KAPA Hyper prep kit used. A clean up step with two 80% ethanol washing steps was performed and samples were purified using KAPA HyperPure Beads and the Amplified Sample Library DNA was eluted in 30 µL of nuclease-free water.

Concentration, Size Distribution, and Quality of the Amplified Sample Library were assessed using Quantus Fluorometer (Promega, Madison, WI, U.S.A.) with the QuantiFluor dsDNA Dye (Promega, Madison, WI, U.S.A.) and the Agilent 4200 TapeStation (Agilent Technologies, Santa Clara, CA, U.S.A.) with the High Sensitivity D1000 ScreenTape for the fragment size analysis as described in the manufacturer's instruction manual. The properly prepared 1:10 diluited Amplified Sample Library should contain ≥ 1000 ng of total DNA and should have an average fragment size distribution at ~320 bp. Example of an amplified library analyzed evaluated using the Agilent 4200 TapeStation and High Sensitivity DNA assay is reported in Figure 4.14.

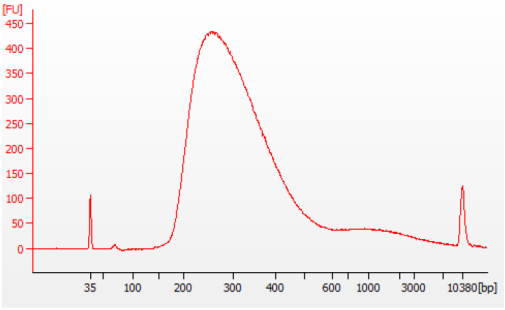


Figure 4.14 Example of an amplified HyperPlus sample library analyzed using the Agilent 4200 TapeStation and High Sensitivity DNA assay.

Following proper evaluation of the abovementioned parameters, multiplex DNA Sample Library was prepared mixing together equal amounts (by mass) of each uniquely indexed DNA sample libraries

to obtain a combined DNA mass of 2 µg. The amount of each sample was calculated using the following formula:

$$\mu L \text{ sample} = \frac{(2000/16 \text{ samples})}{\text{Quantus sample concentration}}$$

PCR Grade water to achieve a final volume 45 µL was added.

20 µL of COT Human DNA and 130 µL of KAPA HyperPure Beads were added to the DNA Sample Library Pool and a one-step 80% ethanol wash was performed to prepare the samples for the hybridization with KAPA Target Enrichment Probes. The DNA Sample Library Pool was eluted in 13.4 µL of Universal Enhancing Oligos. The KAPA Universal Enhancing Oligos Kit contains universal blocking oligonucleotides for use in the hybridization step of the KAPA HyperCap workflow. These oligonucleotides hybridize with specific sites on the DNA library thereby increasing the efficiency of the target enrichment reaction.

A Hybridization Master Mix of 28 µL of Hybridization Buffer, 12 µL of Hybridization Component H and 3 µL of PCR Grade water was added to the bead-bound DNA mixture resuspended in Universal Enhancing Oligos. After incubation at RT, the eluate was transferred into a new well containing the KAPA Target Enrichment Probe.

KAPA Target Enrichment Probes is a complete set of biotinylated 120 bp oligonucleotide probes provided by Roche to perform target enrichment. 4 µL of KAPA Target Enrichment Probe resuspended with recommended volume of Resuspension Buffer (KAPA Probes Resuspension Buffer, Roche) and previously aliquoted to avoid multiple freeze-thawing cycles or potential accidental contamination, were used for each reaction. Hybridization incubation was performed in a thermocycler with the lid temperature set to +105 °C with the following steps (table 4.10).

Table 4.18 Hybridization reaction condition.

Reaction condition	N° of cycles
95 °C	HOLD
5' at 95 °C	X 1
55 °C	16-20 hours

After the overnight hybridization, hybridized DNA was bound to the prepared Capture Beads. Capture beads preparation included the treatment of the capture beads with bead wash buffer. The

Hybridized DNA, after its binding with the capture beads, was heated for 20 minutes at 55 °C. Subsequently, DNA was washed with 3 different Wash Buffer, 10X Stringent Wash Buffer; 10X Wash Buffer I; 10X Wash Buffer II; 10X Wash Buffer III, diluted at 1X concentration and properly added to the library as described in the manufacturer instruction. Finally, the bead-bound DNAs were resuspended in 20 µL of PCR Grade water.

Enriched multiplex DNA sample was amplified using KAPA Hifi Hotstart Readymix (2X) and Post-Capture PCR Oligos. The reaction conditions are described in Table 4.11 using a thermocycler with the lid set at 105 °C.

Table 4.19 Enriched multiplex DNA amplification reaction condition.

Reaction condition	N° of cycles
45'' at 98 °C	X 1
15" at 98 °C	
30" at 60 °C	X 16
30" at 72 °C	
1' at 72 °C	X 1

The number of PCR cycles equal to 16 is based on the capture target size of the experiment.

A clean up step with two 80% ethanol washing steps was performed and samples were purified using KAPA HyperPure Beads. The Amplified Sample Library DNA was eluted in 20 µL of nuclease-free water.

Concentration, Size Distribution, and Quality of the Amplified Sample Library were assessed using Quantus Fluorometer (Promega, Madison, WI, U.S.A.) with the QuantiFluor dsDNA Dye (Promega, Madison, WI, U.S.A.) and the Agilent 4200 TapeStation (Agilent Technologies, Santa Clara, CA, U.S.A.) with the High Sensitivity D1000 ScreenTape for the fragment size analysis as described in the manufacturer's instruction manual. The properly prepared 1:10 diluted Amplified Sample Library should contain ≥ 100 ng of total DNA and should have an average fragment size distribution at ~320 bp.

4.10.4.4 Library Normalization and library sequencing in Miseq instrument

Reagents

- 0.2 N NaOH freshly prepared;

- Phix Control v3 (Illumina);
- Miseq v2 Reagent Kit (box 1 of 2), Illumina. The kit contains the cartridge and the HT1 buffer used for the preparation and dilution of the library to be loaded (Figure 4.15);
- Miseq Reagent Kit V2 / V2 micro / V2 nano (box 2 of 2), Illumina. The kit contains the Incorporation Buffer and the Flow Cell. The Flow cell which could have different capacity sizes (V2, V2 micro, V2 nano) and cycles performed (500, 300, 50 cycles) has to be chosen based on the average size of the library, desired coverage, number of samples analysed, panel size, and on quantity/quality of the starting material (Figure 4.15).

Figure 4.15 Miseq Reagent Kit, Illumina, with cartridge, HT1 buffer, Incorporation Buffer and Flow Cell.

Protocol

The procedure of denaturation and dilution allowed to obtain libraries with a final volume of 600 µL. The recommended loading concentration varies depending on the version of MiSeq Reagent Kit used for the sequencing run, on library preparation methods and quantification methods.

Our NGS analysis involved the concomitant loading of 64 samples to obtain the desired coverage and to optimize the use of reagents. Therefore, 4 libraries of 16 samples each were combined. After quantifying the 4 libraries, it was necessary to convert the concentration from ng/µL to nM using the formula:

$$\mathrm{nM} = \frac{(\mathrm{C2} * 1000000)}{(\mathrm{B2} * 607.4) + 157.9)}$$

where C is the concentration ng/µL and B2 is the average size of the fragments obtained from the TapeStation.

It was necessary to prepare a fresh dilution of 0.2 N NaOH by combining 800 µL of water for injection and 200 µL of stock 1.0 N NaOH as higher concentrations of NaOH in the library inhibit library hybridization to the flow cell decreasing cluster density. In addition, the HT1 present in the MiSeq v2 Reagent Kit cartridge must be thawed.

Each library was then diluted with water for injections to the final concentration of 4 nM. Following the dilution, it is important to quantize the library with the Quantus to verify the fairness of resulted dilutions. The 4 libraries were subsequently pooled in an equivolumetric way and 5 µL of the library obtained were denatured for 5 minutes at RT with 5 µL of 0.2 M NaOH. Finally, a library concentration of 2 nM was obtained. At the end of the incubation, 990 µL of HT1 present in the reagents box were added to obtain 1 mL of 20 pM denatured library.

The experimental setting of the method allowed to evaluate 18 pM as the optimal loading concentration. Therefore, the 20 pM library was further diluted with HT1 to the final expected concentration. Eventually, as per manual, it was necessary to add 1% of the Phix reagent which consists of adapter-ligated library used as a control for Illumina sequencing runs. 594 µL of library at 18 pM were mixed with 6 µL of Phix to obtain a total of 600 µL to be loaded in the Miseq instrument, Illumina. Then, the pool was further denatured at 96 °C for 2 minutes in a block heater. The tube is inverted 1-2 times and immediately placed on ice for 5 minutes.

The denaturation steps enabled the concentration of NaOH to be not higher than 0.001 M (1 mM) in the final solution after diluting with HT1. Higher concentrations of NaOH in the library inhibit the hybridization to the flow cell and decrease cluster density.

For the loading process, it was necessary to prepare the cartridge, the flow cell and design the sample sheet.

The cartridge, present in the Illumina Miseq v2 Reagent Kit (box 1 of 2), must be thawed by placing it for one hour in deionized water. Once thawed, it is necessary to invert the cartridge few times to mix the reagents well and check the effective defrosting state. Once defrosted, if not used immediately, the cartridge must be kept in the fridge until use.

In addition, the flow cell, present in the Miseq Reagent Kit box 2, Illumina, must be prepared. The flow cell must be extracted from its liquid and washed gently under the deionized water flow. Once washed, the flow cell must be air dried avoiding leaving marks.

The final pool was loaded in the position n° 24 of the cartridge and loaded into MiSeq instrument, Illumina. The sequencing analysis was carried out in 25 hours, while the data analysis in 27 hours.

4.10.4.5 Experimental set-up

For each experiment a sample sheet must be prepared. The "Illumina Experiment Manager, IEM" software available in the computer associated with the MiSeq instrument was used to prepare the SampleSheet. The instrument must contain the Manifests SamplePlates and the Reference Genomes data. For the sample sheet preparation, the required data must be entered:

- Sample ID: the sample unique code;
- Sample Name: the same as Sample ID;
- Index1 (I7) and Index2 (I5): the unique UDI indexes used for each sample;
- Manifest: the appropriate manifest that was previously saved in the instrument folder. The manifest, which is provided by the Roche Diagnostic company, was developed based on the reference genome (i.e., hg19) currently present in the Miseq computer.

4.10.4.6 Data analysis

FASTQ data were primarily processed and annotated using Illumina VariantStudio. Illumina VariantStudio data analysis software quickly identifies and classifies disease-relevant variants and reports significant findings in a structured report. It is an easy-to-use application with an intuitive user interface. It is rapid and allows for a complete annotation of the variants with an extensive set of filters and efficient categorization processes which makes it fully customizable.

VCF files for each sample were imported in the program and annotated. Output files included the list of the SNP discovered and, for each SNP, the SNP position (gene, chromosome, codons, cDNA Position, CDS Position Protein Position Amino Acids), the SNP type (e.g., SNP or indel), patients genotype, the variant consequence (e.g., synonymous variants, intron variant, missense variant) and information regarding the Quality and read Depth. Moreover, variant studio includes information on protein function. Specifically, several bioinformatics tools that predict whether nonsynonymous changes are likely to have a deleterious effect on protein function have been developed. In variant studio output data, SIFT and PolyPhen predictions were available. SIFT prediction includes deleterious and tolerated prediction while PolyPhen benign or damaging. Additionally, allele frequencies in world population and from specific databases (ExAc, COSMIC) are listed.

Further informations belong from ClinVar database which is a public archive with free access to reports regarding the relationships between human variations (germline or somatic mutation) and phenotypes, with supporting evidence. Clin var Alleles, ClinVar Allele Type (germline/somatic) and ClinVar Significance (benign/pathogenetic) information were included in the software elaboration data.

4.10.5 Haplotype analysis

A haplotype is a set of DNA variations or SNPs that tend to be inherited together because of their respective close positions in the DNA sequence. During the cell division process, the regions where the SNPs in the same haplotype are present generally do not undergo the crossover and recombination mechanism.

A fundamental parameter when it comes to haplotypes is LD. In the field of genetics, LD means the presence of a non-random statistical association between alleles at 2 or more loci located on the same chromosome.

These specific alleles form ancestral haplotype, which is a haplotype that is widely distributed in any population and transmitted along the descendants. It has been shown that LD is greater in homogeneous populations or those derived from the same nucleus of individuals. The LD is also an important tool in clinical practice to identify chromosomal regions of limited breadth where gene drivers for a particular pathology are located.

LD is influenced by many factors, including selection, rate of genetic recombination, mutation rate, genetic drift, population structure, and genetic linkage. As a result, the pattern of LD in a genome is a strong signal of the population genetic processes that structure it.

The coefficient of LD D', which better explains the deviation of the observed frequency of a haplotype from the expected frequency, is commonly used and can take values between 0 and 1. A value of 1 indicates LD, while 0 indicates LD, and the higher this value, the greater the degree of association.

An alternative to D' is the correlation coefficient between pairs of loci, usually expressed as its square, r2. r2 can take values between 0 and 1 and is a parameter that is adjusted and set to evaluate loci with different allele frequencies. An optimal LD is defined by D = 1 and r = 1, while there are at least 3 haplotypes for D=1 and r < 1.

All the patients analysed by NGS technique were considered for the analysis of the haplotypes. Specifically, among all the genes present in the panel, we focused on the evaluation of the SNPs of the *DPYD* gene. A cut-off of 20X coverage was set to include variants in the analysis. Haplotype frequencies was calculated with PHASE software and LD (expressed as r2 or D') between markers was estimated using LDpop Tool available in ldlink.nci.nih.gov website.

Haplotypes were reconstructed from genotype data using PHASE software. first a containing input file must be generated containing the total number of patients, total number of the SNP analysed and its type (microsatellite or SNP), and patients' genotypes. That information allowed the PHASE program to running. As a result of the entire process, several output files were generated.

The first contains a summary of the individual haplotype estimated for each patient.

Freqs file estimates the sample haplotype frequencies; *Pairs* file assesses the most likely pairs of haplotypes for each case with their probability; *Recom* file contains estimates of recombination parameters across the region using the general model for varying recombination rate from Li and Stephens (2003) and *Hotspot* file contains estimates of recombination parameter.

The main output file was then processed and displayed in a useful for the analysis format where each patient was associated with the different haplotypes generated. Furthermore, the frequency of each haplotype in the whole population was calculated. Distribution of the haplotypes and comparison of the different haplotype with patients' therapy efficacy and toxicity have been performed.

4.11 IMMUNOHISTOCHEMISTRY (IHC) ANALYSIS

IHC analysis is a laboratory technique that exploits specific antibodies for tumor antigens. The IHC technique includes three steps:

1. Preparation of the tumor specimen slides;
2. IHC reaction;
3. Interpretation and quantification of the obtained marker expression.

IHC reactions could be used for histogenetic diagnosis of morphologically undifferentiated neoplasms, for subtyping of neoplasms and for primary neoplasm site characterization. It is also widely used in clinical research for the prognostic markers' identification, the discrimination between benign and malignant cells and the identification of structures and cellular components.

IHC reaction protocol (indirect staging methodology) includes 4 main steps (Figure 4.16):

1. *Tissue fixation*: the fixation method used depends on the tissue type and is based on the programmed experiment. During the fixation step, it is important to maintain the tissue structure intact. For the morphology maintenance, the formalin fixed paraffin embedded (FFPE) method is nowadays the standard method;

2. *Antigen retrieval*: to increase the availability of proteins for detection, a fundamental step is to enhance antigen expression breaking the formalin-induced antigen cross-link and re-exposing the epitopes of the antigen for antibody binding. Heat or enzymatic reaction is exploited: the recovery of the heat-induced epitope involves heating the slides in a homemade or commercial buffer at a different pH, while the enzymatic method involves the use of proteolytic enzymes (pepsin or pronase);

3. *Blocking*: is a crucial step to block and inactivate endogenous enzymes, to minimize background signals, false-positive staining, and non-specific antibody binding;

4. *Antibody labelling and visualization*: indirect or direct staining methodology are available methods for this step. The standard indirect detection methodology includes the incubation of the sample with a primary antibody (1 hour to overnight incubation) followed by a wash for excess unbound primary antibody removal, and an incubation with a secondary labelled antibody (1 hour incubation), whose excess is then washed. Lastly, the amount of secondary antibody label associated with the primary antibody is quantified.

5.

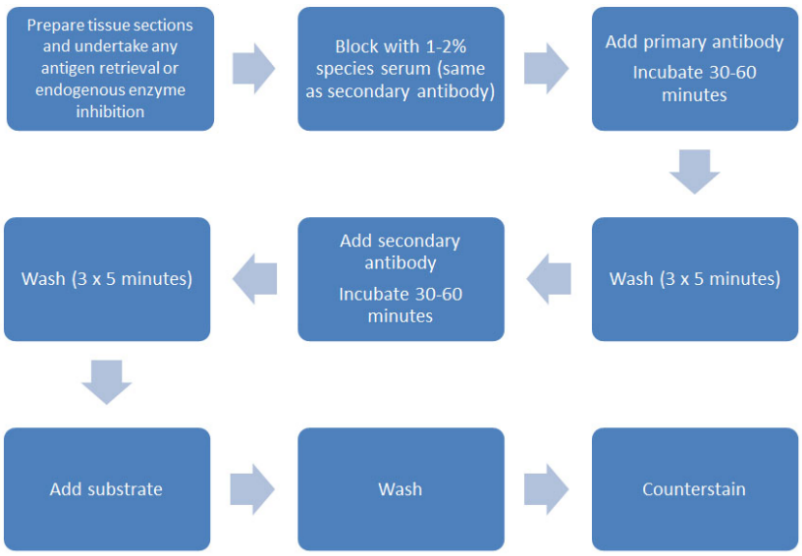


Figure 4.16 IHC reaction steps: Indirect staining methodology.

The Direct Detection method is faster, low-cost, and simpler as it includes only one antibody incubation for marker quantification (Figure 4.17).

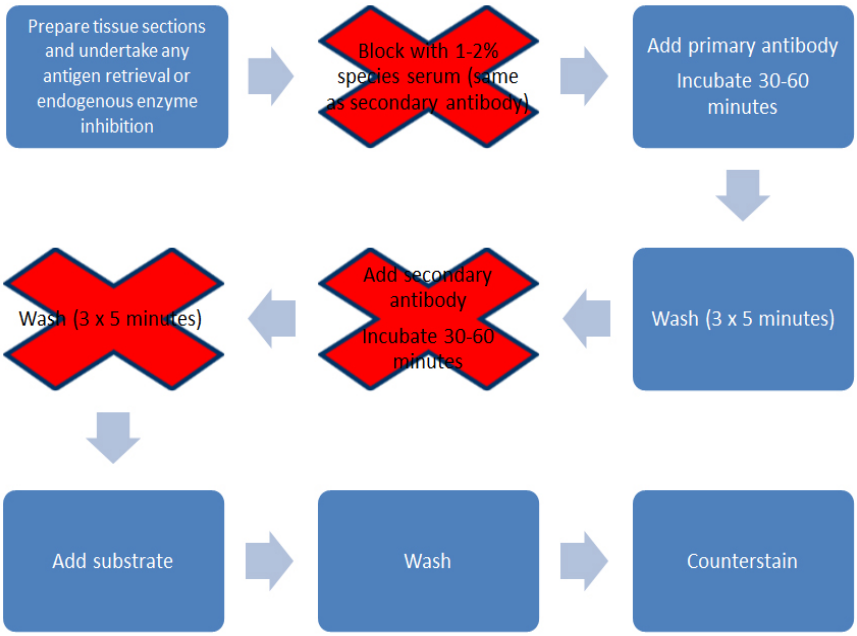


Figure 4.17 IHC reaction steps: Direct staining methodology.

4.11.1 Biomarker selection

For our IHC analysis, a systematic literature data revision was performed to identify a panel of tumor biomarkers for the correlation with the patients' TRG. Proteins involved in the tumor cell growth and proliferation, in the stimulation of tumor immune response, in the DNA repair and in the DNA repair of damage caused by radiotherapy treatment were considered.

The final list of protein analysed included MLH1, GLUT1, Ki67, CA-IX, CXCR4, COX2, CXCL12, HIF1α, VEGF, CD44 and RAD51.

Pre-treatment tumor biopsies collected during staging colonoscopy were fixed in formalin and paraffin embedded. From the tissue biopsy three µm-thick sections were cut. Following haematoxylin and eosin staining, one of the slides was reviewed by a pathologist for cancer diagnosis and two slides were analysed by IHC for the selected biomarkers. Two blinded and independent evaluation were performed by two trained pathologists on the IHC biomarkers expression. None of the experts was aware of the clinical and physiological characteristics of patients of which they had to analyse the

tumor samples, including information regarding sex, radiotherapy and chemotherapy treatment schedule and TRG.

For IHC, 3µm-thick sections were stained on a Dako Omnis platform with the following Primary commercial antibodies list in Table 4.12 for the evaluation of the tumor biomarkers content.

Table 4.20 List of the antibodies used for IHC of selected biomarkers in tumor biopsy.

Tumor biomarkers	Antibodies
MLH1	monoclonal, clone M1, Ventana Medical System, Oro Valley, AZ, USA
Ki67	monoclonal, clone 30-9, Ventana Medical System
RAD51	polyclonal, Santa Cruz, Biotechnology
GLUT1	polyclonal, Cell Marque, Rocklin, CA, USA
COX2	monoclonal, clone SP21, Cell Marque)
CXCL12	na
HIF1alfa	monoclonal, clone H1alpha67, Novus Biological, Centennial, CO, USA
VEGF	polyclonal, Santa Cruz Biotechnology, Dallas, TX
CA-IX	monoclonal, clone EP161, Ventana Medical System
CXCR4	(polyclonal, AbCam, Cambridge, UK
CD44	monoclonal, clone SP37, Ventana Medical System

Tumor biomarkers were evaluated at the nuclear, at the membrane and at the cytoplasmatic level depending on the characteristic protein expression.

MLH1, Ki67, and RAD51 were evaluated at the nuclear level. GLUT1, COX2, CXCL12, HIF1α, VEGF were evaluated at the membrane level. CA-IX, CXCR4, and CD44 were evaluated at the cytoplasmatic level.

4.11.2 Protein expression evaluation

Proteins' expression was evaluated through the assessment of two parameters: the intensity of staining and the proportion of cells presenting staining positivity. The staining intensity was scored from 0 to 4 with the following definition: 0 = absence of intensity, 1 = weak intensity, 2 = moderate intensity, 3 = strong intensity. The proportion of positive cells ranged from 0 to 100%.

A comprehensive H-score, calculated using the method by Huang *et al* was calculated to incorporate the value of intensity and the value of positive percentage [211]. Briefly, the value of intensity was multiplied by the value of percentage previously ranked into 5 categories:

- 0% = 0
- 1-24% = 1
- 25-49% = 2
- 50-74% = 3
- 75-100% = 4

According to the definition of the H-Score [(intensity value) x (positive percentage value)], 0 to 12 H-score values were assessed to the tumor biopsy in the analysis.

4.12 STATISTICAL METHODS

Demographic and clinical data were extracted from the patients' clinical record for all the patients included in the three-case study population. For safety analysis, all toxicities were graded according to the National Cancer Institute CTC-AE (version 5.0). For correlation analysis with therapy efficacy, pathological tumor response was defined according to Mandard's TRG scale.

For the assessment of the 7 SNPs in CRC and LARC case study, toxicity experienced in the first 2 cycles of treatment were defined as "acute toxicity" while the whole ADEs experienced throughout the course of all chemotherapy cycles administered were defined as "cumulative toxicities". Hardy-Weinberg disequilibrium was performed for genotype frequencies using Pearson's chi-squared test.

Association between dichotomous FP-related toxicities and *DPYD* variants status was performed using Fisher exact test. Specifixcally, for rs3918290 IVS14+1G>A (*2A); rs55886062 1679T>G (*13); rs67376798 (c2846A>T) and rs75017182 c.1236C>T /HapB3, statistical analysis was performed using Fisher's Exact Test and Student's t Test. These statistical tests allow to determine the significance of the data expressed by the value p (p-value). Odd ratio (OR) and corresponding 95% CI for G3-4 toxicity according were assessed.

For rs59353118, rs114170368, rs4294451 SNPs, OR and corresponding 95% CI for G4 toxicity according to *DPYD* SNPs were estimated from logistic regression model, adjusted for gender, age, and protocol. Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable and OR is a measure of association between an exposure and an outcome. The OR represents the odds of a selected outcome given a particular exposure, the odds in

the absence of that exposure. Hazard ratio (HR) and corresponding 95% CI for G4 toxicity according to *DPYD* SNPs was estimated from Cox proportional hazards, adjusting for gender, age, and protocol. The HR is a comparison between the probability of events in a treatment group, compared to the probability of events in a control group. Estimates were adjusted for competing risk according to the Fine-Gray model. The Fine-Gray subdistribution hazard model was the default method to estimate the incidence of outcomes over time in the presence of competing risks. A competing risk is an event whose occurrence precludes the occurrence of the primary event of interest and outcomes in medical research are frequently subject to competing risks.

OR and corresponding CI for G4 acute and late toxicity according to *DPYD* SNPs and gender were estimated from univariate unconditional logistic regression model. The results of the sex-analysis were presented graphically, plotting the OR as a blue or red point for man and woman respectively square and the summary OR as a line.

Haplotype correlation with clinical outcome evaluated by TRG classification and with G3-4 any kind of toxicity was assessed with Fisher exact test or Chi square test.as appropriate. OR was calculated where was possible. CR (TRG1) was compared with TRG2-5 while, regarding toxicity assessment, grade 0-2 were compared with grade 3-4.

For correlation analysis with therapy response and IHC biomarkers expression, H-Score according to Huang. *et al* was proposed as an alternative of cellularity percentage and intensity parameters. Additionally, cellularity percentage was correlated with therapy response. The biomarkers' expression level was considered as a continuous variable and difference-s in TRG were evaluated through the Mann-Whitney test. Specifically, complete responders (TRG1) were compared to non-complete responders (TRG=2-4). and TRG1-2 patients to TRG3-4 ones.

All statistical analysis and data visualization were performed using GraphPad Prism version 9 (GraphPad Software, Inc. La Jolla California USA, www.graphpad.com).

Values of $p < 0.05$ (two-sided) were considered statistically significant. The Confidence intervals (CI) were set at 95%.

5 RESULTS

5.1 CASE STUDY 1- COLORECTAL CANCER PATIENTS

5.1.1 Patients' Characteristics and toxicity assessment

A final retrospective case study of 689 CRC patients admitted to the Experimental and Clinical Pharmacology Unit of Centro di Riferimento Oncologico (CRO), Aviano (PN) between 1999 and 2015 were enrolled in case study 1. 7 *DPYD* variants were genotyped: *DPYD* rs3918290, *2A; rs55886062, *13; rs67373798, c.2846A>T, rs56038477 HapB3; rs59353118; rs114170368, and rs4294451.

The baseline characteristics of the entire case study 1 population included are summarized in Table 5.1.

Males represented the 61.1% (n=421) of the patients while females represented 38.9%n (n=268), with a median age of 61 years. Adenocarcinoma of the colon represented the most common malignancy in our cohort including tumor localization in the left (n=192, 26.8%), right (n=146, 21.6%) and transverse colon (n=82, 11.9%). The remaining 269 patients (39.7%) had rectal adenocarcinoma. In 494 (71.1%) patients, 5-FU-based chemotherapy was administered, while 195 (28.3%) patients received a capecitabine-based chemotherapy. The treatment protocol included monotherapy treatment in 78 (11.3%) patients, concomitant radiotherapy in 125 (18.2%) patients, and polytherapy in the remaining 486 patients. The co-administered drugs included irinotecan, oxaliplatin, and monoclonal antibodies (bevacizumab and cetuximab).

Table 5.3 Socio-demographic and clinical characteristic of the entire case study 1 population of 689 CRC patients.

Characteristics	n	(%)
Gender		
Female	268	(38.9)
Male	421	(61.1)
Age (years)		
<55	157	(22.8)
55-64	223	(32.4)
65-69	139	(20.2)
≥70	168	(24.4)
Tumor localization		

Left colon	192	(27.8)
Right colon	146	(21.2)
Transversal colon	82	(11.9)
Rectum	269	(39.1)
Fluoropyrimidine type		
5-FU	494	(71.7)
Capecitabine	195	(28.3)
Protocol		
Monotherapy	78	(11.3)
Monotherapy+Radiotherapy	125	(18.2)
CPT-11	255	(37.0)
OXALI	219	(31.8)
Other	12	(1.7)
Presence of at least one common DPYD variant		
Yes	44	(6.4)
No	645	(93.6)

In Table 5.2 all toxicities are listed according to their severity (CTC-AE grade). Haematological toxicities as mentioned above, includ neutropenia, leukopenia and decreased platelet count, while non-haematological toxicities includ gastrointestinal AE, mucositis, H&F syndrome. All remaining toxicities are included in "other" category. Separate statistical analysis was performed for the listed toxicity categories.

Table 5.4 Distribution of acute and cumulative toxicity among total patients of case study 1 (689 patients) treated with FP-based chemotherapy.

Toxicity type		**G0**	**G1**	**G2**	**G3**	**G4**
Acute toxicity						
Haematological	n	515	79	52	29	14
	%	74.75	11.47	7.55	4.21	2.03
Non Haematological	n	319	198	116	41	15
	%	46.30	28.74	16.84	5.95	2.18

Haematological +Non haematological	n	259	203	139	60	28
	%	37.59	29.46	20.17	8.71	4.06
Other	n	328	240	94	25	2
	%	47.61	34.83	13.64	3.63	0.29
All toxicity	n	171	236	175	79	28
	%	24.82	34.25	25.40	11.47	4.06
Cumulative toxicity						
Haematological	n	374	93	103	80	39
	%	54.28	13.50	14.95	11.61	5.66
Non Haematological	n	209	199	179	81	21
	%	30.33	28.88	25.98	11.76	3.05
Haematological +Non haematological	n	127	163	210	130	59
	%	18.43	23.66	30.48	18.87	8.56
Other	n	162	247	205	59	16
	%	23.51	35.85	29.75	8.56	2.32
All toxicity	n	47	148	260	163	71
	%	6.82	21.48	37.74	23.66	10.30

5.1.2 Genotyping analysis

Genotyping results from 689 CRC patients obtained by Pyrosequencing and Real-time analysis were of good quality, correctly interpretable and therefore usable for correlation analyzes with the toxicity developed by patients.

A total of 44 (6.4%) patients were found to have at least one of the 4 clinically validated variants of *DPYD*: 9 patients with *DPYD* * 2A, 8 patients with *DPYD* c.2846 and 27 patients with *DPYD* c.1236 variant.

Regarding new un-investigated variants, *DPYD* rs59353118 and rs4294451 genotyping was performed for all patients in the study population. The frequency of genotype was described in Table 5.5 in the following paragraph. 62.6% (n=40) and 67.9% (n=438) of the patients were wild-type for rs59353118 and rs4294451 respectively, while heterozygous mutation was reported in 32.1% (n=207) and 28.1% (n=181) of the patients and the homozygous mutation in 5.3% (n=34) and 4% (n=26) of the case population. Additional data are discussed in more detail in the following sections.

5.1.3 Impact Of Clinically Validated *DPYD* rs3918290, *2A; rs55886062, *13; rs67373798, c.2846A>T, rs56038477 c. 1236 HapB3 variants on FP- associated toxicity

The first analysis was performed for the entire case study 1 to assess the impact of the 4 canonical variants of *DPYD*. The presence of at least one of the variants was therefore correlated with the development of haematological, non-haematological, haematological + non-haematological toxicity considered together and with all types of toxicity, including those unrelated to FP treatment.

For the following analysis, correlation with grade 3-4 toxicities and genetic status of patients was assessed. Analysis with the different types of acute toxicity showed a significant correlation between the presence of at least one of the *DPYD* variants and haematological toxicity (OR=25.946, 95% CI: 1.0311-6.5287, p-value = 0.0429) and with non-haematological and haematological toxicity considered together (OR=21.38, 95% CI: 1.0162-4.4980, p-value = 0.0452) (Table 5.3).

Table 5.3 Odd ratio (OR) and corresponding confidence intervals (CI) for Acute G3-4 toxicity according to the presence of at least 1 *DPYD* polymorphisms.

Acute toxicity	**Wild-type**		**At least 1 DPYD variants**				
	n G3-4	(%)	n G3-4	(%)	OR	95% CI	p value
Haematological	37	5.74	6	13.64	25.946	1.0311 - 6.5287	**P = 0.0429**
Non Haematological	49	7.60	7	15.91	23.012	0.9750 - 5.4313	P = 0.0572
Haematological +Non haematological	78	12.09	10	22.73	21.38	1.0162 - 4.4980	**P = 0.0452**
Other	24	3.72	3	6.82	18.933	0.5473 - 6.5501	P = 0.3135
All toxicity	96	14.88	11	25	19.063	0.9316 - 3.9006	P = 0.0774

On the other hand, the evaluation of the cumulative toxicity exhibited by patients throughout the entire course of therapy, showed a statistically significant correlation with all the toxicity classes studied. The presence of at least one variant was associated with a 20 to 24-fold higher risk of developing all types of G3-4 toxicity except for those defined as "other". This last result is consistent

with data from the literature, as adverse effects not directly correlated with the drug studied were included in the "other" toxicity class (Table 5.4).

Table 5.4 Odd ratio (OR) and corresponding confidence intervals (CI) for Cumulative G3-4 toxicity according to the presence of at least 1 *DPYD* polymorphisms.

Cumulative toxicity	**Wild-type**		**At least 1 DPYD variants**				
	n G3-4	(%)	n G3-4	(%)	OR	95% CI	p value
Haematological	105	16.28	14	31.82	24.000	1.2306 - 4.6806	**P = 0.0102**
Non Haematological	90	13.95	12	27.27	23.125	1.1485 - 4.6561	**P = 0.0189**
Haematological + Non haematological	169	26.20	20	45.45	23.471	1.2641 - 4.3581	**P = 0.0069**
Other	67	10.39	8	18.18	19.171	0.8556 - 4.2956	P = 0.1139
All toxicity	212	32.87	22	50.00	20.425	1.1060 - 3.7717	**P = 0.0225**

Following the evaluation of the impact of common variants, the new *DPYD* variants were investigated.

5.1.4 Impact of new un-investigated *DPYD* rs59353118 and rs4294451 variants on FP-associated toxicity

5.1.4.1 Patients' Characteristics and toxicity assessment

To investigate the real impact of the new 3 PGx novel candidate markers (rs59353118 and rs4294451), patients who were homozygous or heterozygous for the *DPYD* *2A, *13, c.2846 and c.1236 mutations were excluded from the analysis as the presence of these variants may influence the development of toxicities. The final study population consisted of 645 patients. Baseline characteristics of the included patients are summarized in Table 5.5. Males and females accounted for 61.2 % and 38.8%, respectively. Patients with tumor located in left (26.8%), right (21.6%) and transverse colon (11.9%) and rectum (39.7%), were all treated with FP-based chemotherapy regimen: 461 (71.5%) patients with 5-FU and 184 (28.5%) patients with capecitabine. Patients were treated with monotherapy (n=73, 11.3%) or with additional administration of radiotherapy, irinotecan, oxaliplatin, bevacizumab and cetuximab (Table 5.5).

In the final study population, no patient with *DPYD* rs114170368 variant was found because it has a high LD with rs75017182 (HapB3).

Table 5.5 Socio-demographic and clinical characteristic of selected 645 CRC patients

Characteristics	n	(%)
Gender		
Female	250	(38.8)
Male	395	(61.2)
Age (years)		
<55	145	(22.6)
55-64	210	(32.7)
65-69	131	(20.4)
≥70	157	(24.4)
Tumor localization		
Left colon	173	(26.8)
Right colon	139	(21.6)
Transversal colon	77	(11.9)
Rectum	256	(39.7)
Fluoropyrimidine type		
5-FU	461	(71.5)
Capecitabine	184	(28.5)
Protocol		
Monotherapy	73	(11.3)
Monotherapy+Radiotherapy	119	(18.5)
CPT-11	239	(37.1)
OXALI	202	(31.3)
Other	12	(1.9)
rs59353118 [a]		
AA	404	(62.6)
AT	207	(32.1)
TT	34	(5.3)
rs4294451[a]		
AA	438	(67.9)
AT	181	(28.1)
TT	26	(4.0)

[a] χ^2 for Hardy-Weinberg disequilibrium: p>0.05.

Haematological toxicities G≥4, including neutropenia, leucopoenia and platelet count decreased, were present in 4.96% of patients while non haematological toxicities G≥4, which included gastrointestinal AE, mucositic, H&F syndrome, were present in 2.7%. Patients which experienced any G≥4 toxicities count for 7.60% of the entire study population (Table 5.6).

Table 5.6 Distribution of late toxicity among 645 CRC patients treated with FP-based chemotherapy.

Toxicity type (tot pz 645)	G 1-5		G 4-5	
	n	(%)	n	(%)
Haematological	288	(44,65)	32	(4.96)
Non Haematological	452	(70,08)	17	(2.64)
Haematological +Non haematological	526	(81,55)	49	(7.60)
Other	491	(76,12)	15	(2.33)
All toxicity	600	(93,02)	60	(9.30)

The molecular genetic background of the CRC patients was profiled to characterize new un-investigated point mutations in intron region of *DPYD* gene by Real time analysis. By classifying toxicity according to patients rs59353118 genotype, 6.76% (14/207) AT patients and 5.88% (2/34) TT patients experienced severe G≥4 haematological toxicity while only 3.96% (16/404) wild type patients reported such toxicity. Regarding non haematological, 8/404 (1.98%) wild-type patients, 8/207 (3.86%) AT patients and 1/44 (2.94%) TT patients experienced G≥4 AEs (Table 5.7).

When looking for the toxicity experienced according to rs4294451 variant, we found that 27/438 (6,16%) wild type patients, 5/181 (2,76%) AT patients and 0/26 (0%) TT patients developed during the entire course of chemotherapy G≥4 haematological AEs. Regarding non haematological ones, 13/438 (2.97%) wild-type patients, 4/181 (2.21%) AT patients and 1/26 (3.85%) TT patients experienced G≥4 AEs (Table 5.8).

Predictive value of *DPYD* variants was next evaluated. Association between SNPs of *DPYD* and toxicities of FP-based chemotherapy outline was analysed in terms of OR and HR. Among patients screened, the onset of G4 haematological + non-haematological toxicities resulted in a significant association with rs593531118 mutated genotype. The presence of the rs593531118 variant exposed patients to 1.92-fold higher risk to develop Haematological & non-haematological toxicity (OR=1.92, 95% CI: 1.07-3.42) (Table 5.7). On the contrary, *DPYD* rs4294451 variant was associated with 0.50

and 0.49-fold reduced risk of Haematological & non-haematological and All toxicities, respectively (OR=0.50, 95% CI: 0.26-0.96; OR=0.49, 95% CI: 0.25-0.94) (Table 5.8).

Table 5.7 Odd ratio (OR) and corresponding confidence intervals (CI) for G4 toxicity according to *DPYD* rs59353118 polymorphisms.

Toxicity type	AA (n=404)	AT (n=207)	TT (n=34)	rs59353118 A>T		
	G4 n (%)	**G4 n (%)**	**G4 n (%)**	**Model**	**OR (95% CI)[a]**	**OR (95% CI)[b]**
Haematological	16 (3,96)	14 (6,76)	2 (5.88)	Dom	1.84 (0.91-3.71)	1.66 (0.80-3.44)
Non-haematological	8 (1.98)	8 (3.86)	1 (2.94)	Rec	1.13 (0.15-8.76)	1.56 (0.47-5.18)
Haematological & non-haematological	24 (5.94)	22 (10.63)	3 (8.82)	Dom	**1.92 (1.07-3.42)**	1.65 (0.87-3.13)
Other	10 (2.48)	5 (2.42)	0 (0)	Add	0.74 (0.29-1.92)	0.69 (0.26-1.85)
All	32 (7.92)	25 (12.08)	3 (8.82)	Dom	1.59 (0.94-2.70)	1.38 (0.77-2.47)

[a]Estimated from logistic regression model. [b]Adjusted for gender, age, and protocol.

Table 5.8 Odd ratio (OR) and corresponding confidence intervals (CI) for G4 toxicity according to *DPYD* rs4294451 polymorphisms.

Toxicity type	AA (n=438)	AT (n=181)	TT (n=26)	rs4294451 A>T		
	G4 n (%)	**G4 n (%)**	**G4 n (%)**	**Model**	**OR (95% CI)[a]**	**OR (95% CI)[b]**
Haematological	27 (6,16)	5 (2,76)	0 (0)	Add	0.45 (0.20-1.05)	0.42 (0.17-1.05)
Non-haematological	13 (2,97)	4 (2,21)	1 (3,85)	Dom	0.64 (0.21-2.00)	1.44 (0.36-5.67)
Haematological & non-haematological	40 (9,13)	9 (4,97)	0 (0)	Add	**0.50 (0.26-0.96)**	0.56 (0.27-1.16)
Other	13 (2.97)	1 (0.55)	1 (3.85)	Add	0.51 (0.16-1.67)	0.35 (0.08-1.57)
All	49 (11.19)	10 (5.52)	1 (3.85)	Dom	**0.49 (0.25-0.94)**	0.54 (0.27-1.10)

We further evaluate the impact of rs59353118 and rs4294451 patients genotype on the different time to toxicity. Cox proportional hazards, adjusting for gender, age, protocol and for competing risk according to the Fine-Gray model highlighted a protective effect of rs4294451 variant in additive model for Haematological, Haematological & non-haematological and All toxicities risk development (HR=0.28, 95% CI: 0.09-0.83; HR =0.40 95%, CI :0.17-0.93; HR= 0.40, 95% CI :0.18-0.90) (Table 5.9). rs4294451 AT/TT patients were significantly less prone to develop G4 all type of toxicity (Fine-Gray test, p-value = 0.0444) suggesting the protective effect of the SNP in mutated patients against early toxicity development with respect to wild type patients. Only a trend for G≥4

haematological toxicity (p-value = 0.0706) emerged (Figure 5.1). No association was reported for rs5935118.

Table 5.9 Hazard ratio (HR) and corresponding confidence intervals (CI) for G4 toxicity according to *DPYD* rs593531118 and rs4294451 polymorphisms.

Toxicity type	**rs593531118 A>T**		**rs4294451 A>T**	
	Model	**HR (95% CI)**	**Model**	**HR (95% CI)**
Haematological	Dom	1.45 (0.69-3.08)	**Add**	**0.28 (0.09-0.83)**
Non-haematological	Add	0.42 (0.08-2.22)	Dom	1.73 (0.22-13.83)
Haematological & non-haematological	Dom	1.32 (0.67-2.62)	**Add**	**0.40 (0.17-0.93)**
Other	Add	0.81 (0.24-2.65)	Dom	0.23 (0.03-1.84)
All	Dom	1.14 (0.62-2.07)	**Dom**	**0.40 (0.18-0.90)**

aEstimated from Cox proportional hazards, adjusting for gender, age, and protocol. Estimates were adjusted for competing risk according to the Fine-Gray model.

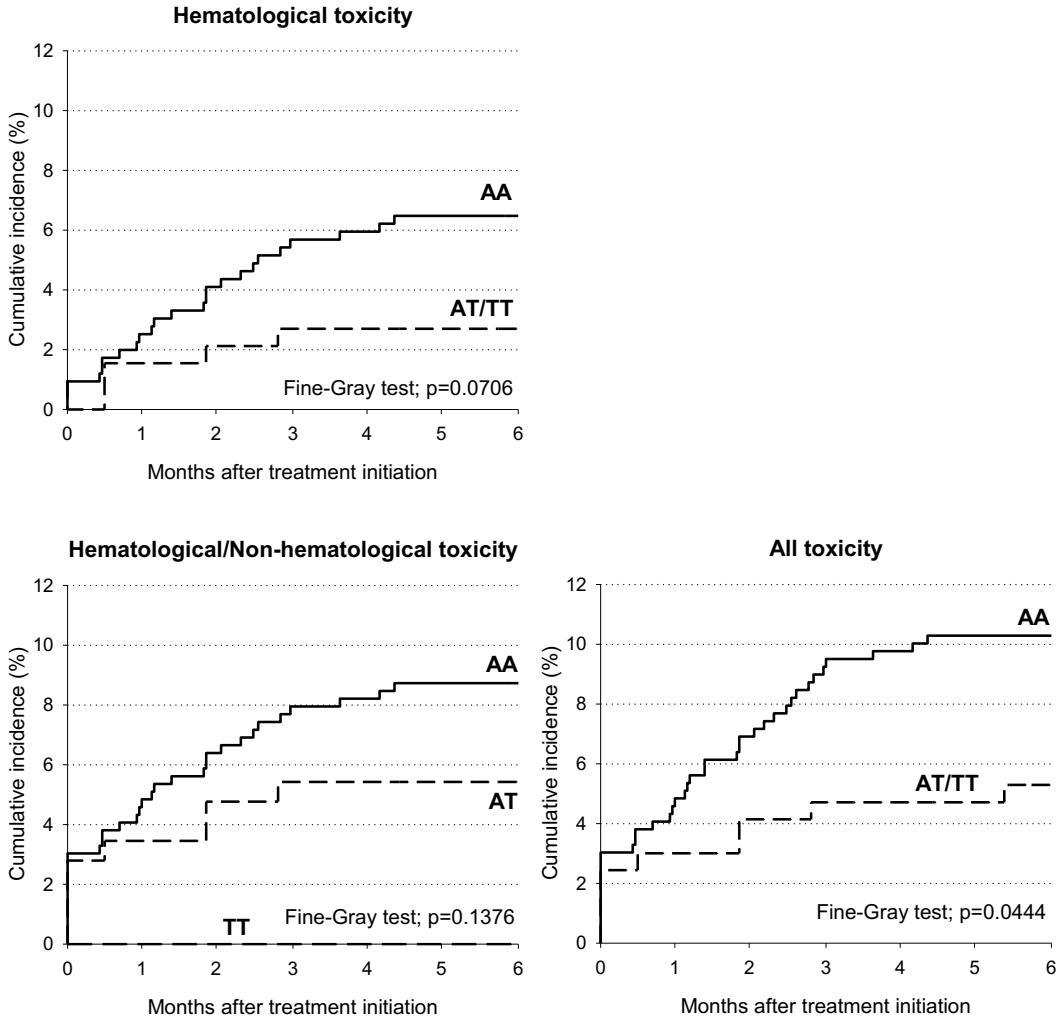


Figure 5.2 Cumulative incidence of toxicity according to rs4294451 A>T polymorphism.

5.1.5 Sex disaggregated data on the impact of *DPYD* variants on FP toxicity

Subsequently, the focus of the analysis shifted to the impact of the presence of these variants on the safety of therapy by assessing the male and female populations separately.

Sex-disaggregated analysis for different new variants distribution, toxicity distribution and for sex-dependent toxicity risk were performed for the different categories of toxicities.

Sex-differences in toxicity distribution was assessed using Fisher exact test. The evaluation was carried out separately considering two subdivisions of the degrees of toxicity (G1-2 vs G3-4 and G1-3 vs G4).

Women appear to be at greater risk of developing haematological, non-haematologicalplus haematological toxicity and all G3-4 toxicities compared to men, in line with previous literature where a worse 5-FU safety profile in women was reported. This greater risk is highlighted both by considering the acute toxicities within the first two cycles of therapy as well as the toxicity developed along the totality of the chemotherapy cycles. If assessing G4 toxicities individually, as the most affecting the clinical course and the patient's life, greater risks of acute non-haematological toxicity and of all toxicities strictly related to FP treatment in cumulative toxicity were noted in women respect to male patients (Table 5.10).

Table 5.10 Distribution of Acute toxicity among 645 patients treated with FP-based chemotherapy by sex.

Acute Toxicity grade	G1-2	G3-4	p-value[$]	G1-3	G 4	p-value[$]
Haematological						
Male n (%)	373 (94.43)	22 (5.57)		387 (97.97)	8 (2.03)	
Female n (%)	235 (94.00)	15 (6.00)	0.86	246 (98.40)	4 (1.60)	0.77
Non haematological						
Male n (%)	376 (95.19)	19 (4.81)		391 (98.99)	4 (1.01)	
Female n (%)	220 (88.00)	300 (12)	**0.0011**	241 (96.40)	9 (3.60)	**0.039**
Haematological +						
Non haematological						
Male n (%)	359 (90.89)	36 (9.11)		383 (96.96)	12 (3.04)	
Female n (%)	208 (83.20)	42 (16.80)	**0.004**	237 (94.80)	13 (5.20)	0.21
Other toxicity						
Male n (%)	383 (96.96)	12 (3.04)		395 (100)	0 (0)	
Female n (%)	238 (95.20)	12 (4.80)	0.28	248 (99.20)	2 (0.8)	0.14
All toxicity						
Male n (%)	347 (87.85)	48 (12.15)		383 (96.96)	12 (3.04)	
Female n (%)	202 (80.80)	48 (19.20)	**0.017**	237 (94.80)	13 (5.20)	0.21

[$] p-value calculated using Fisher exact test

Considering cumulative toxicity, 19.60% (n=49) and 33.20% (n=83) female patients experienced non Heamatological and Haematological + Non haematological G3-4 toxicity respect to only 10.38% (n=41) and 21.775 (n=86) male patients (p-value = 0.0015, p-value = 0.0017). Considering All toxicity,100/250 (40%) female reported a high-grade AE while the same condition was reported only in 112/395 (28,35%) male (p-value = 0.002) (Table 5.11)

Table 5.11 Distribution of Cumulative toxicity among 645 patients treated with FP-based chemotherapy by sex.

Cumulative Toxicity grade	G1-2	G3-4	p-value[$]	G1-3	G 4	P value[$]
Haematological						
Male n (%)	336 (85,06)	59 (14,94)		377 (95,44)	18 (4,56)	
Female n (%)	204 (81,60)	46 (18,40)	0.27	235 (94)	15 (6)	0.46
Non Haematological						
Male n (%)	354 (89,62)	41(10,38)		388 (98,23)	7 (1.77)	
Female n (%)	201 (80,40)	49 (19,60)	**0.0015**	240 (96)	10 (4)	0.13
Haematological +						
Non haematological						
Male n (%)	309 (78,23)	86 (21,77)		370 (93,67)	25 (6,33)	
Female n (%)	167 (66,80)	83 (33,20)	**0.0017**	225 (90)	25 (10)	0.09
Other toxicity						
Male n (%)	359 (90,89)	36 (9,11)		386 (97,72)	9 (2,28)	
Female n (%)	219 (87,60)	31 (12,40)	0.18	244 (97,60)	6 (2,40)	1
All toxicity						
Male n (%)	283 (71,65)	112 (28,35)		363 (91,90)	32 (8,10)	
Female n (%)	150 (60)	100 (40)	**0.002**	221 (88,40)	29 (11,60)	0.16

[$] p-value calculated using Fisher exact test

*5.1.5.1 Impact of Of Clinically Validated DPYD rs3918290, *2A; rs55886062, *13; rs67373798, c.2846A>T, rs56038477 HapB3 variants on toxicity: sex disaggregated analysis*

As far as acute toxicity is concerned, no significant correlation with common genetic variants was highlighted for any category of toxicity, except for the toxicity in the category "other". However, if the frequency of events in the two sexes was considered, it can be noted that a greater percentage of female subjects develop G3-4 acute toxicities in the presence of one of the variants compared to a lower percentage in men (Table 5.12). Specifically, 16.67%, 22.22% and 33.33% of female subjects experienced haematological, non- haematological and all toxicity respect to 11.5%, 11.5% and 19.23% of men patients respectively.

Table 5.12 Odd ratio (OR) and corresponding confidence intervals (CI) for G3-4 acute toxicity according to the presence of at least 1 *DPYD* polymorphisms in male patients.

Acute toxicity	**Male patients**				**Female patients**			
	Wild-type	At least 1 *DPYD* variants			Wild-type	At least 1 *DPYD* variants		
	G3-4 n (%)	G3-4 n (%)	OR (95% CI)	p value	G3-4 n (%)	G3-4 n (%)	OR (95% CI)	p value
Haematological	22 5,57%	3 11,54%	22.115 (0.6163 - 7.9354)	0.2234	15 6,00%	3 16,67%	31.333 (0.8163 - 12.0266)	0.0961
Non Haematological	19 4,81%	3 11,54%	25.812 (0.7117 - 9.3621)	0.1491	30 12%	4 22,22%	20.952 (0.6471 - 6.7838)	0.2172
Haematological +Non haematological	36 9,11%	5 19,23%	23.743 (0.8445 - 6.6758)	0.1011	42 16,80%	5 27,78%	19.048 (0.6447 - 5.6280)	0.2437
Other	12 3,04%	0 0%	0.5834 (0.0336 - 10.1263)	0.7113	12 5%	3 16,67%	39.667 (1.0094 - 15.5875)	**0.0484**
All toxicity	48 12,15%	5 19,23%	17.212 (0.6201 - 4.7779)	0.2972	48 19,20%	6 33,33%	21.042 (0.7517 - 5.8898)	0.1566

When looking for the cumulative toxicity experienced according to selected variants, positive trends for Non Haematological and Haematological + Non haematological toxicities emerged in female population respect to male ones. Moreover the presence of at least one *DPYD* variant was significantly associated with G4 Haematological cumulative toxicities only in female subjects with a p-value of 0.0420 (OR=22.221, 95% CI : 1.0380 - 7.6728) (Table 5.13)

Table 5.13 Odd ratio (OR) and corresponding confidence intervals (CI) for G3-4 cumulative toxicity according to the presence of at least 1 *DPYD* polymorphisms in male patients.

Cumulative toxicity	**Male patients**				**Female patients**			
	Wild-type	At least 1 *DPYD* variants			Wild-type	At least 1 *DPYD* variants		
	G3-4 n (%)	G3-4 n (%)	OR (95% CI)	p value	G3-4 n (%)	G3-4 n (%)	OR (95% CI)	p value
Haematological	59 14,94%	7 26,92%	20.981 (0.8448 - 5.2108)	0.1103	46 18,40%	7 38,89%	28.221 (1.0380 - 7.6728)	**0.0420**
Non Haematological	41 10,38%	5 19,23%	20.557 (0.7357 - 5.7440)	0.1692	49 19,60%	7 38,89%	26.104 (0.9625 - 7.0800)	0.0595*

Haematological +Non haematological	86 21,77%	10 38,46%	22.456 (0.9837 - 5.1 266)	0.0547*	83 33,20%	10 55,56%	25.151 (0.9570 - 6.60 98)	0.0614*
Other	36 9,11%	3 11,54%	13.0007 (0.3723 - 4.5 445)	0.6804	31 12,40%	5 27,78%	27.171 (0.9064 - 8.14 52)	0.0743*
All toxicity	112 28,35%	11 42,31%	18.530 (0.8258 - 4.1 579)	0.1347	100 40%	11 61,11%	23.571 (0.8839 - 6.28 58)	0.0866*

*p-value not significant, trend.

5.1.5.2 Impact of new un-investigated DPYD variants (rs59353118, rs4294451): sex disaggregated analysis

Once looking for the sex-dependant impact of rs59353118, rs4294451, no difference in genotype distribution across the two sexes was observed for both variants (p-value = 0.6045 and p-value = 0.5477) (Table 5.14).

Table 5.14 Distribution of *DPYD* rs59353118, rs4294451among 645 patients treated with FP-based chemotherapy by sex.

rs59353118								
	Tot pt	AA	%	AT	%	TT	%	p-value[$]
M	395	253	64,05%	121	30,63%	21	5,32%	0.6045
F	250	151	60,40%	86	34,40%	13	5,20%	
rs4294451								
	Tot pt	AA	%	AT	%	TT	%	p-value[$]
M	395	271	68,61%	106	26,84%	18	4,56%	0.5477
F	250	167	66,80%	75	30,00%	8	3,20%	

[$] p-value calculated using Fisher exact test

Table 5.15 and Table 5.16 reported the HR and the OR of each toxicity's categories in male and female patients respectively. When considering the OR for the same toxicity type, a trend toward a higher protective effect of rs4294451 variant in female patients than male patients was observed while a slightly higher risk emerged for male population for the presence of rs59353118 variant (Figure 5.2).

Table 5.15 Hazard ratio (HR) and corresponding confidence intervals (CI) for G4 toxicity according to *DPYD* rs563531118 A>T and rs4294451 A>T polymorphisms, by sex.

Toxicity type (G4)	Model	Female (n=250)	Male (n=395)
		HR (95% CI)	HR (95% CI)
rs563531118 A>T			
Haematological	Dom	1,24 (0,42-3,70)	1,67 (0,62-4,51)
Non-haematological	Add	0,92 (0,18-4,85)	-- (--)
Haematological & non-haematological	Dom	1,48 (0,55-4,00)	125 (0,49-3,16)
Other	Add	2,04 (0,57-7,28)	0,35 (0,05-2,73)
All	Dom	1,20 (0,49-2,98)	1,09 (0,48-2,43)
rs4294451 A>T			
Haematological	Dom	0,15 (0,02-1,05)	0,33 (0,07-1,47)
Non-haematological	Dom	1,82 (0,04-95,28)	1,16 (0,07-19,44)
Haematological & non-haematological	Add	0,27 (0,07-1,09)	0,43 (0,14-1,32)
Other	Dom	0,85 (0,08-8,95)	-- (---)
All	Dom	0,34 (0,09-1,21)	0,41 (0,14-1,14)

[a]Estimated from Cox proportional hazards, adjusting for gender, age, and protocol, Estimates were adjusted for competing risk according to the Fine-Gray model

Table 5.16 Odd ratio (OR) and corresponding confidence intervals (CI) for G4 acute and late toxicity according to *DPYD* polymorphisms and sex.

Toxicity type	rs593531118 A>T			rs4294451 A>T		
	Model	Female	Male	Model	Female	Male
		OR (95% CI)[a]	OR (95% CI)[A]		OR (95% CI)[a]	OR (95% CI)[a]
Acute toxicity						
Haematological	Dom	1,54 (0,22-11,09)	3,04 (0,72-12,92)	Dom	0,67 (0,07-6,51)	1,32 (0,32-5,61)
Non-haematological	Dom	1,96 (0,51-7,47)	1,79 (0,2-12,87)	Dom	0,56 (0,12-2,78)	0,73 (0,08-7,05)
Haematological & non-haematological	Dom	1,84 (0,60-5,64)	2,57 (0,80-8,26)	Dom	0,59 (0,16-2,20)	1,10 (0,32-3,71)
Other		---[b]	---[b]		---[b]	---[b]
All	Dom	1,84 (0,60-5,64)	2,57 (0,80-8,26)	Dom	0,59 (0,16-2,20)	1,10 (0,32-3,71)
Late toxicity						
Haematological	Dom	1,81 (0,63-5,16)	1,84 (0,71-4,73)	Add	0,30 (0,07-1,30)	0,58 (0,20-1,62)
Non-haematological	Dom	1,55 (0,44-5,51)	2,41 (0,53-10,95)	Dom	0,86 (0,22-3,40)	0,36 (0,04-3,02)
Haematological & non-haematological	Dom	1,75 (0,76-4,01)	2,02 (0,90-4,56)	Add	0,47 (0,18-1,22)	0,51 (0,21-1,29)
Other	Add	0,69 (0,15-3,18)	0,78 (0,23-2,63)	Add	0,90 (0,19-4,18)	0,28 (0,04-2,08)
All	Dom	1,49 (0,69-3,25)	1,64 (0,79-3,39)	Dom	0,49 (0,19-1,25)	0,48 (0,19-1,20)

[a]Estimated from univariate unconditional logistic regression model, [b]Not accountable: only two events.

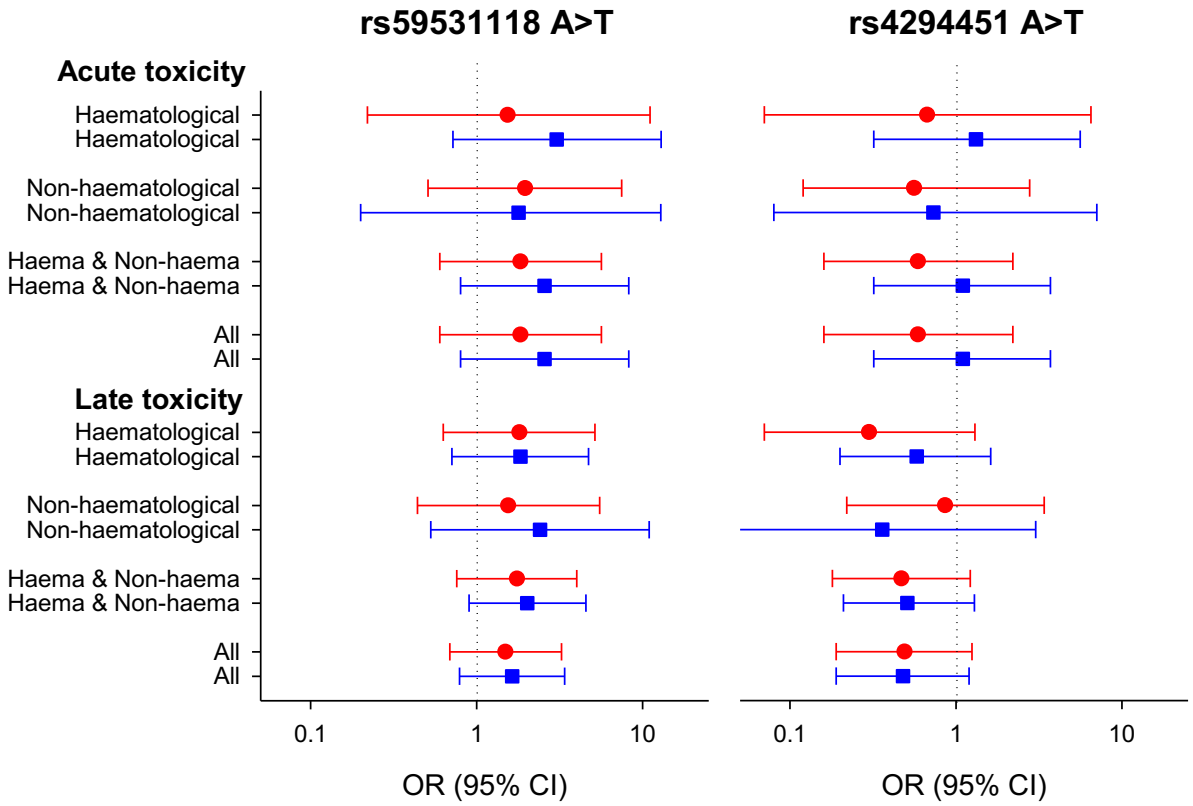


Figure 5.2 Odds ratio (OR) and corresponding 95% confidence interval (CI) for toxicity according to mutation in women (red) and men (blue).

5.2 CASE STUDY 2: LOCALLY ADVANCED RECTAL CANCER PATIENTS

5.2.1 Patients' Characteristics

A total of 229 LARC patients treated with FP-based chemoradiotherapy enrolled in an ongoing prospective clinical study protocol from 1993 to 2020 were included in the study: 212 were evaluable for efficacy analysis while 189 for safety analysis. All patients were homogenously treated with FP in monotherapy with concomitant radiotherapy. In the following paragraphs, the two-case population will be discussed separately.

5.2.2 NGS analysis in Locally Advanced rectal Cancer patients

Beside intronic variants, *DPYD* haplotypes including exonic variants were considered for their predictive effect in FP safety/efficacy profile and for the evaluation of the interaction of patients' sex. A retrospective LARC cohort was selected from the entire study cohort present in the Clinical and Experimental Pharmacology Unit of CRO based on treatment regimen and availability of biological samples and clinical data.

DNAs of patients homogenously treated with FP and concomitant radiotherapy were analysed for the whole exons of 106 genes coding important proteins in the FP pathways using a custom NGS panel. Based on the availability of NGS data, a panel of exonic *DPYD* variants was extracted and corresponding genetic haplotypes were computed for evaluation of the association with toxicity and efficacy profiles.

5.2.2.1 NGS analysis set-up

Fragmentation time Set-up of LARC DNA samples

Of considerable importance for the success of a good sequencing analysis is the length of the fragments generated during the entire sample processing. For this reason, during a new NGS experiment set-up, the evaluation of the correct fragmentation time of the under-study samples was necessary in order to obtain a good quality output for the purpose of the analysis.

The ideal fragmentation timing indicated in the experimental protocols, in fact, could vary by the type and the quality of the genetic starting material. A deep heterogeneous behaviour between DNA from

tissue respect to DNA extracted from blood was reported significantly affecting the correct fragmentation timing. For this reason, it was initially necessary to perform a fragmentation test experiment. A test sample was identified, purified as described above, and fragmented using the timing indicated in the protocol. The addition of the step at 65 °C allows the total blocking of the activity of the fragmentation enzymes. Therefore, the following thermal cycle was carried out (Table 5.17).

Table 5.17 Fragmentation reaction condition for fragmentation time Set-up.

Reaction condition	N° of cycles
4 °C	HOLD
25' at 37 °C	X 1
10' at 65°C	X 1

Following fragmentation step, the determination of the size (length of fragment in bp) was carried out. In the figure below the TapeStation result of the test sample analysed was reported (Figure5.3).

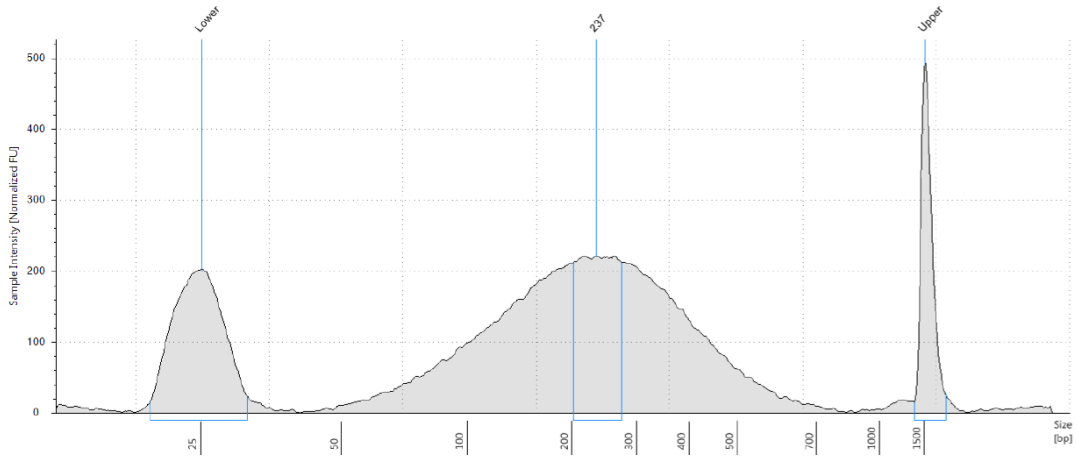


Figure 5.3 Fragmentation profile at Tape-station of the sample used for the time Set-up.

The correct expected TapeStation bp value was around 180-220bp. The fragmentation test conducted with the standard protocol conditions showed a TapeStation profile with a peak at 237bp,

Thanks to the support of the Roche specialist, the fragmentation time was increased from 25 to 27 minutes. The new correct timing was tested during the training analysis and good performance was

reached. In the figure below, the TapeStation of a sample included in the training analysis is reported. In this case, the fragmentation blocking step was not added and the protocol continued up to the first step of evaluating the peaks before the multiplexing process. At this point the peak evaluated at the TapeStation should be around 320bp. The measurement confirmed the correct setting of the fragmentation time (Figure 5.4)

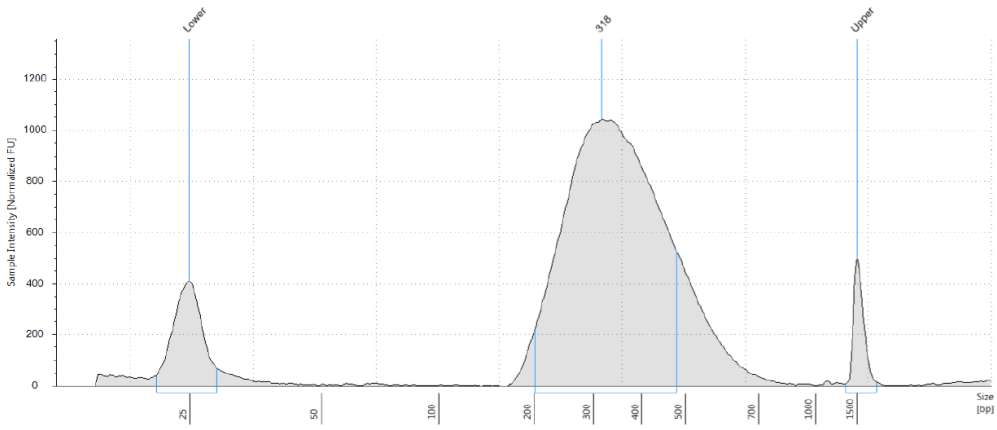


Figure 5.4 Fragmentation profile at Tape-station of one of the training run sample.

Desired coverage set-up

Several specific flow cells could be loaded at Miseq instrument based on the experiment. Moreover, the type of DNA to be analyzed and the corresponding average coverage to be obtained are variables that could impact on the choice of reagents. Using Illumina calculator, various possibilities of reagents and expected coverage were evaluated. Above the alternative options are reported:

- V2 micro flowcell: 20 samples loaded for a coverage of 140X
- V2 micro flowcell: 33 samples loaded for a coverage of 80X
- V2 micro flowcell: 32 samples loaded for a coverage of 87X
- V2 micro flowcell: 27 samples loaded for a coverage of 100X
- V2 flowcell: 96 samples loaded for a coverage of 100X
- V2 flowcell: 80 samples loaded for a coverage of 130X
- V2 flowcell: 72 samples loaded for a coverage of 150X
- V2 flowcell: 64 samples loaded for a coverage of 155X

Therefore, after evaluating the different options, the simultaneous loading of 64 samples per run in a V2 flow-cell was selected to obtain, as previously described, a hypothetical average coverage of 155X.

Only for the training run a V2 nano flow-cell was selected and 10 samples were loaded to obtain a coverage of 60X while for the first run 20 samples (comprising the 10 samples of the training experiment) were anlysed in a V2 micro flow-cell to obtain a coverage of 155X.

Loading pool concentration set-up

Thanks to the advice of the Roche specialist, the load pool concentration was initially set at 10 pM. In the subsequent analysis, the loading concentration was modified and increased up to 18 pM. This adjustment was necessary for low and suboptimal cluster density values obtained in the first runs. The optimal cluster density values were around 1100 k/mm2 while in the experiments conducted it was limited to values between 800 and 900 k/mm2.

NGS Training

The experiment protocol training was carried out in February 2021 with the supervision of the technical support of Roche. A pool of 10 samples was prepared supported by Dr. Marco Messina and finally run in Miseq instrument. The final loading pool concentration was 87 ng/ul with a median size of 363 bp and a loading concentration of 10 pM was used for the analysis (Figure 5.5).

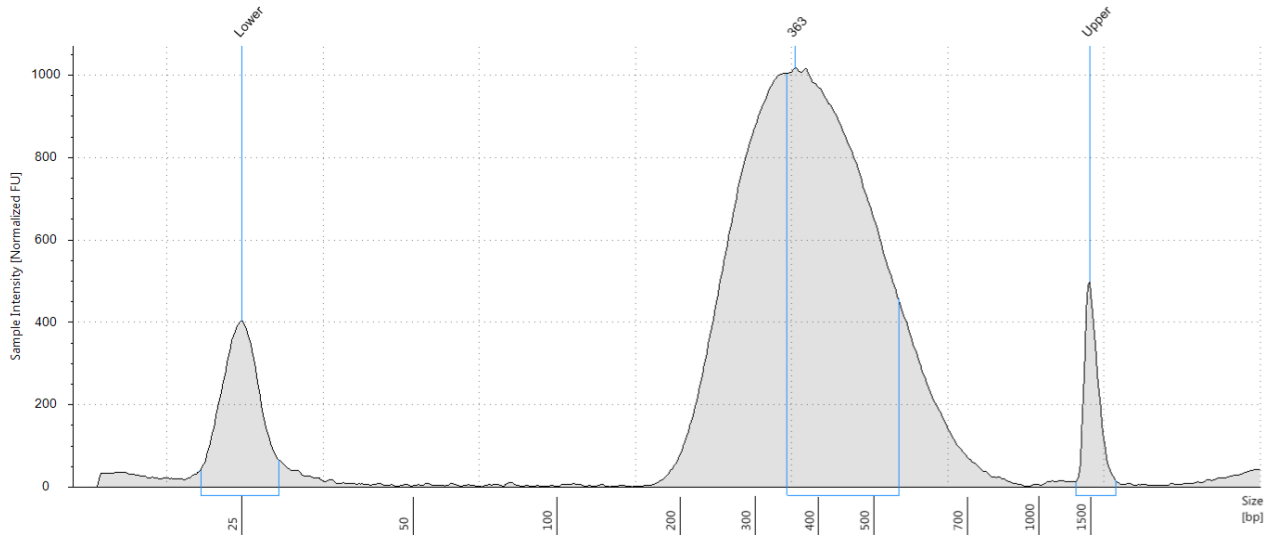


Figure 5.5 Tape-Station profile of the final loading pool for the training run.

The Illumina Enrichment Sequencing Report which reported the technical characteristics and the quality of the run, assessed a consistently low duplicate rate, <1% for all samples with a mean on-

target rate of 77%, a mean coverage of ~38X and a coverage uniformity measure by fold-80 base penalty of 1,4.

The run resulted optimal for the various quality parameters except for the 38X average coverage, lower than expected. The low cluster density (860 K/mm2) could have affected the coverage parameter, and, for this reason, the loading pool concentration was increased to 12pM.

5.2.2.2 Quality analysis of the NGS experiments

First run

In the first run the previous 10 samples were re-loaded with additional 10 samples to final loading condition of 20 samples at 12 pM. A V2 micro flow-cell was used in order to obtain a 14 of expected coverage. The coverage obtained was approximately 100X. The cluster density was of 861 k/mm2, with a clusting passing filter of 94,7% and a Q30 of 96,4%,

Comparing with the training run, with the new set experiment conditions, the median coverage has been significantly improved as shown in Figure 5.6.

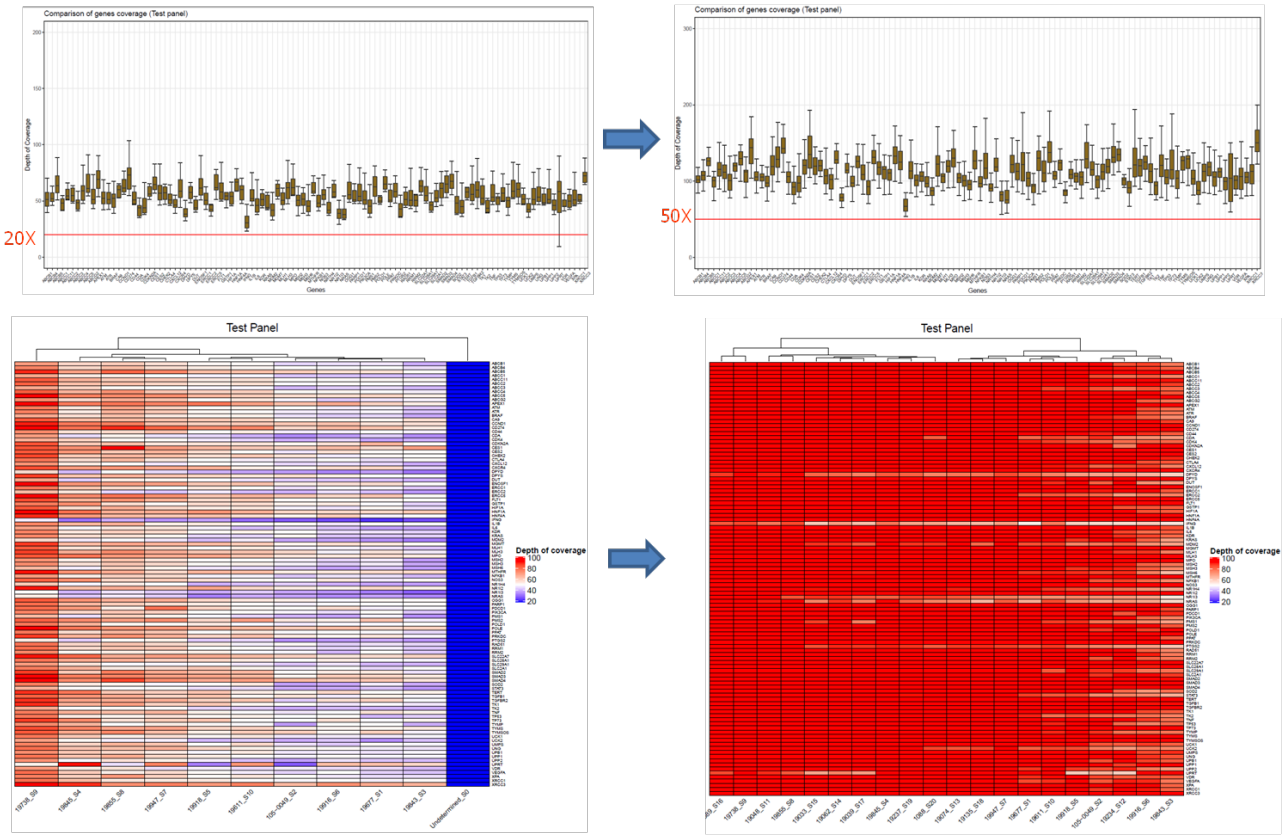


Figure 5.6 Improvement of the coverage per sample and per gene between the training run and the first run.

Second run

In the 2nd run 64 samples at 14 pM were loaded in a V2 flow cell, obtained good quality results. The cluster density, similar to the previous run, was 847 k/mm2, the average coverage was around 120X, less than the expected value of 155X and the clusting passing filter was 95,3% with a Q30 of 96,3%. The used loading conditions were confirmed as the standard for subsequent experiments.

Third run

As in the previously run, the same conditions were mantained apart from the final pool concentration which was increased to 15 pM for cluster density improvement. A cluster density of 861 k/mm2 and a clusting passing filter of 95,5%, with a Q30 of 96,5% were obtained. The cluster density remained lower than the optimal value and the final pool concentration was decided to increase to 16 pM for the fourth run.

Fourth run

With the Fourth run experiment, the involvement of technical support was necessary to overcome issues regarding the obtained cluster density. With a 16pM of loading concentration a cluster density of around 600k / mm2 was achieved.
Carring the library concentration at 20pM, a cluster density of about 800 k / mm2 was obtained.
In-depth analysis with Roche and Illumina technical support, evaluating TapeStation images, outlined a non-optimal peak, which may have compromised the run performance. A double size selection step on the final pool before instrument loading might be helpful in this situation.

Fifth run

Since the TapeStation profiles of the fifth library resulted qualitatively better than the previous ones, a 18pM of pool concentration was used. The run showed good quality results with an improvement in the cluster density to 1118 K / mm2 and a clusting passing filter 92,1% with a Q30 of 95%.

Sixth run

The good results in the previous run and the excellent TapeStation profiles for this pool allowed to reconfirm the experiment conditions and the loading concentration of 18pM. The cluster density obtained was of 1035 K / mm2 and the clusting passing filter of 92,3% with a Q30: 95,1%,

Sixth run, seventh and eighth run

The performance parameters of the 6^{th}, 7^{th} and 8^{th} run were of good quality. The cluster density obtained was of 1035 K / mm2, 1087 K / mm2, K / mm2, respectively with clusting passing filter of 92,3%, 94,1% and 89.9%.

Table 5.18 Loading condition performance of the 8 run.

	N° samples	Flowcell	pM	Coverage attended	Cluster density	Cluster passing filter
Run 01	20	V2micro	12	140X	861 k/mm2	94,7 %
Run 02	64	V2	14	155X	847 k/mm2	95,3 %
Run 03	64	V2	15	155X	861k/mm2	95,5 %
Run 04	64	V2	20	155X	813 k/mm2	95,3%
Run 05	64	V2	18	155X	1118 k/mm2	92,1%
Run 06	64	V2	18	155X	1035 k/mm2	92,3%
Run 07	64	V2	18	155X	1087 k/mm2	94,1%
Run 08	64	V2	18	155X	931 k/mm2	89,9%

5.2.3 Molecular Characterization of LARC patients DNA: Haplotype generation

Of the totality of the genes included in the panel and analysed using the NGS method, the predictive role of specific *DPYD* gene mutations have been identified as the focus of this part of the PhD project. Specifically, VCF files, containing the base alignment information for each variant, were sequence annotated using Variant Studio program. Exon variants were characterized as tolerated or deleterious or benign and damaging by Sift and Polyphen software respectively and their significance status was determined and evaluated as pathogenic or affecting drug response. From NGS analysis in *DPYD* gene, a total of 21 variants were identified. In Table 5.19 the variants identified are listed.

Table 5.19 Characteristics of *DPYD* variants identified by NGS in the case study 2 population of 229 LARC patients.

dbSNP ID and CDS Position	ClinVar Allele Type	ClinVar Significance	Sift	Polyphen
rs1801265 c.85 T>C (*9A)	germline:unknown	pathogenic	tolerated	benign
rs2297595 c.496 A>G	Germline	drug response	deleterious	probably damaging
c.620	unknown	unknown	deleterious	probably damaging
rs754745863 c.771	unknown	unknown	unknown	unknown
rs45589337 c.775	Germline	likely benign	deleterious	benign
c.996	unknown	unknown	unknown	unknown
rs56038477 c.1236	unknown	unknown	unknown	unknown
rs57918000 c.1371	unknown	unknown	unknown	unknown

rs774799003 c.1396	unknown	unknown	tolerated	benign
rs1801158 c.1601 G>A (*4)	unknown	unknown	deleterious	benign
rs142619737 c.1615	unknown	unknown	deleterious	probably damaging
rs762029458 c.1617	unknown	unknown	unknown	unknown
rs1801159 c.1627 A>G (*5)	Germline	benign	deleterious	benign
rs17376848 c.1896 T>C	unknown	unknown	unknown	unknown
rs3918289 c.1905 (*2A)	unknown	unknown	unknown	unknown
rs1801160 c.2194 G>A (*6)	unknown	unknown	tolerated	benign
c.2586	unknown	unknown	unknown	unknown
rs1801267 c.2657 G>A (*9B)	Germline	pathogenic	deleterious	probably damaging
rs67376798 c.2846	germline	drug response	deleterious	possibly damaging
c.3025	unknown	unknown	tolerated	benign
rs773868825 c.3047	unknown	unknown	deleterious	probably damaging

From literature analysis, 3 SNPs emerged as mainly investigated in *DPYD* haplotype analysis even if contradictory results derive from evaluations conducted in high heterogeneous population or in small study cohort. For this reason, 85 T>C (*9A) rs1801265, 496 A>G rs2297595, and 2194 G>A (*6) rs1801160 variants were included in our analysis.

Statistically significant linkage between c.85C/c.496G was previously reported by Hamzic *et al*. using genetic linkage of *DPYD* variants in individual cohorts, combined cohort and the reference EUR-population retrieved from the LD ink database. LD (expressed as r2 or D') between rs2297595 c.496 A>G and rs1801160 c.2194 G>A (*6) and between rs2297595 c.496 A>G and rs1801265 c.85 T>C (*9A) were calculated in European population using ldlink.nci.nih.gov web site and reported in Table 5.20.

Table 5.20 Linkage Disequilibrium (LD) between rs2297595, rs1801160 and rs1801265.

Population	Variant 1	Variant 1 location	Variant 2	Variant 2 location	r^2	D
CEU	rs 2297595 T: 90.91%, C: 9.09%	1:98165091	rs 1801160 C: 93.94%, T: 6.06%	1: 97770920	0.0065	1
CEU	rs 2297595 T: 90.91%, C: 9.09%	1:98165091	rs 1801265 G: 15.66%, A: 84.34%	1:98348885	0.197	0.6048

A total of 6 of different haplotypes were identified including c.2194 G>A (*6), c.85 T>C (*9A) and c.496 A>G variants (Table 5.21).

Table 5.21 List of Haplotype generated and respective allele of rs2297595, rs1801160 and rs1801265.

	2194 G>A (*6)	c.85 T>C (*9A)	c.496 A>G
Hap1	G	T	A
Hap2	G	T	G
Hap3	G	C	A
Hap4	G	C	G
Hap5	A	T	A
Hap6	A	C	A

5.2.4 LARC patients' selection for FP-based therapy efficacy evaluation

Concerning the assessment of the impact of *DPYD* rs59353118 and rs4294451 and generated haplotype in FP efficacy, 212 LARC patients were included in the analysis. Baseline and demographic characteristics for these patients are shown in Table 5.22.

Table 5.22 Socio-demographic and clinical characteristic of selected 212 patients.

Characteristics	n	(%)
Gender		
Female	72	(33.9)
Male	140	(66.1)

5-FU	22	(10.4)
Capecitabine	190	(89.6)
TRG		
1	73	(33.0)
2	119	(17.1)
3	239	(35.8)
4	202	(12.2)
5	12	(1.9)
cT		
2	16	(7.55)
3	185	(87.26)
4	5	(2.36)
n.a.	6	(2.83)
cN		
0	40	(18.87)
1	117	(55.19)
2	28	(13.21)
4	6	(2.83)
pos	10	(4.72)
na	11	(5.19)

Male and female patients represented the 66.1% (n=140) and the 33.9% (n=72) respectively. Oral capecitabine was administered to most patients (n=190; 89.6%), while the remain (n=22; 10.4%) received 5-FU in bolus or in continuous infusion. Oral capecitabine dosage ranged from 1500 to 3800 mg/day based on the patients' conditions. Patients received a total of 45-55 Gy in 23 to 28 radiotherapy fractions. During the collection of patients' clinical data, therapy efficacy assessed in tumor biopsy during the post nCRT surgery and defined as TRG score, was well documented for each patient. Concerning pathological response, 33% (n=73) reached a pCR, while around 1.9% (n=12) had no response (TRG5).

Sex disaggregated analysis was performed on therapy response assessment to outline the prognostic role of sex alone or in combination with patients' genetic make-up. No statistical difference in therapy response in term of TRG emerged across the two sexes while, differently, a slight trend toward a better clinical therapy response in female population have been reported considering pCR (TRG1)

within TRG2, thus including patients with the presence of residual isolated cells scattered through the fibrosis (Table 5.23).

Table 5.23 Odd ratio (OR) and corresponding confidence intervals (CI) for TRG distribution according to sex.

	Male (n=140)	Female (n=72)	OR	95 % CI	p-value
TRG	n (%)	n (%)			
1	47 (33.5)	23 (31.9)			
2-5	93 (66.5)	49 (68.1)	10.767	0.5868 - 1.9756	0.8115
1-2	76 (54.3)	30 (41.6)			
3-5	64 (45.7)	42 (58.4)	16.625	0.9359 - 2.9531	0.0829

5.2.5 Assesment of the *DPYD* rs59353118 and rs4294451 variants on therapy efficacy: overall population and sex-disaggregated data

DPYD rs59353118 and rs4294451 frequencies, including the sex-disaggregated data, are shown in Table 5.24. Of the total 229 patients with efficacy data available, genetic analysis of selected SNP were correctly interpretable in 205 and 202 patients for *DPYD* rs4294451 and rs59353118 respectively. No differences in genotype distribution emerged (Table 5.24).

Table 5.24 Distribution of rs59353118 and rs4294451 in 212 LARC patients according to sex.

rs59353118	Tot pt (n= 202)	Male pt (n=134)	Female pt (n=68)	p-value
	n (%)	n (%)	n (%)	
AA	133 (66)	94 (70)	39 (57)	
AT	61 (30)	34 (25)	27 (40)	
TT	8 (4)	6 (5)	2 (3)	0.1076

rs4294451	Tot pt (n= 205)	Male pt (n=137)	Female pt (n=68)	p-value
	n (%)	n (%)	n (%)	
AA	133 (65)	89 (65)	44 (65)	
TA	67 (33)	45 (33)	22 (32)	
TT	5 (2)	3 (2)	2 (3)	0.9469

[$] p-value calculated using Chi-Square Test for the evaluation of different distribution across sexes.

Table 5.25 reported the correlation analysis of the 2 abovementioned *DPYD* variants with pathological response evaluating thought TRG scale. pCR or TRG1 was compared with TRG2-5. From overall evaluation of the total case study 2 of 212 LARC patients, patients carrying TT allele of rs59353118 variants showed a significant better response compared to AT and wild-type AA patients, with an incidence of TRG 1 of 63% (n=5), 21% (n= 14) and 35% (n=47) respectively (p-value = 0.043, Chi Square test). In recessive model, by considering together AT+TT genotype towards AA patients, no difference in term of OR was assessed. On the contrary, rs4294451 variant was positively correlated with worse therapy response in mutated patients in the analysis of AA genotype vs TA+TT genotype (OR= 21.084; 95 % CI= 1.09-4.06, p-value = 0.026) We further evaluate the impact of rs59353118 and rs4294451 patients genotype separately in men and female cohort. No sex-based difference emerged from the analysis.

Table 5.25 Odd ratio (OR) and corresponding confidence intervals (CI) for TRG according to *DPYD* polymorphisms by sex.

rs59353118		**AA**	**AT**	**TT**	**p-value[$]**	**OR***	**95 % CI**	**p value**
	TRG	n (%)	n (%)	n (%)				
Tot pt	1	47 (35)	14 (21)	5 (63)				
	2-5	86 (65)	47 (79)	3 (37)	**0.043**	14.382	0.761 - 2.719	0.263
Male	1	33 (35)	9 (26)	4 (67)				
	2-5	61 (75)	25 (74)	2 (33)	0.154	11.236	0.512-2.465	0.771
Female	1	14 (36)	5 (19)	1 (50)				
	2-5	25 (74)	22 (81)	1 (50)	0.254	21.467	0.706-6.5233	0.118

rs4294451		**AA**	**TA**	**TT**	**p-value[$]**	**OR***	**95 % CI**	**p value**
	TRG	n (%)	n (%)	n (%)				
Tot pt	1	50 (38)	15 (23)	1 (20)				
	2-5	83 (62)	52 (77)	4 (80)	0.079	**21.084**	**1.09 -4.06**	**0.026**
Male	1	34 (38)	11 (24)	1 (33)				
	2-5	55 (62)	34 (76)	2 (67)	0.281	18.545	0.849 -4.049	0.121
Female	1	16 (36)	4 (18)	0 (0)				
	2-5	28 (64)	18 (82)	2 (100)	0.202	26.571	0.829-9.842	0.096

[$] p-value calculated using Chi-Square Test; * OR calculated in recessive model (wild-type vs heterozygous + homozygous mutated).

5.2.6 Assessment of Response to nCRT in LARC patients: Analysis of *DPYD* Haplotype and Sex Impact

Among selected 229 LARC patients, genotyping using the NGS technique allowed the identification of *DPYD* haplotypes for the totality of the study population. Haplotype frequencies, including the sex-disaggregated data, are shown in Table 5.26.

Table 5.26 Distribution of *DPYD* haplotypes including 2194 G>A (*6), 85 T>C (*9A) and 496 A>G variants among 212 LARC patients according to sex.

	n	%	M	%	F	%	p-value[$]
HAP1							
0 allele	16	7.55	11	7.86	5	6.94	
1 allele	80	37.74	54	38.57	26	36.11	
2 allele	116	54.72	75	53.57	41	56.94	0.8924
HAP2							
0 allele	202	95.28	133	95.0	69	95.83	
1 allele	10	4.72	7	5.0	3	4.17	0.7864
HAP3	n	%	M	%	F	%	
0 allele	168	79.25	109	77.86	59	81.94	
1 allele	40	18.87	28	20.0	12	16.67	
2 allele	4	1.89	3	2.14	1	1.39	0.7697
HAP4	n	%	M	%	F	%	
0 allele	180	84.91	116	82.86	64	88.89	
1 allele	31	14.62	23	16.43	8	11.11	
2 allele	1	0.47	1	0.71	0	0	0.441
HAP5	n	%	M	%	F	%	
0 allele	193	91.04	130	92.86	63	87.50	
1 allele	18	8.49	10	7.14	8	11.11	
2 allele	1	0.47	0	0	1	1.39	0.2258
HAP6	n	%	M	%	F	%	
0 allele	211	99.53	140	100	71	98.61	
1 allele	1	0.47	0	0	1	1.39	0.1622

[$] p-value calculated using Chi-Square Test or Fisher exact test

As reported in Table 5.26, no differences in haplotype distribution across male and female population separately emerged.

Separate correlation analysis of the haplotypes was performed differently by comparing TRG1 vs TRG 2-4. In the entire study population, no significant differences in therapy response respect to the presence of any haplotype have been identified (Table 5.27).

Table 5.27 Haplotype correlation with TRG in 212 LARC patients.

		0 allele	**1 allele**	**2 allele**	**p-value[$]**
	TRG	n (%)	n (%)	n (%)	
HAP1					
	1	5 (31.2)	20 (25.0)	45 (38.8)	
	2-5	11 (68.8)	60 (75.0)	71 (61.2)	0.13
HAP2					
	1	67 (33.2)	3 (30.0)		
	2-5	135 (66.8)	7 (70.0)		1.00
HAP3					
	1	58 (34.5)	10 (25.0)	2 (50.0)	
	2-5	110 (65.4)	30 (75.0)	2 (50.0)	0.39
HAP4					
	1	62 (34.4)	8 (25.8)	0 (0)	
	2-5	118 (65.6)	23 (74.2)	1(100)	0.49
HAP5		0	1	2	
	1	65 (33.7)	5 (27.8)	0 (0)	
	2-5	128 (66.3)	13 (72.2)	1(100)	0.68
HAP6					
	1	70 (33.2)	0 (0)		
	2-5	141 (66.8)	1 (100)		1.00

[$] p-value calculated using Chi-Square Test or Fisher exact test

On the other hand, in the sex-disaggregated assessment of the impact of the haplotypes in therapy efficacy between the sexes, the HAP1, composed by wipe-type C allele for c.2194 (*6), A allele for c.85 (*9A) and T allele for c.496 A>G, emerged impacting, even if the correlation was not statistically significant, on response to treatment only in male population while the same trends have not been reported in female patients (p-value = 0.07) (Table 5.28).

Table 5.28 Haplotype correlation with TRG in 212 LARC patients by sex.

		0 allele	1 allele	2 allele	p-value[$]
	TRG	n (%)	n (%)	n (%)	
HAP1					
Male	1	5 (45.5)	12 (22.2)	30 (40.0)	
	2-5	6 (54.6)	42 (77.8)	45 (60.0)	0.074
Female	1	0 (0)	8 (30.8)	15 (39.6)	
	2-5	5 (100)	18 (68.2)	26 (63.4)	0.25
HAP2					
Male	1	45 (33.8)	2 (28.6)		
	2-5	88 (66.2)	5 (71.4)		0.77
Female	1	22 (31.9)	1 (33.3)		
	2-5	47 (68.1)	2 (66.7)		0.95
HAP3					
Male	1	37 (33.9)	8 (28.6)	2 (66.7)	
	2-5	72 (66.1)	20 (71.4)	1 (33.3)	0.41
Female	1	21 (35.6)	2 (16.7)	0 (0)	
	2-5	38 (64.4)	10 (83.3)	1 (100)	0.35
HAP4					
Male	1	42 (36.2)	5 (21.7)	0 (0)	
	2-5	74 (63.8)	18 (78.3)	1 (100)	0.31
Female	1	20 (31.3)	3 (37.5)		
	2-5	44 (68.8)	5 (62.5)		0.71
HAP5					
Male	1	44 (33.9)	3 (30)		
	2-5	86 (66.15)	7 (70)		1
Female	1	21 (33.3)	2 (25)	0 (0)	
	2-5	42 (66.7)	6 (75.0)	1 (100)	0.70
HAP6					
Male	1	-	-	-	
	2-5	-	-	-	-
Female	1	23 (32.4)	0 (0)		
	2-5	48 (67.6)	1 (100)		0.49

[$] p-value calculated using Chi-Square Test or Fisher exact test

With respect to HAP1 in the male population, it appeared to be involved in a worse therapy response in heterozygous with respect to wild-type and homozygous mutated patients. Additionally, analysing the distribution of TRG in male and female based on the haplotype composition, opposite trends emerged. For HAP1 in female population worse therapy response was reached in patients with 0 allele while in male patients, the heterozygous genotype conferred more risk to have a non-pCR. Opposite trends were highlighted for HAP2 and for HAP4 when male with 0 allele were more prone to respond respect to female. Regarding HAP6 no consideration could be deducted as all male patients were carried of 0 allele (Table 5.28).

HAP1 correlation with therapy response was performed in accord with the over-dominant SNP model, a condition wherein heterozygote shows different phenotype, by comparing separately the heterozygote with the wild-type and the homozygote mutated.

Evaluating the totality of the patients included in the efficacy analysis, no difference was reported in the risk of not responding to the FP-based therapy. The same result was obtained in the same analysis conducted only in female patients. Conversely, male patients bearing the identified DPYD haplotype with C-A-T allele in heterozygosis showed a 2.4-fold higher risk of not responding twith respect to male patients with 0 or 2 allele regarding HAP1 (Table 5.29).

Table 5.29 Odd ratio (OR) and corresponding confidence intervals (CI) for TRG according to *DPYD* haplotypes by sex.

HAP1	**TRG 1**		**TRG 2-5**				
	n	**(%)**	**n**	**(%)**	**OR**	**95% CI**	**p value**
Tot Pz (n=212)							
0-2 allele	50	(37.9)	82	(62.1)			
1 allele	20	(25.0)	60	(75.0)	1.82	0.988-3.33	0.27
Male (n=140)							
0-2 allele	35	(40.1)	51	(59.9)			
1 allele	12	(22.2)	42	(77.7)	**2.40**	**1.109-5.2**	**0.013**
Female (n=72)							
0-2 allele	15	(32.6)	31	(67.4)			
1 allele	8	(30.8)	18	(69.2)	1.088	0.368-3.068	0.436

[$]Estimated from univariate unconditional logistic regression model

5.2.7 Assessment of Safety profile of nCRT in LARC patients: Analysis of *DPYD* Haplotype And Sex Impact

Concerning the assessment of FP toxicity profile according to *DPYD* haplotype, 189 LARC patients enrolled at CRO Aviano between 2002 and 2019 were included in the analysis with available clinical data for ADRs development. Baseline and demographic characteristics for these patients are shown in Table 5.31. Similarly, patient sex, cT, cN and FP treatment scheduley distributions were reported as the previous case population for the safety assessment. Oral capecitabine was administered at a dosage between 1500 to 3800 mg/day based on patiens' conditions with concomitant 23 to 28 fractions of radiotherapy with a total dose ranged from 46 to 55Gy.

Table 5.31 Distribution of late toxicity among 189 LARC patients treated with FP-based chemotherapy.

Characteristics	**n**	**(%)**
Gender		
Female	64	(33.9)
Male	125	(66.1)
Fluoropyrimidine type		
5-FU	22	(11.6)
Capecitabine	167	(88.4)
cT		
2	14	(7.4)
3	159	(84.2)
4	9	(4.7)
n.a.	7	(3.7)
cN		
0	38	(20.1)
1	98	(51.8)
2	31	(16.4)
4	4	(2.1)
pos	10	(5.2)
na	7	(3.4)

Table 5.32 showed the distribution across the study population of the toxicities experienced by the patients. 8/189 (4.3%) patients developed G3-4 (4.8%) haematological toxicities, 9/189 (3.7%) non haematological toxicities, 7/189 toxicities not closely related to FP administration with a total of 20/189 (10.5%) patients which clinical data collected reported the development of any type of G3-4 toxicity.

Table 5.32 Distribution of toxicity among 189 LARC patients treated with FP-based chemotherapy.

Toxicity type (tot pz 189)	**G 1-4**		**G 3-4**	
	n	(%)	n	(%)
Haematological	58	(30.7)	8	(4.3)
Non Haematological	84	(44.4)	9	(4.8)
Haematological + Non haematological	115	(60.8)	16	(8.4)
Other	102	(76,12)	7	(3.7)
All toxicity	160	(84.6)	20	(10.5)

By classifying toxicity according to the class mentioned above, female patients emerged experienced more frequently non-haematological, haematological & non-haematological, Other and All toxicity with respect to male patients (Table 5.33). 10/72 (15.6%) female developed Haematological & non-haematological G3-4 toxicity while only 6/140 (4.8%) experienced the same adverse reaction type (p-value = 0.011). Additionally, considering all the toxicity type, thus incidence in female patients increased to 18.8% (12/64) while in male patients remain lower than 7% (8/125, 6.8%). No statistically significant results were obtained for only the assessment of haematological toxicity incidence since the percentage of toxicity in female population was reported double than in male ones.

Table 5.33 Distribution of toxicity among 189 LARC patients treated with FP-based chemotherapy.

	Male (n=125)		**Female (n=64)**		**p-value[$]**
Toxicity Type and Grade	G 0-2 n (%)	G 3-4 n (%)	G 0-2 n (%)	G 3-4 n (%)	
Haematological	121 (96.8)	4 (3.2)	60 (93.7)	4 (6.3)	0.32
Non-haematological	123 (98.4)	2 (1.6)	57 (89.1)	7 (10.9)	**0.004**

Haematological & non-haematological	119 (95.2)	6 (4.8)	54 (84.4)	10 (15.6)	**0.011**
Other	123 (98.4)	2 (1.6)	59 (92.2)	5 (7.8)	**0.032**
All	117 (93.6)	8 (6.4)	52 (81.2)	12 (18.8)	**0.009**

[$] p-value calculated using Chi-Square Test or Fisher exact test.

We further evaluate the haplotype frequencies, including the sex-disaggregated distribution. Data are shown in table 5.34. No differences in haplotype distribution in female with respect to males have been reported.

Table 5.33 Distribution of *DPYD* haplotypes among 189 LARC patients treated with FP-based chemotherapy.

	n	**%**	**M**	**%**	**F**	**%**	**p-value[$]**
HAP1							
0 allele	17	8.99	11	8.80	6	9.38	
1 allele	74	39.15	53	42.40	21	32.81	
2 allele	98	51.85	61	48.80	37	57.81	0.4334
HAP2							
0 allele	179	94.71	119	95.20	60	93.75	
1 allele	10	5.29	6	4.80	4	6.25	0.6734
HAP3							
0 allele	150	79.37	95	76.00	55	85.94	
1 allele	36	19.05	28	22.40	8	12.50	
2 allele	3	1.59	2	1.60	1	1.56	0.2586
HAP4							
0 allele	154	81.48	99	79.20	55	85.94	
1 allele	34	17.99	25	20.00	9	14.06	
2 allele	1	0.53	1	0.80	0	0	0.4545
HAP5							
0 allele	171	90.48	115	92.00	56	87.50	
1 allele	17	8.99	10	8.00	7	10.94	
2 allele	1	0.53	0	0	1	1.56	0.2932
HAP6							
0 allele	188	99.47	125	100.00	63	98.44	
1 allele	1	0.53	0	0.00	1	1.56	0.1611

[$] p-value calculated using Chi-Square Test or Fisher exact test

In table 5.34 the assessment of the predictive impact of the different haplotypes on toxicity event development have been reported. Toxicity grade 0-2 was compared with grade equal or higher of 3. By classifying toxicity according to patients HAP1, a significant correlation was reported for non-haematological toxicity: no heterozygous patients experienced any ADEs while 5.88% and 8.2 % of wild-type and homozygous mutated patients respectively reported at least one G3-4 events. Even if no significant, the same trend have been highlighted in the other toxicity class, with the except of haematological one, where a lower percentage of heterozygous patients reported severe ADEs compared to carriers of 0 or 2 alleles.

When looking for the toxicity experienced according to HAP2, any types of toxicity were reported with higherfrequency in heterozygous patients with respect to wild-type. A significant statistic correlation with toxicity defined as "other" and "all" with p-value = 0.005 and p-value = 0.04 respectively was assessed. No other significant results emerged for the remain haplotypes. No patients with 2 alleles for HAP3, HAP4, HAP5 and HAP6 experienced G3-4 toxicity.

Table 5.34 Distribution of toxicity among 189 patients treated with FP-based chemotherapy according to *DPYD* haplotypes.

	0 allele	**1 allele**	**2 allele**	**p-value[$]**
	G3-4 n (%)	G3-4 n (%)	G3-4 n (%)	
HAP1				
Haematological	1 (5.88)	5 (6.8)	2 (2.0)	0.29
Non-haematological	1 (5.88)	0 (0)	8 (8.2)	**0.04**
Haematological & non-haematological	2 (11.76)	5 (6.8)	9 (9.2)	0.74
Other	1 (5.9)	2 (2.7)	4 (4.1)	0.78
All	2 (11.8)	7 (9.5)	11 (11.2)	0.92
HAP2				
Haematological	7 (3.9)	1 (10.0)	-	0.35
Non-haematological	8 (4.5)	1(10.0)	-	0.42
Haematological & non-haematological	14 (7.8)	2 (20.0)	-	0.17
Other	5 (2.79)	2 (20.0)	-	**0.005**
All	17 (9.5)	3 (30.0)	-	**0.04**
HAP3				
Haematological	6 (4.0)	2 (5.6)	0 (0)	0.85
Non-haematological	9 (6.0)	0 (0)	0 (0)	0.29
Haematological & non-haematological	14 (9.3)	2 (5.6)	0 (0)	0.66

Other	6 (4.0)	1 (2.7)	0 (0)	0.88
All	17 (11.3)	3 (8.3)	0 (0)	0.72
HAP4				
Haematological	6 (3.9)	2 (5.9)	0 (0)	0.85
Non-haematological	8 (5.2)	1 (3.0)	0 (0)	0.83
Haematological & non-haematological	13 (8.5)	3 (8.8)	0 (0)	0.95
Other	7 (4.5)	0 (0)	0 (0)	0.43
All	17 (11.1)	3 (8.8)	0 (0)	0.87
HAP5				
Haematological	6 (3.5)	2 (11.8)	0 (0)	0.26
Non-haematological	9 (5.3)	0 (0)	0 (0)	0.61
Haematological & non-haematological	14 (8.2)	2 (11.8)	0 (0)	0.84
Other	6 (3.5)	1 (5.9)	0 (0)	0.86
All	18 (10.5)	2 (11.8)	0 (0)	0.93
HAP6				
Haematological	8 (4.3)	0 (0)	-	0.95
Non-haematological	9 (4.8)	0 (0)	-	0.82
Haematological & non-haematological	16 (8.5)	0 (0)	-	0.76
Other	7 (3.7)	0 (0)	-	0.84
All	20 (10.6)	0 (0)	-	0.73

[$] p-value calculated using Chi-Square Test or Fisher exact test

Next, detailed and disaggregated evaluation for the toxicity outcome have been separately assessed in the two sexes with regards to HAP1 and HAP2, the previously haplotypes emerging having a promising impact. The predictive role of the haplotypes on the occurrence of toxic events is shown in the Table 5.35.

Table 5.35 Distribution of toxicity among 189 patients treated with FP-based chemotherapy according to *DPYD* haplotypes by sex.

	0 allele	**1 allele**	**2 allele**	**p-value[$]**
	G3-4 n (%)	G3-4 n (%)	G3-4 n (%)	
HAP1				
MALE (n=125)				
Haematological	0 (0)	4 (7.5)	0 (0)	0.06
Non-haematological	0 (0)	0 (0)	2 (3.3)	0.34
Haematological & non-haematological	0 (0)	4 (7.5)	2 (3.3)	0.42
Other	0 (0)	0 (0)	2 (3.3)	0.34
All	0 (0)	4 (7.5)	4 (6.7)	0.64
HAP1				
FEMALE (n=64)				
Haematological	1 (16.7)	1 (4.8)	2 (5.4)	0.53
Non-haematological	1 (16.7)	0 (0)	6 (16.2)	0.14
Haematological & non-haematological	2 (33.3)	1 (4.8)	7 (18.9)	0.16
Other	1 (16.7)	2 (9.5)	2 (5.4)	0.59
All	2 (33.3)	3 (14.3)	7 (18.8)	0.57
HAP2				
MALE (n=125)				
Haematological	1 (0.8)	0 (0)	-	0.47
Non-haematological	2 (1.7)	0 (0)	-	0.74
Haematological & non-haematological	6 (5.1)	0 (0)	-	0.57
Other	2 81.7)	0 (0)	-	0.75
All	8 (6.7)	0 (0)	-	0.51
HAP2				
FEMALE (n=64)				
Haematological	3 (5.0)	1 (25.0)	-	0.10
Non-haematological	6 (10.0)	1 (25.0)	-	0.35
Haematological & non-haematological	8 (13.3)	2 (50.0)	-	**0.05**
Other	3 (5.0)	2 (50.0)	-	**0.001**
All	9 (15.0)	3 (75.0)	-	**0.003**

[$] p-value calculated using Chi-Square Test or Fisher exact test

Interesting, a divergent effect of the HAP2 emerged from the analysis while no significant results were obtained in the evaluation of the role of HAP1 in the two sexes separately. Specifically, female patients with 1 allele for HAP2 reported higher incidence of all kind of toxicity, with significant statistic data in the assessment of haematological & non-haematological, Other and All type of toxicity (p-value = 0.05, p-value = 0.0001, p-value = 0.003), conversely of what have been shown in men. Those data required further investigation for a greater understanding of the role that patients' sex plays. To note specifically with regard to HAP1, a similar trend for toxicity incidence emerged in female population, with carries of one allele more prone to develop G3-4 toxicity, as what has been previously reported for the evaluation of the entire case study population. On the contrary, in male study population, patients' carriers of one or two alleles presented a higher incidence of toxicity with respect to wild type.

A deep analysis was performed for assessing the toxicity risk for the haplotype previuously significant associated with toxicity development. Female patients have been associated with a 6.5, 8.7 and 17-fold higher risk of toxicity reported in table 5.36 while in men no significance was reported.

Table 5.36 Odd ratio (OR) and corresponding confidence intervals (CI) for G3-4 toxicity according to *DPYD* haplotypes in the total 189 LARC study population and by sex.

	allele	**Toxicity grade**		**OR**	**95 CI**	**p-value[$]**
		G 0-2 n (%)	G3-4 n (%)			
HAP1						
Non-haematological						
Tot pz (n =189)	0-2	106 (92.0)	9 (8.0)			
	1	74 (100)	0 (0)	-	-	-[a]
Male (n=125)	0-2	70 (97.0)	2 (3.0)			
	1	53 (100)	0 (0)	-	-	-[a]
Female (n=64)	0-2	36 (84.0)	7 (16.0)			
	1	21 (100)	0 (0)	-	-	-[a]
HAP2						
Haematological & non-haematological						
Tot pz (n =189)	0	165 (92.0)	14 (8.0)	2,946	0.57-15.231	0.09
	1	8 (80.0)	2 (20)			
Male (n=125)	0	113 (95.0)	6 (5.0)	-	-	-[a]

	1	6 (100)	0 (0)			
Female (n=64)	0	52 (87.0)	8 (13.0)	**6.5**	**0.79-52.8**	**0.04**
	1	2 (50.0)	2 (50.0)			
Other						
Tot pz (n =189)	0	174 (97.2)	5 (2.7)	**8.7**	**1.45-51.92**	**0.008**
	1	8 (80)	2 (20)			
Male (n=125)	0	117 (98.0)	2 (2)	-	-	- [a]
	1	6 (100)	0 (0)			
Female (n=64)	0	57 (95.0)	3 (5.0)			
	1	2 (50.0)	2 (50.0)	**19**	**1.95-185.39**	**0.005**
All						
Tot pz (n =189)	0	162 (90.5)	17 (9.5)	**4.084**	**0.96-17.26**	**0.028**
	1	7 (70)	3 (30)			
Male (n=125)	0	111 (93.0)	8 (7.0)	-	-	- [a]
	1	6 (100)	0 (0)			
Female (n=64)	0	51 (85.0)	9 (15.0)	**17**	**1.58-182.1**	**0.009**
	1	1 (25.0)	3 (75.0)			

[a] Not accountable: zero event in 1 allele patients with G3-4 toxicity [$]Estimated from univariate unconditional logistic regression model

5.3 CASE STUDY 3: EFFECT OF TUMOR BIOMARKERS IN PRE-TREATMENT TUMOR TISSUE EVALUATING IN IMMUNOHISTOCHEMISTRYOF LARC ON TREATMENT EFFICACY

Moreover, an ancillary analysis was performed in order to identify tumor protein biomarkers that were differentially expressed in responder (pCR) and in non-responder patients in a sex-dependent manner.

5.3.1 LARC patients' selection for FP-based therapy response evaluation of tumor biomarkers

A total of 95 patients were finally included in the IHC analysis according to the inclusion criteria. The median follow-up period was 53,2 months (range: 2,0-147,0) with 16/95 (16.84%) patients lost at follow up. Baseline characteristics were listed in the Table 5.37. The median age at diagnosis in around 65 years, with a minimum age of 25 years in a female patients and 80 years old in a male patient. All LARC patient were enrolled between 2003 and 2014 at CRO Aviano hospital were evaluated for the tumor staging and for the average of the mass distance from the anal verge, which median value was 6 cm with a minimum value of 2 cm and a maximum value of 12 cm.

Table 5.37 Socio-demographic and clinical characteristic of selected 95 LARC patients.

Patients baseline characteristic	N	%
All	95	100
Median age at diagnosis. years (range)	65 (25-85)	
Median distance from anal verge. cm (range)	6 (2-12)	
Imaging pre-treatment staging (tnm)		
Ct2n+	6	6.3
Ct3n0	21	22.1
Ct3n+	67	70.5
N.A.	1	1.1

The overall study population comprehend 68 (71.6%) male patients and 27 (28.4%) female patients (Figure 5.7).

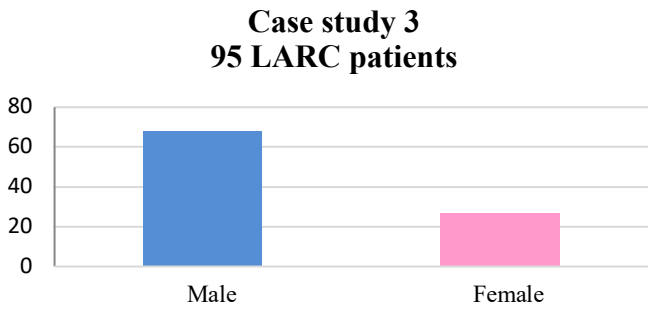


Figure 5.7 Distrubution of 95 LARC patients according to sex.

All patients have completed the scheduled therapy which comprehend several chemo- and radiotherapy scheme. More than 80% of patients were treated with FP-based chemotherapy, in monotherapy (n=45, 57.4%) or with concomitant administration (n=41, 43.2%) of oxaliplatin (n=27), gefitinib (n=9), raltitrexed (n=4) or irinotecan. Concomitant radiotherapy was administered at dosage from 4600 to 5520 cGy. The interval between the end of nCRT and surgery was between 48 and 90 days. Surgery type included low anterior resection for 57 patients, local excision for 13 patients, transanal local excision for 5 patients, transanal endoscopic microsurgery for 4 patients and other type for 16 patients (Table 5.38).

Table 5.38 Treatment characteristic of 95 LARC patients.

Patients treatment characteristic	**N**	**%**
Chemotherapy		
Fluoropyrimidines monotherapy	45	47.4
Fluoropyrimidines + other[1]	41	43.2
No chemotherapy	9	9.5
Radiotherapy		
≤5040cGy	87	91.6
>5040cGy	8	8.4
Type of surgery		
Low anterior resection (LAR)	57	60.0
Local excision (LE)	13	13.7
Transanal local excision (TALE)	5	5.3
Transanal endoscopic microsurgery (TEM)	4	4.2
Others*	16	16.8

* Oxaliplatin (N=27), gefitinib (N=9), raltitrexed (N=4), irinotecan (N=1)

5.3.2 Assessemnt of nCRT response in case study 3 LARC patients

After completing nCRT and surgery intervention, the pathological response was assessed in surgery biobsy: 25/95 (26.32%) patients achieved pCR (ypT0N0) while 70/95 patients (73.68%) patients experienced partial or null tumor response to therapy. The evaluation of patient's tumor response regarding TRG scale highlighted a TRG1 in 25/95 (26.3 % patients), TRG2 in 13/95 (13,7%) patients, TRG3 in 43/95 (45.3%) patients, and TRG4 in 14/95 (15.6%) patients. No patients reported a TRG5, 12/95 (12.63%) patients during the treatment period have developed distant metastasis, especially at level of peritoneum and in the liver. At the better of our knowledge 26/95 (27.37%) patients have received adjuvant FP-based chemotherapy, 31/95 (32.63%) patients experienced distant or local recurrence after surgery (Table 5.39).

Table 5.39 Treatment outcome of 95 LARC patients.

Patients treatment outcome	n	%
Tumor Regression Grade (by Mandard's)		
1	25	26.3
2	13	13.7
3	43	45.3
4	14	15.6
Pathological response evaluation (T)		
0	25	26.32
1	5	5.26
2	23	24.21
3	35	36.84
4	5	5.26
n.a.	2	2.11
Pathological response evaluation (N)		
0	54	56.84
1	15	15.79
1b	1	1.05
2	8	8.42
3	4	4.21
x	11	11.58
n.a.	2	2.11

Pathological response evaluation (M)		
0	83	87.37
1	9	9.47
x	1	1.05
n.a.	2	2.11
Recurrence Status after surgery		
no	62	65.26
yes	31	32.63
n.a.	2	2.10

Separating the assessment of therapy response by sex, data highlighted that TRG1 was achieved in 18/68 (26 %) of male patients and in 7/27 (26%) female patients. No statistical difference emerged comparing male and female treatment response (Table 5.40).

Table 5.40 Distribution of treatment outcome of 95 LARC patients by sex.

Patients sex		**TRG 1 (n=25)**	**TRG 2-5 (n=70)**	**p-value[$]**
Male (n=68)	n	18	118	
	%	26	74	
Female (n=27)	n	7	47	
	%	26	74	
		25	70	1

[$] p-value calculated using Chi-Square Test or Fisher exact test

5.3.3 Impact of the IHC biomarkers on nCRT response in case study 3 LARC patients

Tumor pre-treatment biopsy were available as an inclusion criterion for all the patients included in the study population, Tumor biopsy collected at the baseline timepoint were fixed in formalin (n=60) and the in bouin (n=35).

Expression of MLH1, GLUT1, Ki67, CA-IX, CXCR4, COX2, CXCL12, HIF1alfa, VEGF, CD44 and RAD51 was assessed in each tumor slides and available results were collected. The evaluable biopsy results for the different tumor biomarkers were reported in Table 5.41.

Table 5.41 Evaluable biopsies for selected biomarkers in 95 LARC patients and by sex.

Tumor markers	Cases evaluable (%)	Male cases evaluable (%)	Female cases evaluable (%)
MLH1	87/95 (91.58)	61/68 (89.71)	26/27 (96.30)
GLUT1	95/95 (100)	68/68 (100)	27/27 (100)
Ki67	95/95 (100)	68/68 (100)	27/27 (100)
CA-IX	95/95 (100)	68/68 (100)	27/27 (100)
CXCR4	81/95 (85.26)	59/68 (86.76)	22/27 (81.48)
COX2	95/95 (100)	68/68 (100)	27/27 (100)
CXCL12	95/95 (100)	68/68 (100)	27/27 (100)
HIF1alfa	95/95 (100)	68/68 (100)	27/27 (100)
VEGF	95/95 (100)	68/68 (100)	27/27 (100)
CD44	95/95 (100)	68/68 (100)	27/27 (100)
RAD51	90/95 (94.74)	65/68 (95.59)	25/27 (92.59)

Cellularity percentage (fraction of imminoreactive cell) and immunostaging intensity were evaluated for each biomarker and H-score was calculated. Finally, the H-score and cellularity were correlated with therapt response in the total population and in sisaggregated analysis according to patients' sex.

5.3.3.1 Analysis of IHC biomarkers between TRG1 vs TRG2-4 patients

In the all-study population, evaluating the difference of the biomarkers level comparing TRG1 versus TRG2-4, a significant correlation was assessed for CXCR4 and COX2 (Table 5.42). Regarding CXCR4, pCR was achieved in patients with higher H-score compared to non-responders (H-score 3 vs 2, p-value = 0,010). Similarly, for COX2, the median H-score was higher in responders compared to non-responders (H-score 6 vs 4, p-value =0.030). A trend for Ki67 protein expression highlighted that patient with higher H-score were more likely to get a worse tumor response (H-score 6 vs 9, p-value = 0.059). No correlation between cellularity and TRG response was found significant. A weak tendency was found for VECF cellularity where responders exhibited a lower cellularity value (40% vs 60%, p-value = 0.081).

Table 5.42 Median value and interquartile range (Q1-Q3) of H-score and neoplastic cellularity for selected parameters according to TRG status comparing TRG1 vs TRG2-5.

	Patients		**H-score (Huang F)**			**Cellularity (%)**		
	TRG1	**TRG2-4**	**TRG1**	**TRG2-4**	**p-value**	**TRG1**	**TRG2-4**	**p-value**
MLH1	25	62	6 (1-12)	8 (4-12)	0.126	60 (20-90)	80 (40-90)	0.221
GLUT 1	25	70	6 (4-8)	6 (3-9)	0.903	50 (30-60)	60 (40-70)	0.331
Ki67	25	70	6 (3-9)	9 (6-9)	**0.059**	40 (20-70)	60 (40-70)	0.070
CA IX	25	70	2 (1-2)	2 (0-2)	0.555	5 (2-20)	10 (0-20)	0.721
CXCR4	20	61	3 (2-5)	2 (1-3)	**0.010**	35 (20-50)	20 (10-50)	0.188
COX2	25	70	6 (3-8)	4 (3-6)	**0.030**	70 (60-80)	70 (50-80)	0.649
CXCL12	25	70	2 (2-6)	3 (1-6)	0.778	30 (20-50)	30 (10-50)	0.861
HIF1alfa	25	70	4 (2-6)	6 (2-8)	0.345	70 (40-80)	70 (40-80)	0.910
VEGF	25	70	2 (1-3)	3 (1-6)	0.179	40 (10-60)	60 (20-80)	0.081
CD44	25	70	6 (6-8)	6 (4.8)	0.608	60 (50-80)	70 (40-70)	0.799
RAD51	25	68	4 (1-6)	4 (2-6)	0.254	45 (20-50)	50 (30-65)	0.177

[a]Mann-Whitney test

5.3.3.2 Sex-disaggregated analysis of IHC biomarkers between TRG1 vs TRG2-4 patients

The same Mann-Withney analysis was performed separately for male and female population (Table 5.43 and Table 5.44). In the male population, none of the selected biomarker values were significantly associated with tumor response (Table 5.43), contrary of what has emerged in the total study population analysis. Similarly, with the previous analysis for the cellularity predictive response capacity, a weak tendency remained for the lower VEGF cellularity in responders.

Table 5.43 Median value and interquartile range (Q1-Q3) of H-score and neoplastic cellularity for selected parameters according to TRG status in male patients comparing TRG1 vs TRG2-5.

	Male patients		**H-score (Huang F)**			**Cellularity (%)**		
	TRG1	**TRG2-4**	**TRG1**	**TRG2-4**	**p-value[a]**	**TRG1**	**TRG2-4**	**p-value[a]**
MLH1	18	43	7 (1.75-12)	9 (6-12)	0.349	70 (27,50-90)	80 (40-90)	0,50
GLUT 1	18	50	6 (4-9)	6 (3-8)	0,6157	50 (30-60)	60 (40-70)	0,29
Ki67	18	50	7,5 (5,25-9,75)	9 (8,25-9,75)	0,1369	45 (27,50-72,50)	60 (47,50-72,50)	0,161
CA IX	18	50	1,5 (1-2)	1 (0-2)	0,6354	5 (4,75-20)	10 (5-20)	0,532
CXCR4	16	43	2,5 (2-4)	2 (1-3)	0,1329	30 (12,50-62,50)	20 (10-60)	0,47
COX2	18	50	6 (3-8)	4 (3-6)	0.1642	70 (60-80)	70 (50-80)	0.433
CXCL12	18	50	3 (2-6)	2 (1-6)	0.2590	30 (20-52.50)	30 (10-50)	0.3115
HIF1alfa	18	50	4 (2.75-8)	6 (1.75-8)	0.7638	75 (40-80)	70 (22.75-80)	0.651
VEGF	18	50	2 (0-4)	3 (1-6.5)	0.1253	30 (0-62.50)	55 (20-80)	0.0856
CD44	18	50	6 (6-9.75)	6 (4-8)	0.2924	70 (47.50-80)	70 (40-70)	0.4042
RAD51	17	48	4 (2.5-6)	4 (2.25-6)	0.77	50 (30-55)	50 (30-70)	0.3178

[a]Mann-Whitney test

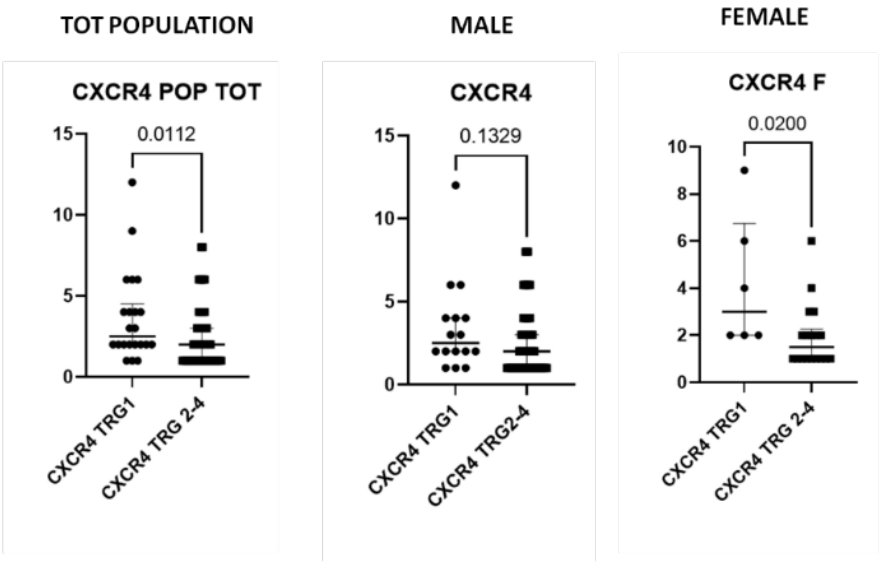
In the analysis performed in the female population, CXCR4, COX2, CXCL12 and RAD51 were found significantly associated with tumor response (Table 5.44). Specifically, higher H-score value of CXCR4 and COX2 were associated with better response (CXCR4, H-score 3 vs 1,5, p-value = 0.02; COX2, H-score 5 vs 4, p-value = 0.038) as emerged in the total study population analysis, while for CXCL12 and RAD51, responders showed a lower H-score value than non-responders (CXCL12 H-score 1 vs 4, p-value = 0.024; RAD51, H-score 1 vs 4, p-value = 0.044). In the female population, no tendency was found for VEGF cellularity value and TRG value while a statistically significant correlation was assessed for CXCL12 cellularity with lower value in responders compared to non-responders (20% vs 35%, p-value = 0.036), This result was in line with the H-score correlation result.

Table 5.44 Median value and interquartile range (Q1-Q3) of H-score and neoplastic cellularity for selected parameters according to TRG status in female patients comparing TRG1 vs TRG2-5.

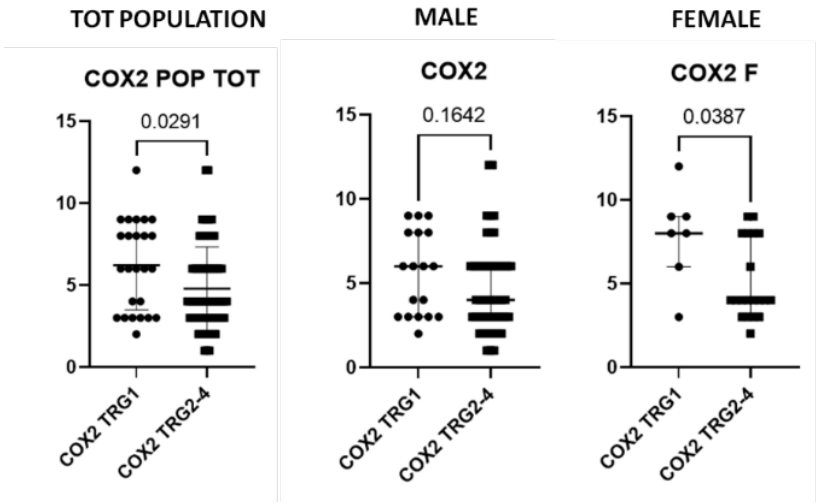
	Female patients		**H-score (Huang F)**			**Cellularity (%)**		
	TRG1	**TRG2-4**	**TRG1**	**TRG2-4**	**p-value[a]**	**TRG1**	**TRG2-4**	**p-value[a]**
MLH1	7	19	2 (1-6)	6 (3-12)	0.1284	40 (20-60)	70 (40-80)	0.199
GLUT 1	7	20	6 (4-8)	6 (4-9)	0.5930	50 (30-70)	50 (40-70)	0.775
Ki67	7	20	6 (3-9)	9 (6-9)	0.2626	30 (20-60)	50 (30-67.50)	0.207
CA IX	7	20	2 (1-4)	2 (1-2.75)	0.6816	5 (5-40)	10 (5-20)	0.815
CXCR4	6	18	3 (2-6.75)	1.5 (1-2.25)	**0.02**	40 (30-52.50)	20 (10-45)	0.170
COX2	7	20	5 (6-9)	4 (3.25-8)	**0.0387**	70 (60-80)	80 (60-80)	0.575
CXCL12	7	20	1 (1-2)	4 (2-6)	**0.0239**	20 (10-20)	35 (15-60)	**0.036**
HIF1alfe	7	20	3 (0-4)	5 (2.25-7.5)	0.1520	70 (0-80)	70 (42.50-80)	0.613
VEGF	7	20	3 (2-3)	3 (1.25-4)	0.7308	50 (30-60)	65 (15-77.50)	0.3654
CD44	7	20	6 (2-6)	6 (4-9)	0.5498	60 (30-70)	70 (32.50-77.50)	0.3747
RAD51	5	20	1 (1-3.5)	4 (2-6)	0.0447	20 (10-40)	40 (22.50-50)	0.1213

[a]Mann-Whitney test

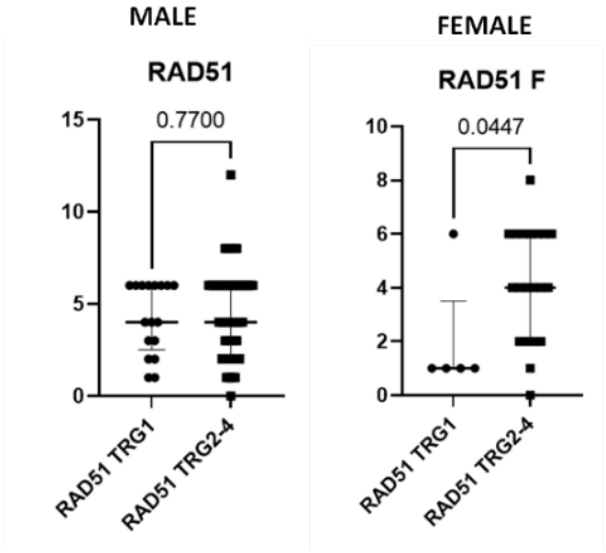
a.



b.



c.



d.

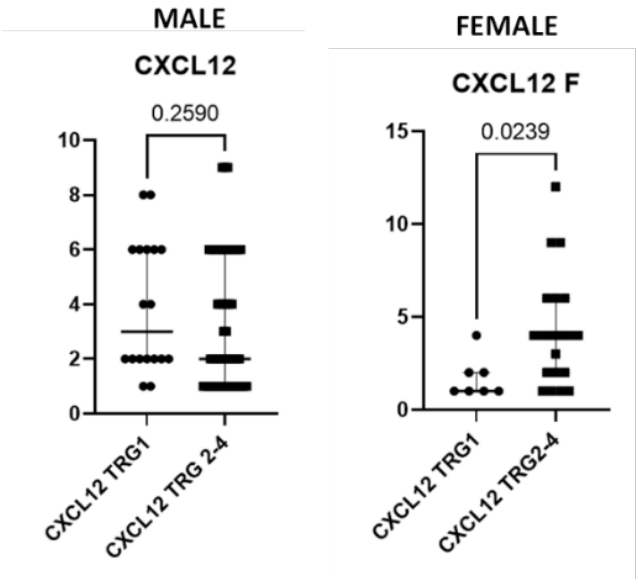


Figure 5.8 Scatter plot of CXCR4, COX2, RAD51 and CXCL12 tumor biomarkers in the totalstudy population and by sex.

5.3.3.3 Analysis of IHC biomarkers between TRG1-2 vs TRG3-4 patients

Patients' biomarkers H-score and cellularity were evaluated and compared in patients with TRG1-2 versus patients with TRG 3-4 (Table 5.45). In the total population, responders showed a significantly higher H-score value for COX2 (H-score 6 vs 4, p-value = 0.011). The association of CXCR4 and TRG1 vs TRG2-4 showed only a trend given by the quartile range (H-score 2 vs 2, p-value = 0.0617). Additionally, a trend was found for Ki67 and fore VEGF H-score (Ki67, H-score 7 vs 9, p-value = 0,072; VEGF, H-score 2 vs 3, p-value = 0.09), with higher values in non responders compared with responders. Considering cellularity parameter, a weak trend was found for the same two biomarkers, Ki67 (45% vs 60%, p-value = 0.086) and VEGF (35% vs 60%, p-value = 0.064).

Figure 5.45 Median value and interquartile range (Q1-Q3) of H-score and neoplastic cellularity for selected parameters according to TRG status comparing TRG1-2 vs TRG3-5.

	Total Patients		**H-score (Huang F)**			**Cellularity (%)**		
	TRG1-2	**TRG3-4**	**TRG1-2**	**TRG3-4**	**p-value[a]**	**TRG1-2**	**TRG3-4**	**p-value[a]**
MLH1	36	51	6 (1-12)	8 (4-12)	0.1802	60 (20-90)	80 (60-90)	0.1958
GLUT 1	38	57	6 (4-9)	6 (3-9)	0.7574	50 (37.50-70)	60 (40-70)	0.5362
Ki67	38	57	7.50 (6-9)	9 (6-9)	0.0720	45 (30-70)	60 (40-70)	0.0865
CA IX	38	57	2 (1-2.250)	1 (0.50-2)	0.6368	5 (4.75-20)	10 (2-20)	0.8820
CXCR4	32	51	2 (2-4)	2 (1-3)	0.0617	40 (12.50-50)	20 (10-60)	0.2223
COX2	38	57	6 (3-8.25)	4 (3-6)	**0.0108**	70 (60-80)	70 (50-80)	0.3351
CXCL12	31	57	2 (2-6)	2 (1-6)	0.5083	30 (20-50)	30 (10-50)	0.5558
HIF1alfe	38	57	4 (1-8)	6 (2.50-8)	0.1958	65 (10-80)	70 (45-80)	0.7153
VEGF	38	57	2 (1-3)	3 (1-6)	0.0907	35 (20-60)	60 (20-80)	0.0640
CD44	38	57	6 (4-8)	6 (6-9)	0.2830	60 (30-80)	70 (50-70)	0.5617
RAD51	34	56	4 (1.75-6)	4 (3-6)	0.4105	50 (20-60)	45 (30-67.50)	0.2795

[a]Mann-Whitney test

5.3.3.4 Sex-disaggregated analysis of IHC biomarkers between TRG1-2 vs TRG3-4 patients

Considering only male patients, no trend was found for Ki67 H-score and TRG, while a trend remained for VEGF H-score and cellularity parameter, with higher H-score and cellularity value in worst responders (H-score 2 vs 3, p-value = 0.065; 30% vs 60%, p-value = 0.082). A new tendency was found for higher CXCL12 H-score correlated with better nCRT response (H-score 6 vs4; p-value

= 0,0612). A significativity remains for higher H-score value in TRG1-2 patients compared with TRG3-4 patients (H-score 6 vs 4, p-value = 0.0498).

Figure 5.46 Median value and interquartile range (Q1-Q3) of H-score and neoplastic cellularity for selected parameters according to TRG status comparing TRG1-2 vs TRG3-5 by sex.

			H-score (Huang F)			Cellularity (%)		
	TRG1-2	**TRG3-4**	**TRG1-2**	**TRG3-4**	**p-value[a]**	**TRG1-2**	**TRG3-4**	**p-value[a]**
Male patients								
COX2	29	39	6 (3-8)	4 (3-6)	**0.0498**	70 (60-80)	60 (50-80)	0.2330
Female patients								
CXCR4	6	18	3 (2-6.75)	1.5 (1-2.25)	**0.02**	40 (30-52.50)	20 (10-45)	0.1703
COX2	9	18	8 (5-9)	4 (3-8)	**0.04321**	80 (65-80)	75 (57.50-90)	0.8902
CXCL12	9	18	2 (1-3)	4 (1.75-6)	**0.0390**	20 (10-30)	35 (10-60)	0.1069

[a]Mann-Whitney test

In female population analysis, TRG1-2 patients showed a significantly higher CXCR4 and COX2 H-score and a lower CXCL12 H-score (CXCR4, H-score 3 vs 1,5, p-value = 0,02; COX2, H-score 8 vs 4, p-value = 0,043; CXCL12, H-score 2 vs 4, p-value = 0,039). Similarly, to the whole population analysis, a trend emerged for Ki67 H-score and cellularity (6 vs 9, p-value = 0,06; 30% vs 50%, p-value = 0.08), and for VEGF cellularity (50% vs 70%, p-value = 0.066, data not shown). Additionally, a weak tendency emerged for HIF1alfa H-score with lower value in patients with a better response compare to whom with worst response (3 vs 6, p-value = 0.057).

6 DISCUSSION AND CONCLUSION

FP, including 5-FU and its prodrug capecitabine, are fluourinated pyrmidine analogues generally used in the treatment of common solid tumors such as CRC as monotherapy or in combination with chemotherapy regimen [61]. The enzyme DPD, encoded by the *DPYD* gene, plays an important role in the catabolism and clearance of 5-FU [72,74,212]. As a matter of fact, it is the main rate limiting enzyme in the catabolic pathway of 5-FU [129,130] and, for this reason, any alteration of enzymatic activity may lead to toxic accumulation of 5-FU [72–74]. T Reduction of enzymatic activity increases the half-life of the drug, resulting in excessive accumulation of the drug and associated toxicity [63–65]. Lack of DPD activity has been shown to result in severe grade 3 or higher 5-FU drug-related toxicities, including grade 3 or higher. In particular, cancer patients with complete or nearly complete DPD deficiency suffered from severe toxicity, which was even life-threatening or fatal after FP therapy [213,214]. The management of suspected acute severe toxicity associated with administration of FP includes 5-FU dose adjustment, delays in chemotherapy administration, or even interruption of further therapy. In addition, in many cases toxicity usually required extensive medical intervention and patients were given grow factor supplementary therapy, especially in cases of haematological adverse reaction [72].

The response to therapy and the toxicological profile of FP have shown a heterogeneous trend in the population, highlighting the need to evaluate diversified therapy for individual patients. Patients' selection for chemotherapy treatment is largely based only on preoperative clinical stage, as determined by imaging techniques, and in most cases does not take into account interindividual differences between patients. The ability to predict response to treatment has not been demonstrated [215]. The wide heterogeneity in clinical outcomes reported in the literature may be due to genetic differences in patients, differences in the genetic nature of the tumor, and subjective and environmental variables.

To date, the identified and clinically recognized toxicity marker for FP based therapy is the enzyme DPD. DPD deficiency is estimated to occurs in 3-5% of the total population [64–66,72] and of the 2 million patients receiving 5-FU annually in the United States, more than 30% experience toxicity associated with 5-FU [216]. It is also noted that among the population group suffering from FP-related toxicity, more than 50% have DPD enzyme deficiency [217]. In Europe, up to 9% of patients carry a DPD gene variant that reduces enzyme activity, and approximately 0.5% of patients lack DPD completely [218]. Studies have shown that there is a significant correlation between the presence of genetic determinants and enzymatic activity, which also is correlated with a different toxicological

profile. Numerous genetic variants in *DPYD* are known to alter protein sequence or mRNA splicing, whereas others are not known to affect DPD activity in a clinically relevant manner, according to current knowledge. In fact, the association between *DPYD* genotype and phenotype has been clearly demonstrated for only a few variants, whereas the functional impact of many rare variants has been assessed only *in vitro*, with limited supporting data. Four *DPYD* variants are widely recognized for their association with severe toxicity, including c.1129-5923 C>G (HapB3), c.1679T>G (*13), c.1905+1G>A (*2A), and c.2846A>T [219]. Therefore, the above genetic variants have been actively included in the FP prescribing process with authoritative PGx guidelines that allow the risk of serious and sometimes fatal AEs to be significantly reduced. Current recommendations for reducing the FP dosage in mutated patients for these variants are based primarily on studies from the Netherlands, in which a 50% reduction in the originally planned FP dose was correlated with a 28% reduction in FP-induced CTCAE grade ≥3 toxicity and a 10% to 0% reduction in mortality [86]. In another analysis, a genotype-based 5-FU dose reduction of 25–50% resulted in a significant reduction in toxicity compared with controls with a relevant risk of 1.31–4.00 depending on genotype [220]. These data support a 50% dose reduction upfront in patients with these variants in heterozygous form [76]. As reported in the PGKB guidelines, and since these clinically relevant variants are predictors of FP-associated toxicity, upfront genetic screening is recommended to improve the safety of cancer patients treated with FP. In addition to *DPYD* genotype-guided dose reduction, dose-titration is strongly recommended based on patients' tolerability to avoid underdosing in patients who could tolerate higher doses. Recently, EMA has included a recommendation to test these variants *a priori* before starting FP-based treatment [87].

However, targeted genotyping is limited to testing known toxicity-associated *DPYD* variants and may not include any additional variants with functional effect that may also be deleterious to DPD function and contribute to severe toxicity [221]. The frequency of deleterious gene SNPs appears to be relatively low, leaving a substantial number of severe 5-FU toxicity cases unexplained [63,131,222–224]. Novel intronic and exonic variants associated with a deficiency of the enzymatic activity have been identified, but characterization of their impact on clinical outcome requires further investigation.

As a result, there is an increasing need for early determination of DPD activity and *DYPD* mutation to better identify patients at high risk, which will ultimately allow clinicians to select more appropriate treatments for patients and improve overall outcomes [63,225]. Particularly, in oncology it would be desirable to significantly improve patient life expectancy and avoid over- or under-treatment of patients. Thus, prognostic and predictive markers could be introduced into clinical practice at the molecular level and in the context of the patients' particular physical condition.

In addition to patient genetics, one of the most important biological characteristics that has recently emerged is the sex of patients. The importance of sex differences in the incidence, aetiology, and treatment of the disease is well known in other medical fields, but remains an underappreciated issue in oncology. *In vitro* and *in vivo* studies do not include separate analysis for the two sexes. Sex contributes to variability in chemotherapy metabolism and dose response between patients which may influence both efficacy and toxicity. However, comparative data on differences between sexes are lacking. Recently, sex differences in the toxicity of 5-FU using human colon cancer cell lines, xenograft mouse models, and Korean patients data were investigated for the first time, highlighting an increased risk of toxicity in women during treatment of CRC [226]. Clinically relevant differences between the sexes include tumour biology, immune system activity, body composition, and drug disposition and effects. A gender gap in pharmacological treatment has been found in the PK and PD parameters of drugs. The largest sex differences are reported in epidemiologic studies examining susceptibility to cancer and survival between the sexes, with males having higher risk and poorer outcome for several non-sex-specific cancers [227]. On the contrary, sex-related differences in therapeutic efficacy and toxicity have been less explored issues underlying unrecognised mechanisms.

In addition, patients' sex has not been included so far among the toxicity risk factors in clinical guidelines. The inclusion of sex is supported by evidence in the literature suggesting differential DPD enzyme activity in males and females and differential toxicity risk with FP-based treatment. Several biological factors may play a role in these differences but a partial explanation may be a lower capacity for 5-FU clearance in women, as shown by at least two pharmacological studies in which lower clearance in women correlates with differences in DPD activity [128,228]. In fact, men have been reported having a 26% higher elimination of 5-FU and a 18% higher apparent elimination of 5FUH2 with respect to women [229]. Among others, Yamashita *et al*., found significantly lower DPD levels in females in a cohort of 97 CRC patients resulting in higher plasma levels of 5-FU causing higher toxicity and better long-term outcome [134]. In addition, Sloan *et al*. observed a more frequent occurrence of stomatitis, leukopenia, alopecia, nausea, vomiting, and diarrhoea in women compared with men [230]. 5-FU toxicity was more extensive in women than in men in terms of average maximum toxicity grade, number of different types of toxicity, and frequency of severe toxicity. Particularly, the incidence of both grade 2 haematological toxicity and moderate to severe non-haematological toxicity were higher in women than in men [231]. Female CRC patients treated with adjuvant capecitabine experience higher dose-limiting toxicity than male patients when the drug is administered according to body surface area [232]. Moreover, the differences between the two sexes in genotype distribution indicate that SNPs may be sex-specific biomarkers [233]. Although there are

no differences in allele frequencies for some other variants, genotypes may have a sex-specific role in toxicity risk.

Given the well-documented higher incidence of toxicity in the female population treated with FP, the identification of SNPsbecame even more important. Nevertheless, the mechanisms responsible for the sex differences in the incidence of AEs caused by treatments with FP and the histologic differences in efficacy are poorly understood.

Recent evidence also suggests that sex differences in tumour biology and molecular markers deserve more attention and systematic investigation in cancer biology and treatment. The assessment of sex differences is even more important in non-sex-specific cancers where men and women should be considered as biologically distinct patient populations. Interventional clinical trials that consider sex as a key variable are needed to improve the balance between drug efficacy and toxicity and to develop personalised dosing regimens [103]. Sex medicine, which considers specific treatment approaches, has recently emerged in response to an unresolved issue.

Similarly, no prognostic markers of response capable of distinguishing responding patients from nonresponding patients were found in the literature. The results of IHC analysis of the biopsy specimen, evaluating marker intensity and cellular positivity, were associated with response to therapy. Evaluation of the expression of specific markers in the tumor biopsy specimen at the time of diagnosis would deserve more attention, as this could help physicians choose the right therapy. For example, patients with high expression of markers for poor prognosis could be treated with more aggressive therapy, while patients with high expression of markers for good response to therapy could be treated with a lower dose.

In this work, novel genetic variants, haplotypes, and tumor characteristics were investigated to better stratify male and female populations as novel determinants of treatment safety and efficacy. Specifically, the clinical context of CRC and LARC patients was identified as a model for evaluating the interplay between patient genetic markers, molecular tumor markers, and patient sex.

Based on these evidences, the genetic analysis has focused particularly on the *DPYD* gene variants because the DPD enzyme is responsible for 80–90% of 5-FU metabolism. Starting with the well-documented *DPYD* variants and then emerging *DPYD* variants, associated with a different enzyme expression in the liver, the possible role of patients' sex has been investigated.

6.1 DPYD NEW UN-INVESTIGATED VARIANTS ANALYSIS IN CRC PATIENTS TO PREDICT NCRT TOXICITY RISK

A correlation analysis between the presence of at least one of the four common variants and the safety profile of the drug in the CRC study population revealed a significantly increased risk of grade 3-4 haematologic and non-haematologic toxicities, as well as all toxicities combined, when considered over the entire course of therapy. These data are consistent with those in the literature, where it is reported that approximately 50% of carriers of *DPYD* variant with decreased function develop severe 5-FU-related toxicity with standard doses [131,223,234,235].
There is substantial evidence of the clinical validity of *DPYD**2A and c.2846A>T, and current guidelines recommend dose reduction of FPs in patients carrying these variants [78,236].

Nowadays, few DPYD*13 mutated patients have been described in a clinical setting, and data from a meta-analysis showed that the risk of global severe toxicity was about four times higher in mutated patients [223,237,238]. Similar functional impairment has been reported for DPYD*2A and for *13, in which a heterozygous genotype resulted in a 40-50% reduction in DPD activity [239].
The clinical validity of c.1236G>A/HapB3 remained unclear for many years until Van Kuilenburg and colleagues reported that normal mRNA production was not completely absent and that DPD activity was reduced by approximately 50% in homozygous patients for c.1236G>A/HapB3 and was not completely impaired [234].
Moreover, the differential impact of sex on toxicity in wild-type and mutated patients demontrated that, in addition to patient genotype, patient sex also influences the risk of developing cumulative haematological toxicities. Of note, mutated female patients tended to develop ADEs more frequently in all toxicity categories, although not significantly. It is well-known from the literature that females are more prone to toxicity after FPs administration, but these data appear to be the first to also show a correlation with the genetic component. These data can also be explained by the well-documented lower 5- FU clearance in females compared with males, which could further enhance the mutational effect. In addition, hormonal status could alter DPD activity. However, since no differences were found between DPD levels in premenopausal and postmenopausal women, the influence of hormonal factors could not be confirmed.

Nowadays, mapping of eQTL in liver is an effective strategy to identify novel PGx marker candidates. Together with genome-wide genotype data from 4 human liver datasets, this enabled the identification of 3 novel cis-eQTL in the *DPYD* gene: rs59353118, rs114170368, and rs4294451 [89]. rs59353118 was the most significant cis-eQTL in a haplotype block associated with *DPYD* expression, with the minor allele associated with decreased expression, whereas the minor allele of rs4294451 was associated with increased enzyme expression that influenced both drug disposition and toxicity risk. A Quantitative Expression Trait Locus (eQTL) is a genetic variant that affects gene expression by

altering gene transcription and transcript stability. eQTL studies in human liver may help develop strategies to maximize the efficacy and safety of clinical interventions by determining how genetic variations influence variability in disease risk and therapeutic outcomes [90]. In addition, sex-specific differences in the expression of metabolic enzymes that metabolize could contribute to sex-specific clinical outcomes. Therefore, a systematic understanding of the role of liver eQTLs in sex-specific traits is of great clinical importance [89].

Our study, conducted in 645 patients with CRC treated with FP as monotherapy or in a combination regimen, demonstrated the impact of rs593531118 and rs4294451 on toxicity risk. Accordingly, when the effects of these variants on DPD expression were evaluated, rs593531118 resulted in an increased risk for ADE when haematological and non-haematological toxicities were considered together. In contrast, rs4294451 showed a protective effect with respect to haematological, non-haematological and all toxicities. rs59353118 is located in *DPYD* intron 14, whose minor allele was associated with decreased enzyme expression, confirming its impact on therapy safety with a 1.92-fold increase in risk. Conversely, rs4294451, previously associated with increased enzyme expression, was associated with a halved risk. In addition, patients with the rs4294451 mutation experienced FP-related toxicities not only less frequently but later than wild-type patients. This further information could be particularly useful in patients for whom prolonged treatment is planned, where FP could gradually accumulate in the body.

To date, no other studies have reported the results of these variants evaluated in a retrospective case population, and further investigation is needed. Nevertheless, these results may provide more information to better stratify the population at risk for developing toxicity. If validated, they could be introduced into clinical practice and added to the 4 *DPYD* variants already included in guidelines for further recommendations on dose adjustment.

Although the study was not significant, it showed a differential effect of the presence of the two variants in the two sexes. In the evaluation of both toxicity and acute toxicity, a slightly greater effect of the two variants was observed in the female population. Indeed, in the presence of both the risk variant and the protective variant, men appear to be at higher risk for toxicity than women. The effect of the protective variant rs4294451 in relation to acute haematological toxicity is a novelty to consider. Although the presence of the variant is protective in the entire population, the effect is maintained only in the female population and it appears to lose intensity in the male.

The molecular mechanisms and interactions underlying this difference in behaviour between males and females remain to be investigated. Certainly, the already confirmed role of differential hormone levels, volume of distribution, and fat body component may influence toxicity risk to some extent.

Nevertheless, the genetic component is known to play a fundamental role in FP and it would be interesting to investigate in further studies how it is related to the sex of patients.

6.2 *DPYD* UN-INVESTIGATED VARIANTS AND HAPLOTYPE ANALYSIS IN LARC PATIENTS TO PREDICT nCRT OUTCOME AND TOXICITY RISK

A second study was then performed on a better selected series of homogeneous patients with LARC uniformly treated with FP in monotherapy and concomitant radiotherapy.

A uniform population entrolled in vclinical studies allows a more precise evaluation and avoids the introduction of potential confounders. In this cohort of LARC patients, an analysis was performed using a fluorescence-based allelic discrimination technique for the evaluation of *DPYD* un-investigated variants and the NGS technique for the multi-SNP analysis by identifying haplotypes. *DPYD* variant data were extrapolated and correlated with patient clinical outcome.

Conflicting results have emerged in the literature regarding the effects of certain additional *DPYD* variants, which that are not so clear and require further investigation [99].

In particular, novel un-investigated *DPYD* variants (rs59353118 and rs4294451) characterized only by their effect at the protein transcriptional level have not yet been aassessed for their direct impact on therapeutic efficacy and toxicity profile.

In this study, which considered 212 LARC patients in terms of responce to therapy, data of considerable interest were highlighted.

The presence of the rs59353118 variant showed a positive association with better clinical outcome compared with wild-type patients. These results were consistent with literature data reporting lower expression of the FP metabolising enzyme in the presence of the variant. As a result of the decreased metabolization rate, plasmatic FP levels were higher in mutated patients, which would explain the better clinical outcome. This finding is also consistent with the previous results on rs59353118 evaluated as part of this PhD project in the CRC case study population where the presence of the variant was associated with greater toxicity, which also resulted directly from high plasma concentrations of the drug. In contrast, evaluation of the rs4294451 variant revealed a positive association with a worse prognosis in terms of TRG. Indeed, it has been described that rs4294451 increases DPD enzyme expression and consequently, increases the metabolic rate of the FP. For this reason, FP remains in the patients' plasma circulation only for a limited period of time, decreasing its anticancer effect and, as expected, affecting the responce of patients to therapy. Moreover, these data

confirmed the result previously obtained in the CRC case population where the presence of the variant protected against grade 4 toxicity, again thanks to the increased metabolism of the drug.

NGS analysis and subsequent SNP identification revealed that *DPYD**9A c.85T>C (rs1801265), c.496A>G (rs2297595), and *DPYD**6 c.2194G > A (rs1801160), both exonic SNPs leadind to amino acid changes in DPD protein structure, were present in the case study population.

DPYD c.85T>C (rs1801265) is a novel missense single nucleotide variant (A>G) on chromosome 1. The substitution of A>G results in a substitution of cysteine for arginine in the coding region at position 29 of the protein. It is associated with altered activity of the *DPYD* enzyme and toxicity of 5-FU [239,240]. The correlation between the *DPYD**9A genotype and the clinical phenotype of DPD deficiency is still controversial. In a study conducted by Khushman *et al*. *DPYD**9A variant was the most common found variant. They reported a correlation between *DPYD**9A genotype and DPD deficiency in patients receiving FPs at full dose, as they all experienced grade 3–4 toxicities (diarrhea) [241]. [241]. In contrast, other studies have not been able to confirm these conclusions. Two recent studies suggest a protective effect for the c.85C allele, indicating that DPD activity may be higher in carriers of c.85C [242,243].

DPYD c.496A>G (rs2297595) is an exonic SNP that results in amino acid changes in which a methionine is replaced with a valine at position 166 of the DPD protein. Mutated allele has been associated with decreased enzymatic activity, but another study reported conflicting data of high activity [240,244,245]. Similar discrepancies have been reported for c.496 A>G as for *DPYD**9A: some studies suggest that c.496G is correlates with FP toxicity, whereas others have failed to demonstrate a relationship [238,245,246]. A protective effect for this variant has been suggested by Kleibl Z *et al* in 2009 [247]. Furthermore, *in vitro* studies have also not provided conclusive data on the effect of this variant on DPD function. The mutated allele has been associated with decreased enzymatic activity and 5-FU toxicity in clinical trial, but another study reported opposite data of high activity [240,244,248]. CPIC has classified it as an allele with normal function, as there is no clear evidence of 5- FU toxicity.

*DPYD**6 c.2194 G>A (rs1801160) is another missense single nucleotide variant (C>T) located on chromosome 1 at position 97305364. The substitution of C>T results in an amino acid change from valine to isoleucine at position 732 of the protein. Decreased DPD activity was observed in mutated patients treated with FP [248,249]. In a comprehensive analysis of 1254 patients, it was associated with 5-FU toxicity along with *DPYD**2A and c.2846A>T. A significant association between the non-synonymous variant c.2194G>A (p.V732I, *DPYD**6) with ADRs by FPs was found [250]. Although the c.2194A allele appears to be relatively common, conflicting results have been reported on its

impact on DPD activity and new studies may provide additional information to the debate on this PGx marker.

With the aim of identifying early markers that could refine the selection of patients with good and poor response, the clinical prognostic role of the *DPYD* haplotype composed of the previously reported 3 SNPs, was investigated in the clinical context of LARC patients receiving nCRT.

Of the 6 haplotypes created by combining the alleles of the three SNPs, HAP1 and HAP2 were associated in two separate case study populations in which the effects on treatment efficacy and toxicity were examined.

Regarding the correlation of efficacy with patient genetics, 212 LARC patients were included in the analysis. No significant impact of a HAP was highlighted in our study cohort. To identify potential prognostic markers for pCR in male and female patients, a sex-differentiated analysis was performed. In our study cohort, which consisted of 66% male and 33% female patients, no difference in treatment response was detected using the TRG score. Despite no significant difference was noted, there was a trend in pathological complete and near-complete CR in favour of male patients (OR =16.625, 95 CI:0.9359- 2.9531, p-value = 0.0829). Although this lack of association may be influenced by the limited number of patients, the trend is consistent with evidence that women have a poorer prognosis when evaluating the therapeutic effect of FPs [251].

Although there were no differences in the distribution of haplotypes between the two sexes, there was a slight trend for the effects of HAP1 in male subjects (p-value = 0.07). This difference appears remarkable when heterozygosity with respect to mutated or wild-type patients is taken into account. The present observation seems to be consistent with the genetic model known as the overdominant SNP model. Specifically, an overdominant model assumes that the heterozygote (mM) has the strongest effect and compares the wild type (MM) plus the homozygous mutant (mm) with mM having a different phenotype than any of the homozygotes. Normally, the heterozygotes are intermediate between the homozygotes, but sometimes molecular overdominance is observed in enzyme kinetics. Overdominance could also be described as heterozygote advantage, where heterozygous individuals have higher fitness than homozygous individuals.

Although the phenomenon of overdominance is not a typical finding in genetic association studies, previous cases were reported [252,253].

A particularly interesting example of heterozygote predominance in humans is sickle cell anemia, in which the disease is determined by a single SNP. While homozygotes have either no protection against malaria or a dramatic predisposition to sickle cell anemia, heterozygotes have fewer

physiological effects and partial resistance to malaria [254]. Apart from this typical example, there are only few well-documented cases of overdominance such as warfarin resistance in rats [255]. Recently, within the overdominant genetic model, both novel SNP rs2535764 was found to be related with response to antipsychotic treatment in the Chinese population and SLC29A1 rs3734703 was found to be associated with response to chemotherapy and survival [254,255]. On this basis, the overdominant genetic model was also applied to the effects of HAP1 on TRG in LARC patients. In the male study population, heterozygous patients for HAP1 had a higher risk of having TRG2-5 (OR: 2.40, 95 CI:1.109-5.2 p-value = 0.013). The HAP1 consisting of wild type alleles 2194 G>A (*6), c.85 T>C (*9A) and c.496 A>G could likely increase DPD activity in the heterozygous form, resulting in lower efficacy.

These results are consistent with the findings obtained from the evaluation of the LARC case study population to assess the effects of HAP1 on the toxicity profile FP.

In a total of 189 patients evaluated for the FP safety profile, HAP1 demonstrated a significant protective effect against the risk of grade 3-4 non-haematological toxicity (p-value = 0.04). The OR for the overdominance model did not take into account that grade 3-4 events did not occur in heterozygous patients. Thus, the presence of HAP1 in heterozygosity and the resulting increased enzymatic activity allows for a reduction in side effects of therapy, but also reduced efficacy. Although not statistically significant, the same protective effect of HAP1 was seen in all toxicity classes assessed except haematological toxicity and toxicity defined as “other”, which was not closely related to FP therapy. No sex difference in the effect of HAP1 was observed in this LARC population. The lack of significance could be due to the small sample size of male and female patients selected for analysis or to possible additional molecular mechanisms not yet described. The PK and PD process and the corresponding efficacy and safety profile of FP could be differentially influenced by additional factors, which could explain the observation made in our analysis.

In addition, our results also support a predictive capacity for HAP2 for “haematological”, “non-haematological”, “other”, and “all” toxicity risks. Specifically, HAP2 consists of the wild-type G and T alleles for 2194 G>A (*6) and c.85 T>C (*9A), respectively, and the mutant G allele for variant c.496 A>G. Analysis revealed a significant correlation between HAP2 with "other" and "all" FP-related toxicities in the entire case study population (p-value = 0.005, p-value = 0.04). Although no patients with 2 alleles for HAP2 were identified, the overdominant genetic model was applied. Heterozygous patients had an 8.7-fold higher risk (95 CI:1.45-51.92, p-value = 0.008) and a 4.08-fold higher risk (95 CI:0.96-17.26, p-value = 0.028) of both grade 3-4 "other" and "all" toxicity, respectively. Regarding sex effects, after confirming a higher incidence of toxicity in female patients,

a disaggregated analysis was performed in accordance with literature data. In the female-only study population, HAP2 showed similar results to the overall study cohort. There was a higher risk of grade 3-4 "haematological and non-haematological," "other," and "all" toxicities with 6.5-, 19-, and 17-fold higher rates in heterozygous patients compared with carriers of the 0 allele. The male population showed an opposite trend, although not significantly, suggesting the role of an additional, as yet unestablished, mechanism. In the context of complex phenotypic traits, the role of sex proved to be of critical importance, especially for redefining treatment dosage. These results suggest a possible role of c.496A>G in increased toxicity risk due to lower enzyme activity. There is conflicting information in the literature regarding its predictive significance, but some studies have confirmed an association with a higher risk of FP toxicity [245,246,249,256].

As mentioned earlier, further studies are needed to better characterize these variants and their role in FP clinical outcomes. As the present study had some limitations, mainly related to the moderate sample size, these preliminary results will be further investigated by expanding the population of homogeneous LARC patients thanks to the continuous collection of clinical data performed in the Department of Experimental and Pharmacology at CRO in collaboration with the Radiation Oncology staff. The main goal of the analysis is to focus more on interindividual variability factors that are still poorly understood, such as novel genetic variants with conflicting evidence and the predominant role of sex in therapy with FP.

6.3 TUMOR BIOMARKERS IHC EVALUATION IN LARC PATIENTS TO PREDICT NCRT OUTCOME

Standard treatment of LARC includes preoperative nCRT and surgery followed by adjuvant treatment. Neoadjuvant therapy involves the use of drugs such as 5-FU and capecitabine with concomitant radiotherapy for a treatment period of 5 weeks. With the aim of identifying early biomarkers that could refine the selection of patients with good and poor responce, the clinical prognostic role of ctDNA was investigated in a study cohort of 95 LARC patients who received nCRT. As a complementary approach, a panel of genetics and protein markers in baseline LARC biopsies was examined to identify potential candidates for implementation of currently available risk stratification algorithms.

In this work, we investigated the IHC expression of twelve candidate proteins with relevant biological significance in LARC to identify their potential application as predictive markers for pCR. Then the differential prognostic value of these markers in male and female patients was investigated separately.

Three markers (Ki67, CXCR4, COX-2) successfully distinguished patients who achieved pCR from non-responders in the overall case population.

Ki-67 protein, encoded by the *MKI67* gene on chromosome 10, is a nuclear protein associated with the regulation of gene expression and is necessary for cellular proliferation. Ki-67 is a well-known proliferation marker whose overexpression is generally recognized as a marker of highly malignant phenotypes in various tumor types. It was previously thought to play an important role in cell cycle regulation, maintenance of heterochromatin, and formation of the perichromosomal layer on mitotic chromosomes [257,258]. Ki67 expression reflects tumor proliferation rate and correlates with the development, progression, metastasis, and prognosis of various tumor types, from breast cancer to renal carcinoma, from uterine sarcoma to CRC [259–261]. Accordingly, in our study cohort, patients with a high Ki67 expression level were more likely to get a worse tumor response after nCRT, in a significantly manner when TRG 1 was compared with TRG 2-4 (p-value = 0.059) and showed only a trend when TRG 1-2 were compared with TRG 3-4 (p-value = 0.0720).

Recently, Luo *et al*. reported that Ki-67 may be a predictive biomarker of poor prognosis in CRC patients, as its high expression significantly correlates with poor OS and DFS [262]. Similarly, Kimura *et al*. and Jakob *et al*. have shown that its overexpression at a given time point is an early marker of poor tumor regression [263,264]. Conversely, other studies showed that LARC patients with a higher rate of Ki67-positive cells had a greater incidence of pCR in biopsies [265]. Stratifying by sex, KI-67 was not significantly correlated with treatment efficacy when considering the two types of TRG comparison groups, except for a slight trend in females in TRG 1-2 vs TRG 3-4 (p- value = 0.0602). Therefore, despite the reported poor prognostic value of the marker, its predictive role for nCRT in LARC is still unclear.

The C-X-C chemokine receptor type 4, CXCR4, also known as Fusin or CD184 (Cluster of Differentiation 184), is a protein encoded by the CXCR4 gene located on the long arm of chromosome 2 at position 2q21. The encoded protein is a CXC-chemokine G protein-coupled receptor consisting of 7 transmembrane regions on the cell surface. It is an α-chemokine receptor specific for stromal factor-1 (SDF-1, also called CXCL12), a molecule with potent chemotactic activity for lymphocytes. CXCR4 is expressed by many cell types, including lymphocytes, hematopoietic stem cells, endothelial and epithelial cells, skeletal and even muscle cells, and tumors. Its specific ligand is the chemokine SDF-1α (Stromal cell Derived Factor-1-alpha; also called CXCL12), which is constitutively secreted by bone marrow stromal cells and by many other cell types in various tissues. CXCR4 and CXCL12 binding is involved in the regulation of numerous biological processes, from cell survival to migration. It enables the initiation of various downstream signalling pathways leading

to a plethora of cellular responses such as intracellular calcium increase, activation of gene transcription, chemotaxis, cell survival and proliferation [266]. Regarding the oncological setting, over-expression of CXCR4 was correlated with unfavourable OS in haematological malignancy, breast cancer, CRC, oesophageal cancer, head and neck cancer, renal cancer, lung cancer, gynaecologic cancer, liver cancer, prostate cancer and gallbladder cancer [267]. Despite the prognostic value of CXCR4 expression in several cancers, few data are available for LARC patients [268]. In a study conducted in 68 LARC patients high CXCR4 expression was correlated with shorter relapse free survival (RFS), and cancer specific survival, and it is clearly also associated with increased risk of death and progression in CRC [269,270].

In contrast, our study found that tumor expression of CXCR4 was increased in patients achieving TRG 1 (p-value = 0.010) and was a borderline marker of TRG1-2 (p-value = 0.0617). The efficacy of therapy could be explained by the high proliferation rate of CXCR4-positive cells, making them more sensitive to chemoradiation treatments.

Further investigations are likely needed to elucidate the biological interplay between CXCR4-mediated signalling pathways and tumour response to nCRT and its predictive role. In addition, the predictive role of CXCR4, which is only conserved in the female population in our study, needs to be better explained. Our analysis revealed that the increased expression of CXCR4 was correlated with TRG1 and TRG1-2 compared to TRG 2-4 and TRG 3-4 respectively, only in the female subgroup (p-value = 0.02 in both analysis). Sex-disaggregated data are available for the clinical context of advanced non-small cell lung cancer, in which overexpression of CXCR4 in females was associated with worse outcome, suggesting a sex-dependent survival difference.

Prostaglandin endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase), also known as cyclooxygenase-2 or COX-2, is an enzyme encoded by the *PTGS2* gene on chromosome 1q31.1. In humans, it is one of the two cyclooxygenases involved in the inflammatory process of prostanoid synthesis, particularly the conversion of arachidonic acid to prostaglandin H2 and including prostaglandin E2, an important mediator of inflammation and angiogenesis. The expression of COX-2 is upregulated in many cancers and plays a critical role in promoting angiogenesis, which directly affects cancer progression. It has been found to be associated with cancer cell resistance to conventional chemotherapy and radiotherapy, as well as progressive tumour growth.

Our results support a predictive potential also for higher COX2 expression, which seems to be associated with TRG1 (p-value = 0.03) and TRG 1-2 (p-value = 0.0108). Literature data are conflicting as to their predictive significance in cancer, particularly in LARC. Smith *et al* reported that overexpression of COX-2 may predict poor response of rectal cancers to RCT, which is consistent

with Edden *et al*, who reported lower tumour regression and less likelihood of T-downstaging in patients with COX-2 overexpression [271,272]. In addition, CRC COX-2 overexpression as determined by IHC appears to be slightly worse OS. However, the prognostic value of COX-2 on survival in CRC remains to be clarified by further large-scale prospective studies [273]. In this regard, studies on the COX2 mechanism in modulating the response to LARC therapy are also required.

When we consider males and females separately, a positive association remains only in the female population with respect to TRG1 (p-value = 0.039) and in both sexes when we consider TRG 1 and TRG2 together (male p-value = 0.049; female p-value = 0.04321). Literature data on CRC cancer report that sex is significantly related to COX-2 expression. From the results of Mahmoud *et al* and Negi *et al*, stronger expression in female patients was associated with CRC [274]. Expression of COX-2 was also increased in female patients with vestibular schwannomas [275].

Limited to our female study cohort, CXCL12 expression proved to be a prognostic marker for treatment efficacy, both when considering the of TRG 1 and TRG 2-4 group separately (p-value = 0.024), and TRG 1-2 and TRG 3-4 (p-value = 0.04). Stromal cell-derived factor 1, also know as CXCL12, is a chemokine protein that in humans is encoded by the CXCL12 gene on chromosome 10 at position 10q11.1. It is expressed in many tissues of mice, including brain, thymus, heart, lung, liver, kidney, spleen, and bone marrow. Its signaling pathway includes migration of hematopoietic cells from fetal liver to bone marrow, formation of large blood vessels, expression regulation of CD20 in B cells, and suppression of osteoclast neogenesis. In adults, it primarily plays a role in the recruitment of endothelial progenitor cells during angiogenesis through a CXCR4-dependent mechanism. Recently, its important role in communication between tumor cells and their surrounding microenvironment, particularly in tumor angiogenesis, tumor cell proliferation, and chemoresistance, has attracted considerable interest [276]. In a meta-analysis conducted by Samarendra *et al*, high CXCL12 expression was associated with lower OS in patients with oesophageal, pancreatic, and lung cancers, whereas in breast cancer patients high CXCL12 expression conferred a OS benefit [277]. Consistent with literature data, our results confirmed the clinical value of CXCL12 as a predictor of poor treatment outcome in female patients. Interestingly, the opposite trend was found, albeit not significantly, in the male cohort, where higher expression correlates with a positive response. As suggested by previous studies, the mechanism by which expression of CXCL12 on the surface of the target could promote dissociation of tumor cells from tumor tissue is mediated by activation of various adhesion molecules and by the secretion induction of matrix metalloproteinase (MMP), an enzyme capable of "cutting" proteins that play a fundamental role in cell invasion and enhancement of tissue remodeling, as well as vascular endothelial growth factor (VEGF) [278].

NA repair protein RAD51 homolog 1 is a protein encoded by the *RAD51* gene on chromosome 15. It plays an essential role in homologous strand exchange, a key step in DNA repair by homologous recombination. In addition, it catalyses strand transfer between a broken sequence and its undamaged homolog allowing resynthesis of the damaged region. Overexpression of *RAD51*, which promotes genomic instability has been found in numerous and different cancers including pancreatic, melanoma, breast, non-small cell lung, prostate cancers, and glioblastoma in which a higher level of the protein has been correlated with poorer prognosis [279,280]. Additionally, decreased RAD51 expression has also been observed in certain cancers, particularly sporadic breast cancer, renal cancer, including clear cell and papillary carcinoma [281]. Overexpressed *RAD51* is associated with a higher resistance to DNA-damaging agents, including radiation and cisplatin, so HR inhibition of *RAD51* may provide another mechanism for a therapeutic target for the chemosensitization and radiosensitization of various cancers. This consideration is consistent with our results showing that higher expression in women correlates with poorer response to therapy.

Hypoxia-inducible factor 1 alpha, also known as HIF-1α, is a subunit of the heterodimeric transcription factor HIF-1 encoded by the *HIF1A* gene on chromosome 14 at position 14q23.2. It is involved in the cellular and systemic homeostatic response to hypoxia by activating the transcription of metabolic, angiogenesis, and apoptosis genes. In addition, HIF-1α plays an essential role in embryonic vascularization, tumor angiogenesis, and the pathophysiology of ischemic diseases. Recent studies have shown that HIF-1α is overexpressed in several cancers, including gastric, breast, prostate, and colorectal, ans directly affects carcinogenesis and disease progression [282]. Moreover, overexpression of HIF-1α had an important impact on the biological effects of tumor cells, recruitment of infiltrated lymphocytes, and angiogenesis in the tumor microenvironment [283,284]. A weak HIF-1α expression was correlated with poor prognosis in resectable pancreatic ductal adenocarcinoma and in CRC [285,286]. From a more recent analysis, upregulation of HIF1A and resulting increased expression correlates with enhanced tumor immune and stromal signatures and aggressive phenotypes in human cancers that respond to more active immunotherapy [287]. To date, few studies have reported conflicting results on the predictive role of HIF1α in LARC biopsies prior to treatment, but in our cohort, and particularly in female subjects, its higher expression showed a slight tendency toward worse TRG.

Regarding VEGF expression and its borderline correlation with TRG 1-2, a trend emerged between a higher H score and TRG 3-4, particularly limited to men. VEGF-A is a heparin-binding protein, a molecule capable of promoting the formation of new vessels. This phenomenon is also known as neo-angiogenesis. It is encoded by the *VEGFA* gene, a member of the *PDGF/VEGF* growth factor family

on chromosome 6. VEGF is expressed in various tissues such as the brain, liver, kidney, and spleen, as well as in the ovaries and uterus. Following hypoxia, several factors are released during the wound healing process under physiological or pathological conditions that are able to mediate important cellular signals for the formation of the walls of new blood vessels. VEGF causes an increase in vascular permeability and in the release of proteases. High VEGF expression has been found in diseases that exhibit increased angiogenesis, such as cancer, atherosclerosis, hemangiomas, skin and mucosal diseases, retinopathy, liver and kidney diseases, and inflammatory diseases. Indeed, neovascularisation is a strategy used by solid tumours to facilitate mass growth and metastasis by ensuring nutrient flow. VEGF expression has been associated with tumour progression, advanced stage, unfavourable survival, and increase in invasion rate and distant metastasis in various solid and non-solid malignancies such as head and neck cancer, non-small cell lung cancer, cervical, gastric, pancreatic, and CRC [288]. Controversy also existed regarding the predictive role of VEGF. VEGF expression correlated significantly with CR in locally advanced mid-to-low rectal cancer [289]. Accordingly, Zlobec *et al* found a significant association between rectal cancer radiation response and VEGF expression, whereas Yu *et al* reported that VEGF mRNA expression was not a predictive biomarker for CR [289,290].

Specifically, a high interesting point no weel investigated but which merit greater attention could be the evaluation of interconnection between the proteins included in the study, i.e. Hif-1a and VEGF. Particularly those two proteins might be used as biomarkers as important results emerged in literature indicating their role in tumor infiltration and poor prognosis in human CRC. In fact, hypoxia is a key signal for the induction of angiogenesis, and one of the key angiogenic factors regulated by hypoxia is VEGF. *VEGF* gene expression was under the control of HIF-1α by binding to the hypoxia-response element in the *VEGF* promoter region. The consideration of the interplaty of the highlighted factors and of additional gene not included in the study could better elucidate the effective role in predicting the clinical outcome of FP treatment.

6.4 CONCLUSIONS AND OUTLOOKS

The aim of this work was to investigate the predictive role of PGx markers, patient gender and their interaction on the outcome of treatment with FP in relation to the safety and efficacy of treatment in CRC. FP are drugs widely used in clinical practice in oncology, and their toxicity and efficacy profiles vary considerably. Literature data suggest that patient genetic component, tumor, and sex are influential factors.

Specifically, this work focused on evaluating i) the predictive role of novel, previously unstudied *DPYD* variants and the role of patient sex on the risk of FP -related toxicity in CRC patients; (ii) the impact of previously unstudied *DPYD* variants and haplotypes and their interaction with patients sex in a subset of clinically homogeneous LARC patients; and (iii) novel IHC biomarkers in tumor biopsy in LARC patients and their sex-specific impact as prognostic predictive factors for treatment efficacy.

The ultimate goal is to identify new markers that can be used to stratify cancer patients for personalized therapy to prevent the occurrence of AEs and treatment failure. Since the occurrence of AEs is often associated with interruption or delay of the patient's treatment, their prevention is synonymous not only with an improvement in the safety profile of the drug, but also with better efficacy and treatment adherence.

The main findings of the present work suggest that the detection of novel eQTLs of intronic *DPYD* variants and exonic haplotypes of the key enzyme in the metabolism of FP may be a useful tool to identify patients at risk of developing toxicity and with variable response to therapy. Specifically, in the study population described here, genetic analysis of novel *DPYD* genetic biomarkers and patient tumor protein expression prior to treatment has demonstrated the ability to be highlighted not only in the overall population but also with differential prognostic value in males and females separately. The reduction in serious toxicities observed in the past by using the PGx approach with the 4 validated *DPYD* variants could be improved by the possible introduction of new biomarkers, and the evaluation of the tumor protein profile could provide clinicians with an important decision support for the most appropriate therapeutic strategy.

Further studies are needed to clarify the prognostic value of these preliminary results, and implementation of the case population is necessary for their validation. These results shed light on a possible application of the haplotype strategy instead of the single mutation approach. These PhD work results could be also proposed to modify clinical trial design where patients' genetic component and sex are not fully integrated in the stratification analysis. With the aim of better understand the specific therapeutic effect and, most importantly, the real side effect of the proposed treatment both sex and genetics aspect deserve more attention in the future. Moreover, these results lead us to reconsider the simple association between patient genotype and sex and to highlight the importance of sex as a new key variable that should be considered and implemented in routine clinical practice.

7 REFERENCES

1. Heidelberger C, Chaudhuri NK, Danneberg P, *et al.* Fluorinated pyrimidines, a new class of tumour-inhibitory compounds. *Nature*. 179(4561), 663–666 (1957).

2. Tiwari M. Antimetabolites: established cancer therapy. *J Cancer Res Ther*. 8(4), 510–519 (2012).

3. Vardanyan R, Hruby V. Chapter 28 - Antineoplastic Agents [Internet]. In: *Synthesis of Best-Seller Drugs*. Vardanyan R, Hruby V (Eds.), Academic Press, Boston, 495–547 (2016) [cited 2022 Jan 20]. Available from: https://www.sciencedirect.com/science/article/pii/B9780124114920000286.

4. Longley DB, Harkin DP, Johnston PG. 5-fluorouracil: mechanisms of action and clinical strategies. *Nat Rev Cancer*. 3(5), 330–338 (2003).

5. Hwang JJ, Marshall JL. Capecitabine: fulfilling the promise of oral chemotherapy. *Expert Opin Pharmacother*. 3(6), 733–743 (2002).

6. Brito RA, Medgyesy D, Zukowski TH, *et al.* Fluoropyrimidines: a critical evaluation. *Oncology*. 57 Suppl 1, 2–8 (1999).

7. Zalcberg J, Kerr D, Seymour L, Palmer M. Haematological and non-haematological toxicity after 5-fluorouracil and leucovorin in patients with advanced colorectal cancer is significantly associated with gender, increasing age and cycle number. Tomudex International Study Group. *Eur J Cancer*. 34(12), 1871–1875 (1998).

8. Comella P, Casaretti R, Sandomenico C, Avallone A, Franco L. Role of oxaliplatin in the treatment of colorectal cancer. *Ther Clin Risk Manag*. 5, 229–238 (2009).

9. Cavanna L, Artioli F, Codignola C, *et al.* Oxaliplatin in combination with 5-fluorouracil (5-FU) and leucovorin (LV) in patients with metastatic gastric cancer (MGC). *Am J Clin Oncol*. 29(4), 371–375 (2006).

10. Bar-Ad V, Palmer J, Yang H, *et al.* Current management of locally advanced head and neck cancer: the combination of chemotherapy with locoregional treatments. *Semin Oncol*. 41(6), 798–806 (2014).

11. Fernández-Martos C, Nogué M, Cejas P, Moreno-García V, Machancoses AH, Feliu J. The role of capecitabine in locally advanced rectal cancer treatment: an update. *Drugs*. 72(8), 1057–1073 (2012).

12. Argiris A, Kotsakis AP, Hoang T, *et al.* Cetuximab and bevacizumab: preclinical data and phase II trial in recurrent or metastatic squamous cell carcinoma of the head and neck. *Ann Oncol*. 24(1), 220–225 (2013).

13. Wong NS, Fernando NH, Nixon AB, *et al.* A phase II study of capecitabine, oxaliplatin, bevacizumab and cetuximab in the treatment of metastatic colorectal cancer. *Anticancer Res*. 31(1), 255–261 (2011).

14. Kim MO, Hong ES, Chai JY, *et al.* Concurrent FP (5-fluorouracil, cisplatin) Chemoradiotherapy for Patients with Esophageal Cancer. *Cancer Res Treat*. 35(4), 330–334 (2003).

15. Miwa M, Ura M, Nishida 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*. 34(8), 1274–1281 (1998).

16. Budman DR, Meropol NJ, Reigner B, *et al.* Preliminary studies of a novel oral fluoropyrimidine carbamate: capecitabine. *J Clin Oncol*. 16(5), 1795–1802 (1998).

17. Sommer H, Santi DV. Purification and amino acid analysis of an active site peptide from thymidylate synthetase containing covalently bound 5-fluoro-2'-deoxyuridylate and methylenetetrahydrofolate. *Biochem Biophys Res Commun*. 57(3), 689–695 (1974).

18. Fang F, Hoskins J, Butler JS. 5-Fluorouracil Enhances Exosome-Dependent Accumulation of Polyadenylated rRNAs. *Mol Cell Biol*. 24(24), 10766–10776 (2004).

19. Blaschke M, Blumberg J, Wegner U, Nischwitz M, Ramadori G, Cameron S. Measurements of 5-FU Plasma Concentrations in Patients with Gastrointestinal Cancer: 5-FU Levels Reflect the 5-FU Dose Applied. *Journal of Cancer Therapy*. 3(1), 28–36 (2012).

20. Gmeiner WH. Novel chemical strategies for thymidylate synthase inhibition. *Curr Med Chem*. 12(2), 191–202 (2005).

21. Kunz C, Focke F, Saito Y, *et al.* Base excision by thymine DNA glycosylase mediates DNA-directed cytotoxicity of 5-fluorouracil. *PLoS Biol*. 7(4), e91 (2009).

22. Thorn CF, Marsh S, Carrillo MW, McLeod HL, Klein TE, Altman RB. PharmGKB summary: fluoropyrimidine pathways. *Pharmacogenet Genomics*. 21(4), 237–242 (2011).

23. Heggie GD, Sommadossi JP, Cross DS, Huster WJ, Diasio RB. Clinical pharmacokinetics of 5-fluorouracil and its metabolites in plasma, urine, and bile. *Cancer Res*. 47(8), 2203–2206 (1987).

24. Bronckaers A, Gago F, Balzarini J, Liekens S. The dual role of thymidine phosphorylase in cancer development and chemotherapy. *Med Res Rev*. 29(6), 903–953 (2009).

25. Amstutz U, Farese S, Aebi S, Largiadèr CR. Dihydropyrimidine dehydrogenase gene variation and severe 5-fluorouracil toxicity: a haplotype assessment. *Pharmacogenomics*. 10(6), 931–944 (2009).

26. Hamzic S, Aebi S, Joerger M, *et al.* Fluoropyrimidine chemotherapy: recommendations for DPYD genotyping and therapeutic drug monitoring of the Swiss Group of Pharmacogenomics and Personalised Therapy. *Swiss Med Wkly*. 150, w20375 (2020).

27. Macdonald JS. Toxicity of 5-fluorouracil. *Oncology (Williston Park)*. 13(7 Suppl 3), 33–34 (1999).

28. Hansen RM, Ryan L, Anderson T, *et al.* Phase III study of bolus versus infusion fluorouracil with or without cisplatin in advanced colorectal cancer. *J Natl Cancer Inst*. 88(10), 668–674 (1996).

29. Iurlo A, Fornier M, Caldiera S, Bertoni F, Foa P. Palmar-Plantar Erythrodysaesthesia Syndrome Due to 5-Fluorouracil Therapy–An Underestimated Toxic Event? *Acta Oncologica*. 36(6), 653–654 (1997).

30. Stein BN, Petrelli NJ, Douglass HO, Driscoll DL, Arcangeli G, Meropol NJ. Age and sex are independent predictors of 5-fluorouracil toxicity. Analysis of a large scale phase III trial. *Cancer*. 75(1), 11–17 (1995).

31. Lichtman SM, Skirvin JA, Vemulapalli S. Pharmacology of antineoplastic agents in older cancer patients. *Crit Rev Oncol Hematol*. 46(2), 101–114 (2003).

32. Scripture CD, Figg WD. Drug interactions in cancer therapy. *Nat Rev Cancer*. 6(7), 546–558 (2006).

33. Wilkinson GR. Drug metabolism and variability among patients in drug response. *N Engl J Med*. 352(21), 2211–2221 (2005).

34. Is This the Drug or Dose for You?: Impact and Consideration of Ethnic Factors in Global Drug Development, Regulatory Review, and Clinical Practice - Huang - 2008 - Clinical Pharmacology & Therapeutics - Wiley Online Library [Internet]. . Available from: https://ascpt.onlinelibrary.wiley.com/doi/full/10.1038/clpt.2008.144.

35. Yang L, Price ET, Chang C-W, *et al.* Gene Expression Variability in Human Hepatic Drug Metabolizing Enzymes and Transporters. *PLoS One*. 8(4), e60368 (2013).

36. Dhingra K. Oncology 2020: a drug development and approval paradigm. *Ann Oncol*. 26(11), 2347–2350 (2015).

37. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J Health Econ*. 47, 20–33 (2016).

38. Patino CM, Ferreira JC. Inclusion and exclusion criteria in research studies: definitions and why they matter. *J Bras Pneumol*. 44(2), 84 (2018).

39. Fisher JA, Kalbaugh CA. Challenging Assumptions About Minority Participation in US Clinical Research. *Am J Public Health*. 101(12), 2217–2222 (2011).

40. Mathur S, Sutton J. Personalized medicine could transform healthcare. *Biomed Rep*. 7(1), 3–5 (2017).

41. van der Wouden CH, Cambon-Thomsen A, Cecchin E, *et al.* Implementing Pharmacogenomics in Europe: Design and Implementation Strategy of the Ubiquitous Pharmacogenomics Consortium. *Clin Pharmacol Ther*. 101(3), 341–358 (2017).

42. Burke W, Brown Trinidad S, Press NA. Essential elements of personalized medicine. *Urol Oncol*. 32(2), 193–197 (2014).

43. Pirmohamed M. Pharmacogenetics and pharmacogenomics. *Br J Clin Pharmacol*. 52(4), 345–347 (2001).

44. Katara P, Yadav A. Pharmacogenes (PGx-genes): Current understanding and future directions. *Gene*. 718, 144050 (2019).

45. Swen JJ, Huizinga TW, Gelderblom H, *et al.* Translating pharmacogenomics: challenges on the road to the clinic. *PLoS Med*. 4(8), e209 (2007).

46. Pinto N, Dolan ME. Clinically Relevant Genetic Variations in Drug Metabolizing Enzymes. *Curr Drug Metab*. 12(5), 487–497 (2011).

47. Uetrecht J. Idiosyncratic drug reactions: current understanding. *Annu Rev Pharmacol Toxicol*. 47, 513–539 (2007).

48. Van Driest SL, Shi Y, Bowton EA, *et al.* Clinically actionable genotypes among 10,000 patients with preemptive pharmacogenomic testing. *Clin Pharmacol Ther*. 95(4), 423–431 (2014).

49. Caudle KE, Dunnenberger HM, Freimuth RR, *et al.* Standardizing terms for clinical pharmacogenetic test results: consensus terms from the Clinical Pharmacogenetics Implementation Consortium (CPIC). *Genet Med*. 19(2), 215–223 (2017).

50. Relling MV, Evans WE. Pharmacogenomics in the clinic. *Nature*. 526(7573), 343–350 (2015).

51. Budnitz DS, Pollock DA, Weidenbach KN, Mendelsohn AB, Schroeder TJ, Annest JL. National surveillance of emergency department visits for outpatient adverse drug events. *JAMA*. 296(15), 1858–1866 (2006).

52. Caraballo PJ, Bielinski SJ, St Sauver JL, Weinshilboum RM. Electronic Medical Record-Integrated Pharmacogenomics and Related Clinical Decision Support Concepts. *Clin Pharmacol Ther*. 102(2), 254–264 (2017).

53. Lazaridis KN. Improving Therapeutic Odyssey: Preemptive Pharmacogenomics Utility in Patient Care. *Clin Pharmacol Ther*. 101(1), 39–41 (2017).

54. Dunnenberger HM, Crews KR, Hoffman JM, *et al.* Preemptive Clinical Pharmacogenetics Implementation: Current programs in five United States medical centers. *Annu Rev Pharmacol Toxicol*. 55, 89–106 (2015).

55. Clinical Pharmacogenetics Implementation Consortium [Internet]. . Available from: https://cpicpgx.org/.

56. Abdullah-Koolmees H, van Keulen AM, Nijenhuis M, Deneer VHM. Pharmacogenetics Guidelines: Overview and Comparison of the DPWG, CPIC, CPNDS, and RNPGx Guidelines. *Front Pharmacol*. 11, 595219 (2021).

57. Swen JJ, Nijenhuis M, van Rhenen M, *et al.* Pharmacogenetic Information in Clinical Guidelines: The European Perspective. *Clin Pharmacol Ther*. 103(5), 795–801 (2018).

58. Swen JJ, Wilting I, de Goede AL, *et al.* Pharmacogenetics: from bench to byte. *Clin Pharmacol Ther*. 83(5), 781–787 (2008).

59. Caudle KE, Klein TE, Hoffman JM, *et al.* Incorporation of Pharmacogenomics into Routine Clinical Practice: the Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline Development Process. *Curr Drug Metab*. 15(2), 209–217 (2014).

60. PharmGKB [Internet]. PharmGKB. Available from: https://www.pharmgkb.org/.

61. Latchman J, Guastella A, Tofthagen C. 5-Fluorouracil toxicity and dihydropyrimidine dehydrogenase enzyme: implications for practice. *Clin J Oncol Nurs*. 18(5), 581–585 (2014).

62. G M, Mc E. Potential importance of dihydropyrimidine dehydrogenase (DPD) in cancer chemotherapy. *Pharmacogenetics* [Internet]. 4(6) (1994). Available from: https://pubmed.ncbi.nlm.nih.gov/7704035/.

63. Amstutz U, Froehlich TK, Largiadèr CR. Dihydropyrimidine dehydrogenase gene as a major predictor of severe 5-fluorouracil toxicity. *Pharmacogenomics*. 12(9), 1321–1336 (2011).

64. Borràs E, Dotor E, Arcusa À, *et al.* High-resolution melting analysis of the common c.1905+1G>A mutation causing dihydropyrimidine dehydrogenase deficiency and lethal 5-fluorouracil toxicity. *Front Genet*. 3, 312 (2013).

65. Morrison GB, Bastian A, Dela Rosa T, Diasio RB, Takimoto CH. Dihydropyrimidine dehydrogenase deficiency: a pharmacogenetic defect causing severe adverse reactions to 5-fluorouracil-based chemotherapy. *Oncol Nurs Forum*. 24(1), 83–88 (1997).

66. Yen JL, McLeod HL. Should DPD analysis be required prior to prescribing fluoropyrimidines? *Eur J Cancer*. 43(6), 1011–1016 (2007).

67. Fleger M, Willomitzer J, Meinsma R, *et al.* Dihydropyrimidine Dehydrogenase Deficiency: Metabolic Disease or Biochemical Phenotype? *JIMD Rep*. 37, 49–54 (2017).

68. van Gennip AH, Abeling NG, Stroomer AE, van Lenthe H, Bakker HD. Clinical and biochemical findings in six patients with pyrimidine degradation defects. *J Inherit Metab Dis*. 17(1), 130–132 (1994).

69. Enns GM, Barkovich AJ, van Kuilenburg ABP, *et al.* Head imaging abnormalities in dihydropyrimidine dehydrogenase deficiency. *J Inherit Metab Dis*. 27(4), 513–522 (2004).

70. Van Kuilenburg AB, Vreken P, Abeling NG, *et al.* Genotype and phenotype in patients with dihydropyrimidine dehydrogenase deficiency. *Hum Genet*. 104(1), 1–9 (1999).

71. Abolmaali K, Balakrishnan A, Stearns AT, *et al.* Circadian variation in intestinal dihydropyrimidine dehydrogenase (DPD) expression: a potential mechanism for benefits of 5FU chrono-chemotherapy. *Surgery*. 146(2), 269–273 (2009).

72. Ezzeldin H, Diasio R. Dihydropyrimidine dehydrogenase deficiency, a pharmacogenetic syndrome associated with potentially life-threatening toxicity following 5-fluorouracil administration. *Clin Colorectal Cancer*. 4(3), 181–189 (2004).

73. Saif MW, Syrigos K, Mehra R, Mattison LK, Diasio RB. DIHYDROPYRIMIDINE DEHYDROGENASE DEFICIENCY (DPD) IN GI MALIGNANCIES: EXPERIENCE OF 4-YEARS. *Pak J Med Sci Q*. 23(6), 832–839 (2007).

74. van Kuilenburg ABP. Dihydropyrimidine dehydrogenase and the efficacy and toxicity of 5-fluorouracil. *Eur J Cancer*. 40(7), 939–950 (2004).

75. Dean L, Kane M. Fluorouracil Therapy and DPYD Genotype [Internet]. In: *Medical Genetics Summaries*. Pratt VM, Scott SA, Pirmohamed M, et al. (Eds.), National Center for Biotechnology Information (US), Bethesda (MD) (2012) [cited 2022 Jan 20]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK395610/.

76. Henricks LM, Lunenburg CATC, Meulendijks D, *et al.* Translating DPYD genotype into DPD phenotype: using the DPYD gene activity score. *Pharmacogenomics*. 16(11), 1277–1286 (2015).

77. Lunenburg CATC, van der Wouden CH, Nijenhuis M, *et al.* Dutch Pharmacogenetics Working Group (DPWG) guideline for the gene-drug interaction of DPYD and fluoropyrimidines. *Eur J Hum Genet*. 28(4), 508–517 (2020).

78. Amstutz U, Henricks LM, Offer SM, *et al.* Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Dihydropyrimidine Dehydrogenase Genotype and Fluoropyrimidine Dosing: 2017 Update. *Clin Pharmacol Ther*. 103(2), 210–216 (2018).

79. Vreken P, Van Kuilenburg AB, Meinsma R, van Gennip AH. Identification of novel point mutations in the dihydropyrimidine dehydrogenase gene. *J Inherit Metab Dis*. 20(3), 335–338 (1997).

80. Raida M, Schwabe W, Häusler P, *et al.* Prevalence of a common point mutation in the dihydropyrimidine dehydrogenase (DPD) gene within the 5'-splice donor site of intron 14 in patients with severe 5-fluorouracil (5-FU)- related toxicity compared with controls. *Clin Cancer Res*. 7(9), 2832–2839 (2001).

81. Van Kuilenburg AB, Vreken P, Beex LV, *et al.* Heterozygosity for a point mutation in an invariant splice donor site of dihydropyrimidine dehydrogenase and severe 5-fluorouracil related toxicity. *Eur J Cancer*. 33(13), 2258–2264 (1997).

82. van Kuilenburg ABP, Dobritzsch D, Meinsma R, *et al.* Novel disease-causing mutations in the dihydropyrimidine dehydrogenase gene interpreted by analysis of the three-dimensional protein structure. *Biochem J*. 364(Pt 1), 157–163 (2002).

83. Collie-Duguid ES, Etienne MC, Milano G, McLeod HL. Known variant DPYD alleles do not explain DPD deficiency in cancer patients. *Pharmacogenetics*. 10(3), 217–223 (2000).

84. Santini J, Milano G, Thyss A, *et al.* 5-FU therapeutic monitoring with dose adjustment leads to an improved therapeutic index in head and neck cancer. *Br J Cancer*. 59(2), 287–290 (1989).

85. Reizine N, Vokes EE, Liu P, *et al.* Implementation of pharmacogenomic testing in oncology care (PhOCus): study protocol of a pragmatic, randomized clinical trial. *Ther Adv Med Oncol*. 12, 1758835920974118 (2020).

86. Deenen MJ, Meulendijks D, Cats A, *et al.* Upfront Genotyping of DPYD*2A to Individualize Fluoropyrimidine Therapy: A Safety and Cost Analysis. *J Clin Oncol*. 34(3), 227–234 (2016).

87. EMA website. EMA recommendations on DPD testing prior to treatment with fluorouracil, capecitabine, tegafur and flucytosine [Internet]. European Medicines Agency (2020). Available from: https://www.ema.europa.eu/en/news/ema-recommendations-dpd-testing-prior-treatment-fluorouracil-capecitabine-tegafur-flucytosine.

88. DIMITROVA EK. EMA recommendations on DPD testing prior to treatment with fluorouracil, capecitabine, tegafur and flucytosine [Internet]. European Medicines Agency (2020). Available from: https://www.ema.europa.eu/en/news/ema-recommendations-dpd-testing-prior-treatment-fluorouracil-capecitabine-tegafur-flucytosine.

89. Etheridge AS, Gallins PJ, Jima D, *et al.* A New Liver Expression Quantitative Trait Locus Map From 1,183 Individuals Provides Evidence for Novel Expression Quantitative Trait Loci of Drug

Response, Metabolic, and Sex-Biased Phenotypes. *Clin Pharmacol Ther*. 107(6), 1383–1393 (2020).

90. Wright FA, Sullivan PF, Brooks AI, *et al.* Heritability and Genomics of Gene Expression in Peripheral Blood. *Nat Genet*. 46(5), 430–437 (2014).

91. rs59353118 RefSNP Report - dbSNP - NCBI [Internet]. . Available from: https://www.ncbi.nlm.nih.gov/snp/rs59353118?horizontal_tab=true.

92. Rosmarin D, Palles C, Pagnamenta A, *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(1), 111–120 (2015).

93. rs114170368 RefSNP Report - dbSNP - NCBI [Internet]. . Available from: https://www.ncbi.nlm.nih.gov/snp/rs114170368?horizontal_tab=true.

94. rs4294451 RefSNP Report - dbSNP - NCBI [Internet]. . Available from: https://www.ncbi.nlm.nih.gov/snp/rs4294451?horizontal_tab=true.

95. Gianfrancesco O, Warburton A, Collier DA, Bubb VJ, Quinn JP. Novel brain expressed RNA identified at the MIR137 schizophrenia-associated locus. *Schizophr Res*. 184, 109–115 (2017).

96. Johnson GC, Esposito L, Barratt BJ, *et al.* Haplotype tagging for the identification of common disease genes. *Nat Genet*. 29(2), 233–237 (2001).

97. Crawford DC, Nickerson DA. Definition and clinical importance of haplotypes. *Annu Rev Med*. 56, 303–320 (2005).

98. International HapMap Consortium. The International HapMap Project. *Nature*. 426(6968), 789–796 (2003).

99. Hamzic S, Schärer D, Offer SM, *et al.* Haplotype structure defines effects of common DPYD variants c.85T > C (rs1801265) and c.496A > G (rs2297595) on dihydropyrimidine dehydrogenase activity: Implication for 5-fluorouracil toxicity. *Br J Clin Pharmacol*. 87(8), 3234–3243 (2021).

100. Gentile G, Botticelli A, Lionetto L, *et al.* Genotype-phenotype correlations in 5-fluorouracil metabolism: a candidate DPYD haplotype to improve toxicity prediction. *Pharmacogenomics J*. 16(4), 320–325 (2016).

101. Mauvais-Jarvis F, Bairey Merz N, Barnes PJ, *et al.* Sex and gender: modifiers of health, disease, and medicine. *Lancet*. 396(10250), 565–582 (2020).

102. Baggio G, Corsini A, Floreani A, Giannini S, Zagonel V. Gender medicine: a task for the third millennium. *Clin Chem Lab Med*. 51(4), 713–727 (2013).

103. Wagner AD, Oertelt-Prigione S, Adjei A, *et al.* Gender medicine and oncology: report and consensus of an ESMO workshop. *Ann Oncol*. 30(12), 1914–1924 (2019).

104. Zucker I, Prendergast BJ. Sex differences in pharmacokinetics predict adverse drug reactions in women. *Biol Sex Differ*. 11, 32 (2020).

105. Dong M, Cioffi G, Wang J, *et al.* Sex Differences in Cancer Incidence and Survival: A Pan-Cancer Analysis. *Cancer Epidemiol Biomarkers Prev*. 29(7), 1389–1397 (2020).

106. Soldin OP, Chung SH, Mattison DR. Sex Differences in Drug Disposition. *J Biomed Biotechnol*. 2011, 187103 (2011).

107. Soldin O, Mattison D. Sex Differences in Pharmacokinetics and Pharmacodynamics. *Clin Pharmacokinet*. 48(3), 143–157 (2009).

108. Mueller S, Saunier K, Hanisch C, *et al.* Differences in fecal microbiota in different European study populations in relation to age, gender, and country: a cross-sectional study. *Appl Environ Microbiol*. 72(2), 1027–1033 (2006).

109. Wen L, Ley RE, Volchkov PV, *et al.* Innate immunity and intestinal microbiota in the development of Type 1 diabetes. *Nature*. 455(7216), 1109–1113 (2008).

110. Yang L, Li Y, Hong H, *et al.* Sex Differences in the Expression of Drug-Metabolizing and Transporter Genes in Human Liver. *J Drug Metab Toxicol*. 3(3), 1000119 (2012).

111. Franconi F, Campesi I. Pharmacogenomics, pharmacokinetics and pharmacodynamics: interaction with biological differences between men and women. *Br J Pharmacol*. 171(3), 580–594 (2014).

112. Libert C, Dejager L, Pinheiro I. The X chromosome in immune functions: when a chromosome makes the difference. *Nat Rev Immunol*. 10(8), 594–604 (2010).

113. Carè A, Bellenghi M, Matarrese P, Gabriele L, Salvioli S, Malorni W. Sex disparity in cancer: roles of microRNAs and related functional players. *Cell Death Differ*. 25(3), 477–485 (2018).

114. Clocchiatti A, Cora E, Zhang Y, Dotto GP. Sexual dimorphism in cancer. *Nat Rev Cancer*. 16(5), 330–339 (2016).

115. Myburgh R, Hochfeld WE, Dodgen TM, Ker J, Pepper MS. Cardiovascular pharmacogenetics. *Pharmacol Ther*. 133(3), 280–290 (2012).

116. Blokland GAM, Grove J, Chen C-Y, *et al.* Sex-Dependent Shared and Nonshared Genetic Architecture Across Mood and Psychotic Disorders. *Biol Psychiatry*. 91(1), 102–117 (2022).

117. Rademaker M. Do women have more adverse drug reactions? *Am J Clin Dermatol*. 2(6), 349–351 (2001).

118. Schmetzer O, Flörcken A. Sex differences in the drug therapy for oncologic diseases. *Handb Exp Pharmacol*. (214), 411–442 (2012).

119. Tran C, Knowles SR, Liu BA, Shear NH. Gender differences in adverse drug reactions. *J Clin Pharmacol*. 38(11), 1003–1009 (1998).

120. Franconi F. Paradigmi della ricerca di genere. *Italian Journal of Gender-Specific Medicine*. 1(1), 6–8 (2015).

121. Franconi F, Campesi I, Colombo D, Antonini P. Sex-Gender Variable: Methodological Recommendations for Increasing Scientific Value of Clinical Studies. *Cells*. 8(5), E476 (2019).

122. U.S. Food and Drug Administration. Guideline for the study and evaluation of gender differences in the clinical evaluation of drugs; notice. *Fed Regist*. 58(139), 39406–39416 (1993).

123. Stein BN, Petrelli NJ, Douglass HO, Driscoll DL, Arcangeli G, Meropol NJ. Age and sex are independent predictors of 5-fluorouracil toxicity. Analysis of a large scale phase III trial. *Cancer*. 75(1), 11–17 (1995).

124. Sloan JA, Loprinzi CL, Novotny PJ, Okuno S, Nair S, Barton DL. Sex differences in fluorouracil-induced stomatitis. *J Clin Oncol*. 18(2), 412–420 (2000).

125. Wang J, Huang Y. Pharmacogenomics of sex difference in chemotherapeutic toxicity. *Curr Drug Discov Technol*. 4(1), 59–68 (2007).

126. Mader RM. The potential of gender-specific tumor pharmacology. *Pharmacogenomics*. 8(3), 271–274 (2007).

127. Wagner AD, Grothey A, Andre T, *et al.* Sex and Adverse Events of Adjuvant Chemotherapy in Colon Cancer: An Analysis of 34 640 Patients in the ACCENT Database. *J Natl Cancer Inst*. 113(4), 400–407 (2020).

128. Milano G, Etienne MC, Cassuto-Viguier E, *et al.* Influence of sex and age on fluorouracil clearance. *J Clin Oncol*. 10(7), 1171–1175 (1992).

129. Etienne MC, Lagrange JL, Dassonville O, *et al.* Population study of dihydropyrimidine dehydrogenase in cancer patients. *J Clin Oncol*. 12(11), 2248–2253 (1994).

130. Milano G, Etienne MC, Pierrefite V, Barberi-Heyob M, Deporte-Fety R, Renée N. Dihydropyrimidine dehydrogenase deficiency and fluorouracil-related toxicity. *Br J Cancer*. 79(3–4), 627–630 (1999).

131. Schwab M, Zanger UM, Marx C, *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*. 26(13), 2131–2138 (2008).

132. Lee AM, Shi Q, Pavey E, *et al.* DPYD variants as predictors of 5-fluorouracil toxicity in adjuvant colon cancer treatment (NCCTG N0147). *J Natl Cancer Inst*. 106(12), dju298 (2014).

133. Loganayagam A, Arenas-Hernandez M, Fairbanks L, Ross P, Sanderson JD, Marinaki AM. The contribution of deleterious DPYD gene sequence variants to fluoropyrimidine toxicity in British cancer patients. *Cancer Chemother Pharmacol*. 65(2), 403–406 (2010).

134. Yamashita K, Mikami Y, Ikeda M, *et al.* Gender differences in the dihydropyrimidine dehydrogenase expression of colorectal cancers. *Cancer Lett*. 188(1–2), 231–236 (2002).

135. Elsaleh H, Joseph D, Grieu F, Zeps N, Spry N, Iacopetta B. Association of tumour site and sex with survival benefit from adjuvant chemotherapy in colorectal cancer. *Lancet*. 355(9217), 1745–1750 (2000).

136. Shah SU, Ahmed T, Badar A, Shafique M, Malik S, Aaqil B. Efficacy of 5-Fluorouracil in the Treatment of Pterygium. *Cureus*. 13(1), e12652.

137. Lech G, Słotwiński R, Słodkowski M, Krasnodębski IW. Colorectal cancer tumour markers and biomarkers: Recent therapeutic advances. *World J Gastroenterol*. 22(5), 1745–1755 (2016).

138. Adlard JW, Richman SD, Seymour MT, Quirke P. Prediction of the response of colorectal cancer to systemic therapy. *Lancet Oncol*. 3(2), 75–82 (2002).

139. Duffy MJ, van Dalen A, Haglund C, *et al.* Clinical utility of biochemical markers in colorectal cancer: European Group on Tumour Markers (EGTM) guidelines. *Eur J Cancer*. 39(6), 718–727 (2003).

140. Pasche B, Mulcahy M, Benson AB. Molecular markers in prognosis of colorectal cancer and prediction of response to treatment. *Best Practice & Research Clinical Gastroenterology*. 16(2), 331–345 (2002).

141. Fernebro E, Bendahl P-O, Dictor M, Persson A, Fernö M, Nilbert M. Immunohistochemical patterns in rectal cancer: application of tissue microarray with prognostic correlations. *Int J Cancer*. 111(6), 921–928 (2004).

142. Huh JW, Kim HC, Kim SH, *et al.* Mismatch Repair Gene Expression as a Predictor of Tumor Responses in Patients With Rectal Cancer Treated With Preoperative Chemoradiation. *Medicine (Baltimore)*. 95(3), e2582 (2016).

143. Duraiyan J, Govindarajan R, Kaliyappan K, Palanisamy M. Applications of immunohistochemistry. *J Pharm Bioallied Sci*. 4(Suppl 2), S307–S309 (2012).

144. Hatanaka Y, Hashizume K, Nitta K, Kato T, Itoh I, Tani Y. Cytometrical image analysis for immunohistochemical hormone receptor status in breast carcinomas. *Pathol Int*. 53(10), 693–699 (2003).

145. Huang M-Y, Lin C-H, Huang C-M, *et al.* Relationships between SMAD3 expression and preoperative fluoropyrimidine-based chemoradiotherapy response in locally advanced rectal cancer patients. *World J Surg*. 39(5), 1257–1267 (2015).

146. Huang F, Shi Q, Li Y, *et al.* HER2/EGFR-AKT Signaling Switches TGFβ from Inhibiting Cell Proliferation to Promoting Cell Migration in Breast Cancer. *Cancer Res*. 78(21), 6073–6085 (2018).

147. Mármol I, Sánchez-de-Diego C, Pradilla Dieste A, Cerrada E, Rodriguez Yoldi MJ. Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer. *Int J Mol Sci*. 18(1), E197 (2017).

148. Recio-Boiles A, Cagir B. Colon Cancer [Internet]. In: *StatPearls*, StatPearls Publishing, Treasure Island (FL) (2022) [cited 2022 Jan 20]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK470380/.

149. Johnson CM, Wei C, Ensor JE, *et al.* Meta-analyses of colorectal cancer risk factors. *Cancer Causes Control*. 24(6), 1207–1222 (2013).

150. Hampel H, Frankel WL, Martin E, *et al.* Feasibility of Screening for Lynch Syndrome Among Patients With Colorectal Cancer. *J Clin Oncol*. 26(35), 5783–5788 (2008).

151. Kawakami H, Zaanan A, Sinicrope FA. Implications of mismatch repair-deficient status on management of early stage colorectal cancer. *J Gastrointest Oncol*. 6(6), 676–684 (2015).

152. Kolligs FT. Diagnostics and Epidemiology of Colorectal Cancer. *Visc Med*. 32(3), 158–164 (2016).

153. Duffy MJ. Carcinoembryonic antigen as a marker for colorectal cancer: is it clinically useful? *Clin Chem*. 47(4), 624–630 (2001).

154. Vodenkova S, Buchler T, Cervena K, Veskrnova V, Vodicka P, Vymetalkova V. 5-fluorouracil and other fluoropyrimidines in colorectal cancer: Past, present and future. *Pharmacol Ther*. 206, 107447 (2020).

155. Folprecht G, Cunningham D, Ross P, *et al.* Efficacy of 5-fluorouracil-based chemotherapy in elderly patients with metastatic colorectal cancer: a pooled analysis of clinical trials. *Ann Oncol*. 15(9), 1330–1338 (2004).

156. Costi R, Leonardi F, Zanoni D, Violi V, Roncoroni L. Palliative care and end-stage colorectal cancer management: The surgeon meets the oncologist. *World J Gastroenterol*. 20(24), 7602–7621 (2014).

157. Glynne-Jones R, Wyrwicz L, Tiret E, *et al.* Rectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Annals of Oncology*. 28, iv22–iv40 (2017).

158. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 68(6), 394–424 (2018).

159. Siegel RL, Miller KD, Fedewa SA, *et al.* Colorectal cancer statistics, 2017. *CA Cancer J Clin*. 67(3), 177–193 (2017).

160. Guida L. NEOPLASIE DEL RETTO E ANO. , 474 (2020).

161. Bosset J-F. Distal rectal cancer: Sphincter-sparing is also a challenge for the radiation oncologist. *Radiotherapy and Oncology*. 80(1), 1–3 (2006).

162. Sauer R, Becker H, Hohenberger W, *et al.* Preoperative versus postoperative chemoradiotherapy for rectal cancer. *N Engl J Med*. 351(17), 1731–1740 (2004).

163. Gérard J-P, Conroy T, Bonnetain F, *et al.* Preoperative radiotherapy with or without concurrent fluorouracil and leucovorin in T3-4 rectal cancers: results of FFCD 9203. *J Clin Oncol*. 24(28), 4620–4625 (2006).

164. Fazeli MS, Keramati MR. Rectal cancer: a review. *Med J Islam Repub Iran*. 29, 171 (2015).

165. Wei EK, Giovannucci E, Wu K, *et al.* Comparison of risk factors for colon and rectal cancer. *Int J Cancer*. 108(3), 433–442 (2004).

166. Oronsky B, Reid T, Larson C, Knox SJ. Locally advanced rectal cancer: The past, present, and future. *Semin Oncol*. 47(1), 85–92 (2020).

167. Yuhara H, Steinmaus C, Cohen SE, Corley DA, Tei Y, Buffler PA. Is Diabetes Mellitus an Independent Risk Factor for Colon Cancer and Rectal Cancer? *Am J Gastroenterol*. 106(11), 1911–1922 (2011).

168. Wang Z-H, Fang J-Y. Colorectal Cancer in Inflammatory Bowel Disease: Epidemiology, Pathogenesis and Surveillance. *Gastrointest Tumors*. 1(3), 146–154 (2014).

169. Kirkegaard H, Johnsen NF, Christensen J, Frederiksen K, Overvad K, Tjønneland A. Association of adherence to lifestyle recommendations and risk of colorectal cancer: a prospective Danish cohort study. *BMJ*. 341, c5504 (2010).

170. Aleksandrova K, Pischon T, Jenab M, *et al.* Combined impact of healthy lifestyle factors on colorectal cancer: a large European cohort study. *BMC Med*. 12, 168 (2014).

171. Murphy N, Norat T, Ferrari P, *et al.* Dietary Fibre Intake and Risks of Cancers of the Colon and Rectum in the European Prospective Investigation into Cancer and Nutrition (EPIC). *PLoS One*. 7(6), e39361 (2012).

172. Sehgal R, Sheahan K, O'Connell PR, Hanly AM, Martin ST, Winter DC. Lynch Syndrome: An Updated Review. *Genes*. 5(3), 497–507 (2014).

173. De Jong AE, Morreau H, Van Puijenbroek M, *et al.* The role of mismatch repair gene defects in the development of adenomas in patients with HNPCC. *Gastroenterology*. 126(1), 42–48 (2004).

174. Ang CW, Dawson R, Hall C, Farmer M. The diagnostic value of digital rectal examination in primary care for palpable rectal tumour. *Colorectal Dis*. 10(8), 789–792 (2008).

175. Catalano OA, Lee SI, Parente C, *et al.* Improving staging of rectal cancer in the pelvis: the role of PET/MRI. *Eur J Nucl Med Mol Imaging*. 48(4), 1235–1245 (2021).

176. Brierley JD, Gospodarowicz MK, Wittekind C. TNM Classification of Malignant Tumours. John Wiley & Sons.

177. Sarma DP. The Dukes classification of colorectal cancer. *JAMA*. 256(11), 1447 (1986).

178. van Gijn W, Marijnen CAM, Nagtegaal ID, *et al.* Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer: 12-year follow-up of the multicentre, randomised controlled TME trial. *Lancet Oncol*. 12(6), 575–582 (2011).

179. Lin J-Z, Peng J-H, Qdaisat A, *et al.* Preoperative chemoradiotherapy creates an opportunity to perform sphincter preserving resection for low-lying locally advanced rectal cancer based on an oncologic outcome study. *Oncotarget*. 7(35), 57317–57326 (2016).

180. Chawla S, Katz AW, Rauh SM, Monson JRT. Can Surgery be Avoided After Preoperative Chemoradiation for Rectal Cancer in the Era of Organ Preservation? Current Review of Literature. *Am J Clin Oncol*. 38(5), 534–540 (2015).

181. Brown CL, Ternent CA, Thorson AG, *et al.* Response to preoperative chemoradiation in stage II and III rectal cancer. *Dis Colon Rectum*. 46(9), 1189–1193 (2003).

182. Dayde D, Tanaka I, Jain R, Tai MC, Taguchi A. Predictive and Prognostic Molecular Biomarkers for Response to Neoadjuvant Chemoradiation in Rectal Cancer. *Int J Mol Sci*. 18(3), 573 (2017).

183. Valentini V, Coco C, Picciocchi A, *et al.* Does downstaging predict improved outcome after preoperative chemoradiation for extraperitoneal locally advanced rectal cancer? A long-term analysis of 165 patients. *Int J Radiat Oncol Biol Phys*. 53(3), 664–674 (2002).

184. Reerink O, Karrenbeld A, Plukker JTM, *et al.* Molecular prognostic factors in locally irresectable rectal cancer treated preoperatively by chemo-radiotherapy. *Anticancer Res*. 24(2C), 1217–1221 (2004).

185. Crane CH, Skibber JM, Feig BW, *et al.* Response to preoperative chemoradiation increases the use of sphincter-preserving surgery in patients with locally advanced low rectal carcinoma. *Cancer*. 97(2), 517–524 (2003).

186. Wheeler JMD, Dodds E, Warren BF, *et al.* Preoperative chemoradiotherapy and total mesorectal excision surgery for locally advanced rectal cancer: correlation with rectal cancer regression grade. *Dis Colon Rectum*. 47(12), 2025–2031 (2004).

187. Sauer R, Liersch T, Merkel S, *et al.* Preoperative versus postoperative chemoradiotherapy for locally advanced rectal cancer: results of the German CAO/ARO/AIO-94 randomized phase III trial after a median follow-up of 11 years. *J Clin Oncol*. 30(16), 1926–1933 (2012).

188. Bonnetain F, Bosset JF, Gerard JP, *et al.* What is the clinical benefit of preoperative chemoradiotherapy with 5FU/leucovorin for T3-4 rectal cancer in a pooled analysis of EORTC 22921 and FFCD 9203 trials: surrogacy in question? *Eur J Cancer*. 48(12), 1781–1790 (2012).

189. Gérard J-P, Azria D, Gourgou-Bourgade S, *et al.* Clinical outcome of the ACCORD 12/0405 PRODIGE 2 randomized trial in rectal cancer. *J Clin Oncol*. 30(36), 4558–4565 (2012).

190. Aschele C, Cionini L, Lonardi S, *et al.* Primary tumor response to preoperative chemoradiation with or without oxaliplatin in locally advanced rectal cancer: pathologic results of the STAR-01 randomized phase III trial. *J Clin Oncol*. 29(20), 2773–2780 (2011).

191. Allegra CJ, Yothers G, O'Connell MJ, *et al.* Neoadjuvant 5-FU or Capecitabine Plus Radiation With or Without Oxaliplatin in Rectal Cancer Patients: A Phase III Randomized Clinical Trial. *J Natl Cancer Inst*. 107(11), djv248 (2015).

192. Schmoll HJ, Stein A, Hofheinz RD, *et al.* Preoperative chemoradiotherapy and postoperative chemotherapy with capecitabine and oxaliplatin vs. capecitabine alone in locally advanced rectal cancer: final analyses. *Annals of Oncology*. 27, vi154 (2016).

193. Bengala C, Bettelli S, Bertolini F, *et al.* Epidermal growth factor receptor gene copy number, K-ras mutation and pathological response to preoperative cetuximab, 5-FU and radiation therapy in locally advanced rectal cancer. *Ann Oncol*. 20(3), 469–474 (2009).

194. Shamseddine A, Zeidan YH, El Husseini Z, *et al.* Efficacy and safety-in analysis of short-course radiation followed by mFOLFOX-6 plus avelumab for locally advanced rectal adenocarcinoma. *Radiat Oncol*. 15, 233 (2020).

195. Hanna CR, O'Cathail SM, Graham JS, *et al.* Durvalumab (MEDI 4736) in combination with extended neoadjuvant regimens in rectal cancer: a study protocol of a randomised phase II trial (PRIME-RT). *Radiat Oncol*. 16, 163 (2021).

196. Ozer L, Yildiz I, Bayoglu V, *et al.* Tailored total neoadjuvant therapy for locally advanced rectal cancer: One size may not fit for all! *Colorectal Dis*. 23(7), 1662–1669 (2021).

197. Bertolini F, Bengala C, Losi L, *et al.* Prognostic and predictive value of baseline and posttreatment molecular marker expression in locally advanced rectal cancer treated with neoadjuvant chemoradiotherapy. *Int J Radiat Oncol Biol Phys*. 68(5), 1455–1461 (2007).

198. Negri FV, Campanini N, Camisa R, *et al.* Biological predictive factors in rectal cancer treated with preoperative radiotherapy or radiochemotherapy. *Br J Cancer*. 98(1), 143–147 (2008).

199. Scully BF, R, C all C, Lee-Kong SA. Complete Clinical Response in Rectal Adenocarcinoma: Review of Treatment Options. *Journal of Clinical Gastroenterology and Hepatology* [Internet]. 1(1) (2016). Available from: https://www.imedpub.com/abstract/complete-clinical-response-in-rectal-adenocarcinoma-review-of-treatmentrnoptions-17745.html.

200. López-Campos F, Martín-Martín M, Fornell-Pérez R, *et al.* Watch and wait approach in rectal cancer: Current controversies and future directions. *World J Gastroenterol*. 26(29), 4218–4239 (2020).

201. Habr-Gama A, Perez RO, Wynn G, Marks J, Kessler H, Gama-Rodrigues J. Complete clinical response after neoadjuvant chemoradiation therapy for distal rectal cancer: characterization of

clinical and endoscopic findings for standardization. *Dis Colon Rectum*. 53(12), 1692–1698 (2010).

202. MacGregor TP, Maughan TS, Sharma RA. Pathological grading of regression following neoadjuvant chemoradiation therapy: the clinical need is now. *J Clin Pathol*. 65(10), 867–871 (2012).

203. Kim SH, Chang HJ, Kim DY, *et al.* What Is the Ideal Tumor Regression Grading System in Rectal Cancer Patients after Preoperative Chemoradiotherapy? *Cancer Res Treat*. 48(3), 998–1009 (2016).

204. Mandard AM, Dalibard F, Mandard JC, *et al.* Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations. *Cancer*. 73(11), 2680–2686 (1994).

205. Dworak O, Keilholz L, Hoffmann A. Pathological features of rectal cancer after preoperative radiochemotherapy. *Int J Colorectal Dis*. 12(1), 19–23 (1997).

206. Ryan R, Gibbons D, Hyland JMP, *et al.* Pathological response following long-course neoadjuvant chemoradiotherapy for locally advanced rectal cancer. *Histopathology*. 47(2), 141–146 (2005).

207. Trakarnsanga A, Gönen M, Shia J, *et al.* Comparison of Tumor Regression Grade Systems for Locally Advanced Rectal Cancer After Multimodality Treatment. *J Natl Cancer Inst*. 106(10), dju248 (2014).

208. Becker K, Mueller JD, Schulmacher C, *et al.* Histomorphology and grading of regression in gastric carcinoma treated with neoadjuvant chemotherapy. *Cancer*. 98(7), 1521–1530 (2003).

209. Rödel C, Martus P, Papadoupolos T, *et al.* Prognostic significance of tumor regression after preoperative chemoradiotherapy for rectal cancer. *J Clin Oncol*. 23(34), 8688–8696 (2005).

210. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol*. 51 Pt 1, 263–273 (1986).

211. Huang D, Sun W, Zhou Y, *et al.* Mutations of key driver genes in colorectal cancer progression and metastasis. *Cancer Metastasis Rev*. 37(1), 173–187 (2018).

212. van Kuilenburg ABP, Häusler P, Schalhorn A, *et al.* Evaluation of 5-fluorouracil pharmacokinetics in cancer patients with a c.1905+1G>A mutation in DPYD by means of a Bayesian limited sampling strategy. *Clin Pharmacokinet*. 51(3), 163–174 (2012).

213. van Kuilenburg AB, Muller EW, Haasjes J, *et al.* Lethal outcome of a patient with a complete dihydropyrimidine dehydrogenase (DPD) deficiency after administration of 5-fluorouracil: frequency of the common IVS14+1G>A mutation causing DPD deficiency. *Clin Cancer Res*. 7(5), 1149–1153 (2001).

214. van Kuilenburg ABP, Meinsma R, Zonnenberg BA, *et al.* Dihydropyrimidinase deficiency and severe 5-fluorouracil toxicity. *Clin Cancer Res*. 9(12), 4363–4367 (2003).

215. Gonçalves-Ribeiro S, Sanz-Pamplona R, Vidal A, *et al.* Prediction of pathological response to neoadjuvant treatment in rectal cancer with a two-protein immunohistochemical score derived from stromal gene-profiling. *Ann Oncol*. 28(9), 2160–2168 (2017).

216. Wu C, Bell CM, Wodchis WP. Incidence and Economic Burden of Adverse Drug Reactions among Elderly Patients in Ontario Emergency Departments. *Drug Saf*. 35(9), 769–781 (2012).

217. Mattison LK, Fourie J, Desmond RA, Modak A, Saif MW, Diasio RB. Increased prevalence of dihydropyrimidine dehydrogenase deficiency in African-Americans compared with Caucasians. *Clin Cancer Res*. 12(18), 5491–5495 (2006).

218. Wörmann B, Bokemeyer C, Burmeister T, *et al.* Dihydropyrimidine Dehydrogenase Testing prior to Treatment with 5-Fluorouracil, Capecitabine, and Tegafur: A Consensus Paper. *Oncol Res Treat*. 43(11), 628–636 (2020).

219. Innocenti F, Mills SC, Sanoff H, Ciccolini J, Lenz H-J, Milano G. All You Need to Know About DPYD Genetic Testing for Patients Treated With Fluorouracil and Capecitabine: A Practitioner-Friendly Guide. *JCO Oncol Pract*. 16(12), 793–798 (2020).

220. Henricks LM, Lunenburg CATC, de Man FM, *et al.* DPYD genotype-guided dose individualisation of fluoropyrimidine therapy in patients with cancer: a prospective safety analysis. *Lancet Oncol*. 19(11), 1459–1467 (2018).

221. Ly RC, Schmidt RE, Kiel PJ, *et al.* Severe Capecitabine Toxicity Associated With a Rare DPYD Variant Identified Through Whole-Genome Sequencing. *JCO Precis Oncol*. 4, PO.20.00067 (2020).

222. Ciccolini J, Gross E, Dahan L, Lacarelle B, Mercier C. Routine dihydropyrimidine dehydrogenase testing for anticipating 5-fluorouracil-related severe toxicities: hype or hope? *Clin Colorectal Cancer*. 9(4), 224–228 (2010).

223. Morel A, Boisdron-Celle M, Fey L, *et al.* Clinical relevance of different dihydropyrimidine dehydrogenase gene single nucleotide polymorphisms on 5-fluorouracil tolerance. *Mol Cancer Ther*. 5(11), 2895–2904 (2006).

224. Ridge SA, Sludden J, Wei X, *et al.* Dihydropyrimidine dehydrogenase pharmacogenetics in patients with colorectal cancer. *Br J Cancer*. 77(3), 497–500 (1998).

225. Mercier C, Ciccolini J. Severe or lethal toxicities upon capecitabine intake: is DPYD genetic polymorphism the ideal culprit? *Trends Pharmacol Sci*. 28(12), 597–598 (2007).

226. Lim H, Kim SY, Lee E, *et al.* Sex-Dependent Adverse Drug Reactions to 5-Fluorouracil in Colorectal Cancer. *Biol Pharm Bull*. 42(4), 594–600 (2019).

227. Radkiewicz C, Johansson ALV, Dickman PW, Lambe M, Edgren G. Sex differences in cancer risk and survival: A Swedish cohort study. *Eur J Cancer*. 84, 130–140 (2017).

228. Port RE, Daniel B, Ding RW, Herrmann R. Relative importance of dose, body surface area, sex, and age for 5-fluorouracil clearance. *Oncology*. 48(4), 277–281 (1991).

229. Mueller F, Büchel B, Köberle D, *et al.* Gender-specific elimination of continuous-infusional 5-fluorouracil in patients with gastrointestinal malignancies: results from a prospective population pharmacokinetic study. *Cancer Chemother Pharmacol*. 71(2), 361–370 (2013).

230. Sloan JA, Goldberg RM, Sargent DJ, *et al.* Women experience greater toxicity with fluorouracil-based chemotherapy for colorectal cancer. *J Clin Oncol*. 20(6), 1491–1498 (2002).

231. Chansky K, Benedetti J, Macdonald JS. Differences in toxicity between men and women treated with 5-fluorouracil therapy for colorectal carcinoma. *Cancer*. 103(6), 1165–1171 (2005).

232. Ilich AI, Danilak M, Kim CA, *et al.* Effects of gender on capecitabine toxicity in colorectal cancer. *J Oncol Pharm Pract*. 22(3), 454–460 (2016).

233. Ruzzo A, Graziano F, Galli F, *et al.* Author Correction: Sex-Related Differences in Impact on Safety of Pharmacogenetic Profile for Colon Cancer Patients Treated with FOLFOX-4 or XELOX Adjuvant Chemotherapy. *Sci Rep*. 10(1), 1918 (2020).

234. Meulendijks D, Henricks LM, Sonke GS, *et al.* Clinical relevance of DPYD variants c.1679T>G, c.1236G>A/HapB3, and c.1601G>A as predictors of severe fluoropyrimidine-associated toxicity: a systematic review and meta-analysis of individual patient data. *Lancet Oncol*. 16(16), 1639–1650 (2015).

235. Sharma BB, Rai K, Blunt H, Zhao W, Tosteson TD, Brooks GA. Pathogenic DPYD Variants and Treatment-Related Mortality in Patients Receiving Fluoropyrimidine Chemotherapy: A Systematic Review and Meta-Analysis. *Oncologist*. 26(12), 1008–1016 (2021).

236. Terrazzino S, Cargnin S, Del Re M, Danesi R, Canonico PL, Genazzani AA. DPYD IVS14+1G>A and 2846A>T genotyping for the prediction of severe fluoropyrimidine-related toxicity: a meta-analysis. *Pharmacogenomics*. 14(11), 1255–1272 (2013).

237. Froehlich TK, Amstutz U, Aebi S, Joerger M, Largiadèr CR. Clinical importance of risk variants in the dihydropyrimidine dehydrogenase gene for the prediction of early-onset fluoropyrimidine toxicity. *Int J Cancer*. 136(3), 730–739 (2015).

238. Loganayagam A, Arenas Hernandez M, Corrigan A, *et al.* Pharmacogenetic variants in the DPYD, TYMS, CDA and MTHFR genes are clinically significant predictors of fluoropyrimidine toxicity. *Br J Cancer*. 108(12), 2505–2515 (2013).

239. Offer SM, Wegner NJ, Fossum C, Wang K, Diasio RB. Phenotypic profiling of DPYD variations relevant to 5-fluorouracil sensitivity using real-time cellular analysis and in vitro measurement of enzyme activity. *Cancer Res*. 73(6), 1958–1968 (2013).

240. Gross E, Ullrich T, Seck K, *et al.* Detailed analysis of five mutations in dihydropyrimidine dehydrogenase detected in cancer patients with 5-fluorouracil-related side effects. *Hum Mutat*. 22(6), 498 (2003).

241. Khushman M, Patel GK, Hosein PJ, *et al.* Germline pharmacogenomics of DPYD*9A (c.85T>C) variant in patients with gastrointestinal malignancies treated with fluoropyrimidines. *J Gastrointest Oncol*. 9(3), 416–424 (2018).

242. Madi A, Fisher D, Maughan TS, *et al.* Pharmacogenetic analyses of 2183 patients with advanced colorectal cancer; potential role for common dihydropyrimidine dehydrogenase variants in toxicity to chemotherapy. *Eur J Cancer*. 102, 31–39 (2018).

243. Etienne-Grimaldi M-C, Boyer J-C, Beroud C, *et al.* New advances in DPYD genotype and risk of severe toxicity under capecitabine. *PLoS One*. 12(5), e0175998 (2017).

244. Offer SM, Fossum CC, Wegner NJ, Stuflesser AJ, Butterfield GL, Diasio RB. Comparative functional analysis of DPYD variants of potential clinical relevance to dihydropyrimidine dehydrogenase activity. *Cancer Res*. 74(9), 2545–2554 (2014).

245. Gross E, Busse B, Riemenschneider M, *et al.* Strong association of a common dihydropyrimidine dehydrogenase gene polymorphism with fluoropyrimidine-related toxicity in cancer patients. *PLoS One*. 3(12), e4003 (2008).

246. Ruzzo A, Graziano F, Galli F, *et al.* Dihydropyrimidine dehydrogenase pharmacogenetics for predicting fluoropyrimidine-related toxicity in the randomised, phase III adjuvant TOSCA trial in high-risk colon cancer patients. *Br J Cancer*. 117(9), 1269–1277 (2017).

247. Kleibl Z, Fidlerova J, Kleiblova P, *et al.* Influence of dihydropyrimidine dehydrogenase gene (DPYD) coding sequence variants on the development of fluoropyrimidine-related toxicity in patients with high-grade toxicity and patients with excellent tolerance of fluoropyrimidine-based chemotherapy. *Neoplasma*. 56(4), 303–316 (2009).

248. van Kuilenburg AB, Haasjes J, Richel DJ, *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*. 6(12), 4705–4712 (2000).

249. Deenen MJ, Tol J, Burylo AM, *et al.* Relationship between single nucleotide polymorphisms and haplotypes in DPYD and toxicity and efficacy of capecitabine in advanced colorectal cancer. *Clin Cancer Res*. 17(10), 3455–3468 (2011).

250. Del Re M, Cinieri S, Michelucci A, *et al.* DPYD*6 plays an important role in fluoropyrimidine toxicity in addition to DPYD*2A and c.2846A>T: a comprehensive analysis in 1254 patients. *Pharmacogenomics J*. 19(6), 556–563 (2019).

251. Yamada Y, Muro K, Takahashi K, *et al.* Impact of sex and histology on the therapeutic effects of fluoropyrimidines and oxaliplatin plus bevacizumab for patients with metastatic colorectal cancer in the SOFT trial. *Glob Health Med*. 2(4), 240–246 (2020).

252. Beckman L, Fröhlander N. Heterozygosity effects in studies of genetic markers and disease. *Hum Hered*. 40(6), 322–329 (1990).

253. Williams J, Spurlock G, Holmans P, *et al.* A meta-analysis and transmission disequilibrium study of association between the dopamine D3 receptor gene and schizophrenia. *Mol Psychiatry*. 3(2), 141–149 (1998).

254. Keeton, W. T., & Gould, J. L. Biological Science (New York, Norton, 1986).

255. Kohn MH, Pelz HJ, Wayne RK. Natural selection mapping of the warfarin-resistance gene. *Proc Natl Acad Sci U S A*. 97(14), 7911–7915 (2000).

256. Falvella FS, Cheli S, Martinetti A, *et al.* DPD and UGT1A1 deficiency in colorectal cancer patients receiving triplet chemotherapy with fluoropyrimidines, oxaliplatin and irinotecan. *Br J Clin Pharmacol*. 80(3), 581–588 (2015).

257. Sobecki M, Mrouj K, Camasses A, *et al.* The cell proliferation antigen Ki-67 organises heterochromatin. *Elife*. 5, e13722 (2016).

258. Sun X, Kaufman PD. Ki-67: more than a proliferation marker. *Chromosoma*. 127(2), 175–186 (2018).

259. Kroeze SGC, Bijenhof AM, Bosch JLHR, Jans JJM. Diagnostic and prognostic tissuemarkers in clear cell and papillary renal cell carcinoma. *Cancer Biomark*. 7(6), 261–268 (2010).

260. Jalava P, Kuopio T, Juntti-Patinen L, Kotkansalo T, Kronqvist P, Collan Y. Ki67 immunohistochemistry: a valuable marker in prognostication but with a risk of misclassification: proliferation subgroups formed based on Ki67 immunoreactivity and standardized mitotic index. *Histopathology*. 48(6), 674–682 (2006).

261. Zhu X, Chen L, Huang B, *et al.* The prognostic and predictive potential of Ki-67 in triple-negative breast cancer. *Sci Rep*. 10(1), 225 (2020).

262. Luo Z-W, Zhu M-G, Zhang Z-Q, Ye F-J, Huang W-H, Luo X-Z. Increased expression of Ki-67 is a poor prognostic marker for colorectal cancer patients: a meta analysis. *BMC Cancer*. 19(1), 123 (2019).

263. Brown DC, Gatter KC. Ki67 protein: the immaculate deception? *Histopathology*. 40(1), 2–11 (2002).

264. Jakob C, Liersch T, Meyer W, Becker H, Baretton GB, Aust DE. Predictive value of Ki67 and p53 in locally advanced rectal cancer: Correlation with thymidylate synthase and

histopathological tumor regression after neoadjuvant 5-FU-based chemoradiotherapy. *World J Gastroenterol*. 14(7), 1060–1066 (2008).

265. Imaizumi K, Suzuki T, Kojima M, *et al.* Ki67 expression and localization of T cells after neoadjuvant therapies as reliable predictive markers in rectal cancer. *Cancer Sci*. 111(1), 23–35 (2020).

266. Ganju RK, Brubaker SA, Meyer J, *et al.* The alpha-chemokine, stromal cell-derived factor-1alpha, binds to the transmembrane G-protein-coupled CXCR-4 receptor and activates multiple signal transduction pathways. *J Biol Chem*. 273(36), 23169–23175 (1998).

267. Zhao H, Guo L, Zhao H, Zhao J, Weng H, Zhao B. CXCR4 over-expression and survival in cancer: a system review and meta-analysis. *Oncotarget*. 6(7), 5022–5040 (2015).

268. Furusato B, Mohamed A, Uhlén M, Rhim JS. CXCR4 and cancer. *Pathol Int*. 60(7), 497–505 (2010).

269. D'Alterio C, Avallone A, Tatangelo F, *et al.* A prognostic model comprising pT stage, N status, and the chemokine receptors CXCR4 and CXCR7 powerfully predicts outcome in neoadjuvant resistant rectal cancer patients. *Int J Cancer*. 135(2), 379–390 (2014).

270. Ottaiano A, Santorsola M, Del Prete P, *et al.* Prognostic Significance of CXCR4 in Colorectal Cancer: An Updated Meta-Analysis and Critical Appraisal. *Cancers (Basel)*. 13(13), 3284 (2021).

271. Smith FM, Reynolds JV, Kay EW, *et al.* COX-2 overexpression in pretreatment biopsies predicts response of rectal cancers to neoadjuvant radiochemotherapy. *Int J Radiat Oncol Biol Phys*. 64(2), 466–472 (2006).

272. Edden Y, Wexner SD, Berho M. The use of molecular markers as a method to predict the response to neoadjuvant therapy for advanced stage rectal adenocarcinoma. *Colorectal Dis*. 14(5), 555–561 (2012).

273. Peng L, Zhou Y, Wang Y, Mou H, Zhao Q. Prognostic significance of COX-2 immunohistochemical expression in colorectal cancer: a meta-analysis of the literature. *PLoS One*. 8(3), e58891 (2013).

274. Negi RR, Rana SV, Gupta V, *et al.* Over-Expression of Cyclooxygenase-2 in Colorectal Cancer Patients. *Asian Pac J Cancer Prev*. 20(6), 1675–1681 (2019).

275. Behling F, Ries V, Skardelly M, *et al.* COX2 expression is associated with proliferation and tumor extension in vestibular schwannoma but is not influenced by acetylsalicylic acid intake. *Acta Neuropathologica Communications*. 7(1), 105 (2019).

276. Meng W, Xue S, Chen Y. The role of CXCL12 in tumor microenvironment. *Gene*. 641, 105–110 (2018).

277. Samarendra H, Jones K, Petrinic T, *et al.* A meta-analysis of CXCL12 expression for cancer prognosis. *Br J Cancer*. 117(1), 124–135 (2017).

278. Zhang J, Liu C, Mo X, Shi H, Li S. Mechanisms by which CXCR4/CXCL12 cause metastatic behavior in pancreatic cancer. *Oncol Lett*. 15(2), 1771–1776 (2018).

279. Klein HL. The consequences of Rad51 overexpression for normal and tumor cells. *DNA Repair (Amst)*. 7(5), 686–693 (2008).

280. Gachechiladze M, Škarda J, Soltermann A, Joerger M. RAD51 as a potential surrogate marker for DNA repair capacity in solid malignancies. *Int J Cancer*. 141(7), 1286–1294 (2017).

281. Richardson C. RAD51, genomic stability, and tumorigenesis. *Cancer Lett*. 218(2), 127–139 (2005).

282. Jin X, Dai L, Ma Y, Wang J, Liu Z. Implications of HIF-1α in the tumorigenesis and progression of pancreatic cancer. *Cancer Cell International*. 20(1), 273 (2020).

283. Stoeltzing O, McCarty MF, Wey JS, *et al.* Role of hypoxia-inducible factor 1alpha in gastric cancer cell growth, angiogenesis, and vessel maturation. *J Natl Cancer Inst*. 96(12), 946–956 (2004).

284. Gort EH, Groot AJ, Derks van de Ven TLP, *et al.* Hypoxia-inducible factor-1α expression requires PI 3-kinase activity and correlates with Akt1 phosphorylation in invasive breast carcinomas. *Oncogene*. 25(45), 6123–6127 (2006).

285. Leppänen J, Helminen O, Huhta H, *et al.* Weak HIF-1alpha expression indicates poor prognosis in resectable pancreatic ductal adenocarcinoma. *World Journal of Surgical Oncology*. 16(1), 127 (2018).

286. Baba Y, Nosho K, Shima K, *et al.* HIF1A overexpression is associated with poor prognosis in a cohort of 731 colorectal cancers. *Am J Pathol*. 176(5), 2292–2301 (2010).

287. Chen B, Li L, Li M, Wang X. HIF1A expression correlates with increased tumor immune and stromal signatures and aggressive phenotypes in human cancers. *Cell Oncol (Dordr)*. 43(5), 877–888 (2020).

288. Costache MI, Ioana M, Iordache S, Ene D, Costache CA, Săftoiu A. VEGF Expression in Pancreatic Cancer and Other Malignancies: A Review of the Literature. *Rom J Intern Med*. 53(3), 199–208 (2015).

289. Yu J, Lee S-H, Jeung TS, Chang H. Expression of vascular endothelial growth factor as a predictor of complete response for preoperative chemoradiotherapy in rectal cancer. *Medicine (Baltimore)*. 98(26), e16190 (2019).

290. Zlobec I, Vuong T, Compton CC, *et al.* Combined analysis of VEGF and EGFR predicts complete tumour response in rectal cancer treated with preoperative radiotherapy. *Br J Cancer*. 98(2), 450–456 (2008).

8 PUBBLICATIONS

1. De Mattia E, Canzonieri V, Polesel J, **Mezzalira S**, Dalle Fratte C, Dreussi E, Roncato R, Bignucolo A, Innocente R, Belluco C, Pucciarelli S, De Paoli A, Palazzari E, Toffoli G, Cecchin E. SMAD3 Host and Tumor Profiling to Identify Locally Advanced Rectal Cancer Patients at High Risk of Poor Response to Neoadjuvant Chemoradiotherapy. *Front Pharmacol.* **2021** Dec 24;12:778781. doi: 10.3389/fphar.2021.778781. PMID: 35002714; PMCID: PMC8740633.

2. **Mezzalira S**, Toffoli G. The effects of sex on pharmacogenetically guided drug treatment. *Pharmacogenomics*. **2021** Oct;22(15):959-962. doi: 10.2217/pgs-2021-0088. Epub 2021 Sep 21. PMID: 34545749.

3. Bignucolo A, Scarabel L, **Mezzalira S**, Polesel J, Cecchin E, Toffoli G. Sex Disparities in Efficacy in COVID-19 Vaccines: A Systematic Review and Meta-Analysis. *Vaccines* (Basel). **2021** Jul 27;9(8):825. doi: 10.3390/vaccines9080825. PMID: 34451950; PMCID: PMC8402482.

4. Dalle Fratte C, **Mezzalira S**, Polesel J, De Mattia E, Palumbo A, Buonadonna A, Palazzari E, De Paoli A, Belluco C, Canzonieri V, Toffoli G, Cecchin E. A panel of tumor biomarkers to predict complete pathological response to neo-adjuvant treatment in Locally Advanced Rectal Cancer. *Oncol Res*. **2021** Jun 9. doi: 10.3727/096504021X16232280278813. Epub ahead of print. PMID: 34108073.

5. Dal Bo M, De Mattia E, Baboci L, **Mezzalira S**, Cecchin E, Assaraf YG, Toffoli G. New insights into the pharmacological, immunological, and CAR-T-cell approaches in the treatment of hepatocellular carcinoma. *Drug Resist Updat.* **2020** Jul;51:100702. doi: 10.1016/j.drup.2020.100702. Epub 2020 Apr 19. PMID: 32371296.

6. **Mezzalira S**, De Mattia E, Guardascione M, Dalle Fratte C, Cecchin E, Toffoli G. Circulating-Free DNA Analysis in Hepatocellular Carcinoma: A Promising Strategy to Improve Patients' Management and Therapy Outcomes. *Int J Mol Sci*. **2019** Nov 5;20(21):5498. doi: 10.3390/ijms20215498. PMID: 31694149; PMCID: PMC6861910.

7. Roncato R, Dal Cin L, **Mezzalira S**, Comello F, De Mattia E, Bignucolo A, Giollo L, D'Errico S, Gulotta A, Emili L, Carbone V, Guardascione M, Foltran L, Toffoli G, Cecchin E. FARMAPRICE: A Pharmacogenetic Clinical Decision Support System for Precise and Cost-Effective Therapy. *Genes* (Basel). **2019** Apr 4;10(4):276. doi: 10.3390/genes10040276. PMID: 30987397; PMCID: PMC6523070.

Acknowledgment

First and foremost, I would like to thank my PhD supervisor, Dr. Erika Cecchin, for giving me the opportunity to work in her research group. From the first moments of my thesis internship in pharmacy, she has given me moral support and encouragement both in my professional career and throughout my studies. I would also like to thank Dr. Giuseppe Toffoli for hosting me in the Clinical and Experimental Pharmacology Unit of CRO Aviano and for all the opportunity give me. In addition, I would like to thank Dr. Elena De Mattia for her scientific support, especially in the NGS analysis, but also for personal speeches.

I am also very grateful to all the clinicians and specialists with whom I had the honour to collaborate. In primis, many thanks to Dr. Elisa Palazzari: our afternoon meetings were not only fruitful from a business point of view, but above all important for my personal growth.

This PhD would not have been possible without the collaboration of all the wonderful people working in this department. Together, they have all contributed to making these years simply unforgettable.

For this reason, a big thank goes to all the current and former members of the Division of Clinical and Experimental Pharmacology Unit of CRO Aviano. Special thanks to Antonella Moro, Franca Sartor and Monica D'Andrea for their experimental and administrative support. Thanks also to those who are no longer with the FCS unit and are crowning their careers elsewhere: thank you Lisa and thank you Francesco for the wonderful people you are and for our friendship that never ends.

And now, last but not least, thank you to my wonderful friend Alessia. She is a very precious friend for her tireless support and for all the conversations we had inside and outside the lab. Thank you for taking me in as if I were part of your family. Thank you to my friends Nicole and Aharon, active members of "stanza borsisti", for all the moments of laughter and sadness we spent together. You were the sun on the darkest days

My personal thanks go first of all to my family. To my wonderful parents who supported my studies, who endured my crises, who were always there for me with valuable advice or with silences that spoke for themselves. Without you, I would not be where I am now! And finally, a special thanks to the person I hope will be by my side forever, Andrea. These years have been testing for us because of the enormous distance that separated us. I just thank you for always being there for me.