INTERNATIONAL DOCTORAL
SCHOOL OF THE USC

Ana Rita
Santos Simões

PhD Thesis

Novel approaches for the discovery of pharmacogenetic biomarkers of chemotoxicity in patients with colorectal cancer

Santiago de Compostela, 2023

Doctoral Programme in Molecular Medicine

DOCTORAL THESIS

NOVEL APPROACHES FOR THE DISCOVERY OF PHARMACOGENETIC BIOMARKERS OF CHEMOTOXICITY IN PATIENTS WITH COLORECTAL CANCER

Ana Rita Santos Simões

INTERNATIONAL PHD SCHOOL OF THE UNIVERSITY OF SANTIAGO DE COMPOSTELA
PHD PROGRAMME IN MOLECULAR MEDICINE.

SANTIAGO DE COMPOSTELA
2023

DECLARATION BY THE THESIS AUTHOR

Mr/Ms. **Ana Rita Santos Simões**

Thesis Title: **Novel approaches for the discovery of pharmacogenetic biomarkers of chemotoxicity in patients with colorectal cancer.**

I submit my thesis, following the appropriate procedure according to the University of Santiago de Compostela Regulations, and declare that:

1) The thesis covers the results of the elaboration of my work.
2) If this is the case, the thesis refers to the collaborations within this work.
3) I confirm that the thesis does not incur in any kind of plagiarism from other authors nor works submitted by me in order to obtain other titles.
4) The thesis is the final version submitted for its defence and both the printed version and the electronic version provide the same content.

And I agree to submit the Documentary Commitment of supervision in the event that the original version is not deposited at the School.

In **Santiago de Compostela**, **17th July 2023**.

Signed by: Ana Rita Simões

SUPERVISOR / TUTOR AUTHORISATION

[Novel approaches for the discovery of pharmacogenetic biomarkers of chemotoxicity in patients with colorectal cancer]

Dr Ceres Fernández-Rozadilla
Prof Ángel Carracedo Álvarez

STATE:

That this thesis corresponds to the work carried out by Ms Ana Rita Santos Simões, under our supervision/tutoring, and hereby authorise its presentation, taking into account that it meets all the relevant requirements stated in the Doctoral Studies Regulations of the USC, and as its directors, it does not incur in the causes of abstention established in the 40/2015 Law.

In accordance with what is stated in the Regulations for Doctoral Studies, it is also declared that this doctoral thesis can be defended based on the modality of *Monographic with reproduction of publications*, in which the participation of the PhD student was decisive for its elaboration and the publications adjust to the Research Plan.

In Santiago de Compostela, 17th July 2023

Signed by: Dr. Ceres Fernández-Rozadilla Signed by: Prof. Ángel Carracedo Álvarez

Funding

This research was supported by *Instituto de Salud Carlos III* (ISCIII, Carlos III Health Institute) with the grant FIS PI 16/01057.

The PhD student has received a grant from ISCIII (PFIS grant FI17/00215).

This project has also been made possible thanks to the collaboration of the Galician Supercomputing Centre (CESGA).

Regarding the guidelines of the International PhD School (EDI) in Health Sciences of the University of Santiago de Compostela (USC), the PhD student Ana Rita Santos Simões declares that she has no conflict of interest in relation to this doctoral thesis.

In Santiago de Compostela, 17th July 2023

Signed by: Ana Rita Simões

Es difícil escribir pocas palabras que reflejen mi total agradecimiento a todas las personas que me ayudaron a lo largo de estos varios años e hicieron posible esta tesis. Efectivamente, sin ellas, sería mucho más difícil.

En primer lugar, a mis directores, Ceres y Ángel. Gracias Ángel por apostar por mí y darme la oportunidad de entrar en la familia de Xenómica; ha sido un honor. Gracias Ceres por todos los debates de martes y no-martes, por tu orientación, tiempo y paciencia. Por transmitirme valiosos conocimientos y espíritu crítico que han contribuido notablemente a la construcción de esta tesis.

Gracias a todos los que participaron en el reclutamiento de pacientes, sin vuestra colaboración y contribución, la investigación no sería posible. A los pacientes. Un besito especial a Goretti que me ha dedicado parte de su tiempo y me inició en la PK/PD; me acordaré siempre de las golosinas 3D.

Al personal de la Fundación/IDIS que de alguna manera me ha ayudado en algún momento de este proyecto.

A Sílvia, llegaste dos semanas antes que yo y me has recibido con los brazos abiertos, me acompañaste durante todo el trabajo, tu ayuda en el lab y en todo lo necesario es inestimable. Dentro y fuera del lab, eres mi hombro amigo y te tomo con inmenso cariño. Graciñas!

A mis compis de despacho, por la buena compañía diaria, los debates de Game of Thrones y los chocolates compartidos. Obrigada ao Vicente, o meu galego/português preferido. A Juan, sin ti estaría perdida en el mundo de la bioinformática.

A los del café, muchas gracias por haberme recibido tan bien en el grupo de la lucha por la sanidad. Las pausas del café y las comidas hicieron más aceptables los días más duros. Me encantó compartir esta experiencia con vosotros.

Thanks to Dr Carlo Largiadèr and the group for receiving me so warmly for four months and sharing the wisdom about *DPYD* and functional studies. Many thanks to Alisa for being so nice and patient with me and for carrying on the work, relentlessly. It was such a pleasure to meet you all. Bern and Switzerland are indeed beautiful.

Obrigada à Dona Lurdes por me tratar tão bem. Ainda bem que há sempre um português em cada canto do mundo.

A mi Trio calavera, mis queridas chicas, no hay forma de agradecer vuestra amistad. Gracias por compartir mis dramas, por todas las llamadas, todo el apoyo y consejos. Santiago es casa para mí. También a Rob, por las pelis de Marvel, los pollos al curry, y por recibirme siempre tan bien en vuestra casa. Os quiero muchísimo!!

Obrigada ao Tiago e à minha Cat, por serem os melhores companheiros de viagem, por serem o meu cantinho português no estrangeiro, por ouvirem os meus desabafos e por todas as palavras de incentivo.

Aos meus amigos de coração. Aos que são 100%, à Irmandade, e à minha Velha Guarda. Agradeço o dia em que nos cruzámos e se tornaram parte da minha vida. Muito obrigada por estarem lá sempre, nos momentos bons e menos bons, porque tudo o que já partilhamos fez-me chegar até aqui e, independentemente da distância, sei que estão para o que der e vier.

À minha Família, as pessoas mais importantes da minha vida. É impossível para mim transpor neste papel o quanto significam para mim. Obrigada pelo vosso amor e apoio incondicional, pela coragem que me dão sempre, por me ensinarem a lutar, a procurar sempre mais e a dar o melhor de mim. Todas as minhas conquistas são vossas também.

Ao meu André. Obrigada por me acompanhares incansavelmente nesta caminhada, pela paciência ao longo deste desafio. Obrigada por estares sempre do meu lado e por seres o meu por de abrigo, por torceres sempre por mim e me dares força, motivação e confiança constantes, mesmo quando eu duvido. Muito obrigada por tudo o que és e fazes por mim.

I extend a heartfelt thank you to all the people who somehow supported me during this work.

Published article related to this thesis

The road so far in colorectal cancer pharmacogenomics: are we closer to individualised treatment?

Ana Rita Simões,[1,2], Olalla Maroñas[1], Ángel Carracedo[1,2,3,4], Ceres Fernández-Rozadilla[1,2]

[1] Grupo de Medicina Xenómica, Universidade de Santiago de Compostela (USC); Santiago de Compostela, Spain.
[2] Instituto de Investigación Sanitaria de Santiago (IDIS); Santiago de Compostela, Spain.
[3] Fundación Pública Galega de Medicina Xenómica; SERGAS, Santiago de Compostela, Spain.
[4] Consorcio Centro de Investigación Biomédica en Red de Enfermedades Raras – CIBERER, Spain.

Contribution of the PhD student: writing—original draft preparation and writing—review and editing.

Impact Factor: 3.508; Q2 (Health Care Sciences & Services)

Contents of this publication are reproduced in the following chapters of this doctoral thesis: Introduction, Materials and Methods

The complete article is included in Annex 5.

Journal of Personalized Medicine. 2020, 10(4), 237. EISSN 2075-442
https://doi.org/10.3390/jpm10040237

Index

Abbreviations

5-FU	5-Fluorouracil
5-FU/LV + irinotecan	FOLFIRI
5-FU/LV + oxaliplatin	FOLFOX
ABC-transporter	Adenosine-triphosphate binding cassette transporter
ADME	Adsorption, distribution, metabolism, excretion
ADR	Adverse drug reaction
BB	Biobank of Santiago de Compostela
BER	Repair base excision repair
bp	Base pair
Capecitabine + oxaliplatin	XELOX
CDA	Cytidine deaminase
CES	Carboxylesterase
CHUO	Complexo Hospitalario Universitario de Ourense
CHUS	Complexo Hospitalario Universitario de Santiago
CIMP	CpG island methylator phenotype
CIN	Chromosomal instability
CRC	Colorectal cancer
CRF	Case report form
CTCAE	Common Terminology Criteria for Adverse Events
DM	Diarrhoea + mucositis
DMN	Diarrhoea + mucositis + neutropenia
DPD	Dihydropyrimidine dehydrogenase
EGA	EGAS00001002763 cohort
EGFR	Epidermal growth factor receptor
EPI	EPICOLON II cohort
eQTL	Expression quantitative trait locus
FDA	Food and Drug Administration
GST	Glutathione S-transferase
GTEx	Genotype-tissue expression
GWAS	Genome-wide association studies
HFS	Hand-foot syndrome
HUGM	Hospital Universitario Gregorio Marañón
HUIS	Hospital Universitario Infanta sofia
ICO	Institut Català d'Oncologia
KOPIE	Institute of Experimental Medicine, Czech Republic
LD	Linkage disequilibrium
LV	Leucovorin
mAb	Monoclonal antibody
MAF	Minor allele frequency
MEX	Technológico in Monterrey, Mexico
MMR	Mismatch repair
MSI	Microsatellite instability
MSI-H	Microsatellite instability-high
MWAS	Methylome-wide association studies
NER	Nucleotide excision
NGS	Next-generation sequencing

non-Finnish European population	NFE
OR	Odds ratios
OXCPN	Chronic oxaliplatin-induced peripheral neurotoxicity
PD-1	Programmed cell death protein 1
Q2	QUASAR2 pharmacogenomics GWAS
SKAT-O	Optimal sequence kernel association test
SLC	Solute carrier protein
SNP	Single nucleotide polymorphism
SPau	Hospital de la Santa Creu i Sant Pau
TNM	Tumour-node-metastasis
TP	Thymidine phosphorylase
TS	Thymidylate synthase
TWAS	Transcriptome-wide association studies
UGT	UDP glucuronosyl-transferase
UTR	Untranslated region
VEGF	Vascular endothelial growth factor
WES	Whole-exome sequencing
WGS	Whole-genome sequencing
wt	Wild-type

Abstract

Colorectal cancer (CRC) is the 3rd most diagnosed cancer and the 2nd cause of cancer-related death. This represents a considerable health and economic problem, and its incidence is expected to increase notably, especially in younger cases. The treatment of choice for CRC is surgical resection, often combined with chemotherapy (and radiotherapy, in the case of rectal cancer). CRC chemotherapy is predominantly based on fluoropyrimidines (5-fluorouracil - 5-FU - or its prodrug capecitabine), which are administered alone or in combination with other drugs, like oxaliplatin or irinotecan. Early diagnosis and the vast advances in chemotherapeutic treatments have allowed for increased survival in CRC patients, but the latter can also be the cause of several adverse drug reactions (ADRs) that might affect the outcome of the treatment and hinder the patient's quality of life. In this context, pharmacogenetics and pharmacogenomics have emerged to study the inherited genetic variation responsible for the inter-variability observed in treatment response and the resulting toxicity. This is crucial to help optimise treatment outcomes by offering more tailored treatment strategies based on molecular information. Hence, our main aim was to identify and validate new germline biomarkers of toxicity to CRC chemotherapeutic treatments, using novel technological approaches.

For the first part of this study, we intended to identify novel toxicity variants/genes by using whole-exome sequencing (WES). For that purpose, we selected CRC patients who had received the most used chemotherapeutic drugs in CRC treatment - fluoropyrimidines, oxaliplatin, and irinotecan - and presented extreme toxicity phenotypes. We performed WES analysis in 163 cases and 52 controls. Secondly, we followed an omic-wide approach, where we focused on common variation using genome-wide association and post-GWAS approaches. As the ultimate purpose of pharmacogenomics is the prospective implementation of genetic biomarkers into clinical guidelines for pre-emptive treatment, we functionally assessed a novel candidate toxicity variant in the *DPYD* gene (rs202212118).

In our WES data, we found 13 cases carrying actionable *DPYD* variants. Moreover, we found 31 novel rare, high-impact variants in 14 of the reported genes. We also found seven patients carrying more than one variant in the same gene or pathway, with one patient hinting at a potential digenic inheritance. Amongst these, we identified a novel *DPYD* variant - c.2071G>T, p.(V691L), rs202212118. Functional assays could however not confirm reduced protein abundance or enzymatic activity derived from this change. Using an exome-wide approach, we discovered and independently validated three novel candidate toxicity genes (*ALDH9A1*, *FAM83A*, and *EPX*). Gene-based analyses also provided 14 genes significantly associated with neuropathy, skin toxicity, and cardiotoxicity. For the polygenic model analyses, we found no significant SNPs in our GWAS, but TWAS approaches identified 20 novel genomic loci that are influencing the toxicity risk through the expression of 24 candidate genes.

Overall, the present work has utilised state-of-the-art approaches to uncover several novel candidate toxicity variants/genes. These results require however further replication in larger cohorts and functional validation. Ultimately, we hope that, by identifying and validating these variants/genes as candidates for chemo-related toxicity, pre-emptive treatment adjustments can be made to avoid life-threatening toxicities, and, in the end, improve the quality of life of CRC patients.

Resumen

El cáncer colorrectal (CCR) es el tercer cáncer más diagnosticado y la segunda causa de muerte relacionada con el cáncer. Representa un problema sanitario y económico considerable, y se prevé que su incidencia aumente notablemente, sobre todo a edades más tempranas. El tratamiento base para el CCR es la resección quirúrgica, a menudo combinada con quimioterapia (y radioterapia, en el caso del cáncer de recto). La quimioterapia del CCR se basa predominantemente en fluoropirimidinas (5-fluorouracilo - 5-FU - o su profármaco capecitabina), que se administran solas o en combinación con otros fármacos, como el oxaliplatino o el irinotecán. El diagnóstico precoz y los grandes avances en los tratamientos quimioterapéuticos han permitido aumentar la supervivencia de los pacientes con CCR, pero también son la principal causa de la aparición de reacciones adversas o tóxicas que pueden afectar a la eficacia del tratamiento y a la calidad de vida del paciente. En este contexto, surgen la farmacogenética y la farmacogenómica como campos de estudio de la variación genética hereditaria responsable de la intervariabilidad observada en la respuesta al tratamiento y la toxicidad. Esto es crucial para ayudar a optimizar los resultados del tratamiento, ofreciendo estrategias terapéuticas más personalizadas basadas en información molecular.

El principal objetivo de nuestro estudio es la identificación y validación de nuevos biomarcadores germinales de toxicidad a los tratamientos quimioterapéuticos usados en CCR, utilizando enfoques tecnológicos novedosos. Para la primera parte del estudio, nos propusimos identificar nuevas variantes/genes de toxicidad mediante la secuenciación de exomas completos (WES). Para ello, seleccionamos 163 casos y 52 controles de entre pacientes con CCR que habían recibido tratamiento con fluoropirimidinas, oxaliplatino e irinotecán y presentaban fenotipos de toxicidad extrema. En segundo lugar, seguimos un abordaje más amplio, en el que nos centramos en la variación común utilizando enfoques de asociación de genoma completo y post-GWAS. Dado que el objetivo final de la farmacogenómica es la implementación prospectiva de biomarcadores genéticos en las guías clínicas para el tratamiento preventivo, realizamos también la evaluación funcional de una nueva variante candidata en el gen *DPYD* (rs202212118).

En nuestros datos WES, detectamos 13 casos portadores de variantes accionables en *DPYD*. Además, encontramos 31 nuevas variantes raras de alto impacto en 14 de los genes descritos. También identificamos siete pacientes portadores de más de una variante en el mismo gen o vía, con un paciente que apuntaba a una posible herencia digénica. Entre estos siete, aparecía un paciente portador de una nueva variante *DPYD* - c.2071G>T, p.(V691L), rs202212118. Sin embargo, los ensayos funcionales realizados no pudieron confirmar la reducción de la abundancia de proteínas o de la actividad enzimática derivada de este cambio. Utilizando un enfoque de exoma completo, descubrimos y validamos de forma independiente tres nuevos genes de toxicidad candidatos (*ALDH9A1*, *FAM83A* y *EPX*). Los análisis basados en genes también proporcionaron 14 genes significativamente asociados con neuropatía, toxicidad cutánea y cardiotoxicidad. En los análisis de modelos poligénicos, no encontramos SNPs significativos en nuestros análisis GWAS, pero los enfoques TWAS identificaron 20 nuevos loci genómicos que están influyendo en el riesgo de toxicidad a través de la regulación de la expresión de 24 genes candidatos.

En general, esta tesis ha utilizado enfoques modernos para descubrir varias variantes/genes candidatos de toxicidad. No obstante, estos resultados requieren una replicación en cohortes

más amplias y una validación funcional. En última instancia, esperamos que mediante la identificación y validación de estas variantes/genes como candidatos de la toxicidad relacionada con la quimioterapia, se puedan realizar ajustes preventivos al tratamiento para evitar toxicidades extremas y, al final, mejorar la calidad de vida de los pacientes con CCR.

Resumo

O cancro colorrectal (CCR) é o terceiro cancro máis diagnosticado e a segunda causa de morte por cancro. Espérase que esta carga aumente máis dun 60%, cun estimado de 3,2 millóns de novos casos e 1,6 millóns de mortes ao ano ata 2040. O CCR representa un problema sanitario e económico considerable, e prevese que a súa incidencia aumente notablemente nos próximos anos, especialmente en idades máis novas. Crese que isto débese a unha serie de cambios ambientais, incluíndo estilos de vida máis sedentarios, maior obesidade, aumento da inxestión de alimentos procesados, alcohol e carne e unha maior lonxevidade xeral. A etioloxía do CCR é complexa, e o seu desenvolvemento está definido por factores xenéticos e ambientais. A susceptibilidade xenética ten un papel importante no desenvolvemento de novos casos de CCR. Cerca do 5% dos CRC considéranse hereditarios e seguen un patrón de herdanza mendeliana. Non obstante, a maioría dos casos de CCR (cerca do 80%) son esporádicos.

A determinación do estadio do tumor é a característica chave que determina o protocolo de tratamento. A localización (colon *vs* recto) tamén é un factor importante na elección do tratamento, xa que hai implicacións relevantes sobre a etioloxía e a anatomía que afectan tanto aos enfoques cirúrxicos como aos resultados do tratamento. De feito, para o cancro de recto, o tratamento varía parcialmente do do colon. A resección cirúrxica é o tratamento de base para o CCR nos estadios I-III, pero a quimioterapia tamén se usa habitualmente en todos os estadios, xa sexa antes (quimioterapia neoadxuvante) e/ou despois da cirurxía (quimioterapia adxuvante). Existen diferentes enfoques para o tratamento do CCR, incluíndo cirurxía, quimioterapia, radioterapia (no caso do cancro de recto), terapia dirixida e inmunoterapia. A quimioterapia do CCR baséase predominantemente en fluoropirimidinas (5-fluorouracilo - 5-FU - ou o seu profármaco oral capecitabina), que se administran sós ou en combinación con outros fármacos citotóxicos, como oxaliplatino ou irinotecán. A quimioterapia non está indicada para os estadios 0-I do CCR. Para o estadio II do cancro de colon, recoméndase o tratamento da quimioterapia con fluoropirimidinas en monoterapia en pacientes con risco intermedio ou alto de recidiva. Tamén podese considerar a adición de oxaliplatino para os pacientes de alto risco. Para o estadio III, a quimioterapia adxuvante adoita administrarse como unha combinación de fluoropirimidinas e oxaliplatino. Para o cancro de colon en estadio IV, utilízase amplamente a quimioterapia paliativa, que consiste en fluoropirimidinas, soas ou, na maioría dos casos, en combinación con oxaliplatino ou irinotecán. O tratamento nesta fase compleméntase xeralmente con anticorpos monoclonais (mAbs): anti-VEGF (bevacizumab) ou anti-EGFR (cetuximab, panitumumab). Dado que se demostrou que a inestabilidade dos microsatélites (MSI) é un bo biomarcador preditivo dun aumento da taxa de resposta á inmunoterapia, os pacientes con CCR en estadio IV con MSI elevada tamén son elixibles para a inmunoterapia con pembrolizumab, un inhibidor do punto de control inmunitario. No caso do cancro de recto, o estándar de atención para os estadios II-III adoita ser o tratamento neoadxuvante con fluoropirimidinas e radiación concomitante (quimiorradiación), seguido da resección cirúrxica e despois quimioterapia adxuvante baseada en fluoropirimidinas con ou sen oxaliplatino. No caso do cancro de recto en estadio IV, o tratamento preferible é a radiación preoperatoria seguida de fluoropirimidinas máis oxaliplatino e cirurxía, ou alternativamente, quimiorradiación e cirurxía.

As fluoropirimidinas son base da quimioterapia do CCR. A súa combinación con outros fármacos, como oxaliplatino e irinotecán, contribúe a maiores taxas de resposta, unha maior supervivencia sen progresión e supervivencia global. As fluoropirimidinas son análogos

sintéticos das pirimidinas, un substrato metabólico natural da célula, e úsanse desde hai máis de medio século para o tratamento de varios cancros. Os tratamentos de fluoropirimidinas inclúen principalmente 5-FU e capecitabina. O 5-FU diríxese e inhibe a timidilato sintase (TS), un encima que xoga un papel fundamental no ciclo do folato e na xeración de purinas e pirimidinas necesarias para a síntese de ADN e ARN. Os metabolitos do 5-FU incorpóranse ao ADN e ao ARN, desestabilizándoos e inhibindo a síntese de proteínas. A maior parte do 5-FU é catabolizado no fígado polo encima dihidropirimidina deshidroxenase (DPD), codificado polo xene *DPYD*. O 5-FU adoita combinarse con leucovorina (LV), un ácido folínico que actúa como un modulador de fármacos potenciando a conexión do 5-FU co TS e atrasando a súa excreción. A capecitabina é un profármaco oral que se metaboliza en 5-FU no fígado. Úsase habitualmente como alternativa pola súa actividade antitumoral aumentada.

Os axentes a base de platino son unha das clases máis grandes de fármacos utilizados no tratamento do cancro. Son axentes alquilantes e citotóxicos que se unen ao ADN, creando inter e intra aductos que perturban a replicación e a transcrición do ADN. Por fin, provocan apoptose e dificultan a proliferación do tumor. Hoxe en día, hai unha multitude de diferentes axentes baseados en platino utilizados na terapia do cancro, incluíndo cisplatino, carboplatino e oxaliplatino. Este último está indicado para o tratamento de varios tumores sólidos, especialmente cancros gastrointestinais, onde mostrou un mellor perfil de seguridade, así como unha maior eficacia que outros fármacos de platino.

O irinotecán é un análogo semisintético e soluble do alcaloide natural camptotecina. É un axente citotóxico e un profármaco do SN-38, que é un metabolito moi activo que se une á ADN-topoisomerase I, evitando o selado do ADN. Isto conduce finalmente á inhibición da replicación e á morte celular. Úsase para o tratamento de moitos cancros, incluíndo CCR metastásico, cancro de pulmón, páncreas, gástrico, ovario e cervical.

As fluoropirimidinas, o oxaliplatino e o irinotecán son axentes citotóxicos, o que significa que matan as células en función de certas propiedades biolóxicas, como as taxas de crecemento e proliferación celular. Estes procesos están aumentados nas células tumorais, o que as fai máis susceptibles a estes fármacos. Non obstante, as células normais tamén poden verse afectadas, e isto pode producir reaccións adversas aos medicamentos. Así, desenvolvéronse axentes dirixidos co obxectivo de matar selectivamente as células tumorais, co que se asocian con menos efectos secundarios. Cetuximab, panitumumab e bevacizumab son os axentes dirixidos máis utilizados na terapia dirixida ao CCR.

Dado que tanto as células tumorais como as normais poden metabolizar os fármacos quimioterapéuticos, a administración de quimioterapia adoita asociarse con reaccións adversas ou tóxicas, especialmente para axentes citotóxicos. Por exemplo, o 94% dos pacientes tratados con 5-FU acaban desenvolvendo reaccións adversas, incluíndo neutropenia, náuseas, e diarreas graves. A capecitabina causa menos toxicidade, aínda que o desenvolvemento da síndrome man-pé e a hiperbilirrubinemia asócianse frecuentemente con este fármaco. Para o oxaliplatino, a neuropatía periférica ocorre en > 90% dos pacientes. Na maioría dos casos, esta toxicidade aguda pódese reverter se se detén o tratamento, pero hai informes de neuropatía crónica causada pola acumulación de compostos de platino nos nervios sensoriais. O irinotecán está ligado a múltiples reaccións adversas, notoriamente diarrea, neutropenia e alopecia. Aínda que nunha proporción reducida en comparación cos

axentes citotóxicos, os anticorpos monoclonais tamén están asociados con reaccións adversas. É probable que os pacientes que reciben cetuximab (>25%) desenvolvan reaccións cutáneas, dor de cabeza, diarrea ou infección, mentres que os pacientes que reciben panitumumab (>20%) poden ter reaccións cutáneas, fatiga, náuseas ou diarrea. Máis do 60% dos pacientes que reciben bevacizumab desenvolverán reaccións adversas, sendo as máis frecuentes hipertensión, proteinuria, hemorraxia das mucosas e impedimento da cicatrización de feridas. Os pacientes que reciben pembrolizumab (>20%) poden ter fatiga, dor musculoesquelética, diminución do apetito, prurito, diarrea, náuseas, erupción cutánea, pirexia, tose, dispnea, estreñimiento e dor abdominal.

Durante moitos anos, houbo a hipótese de que a toxicidade (así como a resposta) aos tratamentos contra o cancro ten un compoñente xenético. A alta incidencia de reaccións adversas fai que sexa de máxima importancia o estudo desta asociación, e é neste contexto no que xurdiu a farmacoxenética hai máis de 50 anos. A farmacoxenética é unha ciencia que se centra no estudo da intervariabilidade hereditaria e no seu impacto tanto no desenvolvemento de reaccións adversas como na resposta ao tratamento. Dado que a maioría dos fármacos utilizados no tratamento do cancro teñen unha ventá terapéutica estreita (é dicir, hai unha pequena marxe para unha terapia eficaz sen reaccións adversas inaceptables), o papel da farmacoxenética é esencial para determinar a idoneidade do tratamento e dosificación para cada paciente. Non obstante, hai unha diferenza fundamental entre a resposta e a toxicidade no que se refire á farmacoxenética. A toxicidade é o resultado dos efectos non desexados da terapia sobre as células normais e, polo tanto, dependería da variación xenómica herdada (xerminal), mentres que a resposta depende da xenómica do propio tumor (somática) e de como os axentes terapéuticos son capaces ou non de matar as células cancerosas.

A arquitectura xenética do desenvolvemento de reaccións adversas aínda está na súa maioría sen descubrir. Inicialmente, considerouse que a susceptibilidade á toxicidade podía estar representada por variantes comúns de alta penetrancia con grandes tamaños de efecto (odds ratio - OR). A razón detrás disto foi que dada a novidade no uso de fármacos contra o cancro no que se refire á historia evolutiva, a selección negativa de variantes xenéticas de susceptibilidade á toxicidade grave non podería ter lugar. Así, os estudos farmacoxenéticos de primeira xeración utilizaban tamaños de estudo moderados e baseáronse no estudo de xenes candidatos que tiñan roles *a priori* relevantes para o proceso en estudo. Dentro da farmacoxenética, estes centráronse en xenes con influencia directa na farmacocinética e farmacodinámica dos fármacos: os farmacoxenes. Nos últimos 20 anos, propuxéronse máis de 200 xenes como farmacoxenes. Estes inclúen principalmente xenes implicados nos procesos ADME, como encimas metabolizadores (DPD, GST e UGT) ou transportadores responsables da entrada e saída do fármaco e os seus metabolitos (familias de proteínas SLC e ABC). Non obstante, os estudos farmacoxenéticos de xenes candidatos non deron evidencia suficiente para vinculalos de forma inequívoca ao desenvolvemento da toxicidade, con só dous xenes - *DPYD* e *UGT1A1* - incluídos ata agora nas guías farmacoxenéticas actuais para o tratamento do CCR.

Aínda que a contribución da xenética ao desenvolvemento das reaccións adversas é irrefutable, as variantes nos farmacoxenes descritas ata agora só foron capaces de explicar un número reducido de casos, e boa parte da variabilidade xenética segue aínda sen explicar. Isto

pode deberse en parte ao feito de que a suposición de que todas as variantes farmacoxenéticas teñen unha alta penetrancia probablemente sexa incorrecta e que a farmacoxenética da toxicidade pode seguir unha herdanza polixénica complexa. A hipótese da "Enfermidade común, variante común" defende que os trazos comúns complexos, como a toxicidade, poden explicarse polo aumento do risco global conferido pola combinación de múltiples variantes xenéticas comúns (frecuencia de alelos menores - MAF≥5%) con baixa penetrancia. Coa finalización do Proxecto Xenoma Humano e do Proxecto HapMap a principios do século XXI, e a expansión acoplada das tecnoloxías de xenotipado e secuenciación de alto rendemento, a farmacoxenética expandiuse rapidamente a todo o xenoma e converteuse en farmacoxenómica.

Usando estes avances, podemos identificar as variantes previstas pola hipótese anterior mediante a realización de estudos de asociación de xenoma completo (GWAS). Os GWAS confían no feito de que hai segmentos do xenoma que se transmiten como un todo: bloques haplotípicos, e a súa tendencia a permanecer xuntos ao longo das xeracións: desequilibrio de enlace (LD). O proxecto HapMap identificou millóns de marcadores (polimorfismos de nucleótido único - SNPs) en todo o xenoma, a súa frecuencia alélica e as correlacións entre eles, xerando o mapa haplotípico completo do xenoma humano. Permitiu a transición do xenotipado de cada SNP de interese a só un pequeno número de SNP representante dos que se pode inferir o xenotipo doutros marcadores por imputación, pódense deseñar matrices de SNP representantes baseados en LD, que permiten investigar a variación SNP no xenoma con respecto a un trazo de interese. Con todo, os GWAS tiveron un modesto éxito ata agora na identificación das variantes xenéticas responsables da toxicidade.

Nos últimos anos, as estratexias de matriz de SNP foron cada vez máis superadas pola secuenciación de próxima xeración (NGS), incluíndo a secuenciación do exoma completo (WES) ou a secuenciación do xenoma completo (WGS). NGS permite unha inspección máis completa das variacións comúns e tamén raras que non serian detectadas por métodos convencionais.

Ademais, a integración de datos transcriptómicos con xenómicos (estudos de asociación a nivel de transcriptoma - TWAS) xurdiu recentemente para mellorar o descubrimento de novos loci de risco. Estes estiman a asociación entre os niveis de expresión preditos xeneticamente e o trazo de interese. Os conxuntos de datos de referencia con variación xenómica e perfís transcriptómicos coincidentes (é dicir, niveis de expresión xénica) utilízanse para construír modelos preditivos de expresión para cada xene. Os TWAS colapsan as variantes xenómicas en rexións (xenes) máis grandes e, polo tanto, estes enfoques reducen a carga de testes múltiples e aumentan a potencia, como é o caso dos estudos farmacoxenómicos.

A pesar de la farmacoxenéticas e farmacoxenómica ser campos de interese para os investigadores dende hai moitos anos, as guías farmacoxenéticas dispoñibles actualmente para o tratamento con fluoropirimidinas no ámbito clínico son aínda demasiado escasas, e só catro variantes de *DPYD* (rs3918290, rs55886062, rs67376798 e rs56038477/HapB3) teñen unha indicación para probas farmacoxenéticas antes do tratamento. A farmacoxenómica da toxicidade xoga un papel fundamental para conseguir un tratamento mellor e máis personalizado. Non obstante, os esforzos realizados ata agora víronse limitados pola falta de

datos homoxeneizados, o complicado e heteroxéneo fenotipado das reaccións adversas, a necesidade de novos enfoques e a necesidade de avaliar a funcionalidade.

Polo tanto, o obxectivo principal do noso estudo é a identificación e validación de novos biomarcadores xerminais de toxicidade para tratamentos quimioterapéuticos máis comunmente usados en CCR, utilizando enfoques tecnolóxicos innovadores. Este obxectivo principal pódese dividir en diferentes subobxectivos específicos, cada un dos cales corresponde a un capítulo desta tese:

- Identificación de novas variantes de toxicidade de penetrancia alta y moderada mediante a análise de secuenciación do exoma completo nunha cohorte de 163 casos e 52 controis con fenotipos de toxicidade extrema (capítulo 1).

- Análises de susceptibilidade á toxicidade a nivel ómico para determinar variantes de toxicidade de baixa penetrancia baixo un modelo de risco polixénico (capítulo 2).

- Avaliación funcional das variantes de toxicidade candidatas, mediante ensaios encimáticos e cuantificación da abundancia de proteínas (capítulo 3).

Para a primeira parte do estudo, propuxemos identificar novas variantes/xenes de toxicidade mediante a secuenciación de exomas completos (WES). Para iso, recrutamos pacientes que recibiran tratamento de primeira liña para o CCR con réximes baseados en fluoropirimidinas (5-FU ou capecitabina, só ou en combinación con oxaliplatino ou irinotecán) co fenotipos de toxicidades extremos. En total, reunimos 163 casos e 52 controis, dos cales realizamos a secuenciación do exoma completo (WES). Na análise dos datos, en primeiro lugar, seguimos un enfoque de xenes candidatos, onde fixemos unha ampla revisión da literatura e revisamos os nosos datos para as variantes de toxicidade xa descritas e as novas variantes nos xenes respectivos. En segundo lugar, ampliamos o noso enfoque nunha análise máis completa, onde buscamos variantes e xenes novos sen hipótese *a priori*, examinando o exoma completo, centrándonos na variación rara. Esta estratexia facilita unha mirada non tendenciosa aos datos e permite o descubrimento de novos xenes que contribúen ao desenvolvemento de reaccións adversas.

Na segunda parte do estudo, tamén seguimos un enfoque máis amplo, onde nos centramos na variación común utilizando estudos de asociación de xenoma completo (GWAS) e post-GWAS. Inicialmente, para os datos WES, realizamos a clásica proba de asociación (Chi cadrado) para as toxicidades agrupadas (globais, dixestivas e hematolóxicas), para os grupos de toxicidade con N>20 (diarrea, neutropenia, astenia e náuseas) e para a síndrome man-pé. Realizamos unha metaanálise cos datos WES e a cohorte adicional EGAS00001002763. Propúxose a metaanálise para aumentar o poder para detectar asociacións para estudos de asociación, en oposición á replicación, particularmente para tamaños de mostra máis baixos. Ademais, complementamos as análises GWAS cun estudo adicional de asociación a nivel transcriptómico para aumentar o noso poder para detectar novos loci de toxicidade comúns.

Dado que o obxectivo final da farmacoxenómica é a implantación prospectiva de biomarcadores xenéticos nas guías clínicas de tratamento preventivo, na terceira parte do

estudio, tamén realizamos a avaliación funcional de unha nova variante candidata no xene *DPYD* (rs202212118).

Nos nosos datos WES, detectamos 13 casos con variantes accionables en *DPYD*. Non puidemos replicar ningunha asociación descrita previamente na literatura, pero en cambio atopamos asociacións significativas con outras reaccións adversas: *ABCB1* rs2032582 e toxicidade cutánea, *SLCO1B1* rs2306283 e neuropatía. Ademais, atopamos 31 novas variantes raras de alto impacto (previsto por polo menos tres de catro preditores *in silico*) en 14 dos xenes descritos. Tamén identificamos sete pacientes portadores de máis dunha variante no mesmo xene ou vía. Por exemplo, un paciente tiña unha variante en *TYMS* e outra en *ENOSF1*, que apuntou a unha posible herdanza dixénica. y que sería interesante investigar máis mediante estudos *in vitro*. Entre estes sete, tamén había un paciente cunha nova variante de *DPYD* - c.2071G>T, p.(V691L), rs202212118, que posteriormente foi analizada funcionalmente. Usando un enfoque do exoma completo, descubrimos e validamos de forma independente tres novos xenes candidatos de toxicidade (*ALDH9A1*, *FAM83A* e *EPX*). As análises baseadas en xenes tamén proporcionaron 14 xenes significativamente asociados coa neuropatía, a toxicidade cutánea y a cardiotoxicidade. Estes resultados precisan validación en cohortes máis grandes e estudos *in vitro* para comprender as consecuencias funcionais destas variantes e o seu papel no desenvolvemento da toxicidade. En conxunto, os resultados das análises de variantes de alta penetrancia poderían explicar potencialmente a toxicidade observada para 58 dos pacientes da nosa cohorte de fenotipos extremos.

Para a análise de modelos polixénicos, non atopamos SNP significativos na nosa análise GWAS. Isto podería estar xustificado polo tamaño reducido da mostra e a baixa potencia para cada reacción adversa independente, destacando a necesidade de cohortes máis grandes. A pesar diso, identificamos unha variante case significativa no GWAS de toxicidade global, a *ITGAX* rs2230424, aínda que non está claro como a variación deste xene podería levar a un aumento da toxicidade. Con todo, a integración de datos de xenómicos coa transcriptómica no enfoque TWAS identificou 20 novos loci xenómicos que antes non estaban relacionados coa toxicidade. Estas rexións están a influír no risco de toxicidade a través da regulación da expresión de 24 xenes candidatos e merecen máis estudos. Máis aló, cinco dos farmacoxenes descritos na literatura - *DPYD*, *TYMP*, *ACYP2*, *UMPS* e *UGT1A1* - tamén foron nominalmente significativos nesta análise, validando así este enfoque.

A caracterización funcional da nova variante *DPYD* - c.2071G>T, p.(V691L), rs202212118, incluíndo a cuantificación da abundancia de proteínas e os ensaios de actividade encimática, non puideron clasificar de forma inequívoca esta variante como patóxena, aínda que a capacidade de metabolismo observada foi do 40% .

En xeral, esta tese utilizou enfoques innovadores para descubrir varias variantes/xenes candidatos de toxicidade. Non obstante, estes resultados requiren replicación en cohortes máis grandes e validación funcional. En definitiva, esperamos que, identificando e validando estas variantes/xenes como candidatos á toxicidade relacionada coa quimioterapia, se poidan facer axustes preventivos ao tratamento para evitar toxicidades extremas e, ao final, mellorar a calidade de vida dos pacientes con CCR.

Introduction

Colorectal cancer

Colorectal cancer (CRC) constitutes a major health problem. It is the 3rd most common tumour in terms of incidence, with over 1.9 million new cases worldwide in 2020, and ranking 2nd in terms of mortality. The distribution of CRC cases worldwide is heterogeneous, with a 4-fold higher incidence in developed countries, but with higher mortality in developing countries[1]. This burden of CRC is expected to increase by more than 60%, with an estimated 3.2 million new cases and 1.6 million deaths per year by 2040[2]. The median age at diagnosis for CRC is currently 67 years, with around 93% of the cases occurring in patients over 45 years, and more than 55% in patients 65 years or older[3]. The reported 5-year survival rates are 99% for stage I, 68–83% for stage II, 45–65% for stage III, and 20-45% for patients with metastasis[4,5]. Interestingly, CRC incidence and mortality in individuals over 50 years have been decreasing, while there has been a notable increase in younger cases[3]. This is believed to be due to a series of environmental changes, including more sedentary lifestyles, greater obesity, increased intake of processed foods, alcohol, and meat, and greater overall longevity[6]. Most CRCs (90-95%) are bowel adenocarcinomas and therefore originate in the bowel epithelium.

CRC aetiology is complex, with its development defined by both genetic, and environmental factors[7]. Genetic susceptibility has a significant role in the development of new cases. Around 5% of CRCs are considered hereditary and follow a Mendelian inheritance pattern. However, most CRC cases (around 80%) are sporadic. The first genetic model of CRC tumorigenesis, known as the adenoma-carcinoma sequence, was proposed by Fearon and Vogelstein in 1990, and involves a series of genetic alterations that lead to chromosomal instability (CIN) (Figure 1)[8]. Later contributions have elucidated that microsatellite instability (MSI) and the CpG island methylator phenotype (CIMP) are also relevant pathways of CRC tumourigenesis (Figure 1)[9,10].

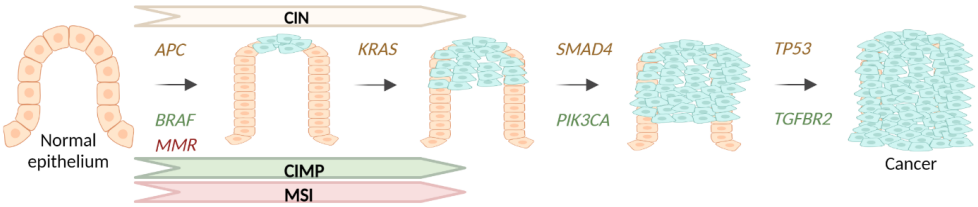


Figure 1. Adenoma–carcinoma sequence model (adapted from Walther et al.)[11]. The classic CIN pathway, responsible for 65-85% of CRC, begins with inactivating mutations in the *APC* tumour suppressor gene. This is the mechanism driving cellular proliferation and formation of early small adenomas, resulting in aneuploidies and a high frequency of loss of heterozygosity. *APC* changes are followed by activating mutations in the *KRAS* oncogene, which results in continuous mitotic signalling and arrest of apoptosis, enabling further growth. Later on, *SMAD4*, *TP53*, and other tumour suppressor genes are deregulated, ultimately empowering the formation of the invasive carcinoma[9,12]. On the other hand, around 15% of CRC cases present microsatellite instability (MSI), which is the result of length variations in highly repetitive regions of the genome (microsatellites). During homeostasis, these variation errors are repaired by the mismatch repair (MMR) genes: *MLH1*, *MSH2*, *MSH6* and *PMS2*. However, the inactivation of the MMR genes (deficient, dMMR) results in an uncontrolled variation of the number of repetitions, leading to MSI and further accumulation of pathogenic variants in important cancer-associated genes[13]. A third pathway driving CRC is the CpG island methylator phenotype (CIMP). CpG islands are genomic regions with a high amount of CpG dinucleotide repeats, usually at gene promoters. The CIMP pathway involves promoter hypermethylation of multiple tumour suppressor genes, leading to their silencing and lower gene expression[14]. *MLH1* is most affected by CIMP, and thus the CIMP phenotype is often associated with MSI[9,10].

Colorectal cancer treatment

The standardised mechanism to quantify cancer and its spread is by staging. Tumour staging provides information about the degree of neoplastic development, allowing the planning of the best treatment strategy, and providing an indication of prognosis and survival. The most widely used staging system comprises a classification where stage ranges from 0 to IV, with 0 being an *in-situ* carcinoma and IV being metastatic disease. CRC may also be classically classified using Dukes' staging, which is a system based on the tumour resectability, measured as the depth of invasion through the mucosa and bowel wall. This was first published in the 1930s for rectal cancer only and is divided into categories A - limited to mucosa, B - invasion through the bowel wall, C - involvement of lymph nodes, and D - widespread metastases. Alternatively, the Tumour-Node-Metastasis (TNM) classification is also used (and even preferred) for the staging of CRC and many other cancers. In this system, T refers to the size and extent of the tumour, N refers to implication of regional lymph nodes, and M refers to the presence of metastases (Table 1)[15].

Tumour staging is the key feature that determines treatment protocol. Surgical resection is the baseline treatment for CRC at stages I-III, but chemotherapy is also commonly used across stages, either before (neoadjuvant) and/or after surgery (adjuvant chemotherapy). There are different approaches to CRC treatment, including surgery, chemotherapy, radiotherapy, targeted therapy, and immunotherapy (Table 2). Location (colon *vs* rectum) is also an important factor in the choice of treatment, as there are relevant implications regarding aetiology and anatomy that affect both surgical approaches and treatment outcomes. For rectal cancer, treatment partially varies from that of the colon, due to the bony constraints of the pelvis limiting surgical access to the rectum, which results in a higher risk of local recurrence[16].

The standard chemotherapy for CRC is based on **fluoropyrimidines**. Fluoropyrimidines can be administered alone or in combination with other cytotoxic drugs, the most important being platinum-based compounds (principally **oxaliplatin**) or **irinotecan**. Chemotherapy is administered on a regular schedule, usually weekly or every 2-4 weeks, where the period of treatment and the corresponding resting period is termed a *cycle*. A particular drug or combination of drugs administered during one or more cycles is referred to as a *line* of chemotherapy. If the first (or subsequent) line of chemotherapy does not yield sufficient results or the patient has intolerable side effects, then the regimen is changed, and another line of chemotherapy starts.

Chemotherapy is not indicated for CRC stages 0-I. For **colon cancer** stage II, chemotherapy treatment with fluoropyrimidines in monotherapy is recommended only in patients with intermediate- or high-risk of recurrence. Addition of oxaliplatin might be considered for the high-risk patients as well. For stage III, adjuvant chemotherapy is given usually as a combination of fluoropyrimidines and oxaliplatin[4]. For stage IV colon cancer, palliative chemotherapy is widely used, consisting of fluoropyrimidines, alone or, in most cases, in combination with oxaliplatin/irinotecan[5]. The treatment at this stage is generally complemented with monoclonal antibodies (mAbs): anti-VEGF (bevacizumab) or anti-EGFR (cetuximab, panitumumab). Because MSI has been shown to be a good predictive biomarker of an increased response rate to immunotherapy, stage IV CRC patients with MSI-High are as

well eligible for immunotherapy with pembrolizumab, an immune checkpoint inhibitor (Table 2)[5,17]. In the case of **rectal cancer**, the standard of care for stages II-III is usually neoadjuvant treatment with fluoropyrimidines and concomitant radiation (chemoradiation), followed by surgical resection, and then adjuvant chemotherapy based on fluoropyrimidines with or without oxaliplatin. In the case of stage IV rectal cancer, the preferable treatment is preoperative radiation followed by fluoropyrimidines plus oxaliplatin and surgery, or alternatively, chemoradiation and surgery (Table 2)[16].

Table 1. Staging correspondence between the different classification systems (adapted from Labianca et al.)[18].

Stage	Definition[19]	T	N	M	Dukes	Location
0	Carcinoma *in situ*: malignant tumour confined to the mucosa, and with no invasion of the submucosa	Tis	N0	M0	-	localised
I	Tumour invades the submucosa or the muscularis propria	T1, T2	N0	M0	A	
IIa	Tumour invades through the muscularis propria into the subserosa or into neighbouring tissues in the intraperitoneal space	T3	N0	M0	B	
IIb	Tumour penetrates the visceral peritoneum and/or directly invades organs or structures in the intraperitoneal space	T4	N0	M0		
IIIa	Tumour invades the submucosa or muscularis propria and has spread to 1-3 regional lymph nodes	T1, T2	N1	M0	C	
		T1	N2a			
IIIb	Tumour invades the subserosa, visceral peritoneum or neighbouring organs, and has spread to 1-3 regional lymph nodes	T1, T2	N2b	M0		
		T2, T3	N2a			
		T3, T4a	N1			
IIIc	Tumour, irrespective of the degree of local invasion, has spread to 4 or more regional lymph nodes	T3, T4a	N2b	M0		
		T4a	N2a			
		T4b	N1, N2			
IV	Tumour has spread to distant organs, irrespective of the degree of local invasion and/or spread to regional lymph nodes	Any T	Any N	M1	D	metastatic

Fluoropyrimidines

Fluoropyrimidines are the backbone of CRC chemotherapy. Their combination with other drugs, like oxaliplatin and irinotecan, contributes to higher response rates, longer progression-free and overall survival[5]. Fluoropyrimidines are synthetic analogues of pyrimidines, a natural metabolic substrate of the cell, and have now been used for over half a century for the treatment of several cancers[20]. CRC fluoropyrimidine-based treatments include majorly **5-FU** and **capecitabine**. 5-FU targets and inhibits thymidylate synthase (TS), an enzyme that plays a key role in the folate cycle and in the generation of both purines and pyrimidines needed for DNA and RNA synthesis. The metabolites of 5-FU are incorporated into DNA and RNA, destabilising them and inhibiting protein synthesis. The majority of 5-FU is catabolised in the liver by the dihydropyrimidine dehydrogenase (DPD) enzyme, encoded by the *DPYD* gene[21]. 5-FU is usually combined with leucovorin (LV), a folinic acid that acts like a drug modulator by enhancing the link of 5-FU with the TS and delaying its excretion. Capecitabine is an oral prodrug that is metabolised into 5-FU in the liver. It is commonly used as an alternative because of its increased antitumour activity (Figure 2)[22,23].

Oxaliplatin

Platinum-based agents are one of the largest classes of drugs used in cancer treatment. They are alkylating and cytotoxic agents that bind to DNA, creating inter and intrastrand adducts that disturb DNA replication and transcription. Ultimately, they cause apoptosis and hinder tumour proliferation (Figure 3)[24,25]. Nowadays, there are a multitude of different platinum-based agents used in cancer therapy, including cisplatin, carboplatin, and oxaliplatin. **Oxaliplatin** was the first platinum-based anticancer drug used for CRC treatment. It was first approved in 1996 in France as a second-line chemotherapy for advanced CRC[26]. In 2002, it was approved by the United States Food and Drug Administration (FDA), as an adjuvant treatment for stage III colon cancer in patients with complete resection of the primary tumour, or for stage IV CRC[27]. Since then, it has been indicated for the treatment of several solid tumours, specially gastrointestinal cancers, where it has shown a better safety profile, as well as higher efficacy than other platinum drugs[28,29].

Irinotecan

Irinotecan (CPT-11) is a semi-synthetic, soluble analogue of the natural alkaloid camptothecin. It is a cytotoxic agent and a prodrug of SN-38, which is a very active metabolite that binds to the DNA-topoisomerase I, preventing DNA resealing. This ultimately leads to the inhibition of replication and cell death (Figure 4)[4]. Irinotecan was first approved in Japan in 1994 for the treatment of lung, cervical and ovarian cancers, and approved in Europe and in the United States in 1995 and 1996, respectively[30]. It is now used for the treatment of many cancers, including metastatic CRC, lung, pancreatic, gastric, ovarian, and cervical cancers.

Table 2. European Society for Medical Oncology (ESMO) guidelines for CRC first line of treatment[4,5,16]. ESMO is a reference organisation that has extensive guidelines regarding diagnosis, treatment, and follow-up for several cancers, including CRC.

Stage	Treatment	
	Surgery	**Pharmacological treatment**
0-I	Surgical resection	No chemotherapy recommended
II	Surgical resection	**Colon** tumours, no chemotherapy for low-risk; adjuvant 5-FU+LV/capecitabine for intermediate-risk; 5-FU/capecitabine ± oxaliplatin (FOLFOX/XELOX) for high-risk **Rectal** cancer, radiation ± neoadjuvant 5-FU+LV/capecitabine chemotherapy AND adjuvant 5-FU+LV/capecitabine ± oxaliplatin
III	Surgical resection	**Colon**: Adjuvant 5-FU+LV + oxaliplatin (FOLFOX) OR capecitabine + oxaliplatin (XELOX) **Rectal** cancer: radiation ± neoadjuvant 5-FU+LV/capecitabine chemotherapy AND adjuvant 5-FU+LV/capecitabine ± oxaliplatin
IV	Surgical removal of primary tumour, if resectable Colon cancer patients with resectable metastases, a favourable prognostic, and a good surgical approach, neoadjuvant systemic treatment may not be needed. Most patients have unresectable metastases. Revaluate after neoadjuvant chemotherapy Rectal cancer, neoadjuvant chemotherapy/chemoradiotherapy + surgery/neoadjuvant radiation + chemotherapy + surgery	5-FU+LV/capecitabine + oxaliplatin (FOLFOX/XELOX) OR 5-FU+LV + irinotecan (FOLFIRI) OR 5-FU+LV + oxaliplatin + irinotecan (FOLFOXIRI) Monoclonal antibodies: - targeted agents, given together with chemotherapy: anti-VEGF (bevacizumab), if *RAS* (*KRAS, NRAS*) and *BRAF* mutated; or anti-EGFR (cetuximab, panitumumab), if *RAS*-wt and *BRAF*-wt - immunotherapy: pembrolizumab

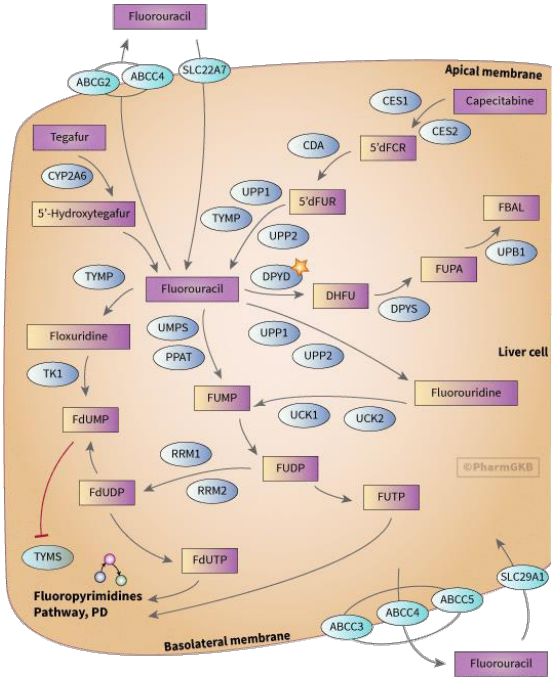


Figure 2. Pharmacokinetics and pharmacodynamics of fluoropyrimidines[21]. Pharmacokinetics refers to the delivery of a drug and/or its metabolites to the target molecules, and includes its absorption, distribution, metabolism, and excretion (ADME) mechanisms, while pharmacodynamics refers to the relationship between the drug concentration and its effect. Capecitabine is metabolised in the liver by carboxylesterases (CES) and cytidine deaminase (CDA) and is then activated into 5-FU by thymidine phosphorylase (TP). 5-FU inhibits the thymidylate synthase (TS, encoded by the *TYMS* gene), crucial for the folate cycle and purine and pyrimidine synthesis. Used with permission from PharmGKB, under a CC license, available at www.pharmgkb.org/pathway/PA150653776 (accessed on 10/02/2023).

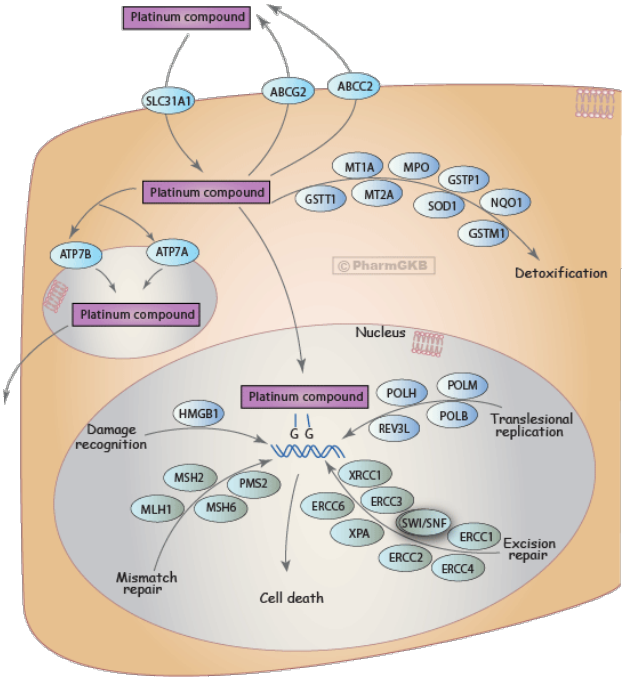


Figure 3. Pharmacokinetics and pharmacodynamics of oxaliplatin[31]. The absorption and excretion of oxaliplatin are mainly carried out by solute carrier proteins (SLCs) and adenosine-triphosphate binding cassette (ABC) transporters. Once inside the cell, the glutathione S-transferases (GSTs) can link actively to oxaliplatin and catalyse its detoxification through the conjugation of glutathione, to inactivate it and facilitate its excretion from the body. Used with permission from PharmGKB, under a CC license, available at www.pharmgkb.org/pathway/PA150642262 (accessed on 10/02/2023).

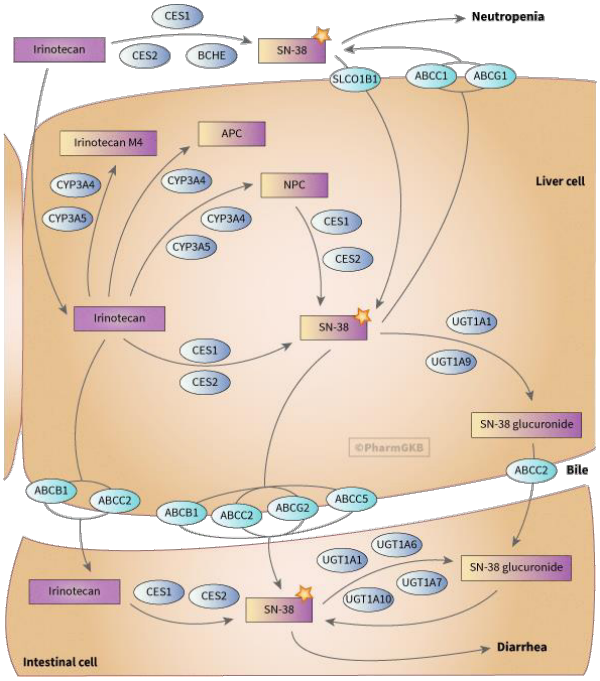


Figure 4. Pharmacokinetics and pharmacodynamics of irinotecan[24]. Irinotecan is hydrolysed in the liver into active SN-38 by CES1 and CES2 and transported in and out of the liver cell by ABC-proteins. SN-38 binds to topoisomerase I, an enzyme required in replication, unwinding of DNA and prevention of lethal strand breaks, its inhibition thereby causing cell death. SN-38 is glucoronised into SN-38G and detoxified by UDP glucuronosyl-transferases (UGTs). SN-38G is inactive, having 1/100 of SN-38´s antitumour activity. Irinotecan can also be oxidised by cytochrome P450 3A into the inactive molecules APC, M4, and NPC. NPC can be hydrolysed by CES into SN-38, increasing its levels in plasma. As for the platinum compounds, irinotecan and its metabolites are uptaken by SLCs and eliminated by ABC-transporters. Used with permission from PharmGKB, under a CC license, available at www.pharmgkb.org/pathway/PA2001 (accessed on 10/02/2023).

Fluoropyrimidines, oxaliplatin and irinotecan are all cytotoxic agents, meaning that they kill cells based on certain biological properties, like cell growth and proliferation rates. These processes are upregulated in tumour cells, making them more susceptible to these drugs. However, normal cells can also be affected, and this can result in adverse drug reactions (ADRs). Targeted agents have been developed with the aim of selectively targeting and killing tumour cells, thus being associated with less side effects.

Monoclonal antibodies

Monoclonal antibodies (mAbs) are laboratory-made proteins that bind to specific antigens on the cell's surface. Therefore, they can be designed to target specific proteins appearing specifically/predominantly in the surface of cancer cells. mAbs can be used to block abnormal growth signalling and the subsequent uncontrolled cell division, which makes them useful as anti-cancer targeted agents, particularly at later stages. Cetuximab, panitumumab, and bevacizumab are the most used mAbs in CRC targeted therapy[5]. Cetuximab and panitumumab target the epidermal growth factor receptor (EGFR), a protein that is constitutively expressed in normal epithelial tissues and overexpressed in some cancers, like CRC. Bevacizumab targets the vascular endothelial growth factor (VEGF), preventing the formation of new blood vessels and inhibiting metastatic progression[32]. Pembrolizumab is another mAb used for CRC immunotherapy[5]. It is an immune checkpoint inhibitor that binds to the programmed cell death protein 1 (PD-1) on the surface of cytotoxic T cells, thereby allowing their activation and the enhancing of the immune response, leading to the death of tumour cells[33].

Pharmacogenetics and Pharmacogenomics

Because both tumour and normal cells can metabolise chemotherapeutic drugs, the administration of chemotherapy is frequently associated with toxic side effects, or **adverse drug reactions (ADRs)**, particularly for cytotoxic agents (Table 3). For instance, 94% of patients treated with 5-FU end up developing ADRs, including severe neutropenia, nausea and diarrhoea[34]. Capecitabine causes less toxicity, yet development of hand-foot syndrome (HFS) and hyperbilirubinemia are frequently associated with this drug[23]. For oxaliplatin, peripheral neuropathy occurs in >90% of patients[35]. In most cases, this acute toxicity can be reversed if the treatment is stopped, but there are reports of chronic neuropathy caused by the accumulation of platinum compounds in the sensory nerves. Irinotecan is linked to multiple ADRs, notoriously diarrhoea, neutropenia, and alopecia[36]. Albeit in a reduced proportion compared to cytotoxic agents, monoclonal antibodies are also associated with ADRs. Patients receiving cetuximab are likely (>25%) to develop cutaneous reactions, headache, diarrhoea, or infection, while patients receiving panitumumab (>20%) may have cutaneous reactions, fatigue, nausea, or diarrhoea[37,38]. More than 60% of patients that receive bevacizumab will develop ADR, the most common being hypertension, proteinuria, mucosal bleeding, and hindered wound healing[5,39]. Patients receiving pembrolizumab (>20%) may have fatigue, musculoskeletal pain, decreased appetite, pruritus, diarrhoea, nausea, rash, pyrexia, cough, dyspnoea, constipation, and abdominal pain[40].

Table 3. The most common toxicity profile of CRC treatments.

Treatment	**Adverse Drug Reactions** (according to FDA labels)[35–39,41]
5-Fluororacil (>30%)	Diarrhoea, neutropenia, mucositis, nausea/vomiting, stomatitis, asthenia, leucopenia, anaemia.
Capecitabine (>30%)	Hand-and-foot syndrome, diarrhoea, nausea/vomiting, abdominal pain, fatigue, hyperbilirubinemia.
Oxaliplatin (>40%)	Peripheral sensory neuropathy, neutropenia, thrombocytopenia, anaemia, nausea/vomiting, increase in transaminases and alkaline phosphatase, diarrhoea, fatigue, stomatitis.
Irinotecan (>30%)	Nausea/vomiting, diarrhoea, neutropenia, alopecia, abdominal pain, constipation, anorexia, leucopenia, anaemia, asthenia, fever, body weight decreasing.
Cetuximab (>25%)	Cutaneous adverse reactions, headache, diarrhoea, infection.
Panitumumab (>20%)	Skin rash, paronychia, fatigue, nausea, diarrhoea.
Bevacizumab (>10%)	Haemorrhage, hypertension, headache, rhinitis, proteinuria, taste alteration, dry skin, lacrimation disorder, back pain, exfoliative dermatitis.

For many years, it has been hypothesised that toxicity (as well as response) to cancer treatments has a genetic component[42]. The high incidence of ADRs makes it therefore of the utmost importance to study this association, and it is in this context that pharmacogenetics emerged over 50 years ago[43]. Pharmacogenetics is a science that focuses on inherited inter-variability and its impact on both development of ADRs and response to treatment. Because most drugs used in cancer treatment have a narrow therapeutic window (i.e. there is a small margin for effective therapy without unacceptable adverse events), the role of pharmacogenetics is essential in determining the appropriateness of treatment and dosage for each patient. However, there is a fundamental difference between response and toxicity

with regard to pharmacogenetics. Toxicity is a result of the unwanted effects of therapy on the normal cells and would therefore depend on **inherited genomic variation**, whereas response is dependent on the genomics of the tumour itself and how the therapeutic agents are able or not to kill cancer cells. It has been estimated that genetic variants explain 26-65% of the heritability of 5-FU-associated toxicity, which reflects on the relevance of genetic variation in the development of ADRs[44]. In the case of response, certain somatic mutations found in the tumour have been correlated with response to chemotherapy (or absence thereof). For instance, patients with somatic *RAS* mutations have a significantly worse or even absent response to cetuximab, and testing for *RAS* mutations is already included in CRC pharmacogenetic guidelines (see below section: *Current pharmacogenetic guidelines for CRC treatment*)[45,46].

The genetic architecture of ADR development is still mostly undiscovered. Initially, it was considered that toxicity susceptibility could be represented by common, high-penetrance variants with large effect sizes (odds ratios - OR). The rationale behind this was that given the novelty in the use of anticancer drugs with regards to evolutionary history, negative selection of genetic variants of susceptibility to severe toxicity could not have possibly taken place. Thus, first-generation pharmacogenetic studies were based on approaches using moderate study sizes. These relied on the study of candidate genes that had roles *a priori* relevant to the process at study. For pharmacogenetics, these were thus focused on genes with a direct influence on drug pharmacokinetics and pharmacodynamics - *pharmacogenes* (Figures 2-4)[42,47]. In the past 20 years, more than 200 genes have been proposed as pharmacogenes. These include mainly genes involved in ADME processes, like metabolising enzymes (DPD, GSTs, and UGTs), or transporters responsible for the influx and efflux of the drug and its metabolites (SLC and ABC protein families). Nonetheless, candidate-gene pharmacogenetic studies have yielded insufficient evidence to link them unequivocally to the development of toxicity, with only 2 genes - *DPYD* and *UGT1A1* - so far included in current pharmacogenetic guidelines for CRC treatment.

DPYD

DPD (encoded by the *DPYD* gene) is the rate-limiting enzyme of 5-FU catabolism (Figure 2). Therefore, a reduction of DPD expression or activity could notably impact the clearance of fluoropyrimidines, and thus result in the accumulation of toxic metabolites in the liver, leading to ADR development. The first reported cases of DPD deficiency go back over 20 years, when it was described as an autosomal recessive disease characterised by an excess of thymine-uracil, and linked to adverse effects after treatment with 5-FU[48–50]. Since then, pathogenic changes in the *DPYD* gene have been identified. The most studied *DPYD* variant is the c.1905+1G>A (rs3918290, *DPYD**2A allele). This variant changes the splice donor site in intron 14 causing exon 14 skipping, and the omission of a 165-bp fragment from the mature mRNA[51]. The resulting enzyme is completely non-functional[52]. An additional three *DPYD* variants have also been established to be deleterious: c.2846A>T, p.(D949V) (rs67376798); c.1679T>G, p.(I560S) (rs55886062, *DPYD**13); and c.1236G>A, p.(E412=) (rs56038477; HapB3). The rs67376798 change affects the structure of the protein and interferes with cofactor binding and electron transport[52]. The enzymatic activity of DPD is impaired resulting in around 40% reduction of protein function when in homozygosity[53]. The rs55886062 G allele destabilises

the structural integrity of the DPD protein[54]. The enzymatic activity is significantly compromised, with up to 75% reduction when in homozygosity[55]. Lastly, the rs56038477 is in complete linkage disequilibrium with the intronic variants c.483+18G>A (rs56276561), c.680+139G>A (rs6668296), c.959-51T>G (rs115349832) and c.1129-5923C>G (rs75017182). The combination of these four intronic variants is known as haplotype B3 (HapB3)[73]. rs75017182 is thought to be the causal variant responsible for the effect on the enzymatic activity, due to aberrant splicing[52]. Data on the enzymatic activity is limited, but one report on two homozygous patients for rs56038477 showed a DPD activity of around 50%[52]. Altogether, the four most studied deleterious variants present a low population frequency - around 3% of people in Caucasian populations carry at least one of these changes (rs56038477 - 2%; rs3918290 - 0.6%; rs67376798 - 0.5%; rs55886062 - 0.06%) and can therefore only explain a small proportion of the toxicity events observed[56]. Various other variants have been described in *DPYD*. However, there is a lack of consensus on their functionality and clinical relevance[53,55].

UGT1A1

UGT1A1 is the major enzyme responsible for the inactivation of irinotecan's metabolite, SN-38. Its encoding gene has more than 100 variants reported to alter its enzymatic activity[57]. The two most studied variants in the *UGT1A1* gene are rs3064744 (*UGT1A1**28) and c.211G>A, p.(G71R) (rs4148323, *UGT1A1**6)[58–62]. The TATA box in the promoter region usually has six thymine–adenine (TA) repeats - $(TA)_6TAA$. The rs3064744 is a change in the number of TA repeats - $(TA)_7TAA$ (seven repeats). It rs3064744 is more frequent in Caucasians than Asians (30% vs 10%, respectively), whereas rs4148323 is more frequent in Asians (14% vs 1%, respectively).

Pharmacogenomic studies

Although the contribution of genetics to the development of ADRs is irrefutable, the variants reported so far in the pharmacogenes have only been able to explain a reduced number of the cases, and much of the genetic variability remains yet unexplained. This may in part be due to the fact that the assumption of pharmacogenomic variants being high penetrance is likely incorrect, and pharmacogenomics of toxicity is more likely to follow a complex polygenic inheritance. The "Common Disease, Common Variant" hypothesis defends that complex common traits, such as toxicity, can be explained by the increased overall risk conferred by the combination of multiple, common (minor allele frequency - MAF≥5%) genetic variants with low penetrance[63].

With the completion of the Human Genome Project and the International HapMap Project at the beginning of the 21st century, and the coupled expansion of high-throughput genotyping and sequencing technologies, pharmacogenetics rapidly expanded to the whole genome and became pharmacogenomics[64,65]. Using these advances, we can identify the variants predicted by the "Common Disease, Common Variant" by performing association studies at a genome-wide scale (**GWAS**). GWASs rely on the fact that across the genome, markers are not inherited independently, but there are segments of the genome that are transmitted as a whole. These are called *haplotype blocks* and their tendency to stay together along generations is known as

linkage disequilibrium (LD)[64]. The HapMap project identified millions of markers (single nucleotide polymorphisms - SNPs) throughout the genome, their allele frequency, and the correlations between them, generating the complete haplotypic map of the human genome. It allowed for the transition from genotyping every single SNP of interest to just a small number of proxy SNPs from which the genotype of multiple other markers can be inferred by *imputation*[66]. From this, SNP arrays can be designed for GWASs using proxies based on LD, which allow us to investigate the extended SNP variation of the genome regarding a trait of interest.

GWASs have improved our knowledge of toxicity risk pharmacogenomic biomarkers[67–70]. In the QUASAR2 trial, Rosmarin et al. analysed over 1,000 stage II/III CRC patients receiving capecitabine with or without bevacizumab[67]. They discovered that variant rs2612091 in the *ENOSF1* gene was in moderate LD with two widely studied SNPs in *TYMS*: rs45445694 and rs11280056 (r^2=0.40 and 0.32, respectively). After testing for dependency, they concluded that it was the rs2612091 G allele alone that increased the risk of toxicity (p=0.0021). They also reported two novel variants in *DPYD*: rs12132152, strongly associated with HFS (OR=6.1, $p=3.6x10^{-8}$), whereas both rs12132152 and rs12022243 affected global toxicity, albeit at a non-significant genome-wide level (OR=1.69, $p=2.55x10^{-5}$; OR=3.83, $p=4.31x10^{-6}$, respectively). These SNPs were independent of each other and of other reported *DPYD* toxicity variants.

Fernandez-Rozadilla et al. similarly used 1,012 patients in a two-stage study to discover novel genetic markers for the prediction of ADRs in patients treated with 5-FU and FOLFOX[68]. They found that associations were very modest, with no novel ones reaching the established genome-wide significance level (best $p=1.076\times10^{-5}$). Nevertheless, moderate evidence of association was found for the rs10876844 intergenic variant (closest gene *METTL7B*) and diarrhoea in patients treated with 5-FU.

Baas et al. included 282 metastatic CRC individuals from the CAIRO2 trial and found some novel SNPs to be moderately associated with toxicity: *ZNF827* rs12646351 (OR=0.04, $p=2.47x10^{-7}$) and rs17806780 (OR=0.04, $p=3.42x10^{-7}$), and *EPHA5* rs7692430 (OR=4.57, $p=4.39x10^{-7}$)[70].

Won et al. completed a GWAS on 343 Korean patients receiving oxaliplatin-based regimens to identify possible genetic markers associated with chronic oxaliplatin-induced peripheral neurotoxicity (OXCPN)[69]. Some of the SNPs with the lowest p-value were intronic or within 100 Kb of neighbour genes related to various neuronal activities, although none were genome-wide significant. Two subsequent and independent studies tried to validate these results. The first included 155 Japanese patients and could only replicate the association between the *FARS2* rs17140129 G allele and OXCPN (OR=6.5, p=0.034). This gene encodes for the mitochondrial phenylalanyl-tRNA synthetase 2, which is crucial for genetic translation, but its relationship with OXCPN is yet to be explained[71]. The second validation study by Terrazzino et al. included 150 Caucasian patients and could not replicate any of the associations initially found with OXCPN[72].

Next-generation sequencing (NGS)

In the past few years, SNP array strategies have increasingly been overtaken by next-generation sequencing (NGS), including **whole-exome (WES)** or whole-genome sequencing (WGS). NGS allows for a more comprehensive inspection of both common and rare variation, albeit at a greater economic cost. Several studies have reported the added value of NGS to identify relevant rare pharmacogenetic variants that would not be identified by other conventional methods[67,73–78]. A study by Kozyra et al. provided sequencing data for 146 pharmacogenes from more than 6,500 individuals, and detected a total of 19,328 single nucleotide variants (62.9% exonic)[75]. Most of these variants were indeed rare (MAF<1%; 92.9%) or very rare (MAF<0.1%; 82.7%), meaning that they would not have been detected or inferred by conventional genotyping methods. The functional impact from rare variants was different across the genes, yet rare variants contributed on average 30-40% of the functional variability in the studied pharmacogenes.

NGS approaches are not only useful to identify rare variants but can be an important asset to reveal copy-number variation as well. An example is the work published by Santos et al., which produced CNV data on 208 ADME genes from 2,504 WGS and 59,898 WES[76]. Of these, 201 (97%) genes had a total 5,589 novel CNVs, where 47% were deletions and 53% were duplications. These novel deletions were responsible for >5% of loss-of-function alleles in a considerable number of genes across different populations (87, 25, 49, 48, 59, and 51 genes in non-Finnish Europeans, Finnish, East Asians, South Asians, Africans, and admixed Americans, respectively). This demonstrates the impact that CNVs might have on ADME genes, and hence the development of ADRs.

Transcriptome-wide association studies (TWAS)

As described above, GWASs have been modestly successful so far in identifying the genetic variants responsible for toxicity. The power of GWASs depends greatly (among other parameters) on the sample size, which in pharmacogenomic studies is smaller than for other heritable traits or diseases like cancer. Furthermore, the fact that most inspected variants are in non-coding regions, makes the results often difficult to interpret[79]. Therefore, researchers are now looking into alternatives that can also fill the gap between the genetic signals and the mechanistic processes driving susceptibility. It has been observed that many of the SNPs associated with a particular trait identified by GWASs are more likely to be expression quantitative trait loci (eQTLs), meaning that the two alleles lead to different levels of gene expression[80]. Therefore, the integration of transcriptome data into GWAS (**transcriptome-wide association studies – TWAS**) has arisen in the past few years to enhance the discovery of new risk loci[81]. TWASs estimate the association between genetically predicted expression levels and the trait of interest[82–84]. Reference datasets with matched genomic variation and transcriptomic profiles (i.e. gene expression levels) are used to build predictive models of expression for each gene. Some of these reference datasets are publicly available; for instance, the Genotype-Tissue Expression (GTEx) Project includes gene expression models for 49 human tissues[85]. Then, these expression models are applied to either individual or summary-level GWAS data to estimate gene expression and its correlation with the analysed trait[82,83]. The advantage of these TWAS approaches is that, while GWASs analyse SNPs one by

one, TWASs collapse the genomic variants into larger regions (genes), and thus these approaches effectively reduce multiple testing burden and increase power at lower sample sizes, as is the case for pharmacogenomic studies. Several recently published TWAS have yielded very interesting results when applied to other phenotypic traits[81,84,86,87]. For instance, for CRC risk predisposition, Fernandez-Rozadilla et al. performed a TWAS and coupled methylome-wide association study (MWAS) using 100,204 CRC cases and 154,587 controls to identify 50 risk loci and 155 candidate target genes that had been undiscovered by GWASs approaches.

Current pharmacogenetic guidelines for CRC treatment

Currently, for treatment with fluoropyrimidines, only the four mentioned *DPYD* variants (rs3918290, rs55886062, rs67376798, and rs56038477/HapB3) have an indication for pharmacogenetic testing before treatment[88–92]. Genotyping of these variants can also be substituted by phenotyping testing for DPD deficiency. This is achieved by measuring blood uracil levels before treatment. Based on these, patients can be classified as normal, intermediate, or poor metabolizers, each having a recommended drug dose (Table 4)[89].

Table 4. *DPYD* genotype-phenotype approach for recommendations on fluoropyrimidines dosing (adapted from Amstutz et al.)[89].

Phenotype	Genotype (considering all four variants)	Dose Recommendation
Normal metabolizer (normal DPD activity)	Normal function alleles	Normal dose
Intermediate metabolizer (decreased DPD activity)	One no-function allele OR One decreased function allele OR Two decreased function alleles (same variant or compound heterozygous)	Dose reduction by 25%-50%
Poor metabolizer (complete DPD deficiency)	Two no-function alleles OR One no-function + one decreased function allele	Avoid 5-FU, or strongly reduce starting dose with early therapeutic drug monitoring

When irinotecan regimens are considered, guidelines from the FDA and the Dutch Pharmacogenetics Working Group (DPWG) suggest dose reduction if the patient is homozygous for the has the *UGT1A1**28/28 allele[90,92]. These are however only recommendations and are not yet implemented as required testing in Spain. Likewise, for response, *RAS* status is to be evaluated before treatment with cetuximab/panitumumab[5,25,92,93].

Limitations of pharmacogenetic & pharmacogenomic studies

From this introduction, we can derive that there is presently a vast amount of data supporting the impact of genetic variability on CRC chemotoxicity based on pharmacogenetic/omic studies. However, the results obtained so far are by no means exceptional, with a small number of inadequately sized studies that have yielded conflicting information about candidate toxicity genes and variants. This could be due to several reasons:

Firstly, ADR recording is not standardised into clinical history records and many physicians do not consistently report toxicity events. This lack of **homogenised data** collection often leads to inherent bias for pharmacogenetic studies and lack of reproducibility among studies. It also makes it challenging to obtain sample sizes from multicentric studies that can offer adequate study power. Standardised Case Report Forms (CRFs) including relevant information like tumour stage, chemotherapy regimen, and adverse events with the respective grade and cycle of appearance, should be used to facilitate the reproducibility and reusability of the data.

Secondly, ADR **phenotyping** is complicated and heterogeneous: some ADRs can be measured directly and quantitatively, like the haematological ones, whereas others are inherently subjective and therefore more prone to bias, like asthenia. This can be mitigated by using well-defined and already established toxicity grading scales, like the Common Terminology Criteria for Adverse Events (CTCAE)[94]. From this, categorisation of case and control groups is often as well problematic, because this classification should consider other variables, like the cycle at which toxicity happens, dosage reductions or delays in the therapeutic administration.

Thirdly, there is an urgent need for **novel approaches** for toxicity pharmacogenomic studies. Association studies, and particularly GWASs, seemed to be a promising method to discover novel pharmacogenes because they inspected genomic-wide variation, but study sample sizes have made the published efforts greatly underpowered to detect the missing pharmacogenomic heritability. Moreover, it is likely that rarer toxicity variants may also contribute to the genomic spectrum of toxicity[95,96]. Novel technologies, such as next-generation sequencing (NGS), allow us to inspect both common and rare variation, plus they offer the possibility of integration of different layers of molecular data, therefore being the basis for novel systems biology approaches.

Lastly, genomic studies have seldom gone further to evaluate the **functionality** of the identified variants or the mechanisms by which these influence the development of ADRs. Functional studies should therefore be standardly used to provide solid evidence for more variants to be included in pharmacogenetic testing as a standard of care.

Justification & Aims

Early diagnosis and advances in treatment options for colorectal cancer (CRC) have allowed for increased survival. However, most anti-cancer drugs have a narrow therapeutic index, meaning that they have a small range of concentration in which they achieve the desired effect without causing toxicity. In practice, this results in the current modest efficacy of chemotherapy and the substantial appearance of ADRs, with nearly 90% of patients ending up developing more than one toxic reaction during the first line of chemotherapy.

As discussed in the introduction, it is in this context that pharmacogenomics of toxicity plays a fundamental role to achieve better and more personalised treatment. However, the efforts made so far have been limited by the lack of homogenised data, the complicated and heterogeneous ADR phenotyping, the need for novel approaches, and the necessity to evaluate the functionality.

Therefore, the main objective of this study was to **identify and validate new germline biomarkers of toxicity** in CRC patients after treatment with chemotherapy, using novel technological approaches. This principal aim can be further divided into different specific sub-objectives, each of which corresponds to a chapter in this thesis:

- Identification of novel high/moderate-penetrance toxicity variants using whole-exome sequencing analysis in a cohort of 163 cases and 52 controls with extreme toxicity phenotypes (Chapter 1). This included a comprehensive assessment of variants in described pharmacogenes, as well as the discovery of novel rare toxicity variants/genes.

- Omic-wide analyses of susceptibility to toxicity to determine low-penetrance toxicity variants under a polygenic risk model (Chapter 2). For this purpose, we performed both GWAS (meta-)analyses, as well as a transcriptome-wide association study of toxicity using the gene expression profiles of 50 human tissue types encompassing over 16,000 thousand samples.

- Functional evaluation of candidate toxicity variants (Chapter 3). We performed the in vitro characterisation of a novel rare *DPYD* missense variant (rs202212118) using enzymatic assays and protein abundance quantification.

Materials & Methods

1. Whole-exome sequencing analysis

Samples & Phenotyping

We recruited stage II-IV CRC patients that had undergone a first line of treatment with one of the following fluoropyrimidine-based regimens: 5-FU+LV, capecitabine, 5-FU+LV with oxaliplatin (FOLFOX), 5-FU+LV with irinotecan (FOLFIRI), or capecitabine with oxaliplatin (XELOX).

Patients were enrolled in collaboration with several recruitment centres in Spain and internationally: the Tecnológico in Monterrey, Mexico (MEX, N=14), Complexo Hospitalario Universitario de Ourense (CHUO, N=36), Complexo Hospitalario Universitario de Santiago (CHUS, N=7), Hospital Universitario Gregorio Marañón (HUGM, N=60), the Institute of Experimental Medicine, Czech Republic (KOPIE, N=4), Hospital Universitario Infanta Sofia (HUIS, N=4), Biobank of Santiago de Compostela (BB, N=6), Institut Català d'Oncologia (ICO, N=21), and Hospital de la Santa Creu i Sant Pau (SPAU, N=33) (Table 5). We also used retrospective patients from the EPICOLON II cohort (EPI, N=30)[68]. EPICOLON II was a multicentric study of the epidemiological features of CRC in Spain. It includes colon cancer patients that had received adjuvant or palliative 5-FU+LV, FOLFOX or FOLFIRI-based chemotherapy, and rectal cancer patients that had received the same regimens alone or in combination with radiotherapy. Altogether, we recruited 163 cases and 52 controls for the discovery phase.

All patients received informed consent (Annex 1), according to the tenets of the Declaration of Helsinki[97], and upon this were included in the collection with reference C.0001387 (pharmacogenetics and pharmacogenomics research), currently incorporated into collection C.0006352 (research into the genetic component of the disease and its response to drugs), PI Dr Ángel Carracedo. Project protocol was approved by the local ethics committee (reference 2017/354, Annex 2). Precisely to avoid the data heterogeneity problem described earlier, we received for each patient the completed Case Report Form (CRF, Annex 3). The CRF included relevant clinical data such as: chemotherapeutic regimen, number of cycles received, treatment lines, ADRs for each cycle, toxicity grade (see below), chemotherapy dose received in each cycle and accumulated, start and end date of treatment, concomitant treatment with radiotherapy, presence and type of recurrence, end of follow-up status, and cause of death, among others.

To overcome the limitation associated with Adverse Drug Reactions (ADRs) phenotyping, these were graded according to the Common Terminology Criteria for Adverse Events (CTCAE) v3.0[94] (Supplementary Table S1), where grade 1 represents mild ADRs, grade 2 represents moderate ADRs, grade 3 represents severe ADRs, grade 4 represents life-threatening or disabling ADRs, and grade 5 represents death related to ADRs. We collected samples from patients with any of the ADRs described in Supplementary Table S1.

We classified the patients from the discovery cohorts into case and control groups, according to the following criteria (Table 5):

- Cases: patients with any extreme (grade 3/4) ADRs on the first three cycles of chemotherapy,
- Controls: patients who received ≥8 cycles and had mild or no ADRs (grade ≤1).

Table 5. Sample counts per phenotype for the discovery phase.

Phenotype	MEX	CHUO	CHUS	HUGM	KOPIE	HUIS	BB	ICO	SPAU	EPI	TOTAL
Diarrhoea	1	13	4	6	1	4	1	6	6	5	47
Neutropenia	2	6		9			4		18	3	42
Asthenia	3	3		3					14	8	31
Nausea	6	7		2		1		1	6	3	26
Mucositis	1	5		4				3		3	16
HFS	1	7	1							2	11
Leucopenia	1							4		3	8
Neuropathy	2	3	1	1							7
Skin toxicity[a]		1		3					1	1	6
Cardiotoxicity[a]		2	2					1	1		6
Anaemia							1	1	1	1	4
Allergy				2					1		3
Hepatotoxicity		1		1				1			3
Thrombopenia	1	1									2
Lymphopenia		1									1
Conjunctivitis		1									1
Enteritis				1							1
Digestive[b]									1		1
Haematological[b]	1										1
Control				31	3			12		6	52

a: any CTCAE ADR in this category; b: unspecified ADR subtype; HFS: hand-foot syndrome; MEX: Mexico; CHUO: Complexo Hospitalario Universitario de Ourense; CHUS: Hospital Clínico Universitario de Santiago; HUGM: Hospital Universitario Gregorio Marañón; KOPIE: Institute of Experimental Medicine, Czech Republic; HUIS: Hospital Universitario Infanta Sofia; BB: Biobank of Santiago de Compostela; ICO: Institut Català d'Oncology; SPAU: Hospital de la Santa Creu i Sant Pau; EPI: EPICOLON II.

Patients were grouped into different hierarchical phenotypic categories. From now on, these will be referred to as toxicity groups:

- Single toxicities (N>5): diarrhoea, neutropenia, asthenia, nausea, mucositis, HFS, leucopenia, neuropathy, skin toxicity, cardiotoxicity.
- Systemic grouping:
  - digestive toxicity (diarrhoea + nausea + mucositis + enteritis, digestive - unspecified),
  - haematological toxicity (neutropenia + leucopenia + anaemia + thrombopenia + lymphopenia, haematological - unspecified), and
  - global toxicity (all toxicities). As *DPYD* variants have been observed to contribute to multiple toxicity phenotypes, we also considered the aggregated toxicities for all patients.

DNA extraction & sequencing

For each recruited patient, a sample of peripheral blood was collected in an EDTA tube during a routine medical appointment and then sent by mail to the Genomic Medicine laboratory of the Health Research Institute of Santiago de Compostela (IDIS). Once the sample was received, it was registered internally and DNA was extracted with the *DNeasy Blood and Tissue* kit (Qiagen, Hilden, Germany). Genomic DNA quantity and quality were evaluated by measuring the DNA concentration with the *Qubit 3.0 Fluorometer* and the *Qubit DNA HS Assay* kit (Thermo Fisher Scientific Inc, Massachusetts, USA). Genomic integrity control was performed with the *Agilent TapeStation 4200* using a *genomic DNA ScreenTape* (Agilent Technologies, California, USA). All samples had a minimum concentration of 13ng/µL and DNA Integrity Number (DIN) >5.

Library preparation and whole exome sequencing (WES) were undertaken using 151-base pairs (bp) paired-end DNA libraries with the *SureSelect V6 + UTRs* kit (Agilent Technologies, California, USA) (Figure 5). This kit is optimised to enrich for genomic coding regions and regulatory untranslated regions (UTRs). It covers approximately 90 Mbp of the genome, 60 of them corresponding to coding regions and 30 Mbp to UTRs, as defined by the ENCODE project[98]. In brief, library preparation started with an enzymatic fragmentation of the gDNA and the addition of adapters to the ends of the fragments resulting from the reaction (Figure 5). The adapter-tagged fragments were purified using *AMPure XP* beads, to discard any contaminants. Libraries were then amplified by PCR and purified again with the *AMPure XP* beads. A quality control was performed to check size distribution of the PCR-enriched fragments, by running a *DNA 1000* chip on an *Agilent 2100 Bioanalyzer*. DNA peaks for this check should be between 245-325-bp, indicating an ideal gDNA fragmentation. The next step was the hybridisation of the library using the *SureSelect All Exon V6 + UTR* kit. The hybridisation used biotinylated oligonucleotide probes, which are captured by streptavidin-coated magnetic beads. At this point, the libraries are amplified by a second PCR using dual-indexing primers for each sample. The PCR products are then purified, and the quality was assessed as before (optimal peaks in this case should be 325-450-bp). Concentration was measured with the *qPCR NGS Library Quantification* Kit (Agilent Technologies, California, USA). Finally, pools for multiplexed sequencing were prepared at normalised concentrations of 10 nM.

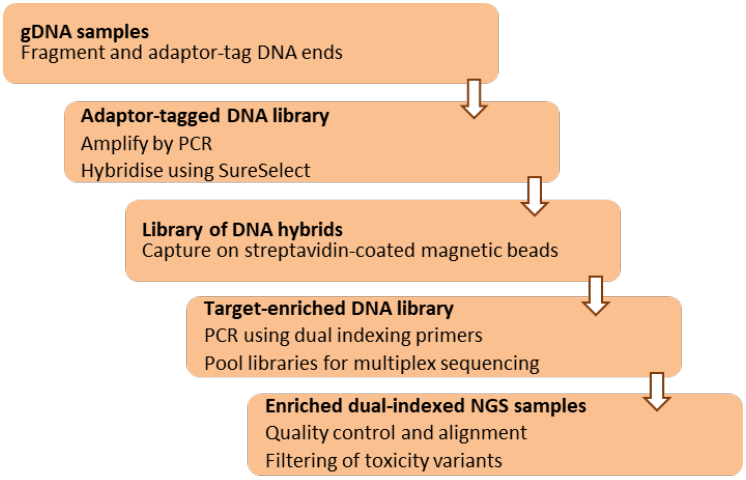


Figure 5. WES sample preparation workflow (adapted from https://www.agilent.com/cs/library/usermanuals/public/G9681-90010.pdf).

The exomes were sequenced on a NovaSeq6000 platform (Illumina Inc, California, USA) to obtain an average depth of 72x (calculated for target regions). After sequencing, the data was processed following the *GATK Best Practices* workflow for germline short variant (SNPs and indel) discovery (Figure 6)[99,100]. The first part of the analysis consisted of pre-processing of the raw FASTQ sequences to produce an analysis-ready Binary Alignment Map (BAM) file. For the trimming of the adapters, we used *Fastp* v0.20[101], an ultra-fast tool that detects adapter sequences by finding the overlap of each pair. The FASTQs were aligned to the reference genome (hg19), with *Burrows-Wheeler Aligner* (BWA) v0.7.17 and *Merge Bam Alignments* from *Picard Tools* v2.24.2. Then, the *MarkDuplicates* implementation in *Picard Tools* was used to clean the duplicated sequences resulting from the library preparation. *GATK* v4.0.10.0 *BaseRecalibrator* was used to reassess the quality of the read bases, eliminating some of the background noise. This process involves the building of a model based on a set of known variants (dbSNP build 138), resulting in a recalibration file, and then it adjusts the quality scores of the bases in the input data based on the model. After this data clean-up, the *GATK HaplotypeCaller* was used to perform the variant calling of SNPs and indels, using the UCSC hg19 genome reference. *GenomicsDBImport* and *GenotypeGVCF* were then run to consolidate and perform joint genotyping for all samples. Lastly, the *Variant Quality Score Recalibration* (VQSR) filter was applied, with the *VariantRecalibrator* and *ApplyRecalibration* tools, which use machine learning to identify annotation profiles of variants that are likely to be real.

Subsequently, the variants were annotated with *ANNOVAR* v2019, which gathers information from different databases: RefGene which specifies the coding and non-coding genes of human proteins from the NCBI RNA reference sequences collection (RefSeq Annotation Release 109); dbNSFP v42a which allows a prediction at a functional level and annotation of potential non-synonymous single nucleotide variants in the human genome through prediction scores; ClinVar v20210501, which gathers information on genomic variation and its relationship with human health and phenotypes; and gnomAD V2.1.1, which provides variant frequency

estimated from exome and genome sequencing of various populations. We used the non-Finnish European population (gnomAD_NFE) as the reference for our analyses.

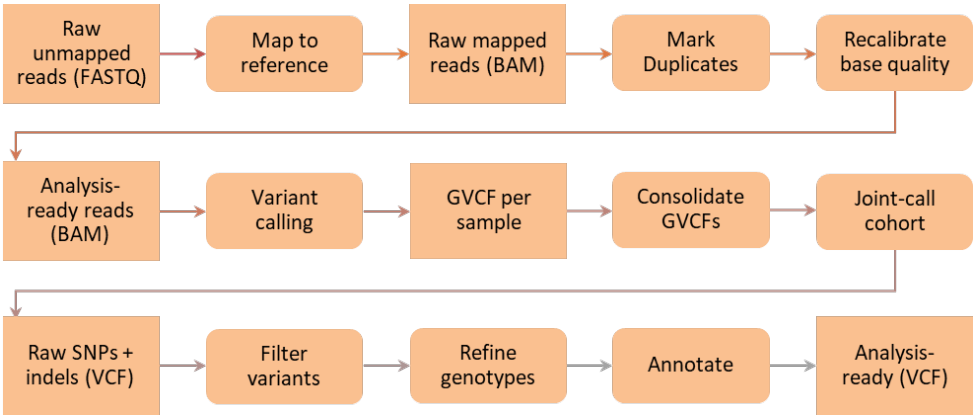


Figure 6. Schematic pipeline from the raw sequence data to analysis-ready data (adapted from GATK Best Practices[100]. The oval shape represents a processing step, and the rectangular shape represents the resulting file.

1.1. Variants in reported toxicity genes

We explored the genetic variation in genes that had already been related to ADR toxicity in the literature. We performed an extensive review of previous works in PubMed using the terms "5-FU", "fluorouracil", "capecitabine", "oxaliplatin", "irinotecan", "CPT-11", "FOLFOX", "FOLFIRI", "XELOX", "XELODA", "bevacizumab", "cetuximab", "panitumumab", "cancer", "colorectal cancer", "colon cancer", "rectal cancer", "gastrointestinal cancer", "digestive cancer", "toxicity", "adverse", "pharmacogenetic", "pharmacogenomics", "pharmacokinetics", "pharmacodynamics". We also reviewed the information available through PharmGKB[24,102], a database with comprehensive information obtained from studies regarding the impact of human genetic variation on drug response and toxicity, and in the pharmacogenetic guidelines from the Clinical Pharmacogenetics Implementation Consortium (CPIC), an international consortium that aims to facilitate pharmacogenetic testing in clinical practice by designing gene/drug guidelines (https://cpicpgx.org/). By these means, we obtained an extended list of toxicity variants and genes (Supplementary Table S2).

We then compiled for each variant a level of evidence (strength of support for their association with toxicity) from PharmGKB[24,102]. These range from 1 to 4 and are based on the integration of p-value, odds-ratio, and replication studies. According to this classification, we observe the following levels:

- Level 1 includes variants with concordant results and clear evidence of association; results must have been replicated in >1 cohort with significant p-values. If there is a variant-specific guideline, the variant is classified as 1A, if not as 1B,
- Level 2 includes variants with moderate evidence of association, including studies with negative association results, and thus must be yet replicated,
- Level 3 variants are reported by a single significant study or have been evaluated in

several studies but with no clear evidence,
- Level 4 variants are based on case-reports, that are biologically plausible but with no significance so far, or without evidence from *in vitro*, molecular, or functional studies.

1.1.1. Variants described in the literature

We first inspected the list of described pharmacogenomic variants. For this, we recovered variants on Supplementary Table S2 and compared their frequencies between cases and controls for each toxicity group using the two-tailed Chi-squared test. We obtained the corrected p-values for each variant and cross-checked the phenotypes and calculated odds-ratios (OR) to compare against those in the reference literature.

1.1.2. Novel variants in reported genes

From the list of reported genes in Supplementary Table S2, we also assessed other variants. Variant selection and prioritisation were focused on variants with higher probability of affecting functionality (similar to *DPYD* variants), and thus we selected rare changes and variation with predicted high impact: truncating variants (nonsense changes and frameshift insertions or deletions), splicing variants (located at canonical sites ±1/2-bp from exons), and high-impact missense variants (predicted by at least 3 out of 4 *in silico* tools: PolyPhen-2 probably/possibly deleterious, CADD_phred≥15, DANN≥0.995, GERP++≥2). PolyPhen-2 is a tool that predicts the impact of an amino acid substitution on both structure and function of a protein[103]. GERP++ analyses the level of nucleotide conservation[104]. CADD integrates different annotations for scoring each variant, including conservation metrics, like GERP++, regulatory and transcript information, and protein-level scores, like PolyPhen-2[105]. DANN uses the same training set and annotation data as CADD, but with a distinct machine learning approach[106].

1.2. Discovery of novel rare toxicity variants/genes

For exome-wide analyses, we used the complete annotated vcf file and applied variant filtering for quality control, as suggested by *Plink 1.9*, where the toolsets are optimised for genotype/phenotype data analysis[107]. Firstly, we performed per-individual filtering to remove individuals with excess missing genotypes (--mind 0.2). Secondly, we undertook per-sample filtering to remove SNPs with >20% missing genotypes (--geno 0.2). After the quality control, 198 patients and 1,397,929 variants remained for analysis.

We defined rare variants as those with a gnomAD frequency NFE≤0.01. Because of the high abundance of these events in exome analysis, we focused our efforts on truncating or splicing variants, which are expected to cause a profound impact on the function of the protein. We prioritised those variants present in ≥4 cases but absent from the controls, as those with the most evidence to be functional.

The list of high-impact variants of interest detected in the discovery cohort was validated in 144 additional samples from CHUS, HUGM, HUIS, SPAU, and EPI (Table 6). These included patients with grade ≥2 toxicities happening up to chemotherapy cycle four. The control group was kept as patients who received ≥8 chemotherapy cycles and had no or grade 1 ADRs.

Table 6. Samples and phenotypes for the validation phase.

Phenotype	CHUS	HUGM	HUIS	SPAU	EPI	TOTAL
Neutropenia	10	12	1	22		45
Diarrhoea	2	13	2	17	10	44
Mucositis	2	4	1	12	8	27
Nausea		5		12	2	19
Anaemia	2	2			7	11
Leucopenia	2	3			4	9
Thrombopenia	3	2			2	7
Neuropathy	2				2	4
Others		2			4	6
Asthenia		2				2
HFS		1			1	2
Cardiotoxicity					1	1
Enteritis		1				1
Controls	4				16	20

HFS: hand-foot syndrome; CHUS: Complexo Hospitalario Universitario de Santiago; HUGM: Hospital Universitario Gregorio Marañón; HUIS: Hospital Universitario Infanta Sofia; SPAU: Hospital de la Santa Creu i Sant Pau; EPI: EPICOLON II.

Genotyping of these variants was carried out at the Santiago node of the Spanish Genotyping Centre (CeGen, Spain) using the *iPLEX Pro* technology (Agena Bioscience, California, USA). This technology offers high performance and sensitivity for low to medium throughput genotyping assays. Briefly, gDNA samples are amplified by PCR and the surplus of nucleotides is dephosphorylated by shrimp alkaline phosphatase. Then, an extension is performed using a mix of oligonucleotide extension primers and mass-modified dideoxynucleotide terminators. The amplification products are finally analysed through mass spectrometry to determine alleles at each position (Figure 7).

Gene-based tests

We evaluated the possibility that toxicity may be caused by different variants in the same gene using the Rare Variant tests (Rvtests) software v2.0.9[108]. Specifically, we used the Optimal Sequence Kernel Association Test (SKAT-O)[109]. Burden tests have more power when most of the variants in a gene are causal and their effects are in the same direction, while SKAT has more power when the majority of the variants in a gene are non-causal, or their effects are in opposite directions. By optimally adapting its behaviour depending on the genetic scenario, SKAT-O allows for a maximised power and analyses the cumulative effect of rare variants within a gene. The different toxicity groups were compared with the controls. To evaluate the effect of the variants' profiles, we performed tests including different types of variants according to their impact:

- rare exonic variants,
- rare exonic variants, excluding synonymous,
- rare truncating variants (nonsense, frameshift, and splicing).

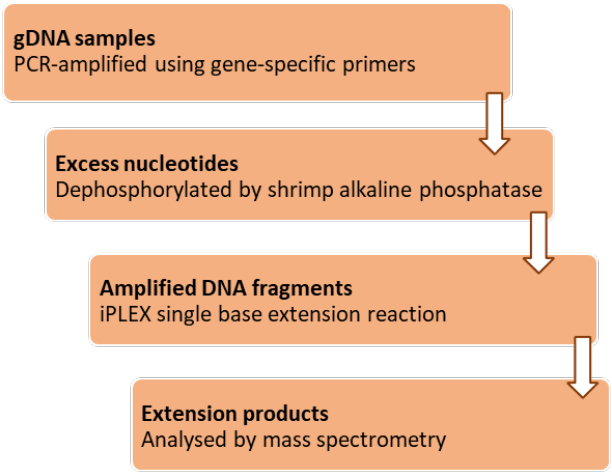


Figure 7. Preparation workflow for the genotyped samples (adapted from https://www.agenabio.com/wp-content/uploads/2015/07/51-20061R1.0-iPLEX-Application-Note_WEB.pdf).

2. Omic-wide analyses of susceptibility to toxicity

2.1. GWAS (meta) analyses

Additional quality control filtering for the GWAS included removal of SNPs that deviated from Hardy-Weinberg equilibrium (--hwe 0.001), and extraction of common variants according to gnomAD frequencies (NFE≥0.05). For the WES data, the classic association test (two-tailed Chi-squared) was applied for the grouped toxicities (global, digestive, and haematological), for all toxicity groups with N>20 (diarrhoea, neutropenia, asthenia, and nausea) and for HFS, with the Bonferroni correction for exome-wide analyses (5.5×10^{-7}). We also calculated the genomic inflation factor and generated Q-Q plots (Supplementary Figure 1) and Manhattan plots, with RStudio[110]. The genomic inflation factor ranged from 1.07 (diarrhoea and asthenia) to 1.21 (HFS).

We ran a meta-analysis with the WES data and the EGAS00001002763 cohort (EGA) (Table 7)[111]. Meta-analysis has been proposed to increase power to detect associations for association studies, as opposed to replication, particularly for lower sample sizes[112]. The EGA dataset included summary statistics data for 504 patients with gastrointestinal cancer, of which 254 were treated with capecitabine and the rest were treated with 5-FU. A subset of 359 patients also received oxaliplatin. Toxicities (diarrhoea, mucositis, and neutropenia - DMN) were recorded jointly after the first four cycles of chemotherapy, and patients were stratified into mild to moderate (grade 0–2) or severe (grade 3-4) subgroups. Also, in the capecitabine-only arm, they analysed diarrhoea and mucositis jointly (DM), and they also recorded HFS, classified as patients with grade 0–2 vs grade 3. Individuals were genotyped on the *Illumina HumanExome Array* (Illumina Inc, California, USA) and variants were filtered with MAF≥0.05.

Table 7. Phenotypes and the respective sample counts for the meta-analysis.

Phenotype	WES*	EGA*	TOTAL
HFS	9/51	13/241	22/292
Diarrhoea+Mucositis	54/51	50/199	104/220
Diarrhoea+Mucositis+Neutropenia	85/51	133/369	218/420

WES: whole-exome sequencing data from our study; EGA: toxicity cohort[111]; *:Ncases/Ncontrols.

Data preparation and quality control guidelines for the EGA dataset were followed using the McCarthy Tools (https://www.well.ox.ac.uk/~wrayner/tools/). We used the HRC-1000G-check-bim-v4.2 to check the quality of the plink files before imputation, including the strand, alleles, position, Ref/Alt assignments and frequency differences. This removed ambiguous A/T and G/C SNPs if MAF>0.4, SNPs with differing alleles, SNPs with >0.2 allele frequency difference, and SNPs not in reference panel - the Haplotype Reference Consortium (HRC) panel v1.1[113]. We then performed imputation – a statistical estimation of missing genotypes from a haplotype reference panel – using the Michigan Imputation Server and the HRC EUROPEAN panel v1.1 as a reference panel. HRC panel combines sequencing data of multiple cohorts into a large panel of human haplotypes. We have followed the imputation quality control guidelines, which indicate that, for common variants, the Rsq value must be ≥0.30. Metal[114] was used for the meta-analysis of the two datasets. The analysis was performed under a fixed effects model, which assumes the same effect size across all studies, thus the assigned weights

are equal, and the combined effect is the estimate of the common effect size. Bonferroni correction was applied for genome-wide analysis (5×10^{-8}).

2.2. Transcriptome-wide association study

We complemented the GWAS analyses with an additional transcriptome-wide association study to increase our power to detect novel common toxicity loci. A locus is considered novel if >1Mbp apart from the literature-reported genes in Supplementary Table S2 or any other GWAS-described loci at a significance level. For this, we used the summary statistics obtained for the QUASAR2 pharmacogenomics GWAS (Q2) (Table 8)[67]. QUASAR2 is a randomised phase III trial that aimed to assess the efficacy of adjuvant capecitabine alone and in combination with bevacizumab in patients with stage III or high-risk stage II CRC. This GWAS included 930 patients of Caucasian ethnicity who received capecitabine, and for which data on diarrhoea, nausea and vomiting, mucositis, neutropenia, thrombopenia, HFS, and global toxicity was available. The ADRs were separated into severe (grade ≥3 according to CTCAE cycles 1 and 2) or mild (grade ≤2). Samples were genotyped using genome-wide SNP panels, from Illumina: Human Hap 370 (N=485), HumanHap 610 (N=352), or Human Omni 2.5 (N=89). Variants were filtered with MAF>1%. The quality control of imputation was assessed, and we kept variants with *info_score*>0.8. Study heterogeneity was calculated and variants with I^2>65% between both arms were excluded.

Table 8. Phenotypes and sample count of Q2.

Phenotype	$N_{cases}/N_{controls}$
Global	142/787
HFS	77/850
Diarrhoea	64/864
Haematological	12/915
Mucositis	10/918
Vomiting	8/918

For the TWAS, we combined the summary statistics of Q2 with gene expression models from different tissues obtained from the following datasets (Supplementary Table S3):

- GTEx v8: with expression models available for 49 human tissues[115]. These included data from 838 healthy donors and 15,201 RNA samples. The majority (85.3%) were of European American origin, and 66.4% were male.

- The Depression Genes and Networks (DGN) project has similarly genetically-predicted expression models available for whole blood from 922 European individuals[116]. These models were also available for download from the PredictDB resources repository.

- The BarcUVa-Seq cohort included 191 individuals with the following features: no personal history of CRC, negative screening colonoscopies[117]. Of these, 93% were of European ancestry and 37% were male. Mucosal biopsies were obtained from the

ascending (n=68), transverse (n=47) or descending (n=76) colon, with a matched peripheral blood sample.

- The Study of Colorectal Cancer in Scotland (SOCCS) is a prospective case-control study with the aim to investigate the associations between colorectal cancer susceptibility and survival outcomes with genetics and environment[118]. Samples were obtained from normal colonic mucosa from freshly resected surgical specimens or rectal biopsies from cancer, non-cancer patients, and healthy individuals. In total, 221 individuals were recruited, mostly of European ethnicity, and 56% male.

- The INTERMPHEN study included 327 samples from 109 European individuals undergoing colonoscopy, where 58% were male[119]. Normal bowel biopsies were obtained from the rectum, distal colon, and proximal colon, as well as a peripheral blood sample.

All genetically-predicted gene expression models were freely available. These had been generated using the Elastic Net method described in the PredictDB pipeline[120]. The GTEx v8 and DGN models were obtained from the PredictDB resources repository (http://predictdb.org/). The BarcUVa-Seq, SOCCS and INTERMPHEN se models were available for download from Zenodo (https://zenodo.org/deposit/6472285).

To run the TWAS analyses, we used the linkage disequilibrium (LD) reference covariance matrix obtained from the GTEx data at the PredictDB resources repository. Then, an individual TWAS was performed for each of the described datasets independently using the S-PrediXcan tool[121]. For cross-tissue and cross-cohort analysis across all 50 datasets, we used SMultiXcan[122]. Results were corrected by applying a Bonferroni correction based on the number of genetically-predicted gene models. We additionally filtered cross-tissue results using a recommended $P_{best} \leq 10^{-4}$ filter, to reduce false positives caused by LD mismatches.

3. Functional characterisation of candidate toxicity variants: rs202212118 in *DPYD*

We observed a novel rare variant in the *DPYD* gene: NM_000110.4:c.2071G>T (p.V691L, rs202212118), that was not reported in PharmGKB or CPIC. Therefore, we ran *in silico* functional prediction of its impact using the *DPYD*-Varifier algorithm, in collaboration with the developing team of this software at the Mayo Clinic[123]. *DPYD*-Varifier is a high-precision predictor that focuses only on *DPYD* and its encoding protein, the DPD enzyme. It generates a training model using random forests using *in vitro* DPD activity data from 156 *DPYD* variants and 15 DPD characteristics including amino acid coordinates, 3D structural information, and changes in amino acid properties. *DPYD*-Varifier has shown higher accuracy (85%) to predict the impact on *DPYD* variants than other tools like PROVEAN, SIFT, PolyPhen-2, FATHMM, PhD-SNP, SNP & GO, UMD-Predictor, and MutantAssessor[123].

Furthermore, we aimed to characterise this novel *DPYD* variant with an *in vitro* enzymatic assay (Figure 8).

Plasmid construction

The pIRESneo3 expression vector with the human *DPYD* sequence was kindly provided by Prof. Dr Robert B. Diasio's group (Mayo Clinic, USA). Vectors included variants p.(V691L) (the novel variant that we intended to characterise), c.1905+1G>A (negative control), and c.2846A>T, p.(I560S) (negative intermediate controls) and were generated using the *Q5® Site-Directed Mutagenesis* Kit (New England Biolabs, Massachusetts, USA) with the primers listed in Supplementary Table S4. The c.1905+1G>A variant causes skipping of exon 14, resulting in a completely non-functional enzyme[70]. The c.2846A>T, p.(I560S) is a missense variant that affects the protein structure and results in a partially impaired enzymatic function[53].

HEK293T cultivation and transfection

We used the human kidney cell line HEK293T/17 (ATCC CRL-11268) for plasmid transfection because it has been shown that these cells do not express the DPD enzyme endogenously[55]. Therefore, we used this cell line to measure DPD expression accurately according to *DPYD* genotype. The vial of frozen cells was thawed and transferred to a 15 mL conical tube with 10 mL of pre-warmed DMEM+/+, then centrifuged for 3 minutes at 1000xg. The supernatant was removed, and the cell pellet was resuspended in 14 mL of medium, transferred to a T75 flask, and incubated in a DMEM+/+ high glucose medium supplemented with 10% foetal bovine serum (FBS), 10 mM HEPES buffer, 1 mM non-essential amino acids, 1 mM sodium pyruvate and 1% Penicillin/Streptomycin at 37°C + 5% CO_2 (Forma™ Steri-Cult™ CO_2 incubator; Thermo Fisher Scientific Inc, Massachusetts, USA) initially for 24 hours, and then split 1:10 when reaching 80% confluency. Cells were passaged 3 times prior to transfection. This included a gentle wash with phosphate-buffered saline (PBS) and detachment by incubation for 5 min at 37°C with 1.5 mL trypsin/EDTA. The trypsin was inactivated with medium. Viable cells were counted using a Neubauer chamber and trypan blue staining (Invitrogen from Thermo Fisher Scientific Inc, Massachusetts, USA), following the manufacturer's protocol. For the transfection, on day 1, $0.5x10^6$ cells were seeded per well in 2 mL of medium in a six-well plate. After 24 hours, transfection was done in triplicates using 2.5 µg plasmid DNA and 250 µL OptiMEM supplemented with 6 µL Lipofectamine®2000 reagent (Life Technologies). On day 3,

the medium was renovated, and cells were harvested on the final day, by removing the medium and washing with PBS, then an incubation for 5 minutes with 100 µL of trypsin/EDTA that was stopped by adding 1 mL of medium. We centrifuged the cells at 900 rpm for 5 minutes, washed with PBS, and repeated the centrifugation.

Protein lysate preparation

Lysis of the cells was achieved physically with beads. The complete cell pellet was resuspended in 250 µL buffer $A^{+/+}$ (35 mM KH_2PO_4, 35 mM KH_2PO_4, 2.5 mM $MgCl_2$, 0.035% ß-mercaptoethanol, 1 PIC tablet, pH=7.4, Annex 4, protocol 1) and added to one volume of cooled soda lime glass beads (0.1 mm diameter). The mix was vortexed three times for 1 minute with 1 minute of cooling in ice in between, and then centrifuged for 30 minutes at 18,400 rcf at 4°C. The supernatant was recovered. Lysates were diluted with buffer $A^{+/+}$ to reach 400 ng/µL (Annex 4, protocol 2).

SDS-polyacrylamide gel electrophoresis and western blotting

For each sample, 4 µg of protein were added to the same volume of sample buffer (4% sodium dodecyl sulphate (SDS), 20% glycerol, 160 mM Tris, 0.002% bromophenol blue) with 5 µL dithiothreitol (DTT), boiled for 5 minutes at 95°C and centrifuged for 2 minutes at 13,000 rpm and 4°C. All samples were then loaded into a 9% running gel with a 3% stacking gel (Supplementary Table S5). The gel ran at 80V, 30 minutes, and then at 120V, 2h. After separation, proteins were transferred onto PVDF membranes by western blotting with a run blocked with an incubation of 1h in a solution containing 5% milk powder with TBS-T buffer 0.1%. The membranes were incubated overnight at 4°C with primary antibodies (1:5000 for mouse-anti-DPD and 1:5000 for mouse anti-β-actin, in blocking solution). The next day, membranes were washed with TBS-T (once for 15 minutes and 3x3 minutes) and incubated for 1h at room temperature with the secondary antibodies (1:7500 anti-mouse, in blocking solution). Again, membranes were washed with TBS-T (once for 15 minutes and 3x3 minutes). Protein signals were detected in the *FUSION SOLO S* imaging system (Vilber, Collégien, France), using the *EVOLUTION-CAPT* software. The loading differences were adjusted using the housekeeping protein β-actin, using the ImageJ v1.54, and the DPD expression of the three mutated proteins was normalised using the wt DPD.

DPD enzymatic conversion of 5-FU and measurement by liquid chromatography-tandem mass spectrometry (LC-MS/MS)

We prepared a master mix of 5-FU (82 µM) in water and NADPH (2 mM) in buffer $A^{+/+}$. We added 80 µL of this master mix to 320 µL of each protein lysate, mixed it by pipetting and centrifuged for 30 minutes at 37°C, 600 rpm. Then, we added 400 µL of saturated ammonium sulphate solution to stop the enzymatic reaction, vortexed for 10s and kept on ice. The enzymatic activity of the DPD enzymes was measured through the conversion of 5-FU to its metabolite 5-FUH2. The end product from the previous step was added to 2.5 mL of ethyl acetate:isopropanol (10:1), spiked with 450 ng/mL of the internal standards 5-fluorouracil-

15N2 (5FU-15N2) and 5-fluorodihydrogenpyrimidine-2,4-dione-13C,15N2 (5FUH2-13C,15N2). After vortexing briefly, the mix was put on a rotary mixer for 10 minutes at 45 rpm. Then, the samples were centrifuged for 10 minutes at 15°C and 3000 rcf. The supernatants were recovered and dried under nitrogen at 40°C. We repeated the previous steps by adding 2 mL of ethyl acetate:isopropanol, then again in the rotary mixer, another centrifugation, and drying. The dried samples were reconstituted in 400 µL acetonitrile:H_2O (8:2), vortexed and sonicated for 1 min. Samples were centrifuged for 5 min at 30,000 rcf to remove any contaminants and the supernatant was recovered in a liquid chromatography-tandem mass spectrometry (LC-MS/MS) vial. The compounds 5-FU and 5-FUH2 were separated and detected on a Waters Xevo TQ-S LC-MS/MS using a normal phase ACQUITY UPLC BEH Amide column, 1.7 µm, 2.1 x 100 mm (Waters Corporation, Massachusetts, USA) and acetonitrile + 0.1% ammonium acetate:H_2O + 0.1% ammonium acetate (1:1), pH=5.3 (eluent A) and acetonitrile + 0.1% ammonium acetate, pH=5.3 (eluent B).

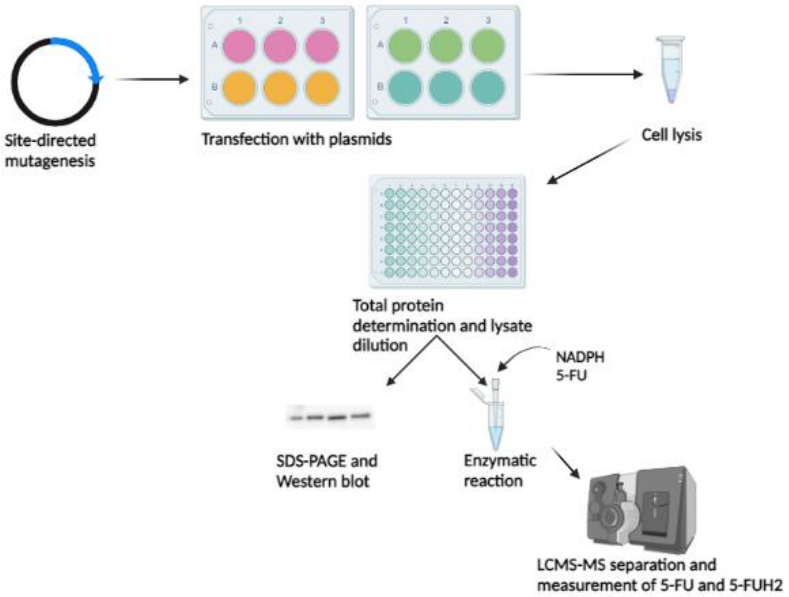


Figure 8. Schematic workflow of the *in vitro* enzymatic assay.

Results

1. Whole-exome sequencing analysis

In this study, we aimed to identify pharmacogenomic variants driving toxicity in CRC patients treated with chemotherapy using NGS approaches. Therefore, we recruited patients who had received first-line treatment for CRC with fluoropyrimidine-based regimens (alone or in combination with oxaliplatin or irinotecan) across 10 different recruitment centres. Cases were identified as patients with severe (CTCAE grades 3 or 4) ADRs on the first three cycles of treatment, whereas controls had received ≥8 chemotherapy cycles and had no or only mild ADRs (grade ≤1). In total, we gathered 163 cases and 52 controls that were subject to whole-exome sequencing (WES). We obtained an average of 32,703,684 variants per exome and a total 35,645,532 variants.

1.1. Variants in reported toxicity genes

1.1.1. Variants described in the literature

We found 13 patients that were heterozygous for the risk alleles of the four clinically actionable *DPYD* variants (rs3918290, rs67376798, rs558860621, and rs56038477). Amongst these, neutropenia and nausea were the most frequent ADRs (4/13, 31%), followed by asthenia and diarrhoea (3/13, 23%), then HFS (2/13, 15%), and lastly neuropathy, mucositis, skin rash, conjunctivitis, and septic shock (1/13, 8%) (Table 9). As these patients' toxicities are well justified by the presence of the risk genotypes, they were removed from further analyses. Interestingly, we also found one control individual carrying the rs56038477 variant. This could be due to the incomplete penetrance of this allele. This patient was also removed from further analyses.

Table 9. Clinically actionable *DPYD* variants. Table presents the patients in our cohort that were heterozygous for these variants, including their phenotype.

rsID	Patient	Phenotype
rs3918290	A313	diarrhoea
	A893	neutropenia + asthenia
rs67376798	A326	asthenia + nausea
	A339	diarrhoea + nausea
	A853	neutropenia
	A856	mucositis
	B1559	neutropenia
rs55886062	A310	HFS + skin rash
rs56038477	1220N	conjunctivitis
	A334	neuropathy
	A338	diarrhoea + HFS
	20011763	nausea + asthenia + septic shock
	20031358	neutropenia + nausea
	293N	control

Two additional patients carried the *UGT1A1* rs4148323 variant in heterozygosity (A363 with asthenia, and A366 with nausea), which is also actionable. Because this variant is described to be associated with toxicity to irinotecan and both patients had received FOLFOX, we could not initially justify the observed toxicity and hence decided to keep them in for further analysis.

Apart from the clinically actionable variants, we also found carriers of 5 other common *DPYD* variants. These had been previously described in the literature, but with conflicting pathogenicity information. rs2297595, rs1801159, and rs1801160 (gnomAD_NFE≥5%), and rs1801158 and rs17376848 (gnomAD_NFE≥1%) are considered level of evidence 3 by PharmGKB (reported by a single significant study or evaluated in several studies but with no clear evidence), and normal function by CPIC[24,89]. The observed frequency of these variants in our cohort was similar to the reported frequency from gnomAD (Table 10).

An additional two previously reported, rare (gnomAD_NFE<1%) *DPYD* variants: rs114096998 and rs776973423, were also detected, which are not described in CPIC or PharmGKB. *DPYD* rs114096998 was found in one control (patient A842, received 11 chemotherapeutic cycles and had no reported toxicity). To our knowledge, this variant has only been mentioned once in a previous study, where the T allele was shown to have an increased enzymatic activity of 138% compared to the wt, with the association being borderline significant (p=0.048)[53]. This variant is not yet included in the pharmacogenetics guidelines but considering its potentially increased enzymatic activity, it may be suggested that carriers of this allele could receive higher 5-FU dosages due to their higher tolerance. This would also explain our phenotyping of this patient as a control. *DPYD* rs776973423 was found in one patient with neutropenia grade 3 after the first chemotherapy cycle (patient A874). The T allele of this variant was previously predicted to be deleterious by *DPYD*-Varifier, which is an algorithm built specifically to predict the functional impact of *DPYD* variants[123]. *In vitro* data (N=4) later showed that this variant reduces DPD enzymatic function by ~50% (unpublished data by Largiadèr et al.). Thus, the toxicity developed by this patient could be explained by this variant.

Besides *DPYD*, the reviewed literature reports a much larger set of genes at varying degrees of evidence regarding their involvement in ADR development. Thus, we performed a focused search for toxicity variants in these genes. As a first step, we considered whether we could validate the association for previously identified toxicity variants. We checked for the presence/absence of such variants in the sequenced exomes and performed statistical analysis using the Chi-squared test comparing cases and controls. We could not replicate any of the reported associations but found significant associations with other ADRs instead: *ABCB1* rs2032582 was associated with skin toxicity ($p_{corrected}$=0.02) and *SLCO1B1* rs2306283 with neuropathy ($p_{corrected}$=1x10^{-3}) (Table 10).

1.1.2. Novel variants in reported genes

We also investigated additional variants in the reported toxicity genes. We prioritised the study of rare changes that would likely affect protein function: truncating variants (nonsense and frameshift), splice variants, and high-impact missense variants (see Materials & Methods for a detailed definition of these). For each gene, we analysed its relationship with any toxicity phenotype (not only with the reported ADR). According to this criteria, 31 rare variants were found in 14 genes, all in heterozygosity (Table 11). Interestingly, one patient (A345) who developed neutropenia carried one novel rare variant in *ABCC5*. Previous reports showed association of variants in this gene with neutropenia[124,125]. *ABCC5* has a missense Z-score>3, indicating that it is extremely intolerant to missense changes.

Table 10. Variants of interest in reported toxicity genes. Table shows the variants previously described in the literature as associated with toxicity, with their frequency in the general population and from our cohort, the odds-ratio and p-value for the reported associated ADR, and the best p-value across all analysed ADRs. We do not mention reported ORs because for most variants, there are multiple, often conflicting ones, depending on the reporting study. OR=Odds-Ratio; CI= Confidence Interval; a: gnomAD_NFE v2.1.1.

Gene	rsID (frequency[a])	MAF (our cohort)	Reported ADR	Nominal p-value for reported ADR	OR (95% CI)	Best corrected p-value (ADR)
ABCB1	rs2032582 (0.64)	0.58	Global[126]	0.42	0.8 (0.52-1.27)	**0.02** (skin toxicity)
ABCC2	rs717620 (0.17) rs3740066 (0.37)	0.23 0.38	Diarrhoea[127]	0.74 0.89	0.9 (0.46-1.66) 1.1 (0.62-1.88)	0.96 (leucopenia) 0.96 (neuropathy)
CDA	rs2072671 (0.28) rs602950 (0.22)	0.35 0.36	Global[128,129]	0.91 0.72	1.1 (0.61-1.52) 0.9 (0.57-1.44)	1 (diarrhoea) 0.88 (leucopenia)
DPYD	rs1801160 (0.05) rs1801159 (0.19) rs1801158 (0.02) rs114096998 ($2x10^{-5}$) rs2297595 (0.09) rs17376848 (0.05) rs776973423 ($1x10^{-4}$)	0.08 0.22 0.05 $2x10^{-3}$ 0.08 0.03 $2x10^{-3}$	Global[67,130–133]	0.63 0.06 0.31 1 0.99 0.15 1	1.5 (0.59-3.68) 1.8 (0.99-3.23) 0.7 (0.27-1.96) NA (NA) 1.2 (0.49-3.05) 4.0 (0.51-2.13) NA (NA)	0.14 (skin toxicity) 0.37 (haematological) 1 (neutropenia) NA 0.69 (nausea) 1 (neutropenia) NA (NA)
ERCC1	rs11615 (0.49)	0.37	Neuropathy[128]	0.24	2 (0.65-6.16)	0.89 (leucopenia)
ERCC2	rs13181 (0.32)	0.32	Haematological[128]	0.77	1.1 (0.64-2.01)	0.98 (digestive)
MTHFR	rs1801133 (0.32) rs1801131 (0.29)	0.35 0.31	Global[128]	0.16 0.62	1.4 (0.88-2.29) 0.9 (0.55-1.42)	1 (skin toxicity) 1 (skin toxicity)
MIR27A	rs895819 (0.34)	0.28	Global[134]	0.53	1.2 (0.73-1.99)	0.15 (HFS)
SLCO1B1	rs2306283 (0.40)	0.38	Digestive[125]	1	1.0 (0.59-1.64)	**$1x10^{-3}$** (neuropathy)
UGT1A1	rs4148323 (0.01)	$4x10^{-3}$	Global[128]	1	NA (NA)	0.95 (nausea)
UGT1A9	rs11692021 (0.35)	0.34	Global[135]	0.64	1.1 (0.71-1.83)	0.81 (leucopenia)
UMPS	rs2279199 (0.47) rs1801019 (0.17)	0.47 0.14	Global[131]	0.82 0.15	1.1 (0.69-1.67) 1.7 (0.85-3.57)	0.98 (cardiotoxicity) 0.21 (cardiotoxicity)

Table 11. Candidate toxicity variants present in the literature-reported genes. For each gene, we include the gnomAD pLI and Z scores, which assign gene-based constraint metrics and intolerance scores for loss of function and missense variants. pLI ranges from 0-1; with pLI≥0.9 being considered extremely intolerant, whereas the Z-score ranges from -5 to 5, with a missense Z-score≥3 considered excessively constrained or intolerant[136,137]. All genes had a pLI≤0.9, meaning that they are predicted to be tolerant to loss of function variation. Only *ABCC5* was predicted to be intolerant to missense changes. In **bold**: values of pLI and missense Z-score that pass the threshold for intolerance; *: patient with >1 candidate toxicity variant.

Gene	pLI	Missense Z-score	Change	rsID	MAF	Patient	Phenotype
ABCB1	0	1.4	p.(Q1107P)	rs55852620	$5x10^{-3}$	A871* 1158N C213	diarrhoea control control
			p.(I261V)	rs36008564	$7x10^{-4}$	339N* A318 A852 C028	diarrhoea neutropenia + diarrhoea control allergic reaction
ABCC2	0	-0.9	p.(S281N)	rs56131651	$7x10^{-3}$	A881 C155 S11028	asthenia control asthenia
			p.(K495E)	rs17222561	$4x10^{-4}$	1243N	diarrhoea + mucositis + leucopenia
			p.(R1079Q)	rs145672804	$7x10^{-4}$	B1550*	neutropenia
			p.(R1100C)	rs142715085	$8x10^{-4}$	B1550*	neutropenia
			p.(R1181L)	rs8187692	$4x10^{-4}$	A330 A366	diarrhoea + HFS nausea
ABCC5	0	**3.6**	p.(K8N)	rs200090959	$4x10^{-5}$	A345	neutropenia
ABCG1	0.1	2.1	p.(R65H)	rs959974308	$1x10^{-5}$	A343	skin toxicity
			p.(S181L)	rs138056067	$1x10^{-3}$	A887	asthenia + allergic reaction
			p.(T628M)	rs199627225	$7x10^{-6}$	A871*	diarrhoea
ABCG2	0	-0.9	p.(M131I)	rs759726272	$4x10^{-5}$	A880*	asthenia
			p.(P112Q)	rs199473672	$6x10^{-4}$	A356*	control
			p.(K652Pfs*3)	rs754196551	$5x10^{-5}$	339N*	asthenia
			p.(G262X)	rs200473953	$3x10^{-5}$	A357	anaemia
			p.(R236X)	rs140207606	$2x10^{-4}$	A880*	asthenia
CES1	0	-2	p.(C87R)	rs149261413	$3x10^{-4}$	A876	neutropenia
			p.(G224R)	rs768635294	$3x10^{-3}$	A356*	control
CES2	0	0.4	p.(R34Q)	rs140461033	$2x10^{-3}$	1268N*	diarrhoea + mucositis

						M085	hepatotoxicity
			p.(G255V)	rs780237910	$7x10^{-5}$	S40588	asthenia + diarrhoea
			p.(P260S)	rs147942040	$1x10^{-4}$	A862 C124	asthenia + allergic reaction + nauseas control
DPYD	0	-0.2	p.(V691L)	rs202212118	$1x10^{-4}$	1268N*	diarrhoea + mucositis
ENOSF1	0	-0.3	p.(S376P)	rs766984806	$3x10^{-5}$	A372	neutropenia
			p.(D75H)	rs377711094	$1x10^{-4}$	A980*	diarrhoea + mucositis
ERCC1	0	0.7	p.(A199T)	rs141811629	$2x10^{-5}$	A847	control
			p.(V623fs*26)	rs756630156	$9x10^{-5}$	A897	neuropathy + HFS
MTHFR	0	0.9	p.(R519C)	rs45496998	$3x10^{-4}$	A842 S270523	control neutropenia
SLC22A7	0	1.2	p.(I213V)	rs369251648	$3x10^{-5}$	A369	thrombopenia
SLCO1B1	0	-0.9	p.(R580X)	rs71581941	$1x10^{-5}$	A868	skin toxicity + diarrhoea
			c.481+1G>T	rs77271279	$9x10^{-5}$	A902	nausea + diarrhoea
TYMS	0.7	2.6	p.(Q32E)	rs1034044510	$2x10^{-5}$	A980*	diarrhoea + mucositis

Remarkably, seven other patients carried two rare, high impact variants (Table 11). Patient A880 carried two variants in *ABCG2* (rs759726272 and rs140207606), and patient B1550 carried two variants in *ABCC2* (rs142715085 and rs145672804). These two genes belong to the ABC-transporter family, which is important in the hepatic influx and efflux of chemotherapeutic drugs.

We also found a few patients with high-impact variants in two genes. Interesting novel variants included the *TYMS* rs1034044510 and the *ENOSF1* rs377711094 changes found in patient A980 who developed severe diarrhoea and mucositis. Additionally, patient 339N carried one novel variant in *ABCB1* and *ABCG2* rs754196551 (stopgain). Patient 1268N had a variant in *CES2* rs140461033 and one in *DPYD* rs202212118. Considering the clinical importance of the well-known pharmacogene *DPYD*, this novel variant is analysed in more detail below (see Chapter 3). Patient A871, had two variants in two ABC genes (*ABCB1* rs55852620 and *ABCG1* rs199627225). Lastly, a control patient (A356) had one variant in each *CES1* rs768635294 and *ABCG2* rs199473672.

1.2. Discovery of novel rare toxicity variants/genes

In addition to the analysis of the variation in the genes described in the literature, we evaluated the contribution of other genes/ variants to toxicity. For this, we first selected rare (gnomAD_NFE≤1%) coding variants expected to have a deep impact on protein function (truncating and splicing variants), and amongst them, we shortlisted those that were present in ≥4 cases with a given phenotype but were absent in the controls. We found 10 variants in 10 different genes, including three stopgain variants, two frameshift insertions, two frameshift deletions, two splicing variants and one startloss variant. These were potentially associated with diarrhoea, digestive, haematological, and global toxicities (Table 12).

Table 12. Rare variants present in ≥4 cases and absent in controls.

SNP	Change	ADR (N≥4)	Gene
rs769388508	NM_001362972:exon4:c.836delA:p.K282Rfs*33	Diarrhoea	*HMGXB4*
		Digestive	
		Haematological	
rs759306300	NM_001304562:exon6:c.508delC:p.T173Rfs*3	Haematological	*INO80E*
rs541080822	NM_182481:exon3:c.178dupA:p.R60Kfs*131	Diarrhoea	*BAGE2*
		Digestive	
rs200971397	NM_145174:exon1:c.164dupT:p.L56Ifs*4	Digestive	*DNAJB7*
rs776852538	NM_017440:exon12:c.1720-1G>T	Diarrhoea	*MDM1*
		Digestive	
rs767564816	NM_001289111:exon6:c.205-1G>T	Digestive	*RP2*
rs201507277	NM_000696:exon1:c.A1C:p.M1?	Diarrhoea	*ALDH9A1*
rs35617692	NM_000502:exon6:c.C771A:p.C257X	Haematological	*EPX*
rs148011353	NM_001288587:exon1:c.G256T:p.G86X	Diarrhoea	*FAM83A*
		Digestive	
rs200143577	NM_004505:exon28:c.A3565T:p.K1189X	Haematological	*USP6*

To validate these findings, we genotyped an additional cohort of 144 samples (20 controls and 124 cases) for these 10 variants using the iPLEX Pro technology (Agena Bioscience, California, USA). The rs769388508 and the rs759306300 variants are short tandem repeats (STRs), thus different allelic variants are not possible to be distinguished by Sequenom. Variants rs200143577 and rs541080822 hybridise in more than one genomic region, so it was not possible to design specific extension primers. Therefore, these four variants were removed from the genotyping assay. Furthermore, variant rs776852538 presented low PCR efficacy, limiting the intensity of the amplified product, making it difficult to correctly assign genotypes.

Three of five remaining genotyped SNPs were identified in nine cases: the *ALDH9A1* rs201507277 was present in two patients with mucositis, one with enteritis, one with diarrhoea, and two with neutropenia; The *FAM83A* rs148011353 SNP was found in a patient with neutropenia, and another with neutropenia and leucopenia; and the *EPX* rs35617692 was found in a patient with mucositis (Table 13). Variants rs200971397 and rs767564816 were not found in any of the cases. All genotyped controls were wt.

Table 13. Validation of rare, high impact variants. We genotyped six variants previously found in ≥4 of our cases with a given phenotype and absent in controls in an additional cohort for validation. Three variants were found in nine patients. All respective genes had a pLI=0, which means they are in principle tolerant to truncating variants. Toxicities in **bold** (either single – diarrhoea, or grouped toxicity – digestive, haematological, global) are associated with the variant in the discovery phase.

Gene	SNP	Relevant ADR in discovery	Patient and ADR (discovery)	Patient - phenotype (validation)
ALDH9A1	rs201507277	Diarrhoea Digestive Global	A317 - **diarrhoea** A855 - **diarrhoea** + mucositis A868 - **diarrhoea** + skin toxicity A902 - **diarrhoea** + nausea *4/147 patients (2.7%)*	7B825 - **diarrhoea** 15008 - mucositis V009 - enteritis DP46 - mucositis 2434486 - neutropenia 2144353 - neutropenia *6/124 patients (4.8%)*
FAM83A	rs148011353	Diarrhoea Digestive Global	1243N - **diarrhoea + mucositis** + leucopenia 339N - **diarrhoea** A328 - **diarrhoea** A345 - neutropenia A761 - **diarrhoea + nausea** + hepatotoxicity M085 - hepatotoxicity S60194 - **nausea** *7/147 patients (4.8%)*	8H739 - leucopenia + neutropenia DP52 - leucopenia + neutropenia *2/124 patients (1.6%)*
EPX	rs35617692	Haematological Global	A861 - **neutropenia** A882 - **neutropenia** S51254 - **neutropenia** 1247N - **leucopenia** + diarrhoea A980 - diarrhoea + mucositis S20020714 - diarrhoea + nausea + asthenia *6/147 patients (4.1%)*	DP40 - mucositis *1/124 patients (0.8%)*

For *FAM83A* and *EPX*, this is the first description of a possible association between these genes and toxicity. This warrants further validation in a higher number of patients as well as molecular assays of functionality to establish the role of these variants/genes in ADR development.

Gene-based tests

We also evaluated the possibility that toxicity may be caused by multiple variants in the same gene. Therefore, we performed gene-based SKAT-O tests according to variant subgroups:

- rare exonic variants (141,743 variants);
- rare exonic variants, excluding synonymous (83,275 variants);
- rare truncating variants (7,570 variants).

We found significant associations in 14 genes (Table 14). Of these, *DGKI* and *SCN5A* genes seem the most interesting according to the scores of probability of intolerance to variation. In our cohort, we found two missense variants and one indel in *DGKI* in three cases with neuropathy (one also had HFS, and another also had nausea), whereas only two rare synonymous variants were found in two controls (Table 14). Two of these variants were also present in 5 patients with other phenotypes, and an additional 9 variants in *DGKI* were found in 13 patients with other phenotypes, including diarrhoea, neutropenia, asthenia, nausea, mucositis, and hepatotoxicity.

Three of our cases (50%) with neuropathy (one also had neutropenia and asthenia) and only five controls (9.8%) had missense variants in *SCN5A*. Three of these variants were also present in 5 patients with other phenotypes, and an additional 6 variants in *SCN5A* were found in 8 patients with other phenotypes, including diarrhoea, nausea, mucositis, asthenia, HFS, neutropenia, leucopenia, thrombopenia, skin toxicity, hepatotoxicity.

Variants in *DGKI* and *SCN5A* are present in 50% of the cases with neuropathy (n=3/6), whilst in less than 10% of the controls (n=2/51 and n=5/51, respectively). Interestingly, patient A331 had one variant in each of these two genes (Table 15). In fact, all cases with skin toxicity or neuropathy except one (A859) had rare variants in more than one gene significant for the analysed ADR. Another example is patient A370, who developed neuropathy, had one rare variant in three of the SKAT-O genes associated with neuropathy (*MYBPC3* rs11570082, *PDZD7* rs142301501, and *SCN5A* rs6791924) (Table 15).

Table 14. SKAT-O rare variant test for genes associated with toxicity. We checked the scores of intolerance probability, to strengthen the selection of the genes that are probably more relevant in toxicity susceptibility. A gene with pLI≥0.9 is considered extremely intolerant to truncating variation, and a gene with a missense Z-score≥3 is considered excessively constrained or intolerant to missense variation)[136,137]. In **bold**: values of pLI and missense Z-score that pass the threshold for intolerance; *: the frequency of rare variants in this gene was higher in controls than in cases; syn: synonymous; a: variant present in more than one patient.

Gene	pLI	Missense Z-score	Associated ADR	SKAT-O test (variants in test)	Variants in other ADRs	Corrected P-value
*ITPR3**	0	**4.55**	Cardiotoxicity	Exonic (cases: 1 missense; controls: 2 missense, 9 syn)	asthenia, diarrhoea, nausea, neutropenia, mucositis, HFS, leucopenia, neuropathy, skin toxicity	0.01
INSYN2B	**0.98**	1.78	Cardiotoxicity	Exonic (cases: 2 missense, 1 syn; controls: 1 missense)	asthenia, diarrhoea, neuropathy, nausea, mucositis, leucopenia, anaemia, neutropenia	0.03
SCMH1	0	1.95	Skin toxicity	Exonic (cases: 1 missense[a]; controls: 1 missense, 1 syn)	neutropenia, diarrhoea	0.04
*MYBPC3**	0	1.45	Neuropathy	Exonic (cases: 1 missense; 2 syn; controls: 3 missense, 2 syn)	nausea, thrombopenia, neutropenia, asthenia, HFS, leucopenia, nausea, diarrhoea, cardiotoxicity,	$5x10^{-3}$
DGKI	**0.93**	**3.08**	Neuropathy	Exonic (cases: 1 indel, 2 missense; controls: 2 syn)	diarrhoea, nausea, HFS, asthenia, neutropenia, hepatotoxicity, mucositis, leucopenia	0.04
STK31	0.01	0.26	Skin toxicity	Exonic excl. syn (cases: 1 stopgain, 1 missense[a]; controls: 3 missense)	nausea, asthenia, leucopenia, cardiotoxicity, diarrhoea, neutropenia	$5x10^{-4}$
ARAP3	0	1.82	Skin toxicity	Exonic excl. syn (cases: 1 indel, 2 missense; controls: 3 missense)	diarrhoea, neutropenia, cardiotoxicity, nausea, asthenia, HFS, linfopenia, neuropathy	$8x10^{-3}$
CLPB	0	1.05	Skin toxicity	Exonic excl. syn (cases: 1 missense[a])	asthenia, nausea, neutropenia, mucositis, asthenia, diarrhoea	0.03
PARVB	0	0.1	Skin toxicity	Exonic excl. syn (cases: 1 missense[a])	diarrhoea, nausea	0.03
CPT1A	0	1.64	Skin toxicity	Exonic excl. syn (cases: 2 missense[a]; controls: 1 missense)	neutropenia, asthenia, diarrhoea, HFS, neutropenia, asthenia	0.03
ZBTB45	0	1.36	Neuropathy	Exonic excl. syn (cases: 3 missense; controls: 1 missense)	neutropenia, asthenia, nausea, cardiotoxicity, diarrhoea, mucositis,	$3x10^{-3}$
PDZD7	0	-0.37	Neuropathy	Exonic excl. syn (cases: 1 missense, 1 missense[a]; controls: 1 indel, 1 missense)	diarrhoea, mucositis, skin toxicity, asthenia, HFS, hepatotoxicity, neutropenia, nausea	0.01
SCN5A	**0.91**	**2.75**	Neuropathy	Exonic excl. syn (cases: 1 missense, 1 missense[a]; controls: 4 missense)	HFS, asthenia, diarrhoea, nausea, leucopenia, mucositis, skin toxicity, hepatotoxicity, neutropenia, thrombopenia	0.02
COL4A2-AS2	0	2.19	Skin toxicity	Truncating (cases: 1 indel; 1 stopgain)	diarrhoea, nausea, mucositis, leucopenia, neutropenia	0.01

Table 15. Cases with variants in more than one significant SKAT-O gene. In red: stopgain variants; in pink: indels; in blue: synonymous variants. The rest of the variants described here are missense.

Patient (phenotype)	Gene and variant
A370 (neuropathy)	*MYBPC3* rs11570082, *PDZD7* rs142301501, *SCN5A* rs6791924
A331 (neuropathy)	*DGKI*:p.P975Lfs*39, *ZBTB45* rs201286708, *PDZD7* rs143414291, *SCN5A* rs199473320
A336 (nausea, neuropathy)	*DGKI*:p.L600Q; *ZBTB45* rs777678588
A897 (neuropathy, HFS)	*DGKI* rs35245703, *PDZD7* rs143414291
A362 (neutropenia, asthenia, neuropathy)	*MYBPC3* rs768065513, *ZBTB45* rs767686852, *SCN5A* rs6791924
A868 (skin toxicity, diarrhoea)	*ARAP3* rs557993156, *PARVB* rs56194750, *CPT1A* rs140958507, *SCN5A* rs41313691
A343 (skin toxicity)	*SCMH1* rs114233776, *STK31* rs41273999
C136 (skin toxicity, nausea)	*ARAP3* rs140445106, *STK31* rs138652787, *CLPB* rs150343959, *PARVB* rs56194750, *COL4A2-AS2* rs768746402
C144 (anorexia, skin toxicity)	*STK31* rs41273999, *CPT1A* rs140958507, *COL4A2-AS2* rs113122708
S91219 (skin toxicity)	*SCMH1* rs114233776, *ARAP3* rs147992246, *CLPB* rs150343959

2. Omic-wide analyses of susceptibility to toxicity

2.1. GWAS (meta) analyses

We performed a GWAS on common variants for all toxicity groups with N>20 (diarrhoea, neutropenia, asthenia, and nausea) and HFS, individually, and also for grouped toxicities (global, digestive, and haematological). We did not find any exome-wide significant associations in any of the comparisons. However, the top variant identified from the global toxicity GWAS (*ITGAX* rs2230424) was borderline significant (Table 16). The Manhattan plots for each analysis are shown in Supplementary Figure 2.

We also performed a meta-analysis of our cohort with the EGAS00001002763 dataset, which included summary statistics data for 504 patients with gastrointestinal cancer treated with fluoropyrimidine-based chemotherapy[111]. The meta-analysis was performed for the phenotypic categories diarrhoea-mucositis-neutropenia (DMN), diarrhoea-mucositis (DM), and HFS. We observed no exome-wide significant SNPs in these analyses either (Table 17).

Table 16. Results of the ADR toxicity GWAS. For each of the eight phenotypic categories, the top associated variant is shown. Only rs2230424 in the *ITGAX* gene was borderline exome-wide significant.

Phenotype (Ncases/Ncontrols)	**Variant**	**Gene**	**P-value**
Global (147/51)	rs2230424	*ITGAX*	$9x10^{-7}$
Digestive (71/51)	rs10985765	*GABBR2*	$1x10^{-5}$
Haematological (52/51)	rs480963	*MROH7*	$1x10^{-4}$
Diarrhoea (44/51)	rs10985765	*GABBR2*	$2x10^{-5}$
Neutropenia (37/51)	rs11568187	*MLC1*	$1x10^{-4}$
Asthenia (28/51)	rs4577273	*CROCC2*	$3x10^{-5}$
Nausea (21/51)	rs11230346	*OR4C11*	$1x10^{-5}$
HFS (9/51)	rs7853758	*SLC28A3*	$4x10^{-6}$

Table 17. Results of the ADR toxicity meta-analysis. For each of the three phenotypic categories, the top associated variant is shown.

Phenotype	**Variant**	**Gene**	**P-value**
DMN	rs3739672	*ANP32B*	$1.3x10^{-5}$
DM	rs59366579	*FLII*	$9.3x10^{-5}$
HFS	rs73264111	*BPIFB4*	$3.6x10^{-5}$

2.2. Transcriptome-wide association study of toxicity

Given the non-significant results of the toxicity GWAS, which may be derived from the small sample sizes in our phenotypes, we decided to complementarily perform a transcriptome-wide association study (TWAS) to detect novel common toxicity loci. For this, we used the GWAS from the QUASAR2 (Q2) study, which included data for global toxicity, haematological toxicity, HFS, diarrhoea, mucositis, and vomiting, together with transcriptomic data from 50 human tissues (16,862 samples) from the GTEx, DGN, BarcUVa-Seq, SOCCS, and INTERMPHEN datasets[67,115–119].

We initially checked the association for the literature-reported pharmacogenes (as summarised in Supplementary Table S2). Five genes: *DPYD*, *TYMP*, *ACYP2*, *UMPS*, and *UGT1A1*, reassuringly had a nominally significant p-value (Table 18). For *DPYD* and *UGT1A1*, the associated TWAS phenotype in our analysis matched the reported phenotype (*DPYD* with global toxicity and mucositis, and *UGT1A1* with mucositis), whereas for the others, the TWAS phenotype was different from the reported one (*TYMP* was associated with vomiting, while the reported phenotype was global toxicity; *ACYP2* was associated with vomiting in our analysis but had been reported with neuropathy; *UMPS* was associated with mucositis, but reported phenotypes were global toxicity, asthenia and nausea). The expression of *TYMP* and *UGT1A1* was higher in cases compared to controls (Z_mean 0.80 and 1.73, for vomiting and mucositis, respectively). *DPYD, UMPS* and *ACYP2* expression was lower in cases than in controls (*DPYD*: Z_mean -3.12 for mucositis, Z_mean -2.76 for global toxicity; *UMPS*: Z_mean -1.01 for mucositis; and *ACYP2*: Z_mean -1.38 for vomiting).

Most importantly, we discovered 20 genomic loci significantly associated with toxicity risk (see definition of "novel" loci in *Materials and Methods* - page 36) (Table 19). At four of these loci, two candidate genes were identified (*RP11-785H5.1* and *PRB2*; *AC011526.1* and *CEACAM3*; *TFPT* and *LILRA2*; *SMARCB1* and *SUSD2*). One locus (*MYOM2*) was only linked to global toxicity. Six loci (*SSPO*, *SLC20A2*, *RP11-434C1*.1, *RBFADN*, *ZNF112*, and *COMMD7*) were associated with just one ADR, whereas the remaining 13 loci were linked to multiple ADRs. Of these 13, eight genes were linked to two ADRs including global toxicity *CPSF1, PCDHB9*, *TFPT* + *LILRA2, RP11-114G11.5*, *ZNF117*, *SLC52A2*, *GPR179*, *SMARCB1* + *SUSD2*) while the other five (*C1orf86*, *SRGAP2C*, *MYO19*, *RP11-785H5.1* + *PRB2*, *AC011526.1* + *CEACAM3*) were associated with more than one ADR but not with global toxicity.

These candidate genes have been described to participate in several cellular processes: *COMMD7*, *ZNF112*, *ZNF117*, *TFPT*, *PRB2*, *SMARCB1* are involved in transcription and chromatin regulation, whereas *CEACAM3*, *PCDHB9* and *SSPO* are implicated in cell adhesion. According to Gene Ontology Enrichment Analysis however, we could not identify any significant shared pathways Moreover, amongst the 24 total genes, five code for unannotated long non-coding RNAs (lncRNAs), which could hint at a role in gene expression regulation[138].

Furthermore, we inspected whether the specific associations were more significant in the matching tissue (for example, haematological toxicity with whole blood, diarrhoea, mucositis or vomiting with gastrointestinal tissues), by investigating the most significant single tissue for each signal. However, we could not identify any relevant pattern. Arguably, it may be the case that the analyses including only single tissues are less powered than the multi-tissue analyses to detect associations due to sample size.

Table 18. TWAS associations for the literature-reported genes. Only associations that achieve nominal significance are reported. *: phenotype for any reported variant in the literature in this gene; a: this gene has been widely studied and variants in these genes have been associated with different toxicities besides the mentioned ones.

Gene	TWAS phenotype	P-value	Z_mean	Reported phenotype*
DPYD	mucositis	0.003	-3.12	global, mucositis, other[a]
	global	0.009	-2.76	
TYMP	vomiting	0.017	0.80	global
ACYP2	vomiting	0.025	-1.38	neuropathy
UMPS	mucositis	0.026	-1.01	global, asthenia, nausea
UGT1A1	mucositis	0.036	1.73	global, mucositis, other[a]

Table 19. TWAS significant gene-phenotype associations. Table shows 24 novel genes significantly associated with different ADRs, and their function. Gene start and end positions are based on the GRCh38.p14 version of the genome reference. #: number of novel TWAS locus. A positive or negative Z-score means that the gene expression is higher or lower (respectively) than the mean expression in controls.

Locus (#)	Gene	Chr	Start (bp)	End (bp)	Function	Phenotype	P-value	Z_mean
1	*C1orf86*	1	2,184,477	2,212,720	DNA damage, DNA repair	Mucositis	6×10^{-25}	-0.01
						Haematological	1×10^{-26}	-1.11
						HFS	3×10^{-47}	0.36
2	*SRGAP2C*	1	121,184,975	121,392,874	Neurogenesis	Mucositis	2×10^{-48}	2.26
						Haematological	4×10^{-13}	1.23
						Diarrhoea	8×10^{-16}	-1.31
3	*PCDHB9*	5	141,187,161	141,191,541	Cell adhesion	Global	2×10^{-95}	-1.39
						Mucositis	2×10^{-82}	-1.64
						Haematological	4×10^{-9}	0.71
						Vomiting	5×10^{-61}	-1.34
						HFS	3×10^{-77}	-1.38
						Diarrhoea	5×10^{-10}	-0.18
4	*SSPO*	6	48,423,164	48,478,196	Cell adhesion, catalytic activity, regulation of peptidase activity	Vomiting	3×10^{-17}	3.69
5	*RP11-114G11.5*	7	57,404,767	57,412,147	Long noncoding RNA	Global	6×10^{-165}	-6.22
						Mucositis	3×10^{-98}	4.81
						Haematological	3×10^{-11}	1.54
6	*ZNF117*	7	64,971,772	64,991,036	Transcription regulation	Global	1×10^{-10}	-1.15
						Haematological	2×10^{-235}	-0.68
7	*MYOM2*	8	2,045,046	2,145,456	Binds myosin, titin, and light meromyosin	Global	1×10^{-8}	2.57
8	*CPSF1*	8	14,439,323	144,409,335	mRNA processing	Global	8×10^{-8}	-3.75
						HFS	5×10^{-17}	-5.65
9	*SLC20A2*	8	42,416,475	42,541,954	Phosphate and sodium transport	Haematological	3×10^{-10}	-0.4
10	*SLC52A2*	8	144,358,552	144,361,272	Riboflavin transport	Global	5×10^{-48}	4.05
						Mucositis	1×10^{-28}	3.17

						Haematological	$2x10^{-34}$	3.51
						HFS	$2x10^{-65}$	5.02
11	*RP11-785H5.1*	12	11,171,194	11,176,016	Long noncoding RNA	Haematological	$3x10^{-15}$	0.96
						HFS	$4x10^{-6}$	-0.51
	PRB2	12	11,391,540	11,395,567	Protein and chromatin binding	Haematological	$7x10^{-12}$	0.7
						Vomiting	$9x10^{-17}$	1.17
						HFS	$9x10^{-22}$	-0.96
12	*RP11-434C1.1*	12	11,548,030	11,564,403	Long noncoding RNA	Vomiting	$2x10^{-8}$	1.34
13	*MYO19*	17	36,495,636	36,544,815	Actin-binding, Motor protein, Myosin	Mucositis	$8x10^{-124}$	-4.45
						Haematological	$2x10^{-28}$	-2.48
						HFS	$5x10^{-29}$	-1.91
14	*GPR179*	17	38,324,571	38,343,956	Signal transduction through retinal depolarizing bipolar cells	Global	$1x10^{-23}$	-1.37
						HFS	$2x10^{-8}$	-0.5
						Diarrhoea	$2x10^{-13}$	-1.15
15	*RBFADN*	18	80,067,256	80,081,304	Long noncoding RNA	Haematological	$7x10^{-184}$	7.41
16	*AC011526.1*	19	41,454,169	41,500,649	Long noncoding RNA	Haematological	$1x10^{-39}$	-1.64
						HFS	$3x10^{-36}$	-0.04
	CEACAM3	19	41,796,587	41,811,554	Cell adhesion, innate immune system	Mucositis	$1x10^{-12}$	2.7
17	*ZNF112*	19	44,326,553	44,367,217	Transcription regulation	Haematological	$1x10^{-23}$	-2.02
18	*TFPT*	19	54,107,020	54,115,657	Transcription regulation, apoptosis	Global	$1x10^{-28}$	2.75
						Haematological	$2x10^{-15}$	2.06
	LILRA2	19	54,572,988	54,590,287	Innate immune responses against microbial infection	Diarrhoea	$1x10^{-12}$	-2.36
19	*COMMD7*	20	32,702,699	32,743,467	Transcription regulation	Vomiting	$9x10^{-17}$	2.23
20	*SMARCB1*	22	23,786,966	23,838,009	Chromatin and transcriptional regulation, DNA-binding	Vomiting	$1x10^{-22}$	0.06
	SUSD2	22	24,181,487	24,189,106	Regulation of cell cycle and division	Global	$2x10^{-95}$	-6.63

3. Functional characterisation of candidate toxicity variants: rs202212118 in *DPYD*

As described in Chapter 1, we found a novel rare, high-impact, missense variant in the *DPYD* gene: c.2071G>T, p.(V691L) (rs202212118). This variant was observed in a patient (1268N) that had developed grade 3 diarrhoea and mucositis in the first cycle of chemotherapy with 5-FU. It is a very rare (gnomAD_NFE=0.0001) variant, which was predicted as likely deleterious by four *in silico* classifiers: PolyPhen-2 (probably damaging), CADD=29.3, DANN=0.997, and GERP++=5.9.

We further validated this variant *in silico*, using the *DPYD*-Varifier algorithm[123]. By design, this algorithm uses a weighted metric to derive a likelihood score trained on multiple *DPYD* gene variants. The variant was predicted to be neutral, although the prediction score was relatively borderline, with a 55% likelihood of being neutral (and complementarily a 45% likelihood of being deleterious). Therefore, we proceeded to validate this variant functionally and performed an enzymatic *in vitro* assay to characterise it.

We transfected HEK293T/17 cells with four different plasmids (wt; c.1905+1G>A - negative control; c.2846A>T, p.(D949V) - intermediate enzymatic function; and c.2071G>T, p.(V691L). DPD expression in the four protein lysates was confirmed using western blotting, and representative blots are shown in Figure 9. After accounting for the loading differences using the values of β-actin expression, the DPD expression of the three mutated proteins was normalised using the wt DPD expression. Variants c.1905+1G>A and c.2846A>T, p.(D949V) showed decreased DPD expression, whereas no significant difference was observed for c.2071G>T, p.(V691L) (Table 20).

In parallel, we also measured by LC-MS/MS the DPD enzymatic activity through the conversion of 5-FU to its metabolite 5-FUH2, by LC-MS/MS (Figure 10). Interestingly, the novel variant c.2071G>T, p.(V691L) showed conversion values of only ~40%. However, these were not significantly different than those observed for the wt.

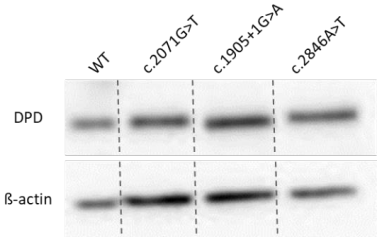


Figure 9. Western blotting of DPD expression for the four assayed plasmids. We can see that the four DPD enzymes are detectable. β-actin was used as a loading control. The measurement of the DPD and β-actin expression for each plasmid was done with ImageJ, a software that quantifies protein bands. We considered increased/decreased expression if a difference of 25% DPD was observed when compared to the wt.

Table 20. Expression of the different DPD proteins. The DPD expression of the three mutated proteins was normalised using the wt DPD. *Mean of the values obtained from the three assay runs; s.d: standard deviation.

Assay	wt	c.2071G>T, p.(V691L)	c.1905+1G>A	c.2846A>T, p.(D949V)
Mean±s.d.*	1	0.81±0.31	0.69±0.55	0.73±0.79

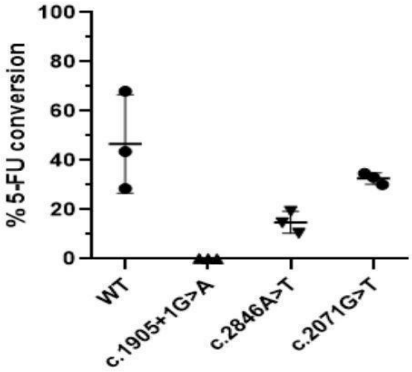


Figure 10. Relative enzymatic activities of the different DPD proteins. We could not normalise the relative activities using the quantity of expressed protein, due to insufficient sample to run both the western blot and the LCMS-MS within the same assay. As expected, the conversion values for the negative controls were null for the c.1905+1G>A, as this variant results in a completely non-functional enzyme, and <20% for the c.2846A>T, p.(D949V), as this variant reduces moderately the DPD enzymatic activity.

Discussion

Adverse Drug Reactions (ADRs) are a frequent side effect of the administration of chemotherapeutic agents in cancer. Not only do they limit or require changes in treatment regimens, but they also have a great negative impact on the patient's quality of life. Consequently, pharmacogenetic studies have become of great interest to decipher the genetic variants that influence the development of toxicity. Knowledge on these factors could potentially help tailor treatment strategies according to a patient's genetic profile, thereby decreasing the incidence of toxic events. However, to this date, the results obtained in pharmacogenomic studies cannot offer a plausible explanation for the toxicity observed in a large proportion of the patients. The emergence of novel high-throughput technologies, such as NGS, has opened up the field to novel strategies that may be helpful in unravelling this missing genetic component. Along these lines, the major aim of this thesis was to carry out pharmacogenomic studies using novel technologies and analytic approaches, to identify and validate new germline biomarkers of susceptibility to chemo-related toxicity in CRC patients.

In Chapter 1 of our study, we selected 215 CRC patients that had received a first-line treatment with fluoropyrimidine-based regimens (alone or in combination with oxaliplatin or irinotecan) and had developed extreme toxicity phenotypes. We then sequenced their whole exome. Initially, we followed a candidate-gene approach and investigated the variants that had already been reported in the literature. Amongst these are the handful of genetic variants for which there are pharmacogenetic guidelines for CRC treatment to date: the *DPYD* rs3918290, rs67376798, rs55886062, and rs56038477 for fluoropyrimidines, and the *UGT1A1* rs3064744 and rs4148323 for irinotecan. We identified 14 patients with the four actionable *DPYD* variants. These were previously shown to be associated with global toxicity. Accordingly, 13 of these patients had presented extreme toxicities (diarrhoea, mucositis, nausea, neutropenia, HFS, skin toxicity, conjunctivitis, neuropathy, asthenia, and septic shock). rs56038477 was also present in one control, potentially reflecting an incomplete penetrance of this variant. Altogether, the frequency of the four clinically actionable *DPYD* variants in our cohort was 6.6%, higher than the reported population frequency (3%). This is unsurprising considering our selection criteria based on extreme phenotypes. This discovery emphasises the need for a meaningful implementation of pre-emptive testing, which could help identify carriers of *DPYD* risk variants prior to treatment, which could have prevented the development of chemo-related toxicity in these patients.

Additional variants in *DPYD* (rs2297595, rs1801159, rs1801160, rs1801158 and rs17376848) were also found in our cohort that have conflicting interpretations based on the literature and they have been declared as of normal function by CPIC. These variants were found in both cases and controls, and they were not significantly associated with any toxicity, thus it seems that they do not have a profound effect on DPD function. It remains feasible however that these may still be responsible for the toxicity via lower penetrance effects.

Besides *DPYD*, we also investigated associations in other literature-reported genes. We did not replicate any of the previously reported associations but found significant results for skin toxicity (*ABCB1* rs2032582) and neuropathy (*SLCO1B1* rs2306283), which are the smaller test groups. Therefore, these results should be interpreted with caution.

Complementarily, we checked for the presence of novel rare, high-impact variants present in the literature-reported genes. We have identified 31 rare variants in 14 genes, all in

heterozygosity. For instance, we found a patient (A345), treated with FOLFIRI (5-FU+LV+irinotecan), carrying one candidate variant in *ABCC5*, who had developed neutropenia. Previous studies had indeed linked variants in this gene with irinotecan-related neutropenia[124,125]. According to its missense Z-score, *ABCC5* is intolerant to missense changes, making this variant a good candidate to explain the toxicity in this patient.

Interestingly, some patients carried more than one rare variant in the same gene, or in genes within the same pathway, specifically in the family of ABC-transporters. Patient A880 (treated with FOLFOX - 5-FU+LV+oxaliplatin) carried two variants in *ABCG2,* whereas patient B1550 (treated with FOLFOX) carried two variants in *ABCC2*. Both these genes belong to the ABC-transporters family, crucial to drug influx and efflux. As oxaliplatin is eliminated from the liver cells largely by *ABCG2* and *ABCC2*, this could increase the toxic accumulation of oxaliplatin inside the cell[31]. According to the gnomAD *Variant Co-Occurrence* tool, the two variants in *ABCC2* are probably within different haplotypes, whilst the two variants in *ABCG2* are probably in the same haplotype. Hence, patient B1550 may be a compound heterozygote for *ABCC2*. Functional studies, for instance gene knockdowns, would be useful to determine if the combination of these two variants could explain the toxicity in this patient. Other interesting examples include patient 339N, who carried one novel variant in each *ABCB1* and *ABCG2* (again both ABC-transporters). As this patient was treated with 5-FU, the stopgain variant in *ABCG2* could impair drug efflux, increasing its concentration inside the cell. Nonetheless, the *ABCG2* rs754196551 is a stopgain variant located in the last exon (exon 16), reason why it might not modify the protein function and it should be interpreted with caution[139]. *ABGB1* however, has not been reported to have a role in 5-FU pharmacokinetics. Therefore, its effect on this patient remains unclear. Patient A871 had two variants in ABC genes *ABCB1* (rs55852620) and *ABCG1* (rs199627225). Still, this patient was treated with 5-FU+LV, whereas neither of these two genes had been previously linked to 5-FU pharmacokinetics, which again makes this data difficult to interpret in the context of the received treatment, albeit it possibly having consequences for toxicity for prospective treatment rounds.

Additionally, we found three more patients with novel rare, high impact variants in two genes, hinting at a potential digenic inheritance. Patient A980, with severe diarrhoea and mucositis, carried the *TYMS* rs1034044510 and the *ENOSF1* rs377711094. *TYMS* encodes for the thymidylate synthase (TS), the main target of 5-FU. Two variants in this gene have been extensively studied: the rs45445694 (5'VNTR 28bp-repeat) and the rs11280056 (3'UTR 6bp-indel)[67,140–142]. In 2015, a study reported an association of an intronic variant located in the *ENOSF1* (which overlaps in part with *TYMS*) capable of explaining the toxicity attributed to these two *TYMS* polymorphisms. The *ENOSF1* rs2612091 and *TYMS* rs45445694 and rs11280056 are in moderate LD (r2=0.40 and 0.32, respectively), but after testing for dependency, they concluded that it was the rs2612091G allele alone that increased the risk of toxicity (p=0.002). In 2020, another study showed the independent association of *ENOSF1* rs2612091 and *TYMS* rs45445694 with toxicity[143]. Due to the LD, the assessment of the exact impact of each individual variant is not straightforward, but the presence of more than one variant could potentially confer a higher risk. Moreover, the biological function of *ENOSF1* is still unclear, but it has been suggested that it can modify *TYMS* expression[67,144]. Therefore, further studies on these two variants would be appealing to help understand the potential influence on each other and the interaction between both genes, particularly in the context of this specific patient carrying both. A control patient (A356) carried one variant in each *CES1*

and *ABCG2*. This patient was treated with FOLFOX, and while changes in the CES1 protein are not expected to influence FOLFOX's concentration in the cell, changes in the ABCG2 protein might impair both 5-FU and oxaliplatin efflux. However, this patient did not present any significant toxicity. Lastly, patient 1268N, who received 5-FU+LV, had a variant in *CES2* and one in *DPYD* rs202212118. CES2 has not been previously described in the literature to participate in the pharmacokinetics or pharmacodynamics of 5-FU nor oxaliplatin. *DPYD* however is, as we have seen extensively, an extremely important pharmacogene, hence, we decided to functionally characterise this novel variant. In the end, by looking in detail at these variants and genes, we can only hypothesise about a potential digenic inheritance for the patient with a variant in *TYMS* and another in *ENOSF1*. This warrants further studies, for instance, it would be interesting to see whether these variants are in *cis* or *trans*.

Irrespective of the inferences and hypotheses we can make based on the *in silico* analysis of the data, functional characterisation of candidate toxicity variants is fundamental to evaluate their true functionality and ultimately assign toxicity and understand the molecular mechanisms behind the development of ADRs. In general, functional studies aim to evaluate the enzymatic activity of the candidate variants, by measuring the drug level and its metabolites, thus assessing their bioavailability and their role in the development of toxicity. In pharmacogenomics, most functional assays have been focused on *DPYD* and the DPD enzyme, given its decisive role in 5-FU metabolism and its confirmed influence on toxicity. The novel *DPYD* variant - c.2071G>T, p.(V691L), rs202212118 - mentioned above, was found in a patient that had presented grade 3 diarrhoea and mucositis in the first cycle of 5-FU chemotherapy. This variant was predicted to be deleterious by four *in silico* classifiers and had a 45% likelihood of being deleterious) by *DPYD*-Varifier[123]. Hence, we performed an *in vitro* enzymatic assay to characterise this variant and help us determine its real impact on the DPD function. Our results showed that this modification in the DPD does not affect its expression, but it seems to produce an enzyme with conversion values of the 5-FU of 40%. Nevertheless, these were not significantly different than those observed for the wt. Ultimately, if this variant is indeed proven to have a critical impact on the DPD function, it could be included in the pharmacogenetic guidelines.

Although the candidate gene approach produced some interesting results, particularly for rare variation, its view is very limited within a set of genes that have been previously described to be involved in the pharmacokinetics/pharmacodynamics of chemotherapeutic drugs. Therefore, we expanded our analysis to an exome-wide, hypothesis-free approach. We discovered and validated in an additional set of 144 patients three rare, high impact variants (*ALDH9A1* c.A1C:p.M1?, rs201507277; *FAM83A* c.G256T:p.G86X, rs148011353; *EPX* c.C771A:p.C257X, rs35617692), potentially associated with diarrhoea, digestive, haematological, and global toxicities. ALDH9A1 belongs to the aldehyde dehydrogenase family of enzymes (ALDHs) that are responsible for the oxidation of toxic aldehydes and are upregulated in many cancers. It has been shown that cancer treatments, such as chemo and radiotherapy are related to the production of reactive oxygen species (ROS) that lead to aldehyde accummulation[145,146]. Therefore, changes in this enzyme could impair the detoxification of cellular aldehydes and the protection against the oxidative stress caused by chemotherapy, ultimately leading to toxicity. The *FAM83A* gene belongs to a family of eight oncogenes, which function as intermediaries in the EGFR/RAS signalling pathway, having a significant role in cell survival, proliferation, invasion and immune evasion[147]. The FAM83A

protein has been identified as a diagnostic and prognostic biomarker for lung, breast, and pancreatic cancer[148,149], but no prior relationship to toxicity has been described. The *EPX* gene encodes the eosinophil peroxidase. This enzyme has an oxidative inflammatory function and acts against parasites and bacteria[150,151]. For *FAM83A* and *EPX*, this is the first description of a possible association between these genes and toxicity. Ultimately, *in vitro* studies should be used to confirm the actual functional consequences of these two stopgain variants possibly associated with the development of toxicity.

Since it is difficult to identify novel toxicity loci at the variant level due to the low frequency of these high-impact changes, we also performed gene-based tests for the exome-wide data. These assess the cumulative effects of multiple variants in the same gene, considerably reducing the multiple testing burden and increasing the power to detect novel toxicity genes. We identified 14 genes significantly associated with neuropathy (5), skin toxicity (7), and cardiotoxicity (2). Two genes associated with neuropathy (*DGKI* and *SCN5A*) had borderline significant pLI and missense Z-score values, which represent their intolerance to truncating and missense variation, respectively. *DGKI* is expressed in the brain and retina and encodes for an enzyme with the same name that belongs to the family of diacylglycerol (DAG) kinases[152]. These convert DAG into phosphatidic acid regulating its cellular levels[153,154]. The DGKI enzyme has synaptic functions and is involved in neural responses[153]. *SCN5A* encodes for the sodium voltage-gated channel alpha subunit 5 that can be found in neuronal cells[155]. Indeed, sodium voltage-gated channels have already been implicated in chemotherapy-related neuropathy[156,157]. It has been shown that oxaliplatin interferes with the activity of sodium voltage-gated channels, and the altered flux of ions can lead to neuropathy. Hence, it is reasonable to argue that both *DGKI* and *SCN5A* might have a role in the development of neuropathy. Interestingly, variants in either *DGKI* and/or *SCN5A* are present in 50% of the cases with neuropathy, whilst in less than 10% of the controls. Only one patient with neuropathy (A331) had one variant in each of these two genes, and the rest of the cases only had one variant in one of the genes. We must however consider that our SKAT-O analysis only identified significant genes in the smaller phenotypic groups (N<10), thus these results should be further validated in larger datasets.

Ultimately, our aim in this thesis was to explain the toxicity developed by each patient based on their genetic profile. Table 21 shows the variants from the patients in our cohort identified in the different analytic strategies discussed so far. Taking all of these into account, we suggest that toxicity susceptibility is likely being influenced by a spectrum of different types of variants and we propose a model for the genetic susceptibility to toxicity caused by chemotherapeutic drugs (Figure 11). For instance, the toxicity of patient A328 might be due to the *FAM83A* stopgain variant identified in seven patients. Patient A980 carried two rare, high-impact variants in the toxicity genes *TYMS* and *ENOSF1*, and had a potentially digenic inheritance, whereas patient A345 carried one variant in *ABCC5* (reported gene) and one in the novel candidate gene *FAM83A*. Additionally, we detected a possible oligogenic inheritance for other patients: patient A331 carried variants in more than one SKAT-O gene (*DGKI* p.P975Lfs*39, *ZBTB45* rs201286708, *PDZD7* rs143414291, *SCN5A* rs199473320, *ARAP3* rs139433211) in a model where the cumulative influence of numerous variants ultimately leads to ADRs. Altogether, from the analyses described above, we would be able to propose candidate variants/genes for 58 of our patients. However, for the remaining 102 cases, we did not find any strong candidate variant/gene. We argue that toxicity in these patients may therefore be

due to the polygenic contribution of multiple variants, each conferring a small increase in toxicity.

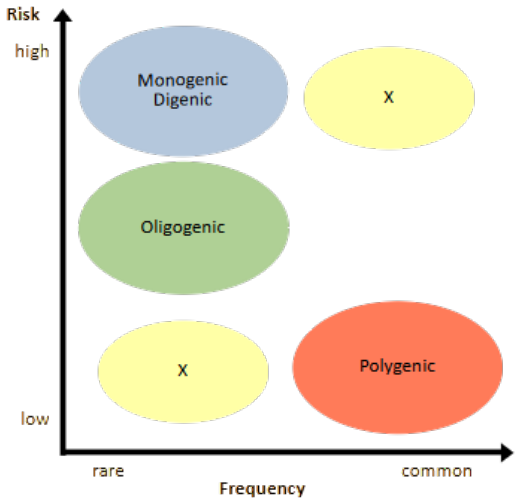


Figure 11. Proposed model for pharmacogenomic susceptibility. The genetic architecture of ADRs reflects on a spectrum from one or a few rare high penetrance variants (monogenic, digenic, oligogenic inheritance) to common low penetrance variants (polygenic inheritance) with a cumulative risk. For instance, for monogenic/digenic cases, we would have the patients with the clinically actionable *DPYD* variants, patient A980 with one variant in *TYMS* and another in *ENOSF1*, and patient B1550 with two variants in *ABCC2*. For oligogenic cases, we would have patient A331 carrying variants in *DGKI*, *ZBTB45*, *PDZD7*, *SCN5A*, and *ARAP3.* For polygenic cases, we would have patients A311 and A341, with only common variants in literature-reported genes; this category also includes variants/genes from GWAS/TWAS. The X includes variants that have not been observed (rare variants with low penetrance, and common variants with high penetrance).

GWAS approaches have been greatly successful in the discovery of low-penetrance susceptibility alleles for different traits, including CRC risk itself[158]. However, pharmacogenomic GWAS have been scarce, with few positive findings. Based on the "Common Disease, Common Variant", we performed toxicity GWASs for common variants for the most frequent toxicity groups (diarrhoea, neutropenia, asthenia, nausea, HFS, digestive, haematological, and global). However, our study did not detect any significant associations. The most likely candidate variant from our analyses was the top variant from the global toxicity GWAS (*ITGAX* rs2230424), which was borderline exome-wide significant. The *ITGAX* gene encodes for the integrin alpha X chain protein, which, together with the beta 2 chain (*ITGB2*), forms a leukocyte-specific integrin (CR4). CR4 was shown to have an important role in immune responses in cancer patients and it is a biomarker for prognosis in gastric cancer patients, and infiltration in stomach cancer patients[159]. Nonetheless, there is no reported direct link between this gene function and the development of toxicity. We also performed a meta-analysis with publicly available data from an additional toxicity cohort but found no significant SNPs for the inspected ADRs (diarrhoea plus mucositis plus neutropenia, diarrhoea plus mucositis, and HFS). These results are probably justified by the reduced sample size and low power, stressing the need for larger pharmacogenomic cohorts of individual ADRs.

In the past few years, novel omic data integration strategies have emerged that can help alleviate the power restrictions of GWASs. These approaches, including TWAS, are increasingly being favoured due to the growth of reference expression datasets, such as GTEx. TWAS approaches are thus becoming a valuable resource for the discovery of novel susceptibility genetic loci as well as in the prioritisation of candidate target genes, leading to more interpretable results. Extending this approach to pharmacogenomics seems to be a promising strategy to identify novel toxicity genes. Although genes identified by TWASs are not directly causal, it allows us to identify potential candidate targets for which the genetically-predicted expression contributes to the variation in the phenotype of interest. In our study, we combined genomic data from one of the largest toxicity GWAS reported to date (QUASAR2 N=930), with transcriptomic data from 50 human tissues from the GTEx, DGN, BarcUVa-Seq, SOCCS, and INTERMPHEN datasets (16,862 samples) to identify novel loci and candidate genes for global toxicity, haematological toxicity, HFS, diarrhoea, mucositis, and vomiting[67,115–119].

Five of the literature-reported toxicity genes were significantly (nominally) replicated in our analysis: *DPYD*, *TYMP*, *ACYP2*, *UMPS*, and *UGT1A1*. Others, like *TYMS*, were notably absent from our results. Curiously, we only identified the same correlation between gene and a previously reported toxicity for *DPYD* (global toxicity and mucositis), and *UGT1A1* (mucositis). Nonetheless, *TYMP* and *UMPS* were previously linked to global toxicity, thus it is not surprising to find that, in our study, they were associated with other ADRs. In our study, *TYMP* expression was upregulated in cases. As the TYMP enzyme is responsible for the activation of capecitabine into 5-FU and for the conversion of 5-FU into metabolites that inhibit the thymidylate synthase (TS), higher *TYMP* expression would lead to higher activation of capecitabine into 5-FU and a higher DNA disruption, which would be concordant with the observed direction of effect. Conversely, *DPYD* and *UMPS* expressions were downregulated in cases. Since the DPD enzyme converts 5-FU into inactive metabolites, downregulation of this gene is very likely responsible for the development of toxicity. As for 5-FU activation, it can occur through two pathways: by TYMP, as described above, or by UMP synthase. This latter enzyme converts 5-FU into active metabolites. Hence, *UMPS* downregulation might lead to a reduction in the production of active metabolites, which seems to be contradictory with *a priori* expectations. Interestingly, the two other associated genes, *UGT1A1* and *ACYP2* were previously only associated with toxicity caused by irinotecan and platinum drugs, respectively. UGT1A1 is the main enzyme responsible for SN-38 inactivation and the ACYP2 enzyme influences the homeostasis of the Ca^{2+}/Mg^{2+}-ATPase pump in the muscle and in the cochlea (part of the ear involved in hearing) but its exact effects in toxicity are still unclear. Since the patients in the Q2 cohort were treated exclusively with capecitabine (alone or with bevacizumab), it is surprising to see that these both come up in our TWAS, which could be indicative that there may be other processes in which these genes are also involved in the pharmacokinetics and/or pharmacodynamics of 5-FU.

Importantly, TWASs bring an added value to the study of pharmacogenomic susceptibility, as they further estimate the effect of genomic variation on gene expression levels and its association with toxicity. Indeed, our TWAS analysis discovered 20 novel genomic loci associated with six toxicity phenotypes (global toxicity, haematological toxicity, HFS, diarrhoea, mucositis, and vomiting). These loci define genomic regions that are influencing toxicity risk through regulation of gene expression and propose 24 candidate genes as targets for these effects. One gene was only linked to global toxicity, possibly because it is our largest

phenotype group and associations in smaller phenotypes could not be detected due to power restrictions. Six of the genes were associated with a single toxicity phenotype (vomiting and haematological), making them "ADR-specific". Eight loci were linked to more than one ADR, including global toxicity, meaning that the specific phenotypes might be driving the global toxicity signal (for example, the *CPSF1* was associated with HFS and global toxicity, thus HFS might be driving the global association). The remaining five loci were also linked to more than one phenotype, but not with global toxicity. A possible explanation is that the SNPs within the models for these genes have effects in opposite directions in the different ADRs. This effect is biologically difficult to interpret but could however be the result of large confidence intervals in the GWAS assigning effect sizes in opposite directions. Interestingly, we observe three genomic loci for which two genes are associated in the region. This observation of co-regulation is similar to LD patterns for SNPs, and it can occur due to pleiotropy, where a causal variant(s) affects multiple traits, as previously observed, or two different causal variants in LD affecting the expression of each gene[81]. This shared expression makes it more difficult to establish the causal gene, and therefore colocalisation analyses would be needed to untangle the downstream mechanisms driving toxicity at these three loci.

It has been argued that using multiple tissues instead of selecting one that is *a priori* more relevant could be a more efficient strategy to discover novel genetic loci[122]. In our study, which is greatly constricted by the power conferred by the small sample sizes of the GWAS, we find that selecting a single or a few tissues that are *a priori* causal seems to narrow the possibility of finding relevant candidate genes. Indeed, when we investigated the relationship between the type of toxicity and the tissue from the best association (that with the lowest p-value from the single tissue analysis), we did not identify any "evident" tissue-ADR relationship. Moreover, pharmacogenes have often been related to global toxicity phenotypes (e.g. *DPYD*), which would also argue for a combined multi-tissue analysis, at least until pharmacogenomic study sample sizes are large enough to assess each tissue type independently.

Although the field of pharmacogenomics has been on the rise these last decades, there is still a long way to go to resolve the missing heritability problem, to characterise candidate toxicity variants/genes, to fully understand their mechanism of action and influence of the toxicity, and ultimately, translate the true impact findings into clinical practice. We believe that important efforts need to be put into the creation of large, comprehensive cohorts. Data sharing in the context of Open Science practices, as well as the integration of different approaches and several omics will as well facilitate a quicker development of the field, which still stands considerably beyond the achievements of genomic studies for other heritable traits.

In this context, we are currently looking into the recruitment of more patients, to create larger cohorts in which to further validate our results. Besides, it would also be interesting to functionally validate the rare high penetrance variants/genes that we have discovered in this study, possibly with enzymatic assays. Regarding the common, low penetrance variants, it is not as straightforward to study them *in vitro*, as their effects would be small, and their impact would only be evident through the cumulative effect of multiple variants. However, there have been many studies in the past 10 years supporting the utility of using polygenic risk scores (PRSs) as a means to stratify the population according to their risk[160–162]. PRSs aggregate the weighted effect of numerous trait-associated alleles to estimate the susceptibility to that trait

of a certain individual[163]. Researchers are now starting to extend this approach into transcriptional risk scores (TRSs), by using the genetically-predicted gene expression[164–166]. After validation of our TWAS results, the translation into a TRS seems to be an interesting next step, to help to categorise patients based on their individual risk of developing toxicity.

Table 21. Summary of all toxicity variants found in the different analyses from Chapter 1, including reported and novel variants in literature-reported genes, as well as novel rare variants/genes. a: corresponds to section 1.1.1. Variants described in the literature; b: corresponds to section 1.1.2. Novel variants in reported genes; c: corresponds to section 1.2. Discovery of novel rare toxicity variants/genes; d: corresponds to section gene-based tests; in red: stopgain and start-loss variants; in pink: indels; in purple: splicing variants; in blue: synonymous variants; in orange: other variants (upstream, 5' UTR, non-coding RNA). The rest of the variants described here are nonsynonymous. *: patients for whom we believe that we have found the cause of toxicity. Note: We do not consider that we have sufficient evidence to justify the toxicity of patients 339N, A871, A880.

Patient	Phenotype	Variants in reported toxicity genes		Discovery of novel rare toxicity variants/genes	
		Variants described in the literature[a]	Novel rare variants[b]	Novel rare variants[c]	Gene-based tests[d]
1158N	control	*DPYD* rs114096998, rs1801160; *ERCC1* rs11615; *MTHFR* rs1801131, rs1801133; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			
1172N	diarrhoea	*DPYD* rs1801159; *UGT1A7* rs11692021; *ERCC2* rs13181; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCB1* rs2032582; *UMPS* rs2279199			*ITPR3* rs41267659
117N	diarrhoea, nausea, mucositis, leucopenia, anaemia	*DPYD* rs1801159; *ERCC1* rs11615; *UGT1A7* rs11692021; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *MIR27A* rs895819; *ABCB1* rs2032582			*COL4A2-AS2* rs761923541
1243N*	diarrhoea, mucositis, leucopenia	*UGT1A7* rs11692021; *ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801133; *DPYD* rs2297595; *ABCC2* rs3740066; *UMPS* rs2279199	*ABCC2* rs17222561	*FAM83A* rs148011353	
1247N*	diarrhoea, leucopenia	*DPYD* rs17376848, rs1801159; *ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801133; *SLCO1B1* rs2306283; *ABCB1* rs2032582; *UMPS* rs2279199		*EPX* rs35617692	
1256N	control	*ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *MIR27A* rs895819; *UMPS* rs2279199			*INSYN2B* rs149156890, *STK31* rs35995607
1268N	diarrhoea, mucositis	*ERCC1* rs11615; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *ABCC2* rs3740066; *MIR27A* rs895819; *ABCB1* rs2032582	*CES2* rs14046103,		

			DPYD rs202212118		
139N	control	*UGT1A7* rs11692021; *ERCC2* rs13181; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *ABCC2* rs3740066, rs717620; *UMPS* rs2279199			
192N	control	*ERCC1* rs11615 *MTHFR* rs1801133; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			
202N	control	DPYD rs1801159, rs2297599; ERCC1 rs11615; MTHFR rs1801131; ABCC2 rs3740066, rs717620; CDA rs602950; ABCB1 rs2032582; UMPS rs2279199			
226N*	cardiotoxicity	*DPYD* rs17376848; *ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			*INSYN2B* rs187220094
242N	hepatotoxicity	*DPYD* rs1801159; *UGT1A7* rs11692021; *ERCC2* rs13181; *MTHFR* rs1801133; *ABCC2* rs3740066; *MIR27A* rs895819; *ABCB1* rs2032582; UMPS rs2279199, rs1801019			*PDZD7* rs397516633
253N	control	*DPYD* rs1801159; *ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801131; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582; UMPS rs2279199			*ITPR3* rs370555491
323N	control	*MTHFR* rs1801133; *CDA* rs2072671, rs602950; *ABCC2* rs3740066, rs717620			
326N	control	*DPYD* rs1801158, rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			*SCN5A* rs539877292
339N*	diarrhoea	*DPYD* rs1801159, rs1801160; ERCC1 rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801131, rs1801133; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *CDA* rs602950; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019	*ABCB1* rs36008564, *ABCG2* rs754196551	*FAM83A* rs148011353	*ITPR3* rs76184558, *CPT1A* rs61731903
345N	control	DPYD rs1801159; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582; *UMPS* rs2279199			*SCN5A* rs45620037, *ARAP3* rs150744263
420N	control	*DPYD* rs1801158, rs1801159, rs1801160; UGT1A7 rs11692021; *ERCC2* rs13181; *MTHFR* rs1801133; CDA rs2072671, rs602950; *SLCO1B1* rs2306283; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			*SCMH1* rs758581293
475N	control	*CDA* rs2072671, rs602950; *ABCB1* rs2032582; *UMPS* rs2279199			

621N	control	*SLCO1B1* rs2306283; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			*MYBPC3* rs397516010
A311	HFS	*DPYD* rs2297595; *ERCC1* rs11615; *MTHFR* rs1801131, rs1801133; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			
A312	HFS	*ERCC2* rs13181; *MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			*ARAP3* rs138728743
A314	thrombopenia	*DPYD* rs1801159; *MTHFR* rs1801131, rs1801133; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			
A315	cardiotoxicity, nausea	*DPYD* rs1801158; *ERCC1* rs11615; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			*ARAP3* rs145726375
A316	mucositis	*UGT1A7* rs11692021; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			*PDZD7* rs200730376
A317*	diarrhoea	*DPYD* rs1801160; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *MIR27A* rs895819; *ABCB1* rs2032582		*ALDH9A1* rs201507277	*SCN5A* rs199473112, *ARAP3* rs139441796
A318*	neutropenia, diarrhoea	*DPYD* rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *ABCC2* rs3740066; *ABCB1* rs2032582; *UMPS* rs2279199	*ABCB1* rs36008564		
A319	neutropenia	*DPYD* rs1801159; *MTHFR* rs1801131, rs1801133; *SLCO1B1* rs2306283; *ABCB1* rs2032582; *UMPS* rs2279199			
A320	diarrhoea	*UGT1A7* rs11692021; *ERCC2* rs13181; *MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *ABCB1* rs2032582			*PDZD7* rs556216049
A321	diarrhoea	*DPYD* rs1801159; *ERCC1* rs11615; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			*SCN5A* rs199473320
A322	nausea	*MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582			*DGKI* rs1385154556
A323	neutropenia	*DPYD* rs1801159, rs2297595; *UGT1A7* rs11692021; *ERCC2* rs13181; *ABCC2 rs3740066*, rs717620; ABCB1 rs2032582; *UMPS* rs2279199			

A324	HFS	*ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			
A325	lymphopenia	*DPYD* rs1801158, rs1801159; *UGT1A7* rs11692021; *ERCC2* rs13181; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582; *UMPS* rs2279199			
A327	diarrhoea	*UGT1A7* rs11692021; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *ABCC2* rs3740066; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			
A328*	diarrhoea	*ERCC1* rs11615; *MTHFR* rs1801133; *SLCO1B1* rs2306283; *ABCB1* rs2032582		*FAM83A* rs148011353	
A329	mucositis	*ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			*PDZD7* rs911911177
A330*	diarrhoea, HFS	*DPYD* rs1801159, rs2297595; *ERCC2* rs13181; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *MIR27A* rs895819	*ABCC2* rs8187692		*SCN5A* rs6791924, *CPT1A* rs61731903
A331*	neuropathy	*DPYD* rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			*DGKI* p.P975Lfs*39, *ZBTB45* rs201286708, *PDZD7* rs143414291, *SCN5A* rs199473320, *ARAP3* rs139433211
A332	nausea, asthenia	*DPYD* rs17376848; *ERCC2* rs13181; *MTHFR* rs1801133; *CDA rs2072671*, rs602950; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			*SCN5A* rs199473211, *ITPR3* rs111882381, *STK31* rs140457583
A333	neutropenia	*DPYD* rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCB1* rs2032582; *UMPS* rs2279199			*ITPR3* rs2229633

A335	neutropenia, diarrhoea, nausea	*ERCC1* rs11615; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *ABCB1* rs2032582; *UMPS* rs2279199			*PDZD7* rs200592310
A336*	nausea, neuropathy	*DPYD* rs1801160; *UGT1A7* rs11692021; *MTHFR* rs1801131; *SLCO1B1* rs2306283; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			*DGKI* p.L600Q; *ZBTB45* rs777678588, *ITPR3* rs146844329
A337	mucositis, HFS	*DPYD* rs1801158; *UGT1A7* rs11692021; *MTHFR* rs1801131; *ABCC2* rs3740066; *ABCB1* rs2032582			*SCN5A* rs41313691
A340*	cardiotoxicity	*UGT1A7* rs11692021; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *ABCC2* rs3740066; *MIR27A* rs895819; *ABCB1* rs2032582			*INSYN2B* rs115307165, *MYBPC3* rs200119454, *ZBTB45* rs200338006
A341	diarrhoea	*DPYD* rs2297595; *ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801131; *SLCO1B1* rs2306283; ABCC2 rs3740066, rs717620; *UMPS* rs2279199, rs1801019			
A342	HFS	*DPYD* rs2297595; *ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801131; *CDA* rs2072671; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582			*ITPR3* rs146338194, rs148020704
A343*	skin toxicity	*DPYD* rs1801159, rs2297595; *ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *ABCC2* rs3740066; *ABCB1* rs2032582	*ABCG1* rs959974308		*SCMH1* rs114233776, *STK31* rs41273999
A344	control	*DPYD* rs1801158; *ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801133; *ABCC2* rs3740066; *CDA* rs602950, rs717620; *UMPS* rs2279199			
A345*	neutropenia	*DPYD* rs17376848; *UGT1A7* rs11692021; *ERCC2* rs13181; *MTHFR* rs1801131; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *ABCB1* rs2032582; *UMPS* rs2279199	*ABCC5* rs200090959	*FAM83A* rs148011353	
A346	diarrhoea	*ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801133; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *UMPS* rs2279199, rs1801019			
A347	neutropenia	*DPYD* rs1801159, rs1801160; *ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			*ARAP3* rs201450767

A349	HFS	*DPYD* rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801133; *SLCO1B1* rs2306283; *CDA* rs602950; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			
A350	leucopenia	*DPYD* rs17376848, rs2297595, rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			*MYBPC3* rs202088839, *ITPR3* rs41267659
A351*	leucopenia	*DPYD* rs1801159, rs2297595; *UGT1A7* rs11692021; *MTHFR* rs1801131, rs1801133; *MIR27A* rs895819	*ABCB1* rs55852620		*MYBPC3* rs200352299
A352	diarrhoea	*DPYD* rs1801159; *MTHFR* rs1801133; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582; *UMPS* rs2279199			
A353	nausea	*DPYD* rs17376848; *ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801133; *ABCC2* rs3740066; *MIR27A* rs895819; *ABCB1* rs2032582			*ITPR3* rs111882381
A354	control	*ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950 *02950*; ABCB1 rs2032582			
A355	asthenia	*DPYD* rs1801160; *ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801133; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582			
A356	control	*ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *ABCC2* rs3740066; *MIR27A* rs895819; *UMPS* rs2279199	*ABCG2* rs19947367, *CES1* rs768635294		
A357*	anaemia	*DPYD* rs1801159; *UGT1A7* rs11692021; *ERCC2* rs13181; *MTHFR* rs1801131, rs1801133; *MIR27A* rs895819; *UMPS* rs2279199	*ABCG2* rs200473953		
A358	control	*DPYD* rs1801159, rs2297595; *ERCC2* rs13181; *MTHFR* rs1801131; *UMPS* rs2279199, rs1801019			*INSYN2B* rs149156890
A359	asthenia	*DPYD* rs1801160; *ERCC1* rs11615; *MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *ABCB1* rs2032582; *UMPS* rs2279199			*ARAP3* rs575984491
A360	control	*DPYD* rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582			
A361	haematological, esophagitis, vomiting	*DPYD* rs17376848; *ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801133; *SLCO1B1* rs2306283; *MIR27A* rs895819; *ABCB1* rs2032582; UMPS rs2279199			
A362*	neutropenia, asthenia, neuropathy	*DPYD* rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCB1* rs2032582; UMPS rs2279199			*MYBPC3* rs768065513, *ZBTB45*

					rs767686852, *SCN5A* rs6791924
A363	asthenia	*DPYD* rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; *UGT1A1* rs4148323; *UGT1A7* rs11692021; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			*SCN5A* rs41315493
A364	diarrhoea, nausea	*DPYD* rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801133; *SLCO1B1* rs2306283; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			*DGKI* rs35245703
A365	nausea, leucopenia	*ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *UMPS* rs2279199			*SCN5A* rs41310765, *STK31* rs140539262
A366*	nausea	*ERCC1* rs11615; *ERCC2* rs13181; *UGT1A1* rs4148323; *UGT1A7* rs11692021; *MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *UMPS* rs2279199	*ABCC2* rs8187692		*DGKI* rs61757580, *MYBPC3* rs11570112
A367	neutropenia	*DPYD* rs17376848, rs1801159; *MTHFR* rs1801133; *SLCO1B1* rs2306283; *ABCB1* rs2032582			
A368	asthenia	*ERCC1* rs11615; *MTHFR* rs1801133; *CDA* rs2072671 rs602950; *ABCC2* rs3740066; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			*DGKI* rs61757580
A369*	thrombopenia	*DPYD* rs1801159; *ERCC1* rs11615; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019	*SLC22A7* rs369251648		*SCN5A* rs6791924, *MYBPC3* rs200162906
A370*	neuropathy	*ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801131, rs1801133; CDA rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			*MYBPC3* rs11570082, *PDZD7* rs142301501, *SCN5A* rs6791924
A371	mucositis	*MTHFR* rs1801133; *CDA* rs2072671, rs602950; *ABCC2* rs3740066, rs717620; *UMPS* rs2279199			
A372*	nausea	*DPYD* rs17376848, rs1801159; *ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801133; SLCO1B1 rs2306283; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019	*ENOSF1* rs766984806		*ITPR3* rs529663156, *ARAP3* p.Q391Pfs*31, rs575984491

A373	nausea	*DPYD* rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801133; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			*DGKI* rs617575807, *MYBPC3* rs769101292
A374	HFS	*DPYD* rs1801160; *MTHFR* rs1801133; *CDA* rs2072671, rs602950			
A833	control	*UGT1A7* rs11692021; *ERCC2* rs13181; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582; *UMPS* rs2279199			
A834	control	*ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *UMPS* rs2279199			
A835	control	*ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801131; *ABCB1* rs2032582; *UMPS* rs2279199			*MYBPC3* rs35736435, *INSYN2B* rs757980684
A836	control	*UGT1A7* rs11692021; *ERCC2* rs13181; *MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283			*ARAP3* rs141698932
A837	control	*DPYD* rs1801158; *ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			
A838	control	*DPYD* rs1801160; *ERCC2* rs13181; *MTHFR* rs1801131, rs1801133; *ABCC2* rs3740066, rs717620; ABCB1 rs2032582; *UMPS* rs2279199			*ITPR3* rs41271255
A839	control	*ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *ABCB1* rs2032582			*MYBPC3* rs35736435, *ITPR3* rs76184558
A840	control	*ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801131, rs1801133; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			*MYBPC3* rs200224422, *SCN5A* rs41276525, *ITPR3* rs146338194, rs148020704
A841	control	*ERCC2* rs13181; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582			*PDZD7* rs762000985,

					ITPR3 rs146844329, rs146844329
A842	control	*DPYD* rs1801159; *ERCC2* rs13181; *SLCO1B1* rs2306283; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019	*MTHFR* rs45496998		*ITPR3* rs111882381, rs41271253
A843	control	*DPYD* rs1801159; *ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801133; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			*SCN5A* rs41315493
A844	control	*DPYD* rs1801160, rs2297595; *ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801131; *ABCC2 rs3740066*, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			*DGKI* rs142433066, *SCN5A* rs6791924
A845	control	*CDA* rs2072671, rs602950; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			
A846	control	*ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs717620; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			
A847	control	*ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *MIR27A* rs895819; *ABCB1* rs2032582	*ERCC1* rs141811629		
A848	control	*DPYD* rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; MTHFR rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066			*STK31* rs33998018
A849	control	*DPYD* rs1801159; *UGT1A7* rs11692021; *MTHFR* rs1801131, rs1801133; *ABCC2* rs3740066; *UMPS* rs2279199, rs1801019			*ITPR3* rs73408284
A850	control	*ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582; *UMPS* rs2279199			
A851	control	*ERCC1* rs11615; *UGT1A7* rs11692021; *ABCB1* rs2032582; *UMPS* rs2279199			
A852	control	*ERCC1* rs11615; *ERCC2* rs13181; *CDA* rs2072671, rs602950; *ABCB1* rs2032582; *UMPS* rs2279199	*ABCB1* rs36008564		
A854	diarrhoea	*DPYD* rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801131; *SLCO1B1* rs2306283; *MIR27A* rs895819; *UMPS* rs2279199, rs1801019			
A855*	diarrhoea, mucositis	*ERCC1* rs11615; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *ABCB1* rs2032582		*ALDH9A1* rs201507277	*ZBTB45* rs145996316,

					ITPR3 rs202114409
A857	diarrhoea	*DPYD* rs2297595; *ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801131; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582			
A858	diarrhoea	*ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801131, rs1801133; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *ABCB1* rs2032582			*DGKI* rs35245703
A859*	neuropathy	*ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801131, rs1801133; *SLCO1B1* rs2306283; *ABCB1* rs2032582			*MYBPC3* rs568935618
A860	neutropenia	*DPYD* rs1801159, rs2297595, rs1801160; *ERCC2* rs13181; *MTHFR* rs1801133; *SLCO1B1* rs2306283; *ABCB1* rs2032582; *UMPS* rs2279199			*ARAP3* rs139441796
A861	neutropenia	*UGT1A7* rs11692021; *ERCC2* rs13181; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582		*EPX* rs35617692	
A862*	asthenia, allergy, nausea	*UGT1A7* rs11692021; *MTHFR* rs1801133; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199	*CES2* rs147942040		
A863	nausea	*DPYD* rs1801160; *ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *ABCB1* rs2032582; *UMPS* rs2279199			*ITPR3* rs200523487
A864	neutropenia	*DPYD* rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *MIR27A* rs895819; *ABCB1* rs2032582			
A865	diarrhoea	*DPYD* rs17376848; *UGT1A7* rs11692021; *ERCC2* rs13181; *MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			
A866	mucositis	*MTHFR* rs1801131; *SLCO1B1* rs2306283; *ABCB1* rs2032582; *UMPS* rs2279199			
A867	mucositis	*DPYD* rs1801159, rs2297595; *ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801133; *CDA* rs2072671; *MIR27A* rs895819; ABCB1 rs2032582			
A868*	skin toxicity, diarrhoea	*DPYD* rs1801158, rs1801160; *ERCC2* rs13181; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *UMPS* rs2279199	*SLCO1B1* rs71581941	*ALDH9A1* rs201507277	*ARAP3* rs557993156, *PARVB* rs56194750, *CPT1A* rs140958507,

					SCN5A rs41313691
A869	neutropenia	*ERCC1* rs11615; *UGT1A7* rs11692021; *CDA* rs2072671, rs602950; *MIR27A* rs895819			
A870	asthenia	*DPYD* rs1801158; *ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *ABCC2* rs3740066; *ABCB1* rs2032582			*ITPR3* rs75810069
A871	diarrhoea	*ERCC1* rs11615; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019	*ABCB1* rs55852620, *ABCG1* rs199627225		
A872	asthenia	*DPYD* rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801131; *CDA* rs2072671; *SLCO1B1* rs2306283; *ABCB1* rs2032582; *UMPS* rs2279199			*SCN5A* rs6791924
A873	neutropenia, enteritis	*DPYD* rs2297595; *ERCC1* rs11615; *MTHFR* rs1801133; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582; *UMPS* rs2279199			
A874	neutropenia	*DPYD* rs1801159, **rs776973423**; *ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			*ARAP3* rs770648102
A875	neutropenia	*ERCC2* rs13181; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			*MYBPC3* rs11570046, *CPT1A* rs151271754
A876*	neutropenia	*DPYD* rs1801159; *UGT1A7* rs11692021; *ERCC2* rs13181; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019	*CES1* rs149261413		*PDZD7* rs757518712, *SCMH1* rs758581293
A877	asthenia	*ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *ABCB1* rs2032582; *UMPS* rs2279199			
A878	asthenia	*DPYD* rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801131, rs1801133; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *CDA* rs602950; *ABCB1* rs2032582			*ITPR3* rs763520343
A879	asthenia, nausea	*DPYD* rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			*ITPR3* rs35394497

A880*	asthenia	*DPYD* rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; *ABCC2* rs3740066; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199	*ABCG2* rs75972627, rs140207606		
A881*	asthenia	*DPYD* rs2297595; *UGT1A7* rs11692021; *ERCC2* rs13181; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *UMPS* rs2279199, rs1801019	*ABCC2* rs56131651		*ITPR3* rs759538163, rs765123685
A882*	neutropenia	*DPYD* rs1801159, rs1801160; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019		*EPX* rs35617692	*DGKI* rs61751966
A883	nausea, diarrhoea, asthenia	*UGT1A7* rs11692021; *ERCC2* rs13181; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			*SCN5A* rs137854609
A884	diarrhoea	*DPYD* rs1801159, rs2297595; *MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			*SCMH1* rs780756941
A885	diarrhoea	*ERCC2* rs13181; *MTHFR* rs1801133; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582			*SCN5A* rs41261344, *ARAP3* rs575984491
A886	leucopenia	*DPYD* rs1801160; *ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801133			*DGKI* rs142433066
A887*	asthenia, allergy	*DPYD* rs1801158, rs1801160; *UGT1A7* rs11692021; *ERCC2* rs13181; *MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199	*ABCG1* rs138056067		
A888	neutropenia	*DPYD* rs2297595; *ERCC2* rs13181; *MTHFR* rs1801133; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582			*STK31* rs41273999
A889	neutropenia	*DPYD* rs1801160; *ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801133; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *MIR27A* rs895819; *UMPS* rs2279199			*ITPR3* rs184074785
A890	neutropenia	*ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *MIR27A* rs895819; *UMPS* rs2279199			*ITPR3* rs111882381, rs41271253
A891	anaemia, neutropenia, diarrhoea	*DPYD* rs1801160; *ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			

A892	neutropenia	*DPYD* rs2297595; *ERCC1* rs11615; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *UMPS* rs2279199			
A894	cardiotoxicity	*DPYD* rs2297595; *ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582			
A895	mucositis	*DPYD* rs1801158; *ERCC1* rs11615; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			*DGKI* rs1185883005, *PDZD7* rs1415971384, *COL4A2-AS2* rs753335766
A896	diarrhoea	*DPYD* rs1801158; *ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801133; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582; *UMPS* rs2279199			*SCMH1* rs143365597
A897*	neuropathy, HFS	*DPYD* rs1801158, rs1801159; *ERCC1* rs11615; *UGT1A7* rs11692021; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582	*ERCC1* rs756630156		*DGKI* rs35245703; *PDZD7* rs143414291
A898	nausea, diarrhoea	*UGT1A7* rs11692021; *ERCC2* rs13181; *MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *MIR27A* rs895819; *ABCB1* rs2032582			*ITPR3* rs146844329
A899	nausea, diarrhoea	*DPYD* rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCB1* rs2032582; *UMPS* rs2279199			*PARVB* rs902926441
A900*	cardiotoxicity	*DPYD* rs1801160; *ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801133; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			*INSYN2B* rs187220094
A901	diarrhoea	*DPYD* rs1801159; *ERCC1* rs11615; *UGT1A7* rs11692021; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			
A902*	nausea, diarrhoea	*DPYD* rs1801158; *UGT1A7* rs11692021; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *MIR27A* rs895819; *UMPS* rs2279199	*SLCO1B1* rs77271279	*ALDH9A1* rs201507277	

B1540	anaemia	*ERCC2* rs13181; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *MIR27A* rs895819; *UMPS* rs2279199, rs1801019			
B1550*	neutropenia	*DPYD* rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *MIR27A* rs895819; ABCB1 rs2032582; *UMPS* rs2279199	*ABCC2* rs14567280, rs142715085		*DGKI* rs527683728
B1558	neutropenia	*ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *ABCC2* rs3740066, rs717620			
B1566	neutropenia	*DPYD* rs1801159; *UGT1A7* rs11692021; *ERCC2* rs13181; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *ABCC2* rs3740066; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			
A741	neutropenia	*DPYD* rs1801158, rs1801160; *ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCB1* rs2032582			*SCMH1* rs114233776, *CPT1A* rs140958507
A761*	nausea, diarrhoea, hepatotoxicity	*ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019		*FAM83A* rs148011353	
AA843	asthenia	*DPYD* rs1801160; *ERCC2* rs13181; *MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			
AA869*	cardiotoxicity	*DPYD* rs1801159; *UGT1A7* rs11692021; *ERCC2* rs13181; *MTHFR* rs1801133; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			*ITPR3* rs2229633
A980*	diarrhoea, mucositis	*ERCC1* rs11615; *MTHFR* rs1801131, rs1801133; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582	*ENOSF1* rs37771109, *TYMS* rs1034044510	*EPX* rs35617692	*DGKI* rs142433066, *ZBTB45* rs200338006
CHUS7	diarrhoea	*DPYD* rs1801159, rs1801160; *UGT1A7* rs11692021; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *ABCB1* rs2032582			*MYBPC3* rs780907679, *COL4A2-AS2* rs753335766
CRC392	control	*ERCC1* rs11615; *UGT1A7* rs11692021; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *ABCB1* rs2032582; *UMPS* rs2279199			*MYBPC3* rs34580776
CRC404	control	*DPYD* rs17376848; *ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *ABCB1* rs2032582; *UMPS* rs2279199			*SCMH1* rs150272371,

					PDZD7 rs794727142
CRC481	control	*DPYD* rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *ABCB1* rs2032582; *UMPS* rs2279199			
HUIS4	diarrhoea	*ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			*MYBPC3* rs752200396
HUIS7	diarrhoea	*DPYD* rs1801159; *ERCC2* rs13181; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			
C001	control	*UGT1A7* rs11692021; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			
C028*	allergy	*UGT1A7* rs11692021; *ERCC2* rs13181; *MTHFR* rs1801133; *ABCB1* rs2032582; *UMPS* rs2279199	*ABCB1* rs36008564		
C047	control	*UGT1A7* rs11692021; *MTHFR* rs1801131; *UMPS* rs2279199			*MYBPC3* rs200663253
C120	control	*DPYD* rs1801159, rs2297595; *ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801131; *SLCO1B1* rs2306283; *ABCB1* rs2032582			
C124	control	*DPYD* rs2297595; *UGT1A7* rs11692021; *ERCC2* rs13181; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283	*CES2* rs147942040		*DGKI* rs1436918187
C136*	skin toxicity, nausea	*DPYD* rs1801159; *ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *MIR27A* rs895819; *ABCB1* rs2032582			*ARAP3* rs140445106, *STK31* rs138652787, *CLPB* rs150343959, *PARVB* rs56194750, *COL4A2-AS2* rs768746402
C144*	anorexia, skin toxicity	*DPYD* rs1801160; *ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801133; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819			*STK31* rs41273999, *CPT1A* rs140958507,

					COL4A2-AS2 rs113122708
C155	control	*DPYD* rs1801158, rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *MIR27A* rs895819; *ABCB1* rs2032582	*ABCC2* rs56131651		*ITPR3* rs2229633
C170	control	*DPYD* rs2297595; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			*PDZD7* rs1324603673
C175	control	*DPYD* rs1801159, rs2297595; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *MIR27A* rs895819; *ABCB1* rs2032582			*ARAP3* p.R853W
C177	control	*DPYD* rs1801160; *ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801131; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582			
C207	hepatotoxicity	*ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582; *UMPS* rs2279199			
C213	control	*DPYD* rs1801158; *ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801131, rs1801133; *SLCO1B1* rs2306283; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199	*ABCB1* rs55852620		*ZBTB45* rs1476542660
M053	control	*ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			*ITPR3*:p.N1242K
M085*	hepatotoxicity	*UGT1A7* rs11692021; *ERCC2* rs13181; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199	*CES2* rs140461033	*FAM83A* rs148011353	*DGKI* p.Q491Q, *SCN5A* rs41313691
M128	control	*DPYD* rs1801159; *ERCC2* rs13181; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			*CPT1A* p.N538S; *STK31* rs143223805
S11028*	asthenia	*DPYD* rs1801160; *UGT1A7* rs11692021; *MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582; *UMPS* rs2279199	*ABCC2* rs56131651		*DGKI* rs35245703
S11157	asthenia, nausea	*DPYD* rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801131; *SLCO1B1* rs2306283; *MIR27A* rs895819			*MYBPC3* rs397515956

S11622	neutropenia, asthenia	*DPYD* rs1801159; *ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCB1* rs2032582; *UMPS* rs2279199			*ARAP3* rs116475968, *CPT1A* rs61731903
S11693	neutropenia	*DPYD* rs2297595; *ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801133; *SLCO1B1* rs2306283; *MIR27A* rs895819; *UMPS* rs2279199, rs1801019			
S20122	neutropenia	*UGT1A7* rs11692021; *MTHFR* rs1801131; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582			*ITPR3* rs138821995
S20714*	asthenia, diarrhoea, nausea	*MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *ABCB1* rs2032582		*EPX* rs35617692	
S21854	neutropenia	*DPYD* rs1801159, rs1801160; *ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801133; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *UMPS* rs2279199			*ARAP3* rs752178354
S21876	neutropenia, mucositis	*DPYD* rs2297595; *ERCC2* rs13181; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582; *UMPS* rs2279199			*DGKI* rs527683728, *COL4A2-AS2* rs753335766
S22232	asthenia, diarrhoea	*DPYD* rs1801159; *UGT1A7* rs11692021; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *UMPS* rs2279199, rs1801019			*ITPR3* rs757118364
S22378	asthenia	*DPYD* rs2297595; *UGT1A7* rs11692021; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *UMPS* rs2279199			
S30403	diarrhoea	*DPYD* rs1801159, rs1801160; *ERCC1* rs11615; *UGT1A7* rs11692021; *MTHFR* rs1801131; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *UMPS* rs2279199			
S31104	neutropenia	*DPYD* rs1801159, rs2297595; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582			*CLPB* rs143097446
S40588*	asthenia, diarrhoea	*UGT1A7* rs11692021; *MTHFR* rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *MIR27A* rs895819; *ABCB1* rs2032582; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019	*CES2* rs780237910		*MYBPC3* rs750609594
S42688	neutropenia	*ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801131, rs1801133; *CDA* rs2072671, rs602950; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199			

S51254*	neutropenia	*DPYD* rs1801158, rs1801159; *UGT1A7* rs11692021; *SLCO1B1* rs2306283; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019		*EPX* rs35617692	
S51733	asthenia, nausea	*DPYD* rs1801158, rs1801160; *ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801131, rs1801133; *ABCC2* rs3740066, rs717620; UMPS rs2279199			
S51742	asthenia	*DPYD* rs1801159; *MTHFR* rs1801133; *SLCO1B1* rs2306283; *ABCC2* rs3740066, rs717620; *ABCB1* rs2032582			*DGKI* rs61751966
S60194*	nausea	*DPYD* rs1801159; *SLCO1B1* rs2306283; *ABCC2* rs3740066; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019		*FAM83A* rs148011353	*DGKI* p.R259Q
S70523*	neutropenia	*DPYD* rs1801159, rs1801160, rs2297595; *UGT1A7* rs11692021; *MTHFR* rs1801133; SLCO1B1 rs2306283; *ABCC2* rs3740066, rs717620; *MIR27A* rs895819	*MTHFR* rs45496998		*DGKI* p.S419T, *MYBPC3* rs11570052, *ARAP3* rs575984491
S71006	mucositis, asthenia, diarrhoea	*ERCC1* rs11615; *ERCC2* rs13181; *MTHFR* rs1801131; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			*CLPB* p.Y174C
S71910	neutropenia, asthenia	*DPYD* rs1801159, rs2297595; *ERCC1* rs11615; *ERCC2* rs13181; *UGT1A7* rs11692021; *MTHFR* rs1801131, rs1801133; *ABCC2* rs3740066; *ABCB1* rs2032582			*DGKI* rs970263048
S91219*	skin toxicity	*DPYD* rs17376848; *ERCC2* rs13181; *MTHFR* rs1801131; *ABCC2* rs3740066; *MIR27A* rs895819; *ABCB1* rs2032582; *UMPS* rs2279199, rs1801019			*SCMH1* rs114233776, *ARAP3* rs147992246, *CLPB* rs150343959, *ITPR3* rs41267659

Conclusions

From this thesis, and according to the established objectives, we can draw the following conclusions:

1. Identification of novel high/moderate-penetrance toxicity variants using WES analysis in a cohort of 163 cases and 52 controls with extreme toxicity phenotypes (Chapter 1).

We found 13 cases with extreme ADRs carrying actionable *DPYD* variants, emphasising the importance of pre-emptive testing. Moreover, we could not replicate any previously reported associations for the literature pharmacogenes but instead found significant associations with other ADRs: *ABCB1* rs2032582 and skin toxicity, *SLCO1B1* rs2306283 and neuropathy. Additionally, we found 31 novel rare, high-impact variants in 14 of the pharmacogenes. We also found seven patients carrying more than one variant in the same pharmacogene or pathway. For instance, one patient carried one variant in *TYMS* and another in *ENOSF1*, hinting at a potential digenic inheritance, which would be interesting to further investigate *in vitro*.

Using a hypothesis-free exome-wide approach, we discovered and independently validated three rare, high impact variants in novel genes (*ALDH9A1*, *FAM83A*, and *EPX*) that can potentially be responsible for toxicity. This warrants further validation in larger cohorts and *in vitro* studies to understand the functional consequences of these variants and their role in toxicity development. Gene-based analyses also provided a list of 14 genes significantly associated with neuropathy, skin toxicity, or cardiotoxicity. Altogether, the results from the high-penetrance variant analyses could potentially explain the toxicity observed for 58 of the patients in our extreme phenotypes cohort.

2. Omic-wide analyses of susceptibility to toxicity to determine low-penetrance toxicity variants under a polygenic risk model (Chapter 2).

We found no significant SNPs in our GWAS (meta) analyses. This might be justified by the reduced sample size and low power for each independent ADR, stressing the need for larger cohorts. However, we identified one borderline significant variant in the global toxicity GWAS, the *ITGAX* rs2230424, although it is unclear how variation in this gene could ultimately lead to increased toxicity.

GWAS data integration with transcriptomics in the TWAS approach identified however 20 novel genomic loci previously unreported to be linked with toxicity. These regions are influencing the toxicity risk through regulation of the expression of 24 candidate genes and warrant further study. Reassuringly, five of the literature-reported pharmacogenes - *DPYD*, *TYMP*, *ACYP2*, *UMPS*, and *UGT1A1* - were also nominally significant in this analysis, thereby validating the approach.

3. Functional evaluation of candidate toxicity variants (Chapter 3).

Functional characterisation of the novel *DPYD* variant - c.2071G>T, p.(V691L), rs202212118, including Western blotting protein abundance quantification and enzymatic activity assays could not unequivocally classify this variant as pathogenic, even though the observed metabolising capacity was 40%.

References

1. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians* (2021).
2. Morgan, E. *et al.* Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. *Gut* **72**, 338–344 (2023).
3. SEER 21 2014–2018, All Races, Both Sexes. https://seer.cancer.gov/statfacts/html/colorect.html.
4. Argilés, G. *et al.* Localised colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Annals of Oncology* **31**, 1291–1305 (2020).
5. Cervantes, A. *et al.* Metastatic colorectal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Annals of Oncology* **34**, 10–32 (2022).
6. Arnold, M. *et al.* Global patterns and trends in colorectal cancer incidence and mortality. *Gut* **66**, 683–691 (2017).
7. Sameer, A. S. Colorectal cancer: Molecular mutations and polymorphisms. *Frontiers in Oncology* **3**, 1–8 (2013).
8. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).
9. Tariq, K. & Ghias, K. Colorectal cancer carcinogenesis: a review of mechanisms. *Cancer Biology and Medicine* **13**, 120–135 (2016).
10. Grady, W. M. & Carethers, J. M. Genomic and Epigenetic Instability in Colorectal Cancer Pathogenesis. *Gastroenterology* **135**, 1079–1099 (2008).
11. Walther, A. *et al.* Genetic prognostic and predictive markers in colorectal cancer. *Nature Reviews Cancer* **9**, 489–499 (2009).
12. Ahnen, D. J. The American College of gastroenterology emily couric lecture the adenoma-carcinoma sequence revisited: Has the era of genetic tailoring finally arrived. *American Journal of Gastroenterology* **106**, 190–198 (2011).
13. Ewing, I., Hurley, J. J., Josephides, E. & Millar, A. The molecular genetics of colorectal cancer. *Annals of Gastroenterology* **27**, 9–14 (2014).
14. Weisenberger, D. J. *et al.* CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet* **38**, 787–793 (2006).
15. Haq, A. I., Schneeweiss, J., Kalsi, V. & Arya, M. The Dukes staging system: a cornerstone in the clinical management of colorectal cancer. *The Lancet Oncology* **10**, 1128 (2009).
16. Glynne-Jones, R. *et al.* Rectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology* **28**, v22-iv40 (2017).
17. Luchini, C. *et al.* ESMO recommendations on microsatellite instability testing for immunotherapy in cancer, and its relationship with PD-1/PD-L1 expression and tumour mutational burden: a systematic review-based approach. *Annals of Oncology* **30**, 1232–1243 (2019).
18. Labianca, R. *et al.* Early colon cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Annals of Oncology* **24**, vi64–vi72 (2013).
19. Billiau, A. COLORECTAL CANCER: A GUIDE FOR PATIENTS. *European Society for Medical Oncology* (2016).
20. Ansfield, F. J. & Heidelberger, C. Experimental and Clinical Use of Fluorinated Pyrimidines in Cancer Chemotherapy. *Cancer Research* **23**, 1226–1243 (1963).
21. Thorn, C. F. *et al.* Pharm GKB summary: Fluoropyrimidine pathways. *Pharmacogenetics and Genomics* **21**, 237–242 (2011).

22. Longley, D. B., Harkin, D. P. & Johnston, P. G. 5-Fluorouracil: Mechanisms of action and clinical strategies. *Nature Reviews Cancer* **3**, 330–338 (2003).
23. Saif, M. W., Katirtzoglou, N. A. & Syrigos, K. N. Capecitabine: An overview of the side effects and their management. *Anti-Cancer Drugs* **19**, 447–464 (2008).
24. Whirl-Carrillo, M. *et al.* Pharmacogenomics knowledge for personalized medicine. *Clinical Pharmacology and Therapeutics* **92**, 414–417 (2012).
25. Zhao, B. *et al.* Mechanisms of resistance to anti-EGFR therapy in colorectal cancer. *Oncotarget* **8**, 3980–4000 (2017).
26. Baldwin, J. FDA Evaluating Oxaliplatin for Advanced Colorectal Cancer Treatment. *Journal of the National Cancer Institute* **94**, 1191–1193 (2002).
27. Food and Drug Administration. HIGHLIGHTS OF PRESCRIBING INFORMATION. Eloxatin. (2002).
28. Graham, J., Muhsin, M. & Kirkpatrick, P. Oxaliplatin. *Nature Reviews Drug Discovery* **3**, 1–2 (2004).
29. ESMO. ESMO Clinical Practice Guidelines: Gastrointestinal Cancers. https://www.esmo.org/guidelines/guidelines-by-topic/gastrointestinal-cancers.
30. Saijo, N. Clinical Trials of Irinotecan Hydrochloride (CPT, Campto Injection, Topotecin Injection) in Japan. *Annals of the New York Academy of Sciences* **803**, 292–305 (1996).
31. Marsh, S. *et al.* Platinum pathway. *Pharmacogenetics and Genomics* **19**, 563–564 (2009).
32. Lambrechts, D. *et al.* Genetic markers of bevacizumab-induced hypertension. *Angiogenesis* **17**, 685–694 (2014).
33. Wookey, V. & Grothey, A. Update on the role of pembrolizumab in patients with unresectable or metastatic colorectal cancer. *Therap Adv Gastroenterol* **14**, (2021).
34. Food and Drug Administration. HIGHLIGHTS OF PRESCRIBING INFORMATION. Xeloda. (2002).
35. Food and Drug Administration. HIGHLIGHTS OF PRESCRIBING INFORMATION. Oxaliplatin. (2002).
36. Food and Drug Administration. HIGHLIGHTS OF PRESCRIBING INFORMATION. Irinotecan. (2014).
37. Food and Drug Administration. HIGHLIGHTS OF PRESCRIBING INFORMATION. Cetuximab. (2009).
38. Food and Drug Administration. HIGHLIGHTS OF PRESCRIBING INFORMATION. Panitumumab. (2015).
39. Food and Drug Administration. HIGHLIGHTS OF PRESCRIBING INFORMATION. Bevacizumab. (2017).
40. Food and Drug Administration. HIGHLIGHTS OF PRESCRIBING INFORMATION. Pembrolizumab. (2014).
41. Food and Drug Administration. HIGHLIGHTS OF PRESCRIBING INFORMATION. Xeloda. (2015).
42. Roden, D. M. & George, A. L. The genetic basis of variability in drug responses. *Nature Reviews Drug Discovery* **1**, 37–44 (2002).
43. Motulsky, A. G. Drug reactions, enzymes, and biochemical genetics. *Journal of the American Medical Association* **165**, 835–837 (1957).
44. Watters, J. W., Kraja, A., Meucci, M. A., Province, M. A. & McLeod, H. L. Genome-wide discovery of loci influencing chemotherapy cytotoxicity. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 11809–11814 (2004).

45. Lièvre, A. *et al.* KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. *Cancer Research* **66**, 3992–3995 (2006).
46. Taieb, J. *et al.* Prognostic Value of BRAF and KRAS Mutations in MSI and MSS Stage III Colon Cancer. *JNCI: Journal of the National Cancer Institute* **109**, djw272 (2017).
47. Simões, A. R., Fernández-Rozadilla, C., Maroñas, O. & Carracedo, Á. The road so far in colorectal cancer pharmacogenomics: Are we closer to individualised treatment? *Journal of Personalized Medicine* **10**, 1–40 (2020).
48. Berger, R. *et al.* Dihydropyrimidine dehydrogenase deficiency leading to thymine-uraciluria. An inborn error of pyrimidine metabolism. *Clinica chimica acta; international journal of clinical chemistry* **141**, 227–234 (1984).
49. Diasio, R. B., Beavers, T. L. & Carpenter, J. T. Familial deficiency of dihydropyrimidine dehydrogenase. Biochemical basis for familial pyrimidinemia and severe 5-fluorouracil-induced toxicity. *Journal of Clinical Investigation* **81**, 47–51 (1988).
50. Van Kuilenburg, A. B. P. *et al.* Genotype and phenotype in patients with dihydropyrimidine dehydrogenase deficiency. *Human Genetics* **104**, 1–9 (1999).
51. Vreken, P. *et al.* A point mutation in an invariant splice donor site leads to exon skipping in two unrelated Dutch patients with dihydropyrimidine dehydrogenase deficiency. *J Inherit Metab Dis* **19**, 645–654 (1996).
52. Henricks, L. M. *et al.* Translating DPYD genotype into DPD phenotype: using the DPYD gene activity score. *Pharmacogenomics* **16**, 1277–1286 (2015).
53. Offer, S. M. *et al.* Comparative functional analysis of DPYD variants of potential clinical relevance to dihydropyrimidine dehydrogenase activity. *Cancer Research* **74**, 2545–2545 (2014).
54. Van Kuilenburg, A. B. P. *et al.* Novel disease-causing mutations in the dihydropyrimidine dehydrogenase gene interpreted by analysis of the three-dimensional protein structure. *Biochemical Journal* **364**, 157–163 (2002).
55. Offer, S. M., Wegner, N. J., Fossum, C., Wang, K. & Diasio, R. B. Phenotypic profiling of DPYD variations relevant to 5-fluorouracil sensitivity using real-time cellular analysis and in vitro measurement of enzyme activity. *Cancer Research* **73**, 1958–1968 (2013).
56. Henricks, L. M., Opdam, F. L., Beijnen, J. H., Cats, A. & Schellens, J. H. M. DPYD genotype-guided dose individualization to improve patient safety of fluoropyrimidine therapy: Call for a drug label update. *Annals of Oncology* **19**, 1459–1467 (2017).
57. Barbarino, J. M., Haidar, C. E., Klein, T. E. & Altman, R. B. PharmGKB summary: Very important pharmacogene information for UGT1A1. *Pharmacogenetics and Genomics* **24**, 177–183 (2014).
58. Innocenti, F. *et al.* Genetic variants in the UDP-glucuronosyltransferase 1A1 gene predict the risk of severe neutropenia of irinotecan. *Journal of Clinical Oncology* **22**, 1382–1388 (2004).
59. Massacesi, C. *et al.* Uridine diphosphate glucuronosyl transferase 1A1 promoter polymorphism predicts the risk of gastrointestinal toxicity and fatigue induced by irinotecan-based chemotherapy. *Cancer* **106**, 1007–1016 (2006).
60. Iyer, L. *et al.* UGT1A1*28 polymorphism as a determinant of irinotecan disposition and toxicity. *Pharmacogenomics Journal* **2**, 43–47 (2002).
61. Liu, C. Y. *et al.* UGT1A1*28 polymorphism predicts irinotecan-induced severe toxicities without affecting treatment outcome and survival in patients with metastatic colorectal carcinoma. *Cancer* **112**, 1932–1940 (2008).

62. Liu, X., Cheng, D., Kuang, Q., Liu, G. & Xu, W. Association of UGT1A1*28 polymorphisms with irinotecan-induced toxicities in colorectal cancer: A meta-analysis in Caucasians. *Pharmacogenomics Journal* **14**, 120–129 (2014).
63. Pritchard, J. K. & Cox, N. J. The allelic architecture of human disease genes: common disease-common variant. .. or not? *Human Molecular Genetics* **11**, 2417–2423 (2002).
64. Wall, J. D. & Pritchard, J. K. Haplotype blocks and linkage disequilibrium in the human genome. *Nature Reviews Genetics* **4**, 587–597 (2003).
65. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
66. Musunuru, K. & Kathiresan, S. HapMap and mapping genes for cardiovascular disease. *Circ Cardiovasc Genet* **1**, 66–71 (2008).
67. Rosmarin, D. *et al.* A candidate gene study of capecitabine-related toxicity in colorectal cancer identifies new toxicity variants at DPYD and a putative role for ENOSF1 rather than TYMS. *Gut* **64**, 111–120 (2015).
68. Fernandez-Rozadilla, C. *et al.* Pharmacogenomics in colorectal cancer: A genome-wide association study to predict toxicity after 5-fluorouracil or FOLFOX administration. *Pharmacogenomics Journal* **13**, 209–217 (2013).
69. Won, H. H. *et al.* Polymorphic markers associated with severe oxaliplatin-induced, chronic peripheral neuropathy in colon cancer patients. *Cancer* **118**, 2828–2836 (2012).
70. Baas, J. *et al.* Genome wide association study to identify predictors for severe skin toxicity in colorectal cancer patients treated with cetuximab. *PLoS ONE* **13**, e0208080 (2018).
71. Oguri, T. *et al.* Genetic polymorphisms associated with oxaliplatin-induced peripheral neurotoxicity in Japanese patients with colorectal cancer. *International Journal of Clinical Pharmacology and Therapeutics* **51**, 475–481 (2013).
72. Terrazzino, S. *et al.* Genetic determinants of chronic oxaliplatin-induced peripheral neurotoxicity: A genome-wide study replication and meta-analysis. *Journal of the Peripheral Nervous System* **20**, 15–23 (2015).
73. Mizzi, C. *et al.* Personalized pharmacogenomics profiling using whole-genome sequencing. *Pharmacogenomics* **15**, 1223–1234 (2014).
74. Apellániz-Ruiz, M. *et al.* Whole-exome sequencing reveals defective CYP3A4 variants predictive of paclitaxel dose-limiting neuropathy. in *Clinical Cancer Research* vol. 21 322–328 (2015).
75. Kozyra, M., Ingelman-Sundberg, M. & Lauschke, V. M. Rare genetic variants in cellular transporters, metabolic enzymes, and nuclear receptors can be important determinants of interindividual differences in drug response. *Genetics in Medicine* **19**, 20–29 (2017).
76. Santos, M. *et al.* Novel copy-number variations in pharmacogenes contribute to interindividual differences in drug pharmacokinetics. *Genetics in Medicine* **20**, 622–629 (2018).
77. Ingelman-Sundberg, M., Mkrtchian, S., Zhou, Y. & Lauschke, V. M. Integrating rare genetic variants into pharmacogenetic drug response predictions. *Human Genomics* **12**, (2018).
78. Schaller, L. & Lauschke, V. M. The genetic landscape of the human solute carrier (SLC) transporter superfamily. *Human Genetics* **138**, 1359–1377 (2019).
79. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research* **47**, D1005–D1012 (2019).

80. Nicolae, D. L. *et al.* Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. *PLoS Genet* **6**, e1000888 (2010).
81. Fernandez-Rozadilla, C. *et al.* Deciphering colorectal cancer genetics through multi-omic analysis of 100,204 cases and 154,587 controls of European and east Asian ancestries. *Nat Genet* **55**, 89–99 (2023).
82. Wainberg, M. *et al.* Opportunities and challenges for transcriptome-wide association studies. *Nat Genet* **51**, 592–599 (2019).
83. Li, B. & Ritchie, M. D. From GWAS to Gene: Transcriptome-Wide Association Studies and Other Methods to Functionally Understand GWAS Discoveries. *Frontiers in Genetics* **12**, (2021).
84. Guo, X. *et al.* Identifying Novel Susceptibility Genes for Colorectal Cancer Risk From a Transcriptome-Wide Association Study of 125,478 Subjects. *Gastroenterology* **160**, 1164–1178 (2021).
85. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
86. Yuan, Y. *et al.* Multi-omics analysis to identify susceptibility genes for colorectal cancer. *Human Molecular Genetics* **30**, 321–330 (2021).
87. Bien, S. A. *et al.* Genetic variant predictors of gene expression provide new insight into risk of colorectal cancer. *Hum Genet* **138**, 307–326 (2019).
88. Agencia Española de Medicamentos y Produtos Sanitarios. Fluorouracilo, capecitabina, tegafur y flucitosina en pacientes con déficit de dihidropirimidina deshidrogenasa. https://www.aemps.gob.es/informa/notasinformativas/medicamentosusohumano-3/seguridad-1/2020-seguridad-1/fluorouracilo-capecitabina-tegafur-y-flucitosina-en-pacientes-con-deficit-de-dihidropirimidina-deshidrogenasa/ (2020).
89. Amstutz, U. *et al.* Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Dihydropyrimidine Dehydrogenase Genotype and Fluoropyrimidine Dosing: 2017 Update. *Clinical Pharmacology & Therapeutics* **13**, 210–216 (2017).
90. The Pharmacogenetics Working Group. UGT1A1: irinotecan. (2018).
91. European Medicines Agency. EMA recommendations on DPD testing prior to treatment with fluorouracil, capecitabine, tegafur and flucytosine. (2020).
92. Food and Drug Administration. Table of Pharmacogenomic Biomarkers in Drug Labeling 2021. (2021).
93. European Medicines Agency. Erbitux (cetuximab).
94. Trotti, A. *et al.* CTCAE v3.0: Development of a comprehensive grading system for the adverse effects of cancer treatment. *Seminars in Radiation Oncology* **13**, 176–181 (2003).
95. Singleton, A. B., Hardy, J., Traynor, B. J. & Houlden, H. Towards a complete resolution of the genetic architecture of disease. *Trends Genet* **26**, 438–442 (2010).
96. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
97. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *Clinical Review & Education* **310**, 2191–2194 (2013).
98. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
99. Van der Auwera GA & O'Connor BD. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. (O'Reilly Media, 2020).

100. Depristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43**, 491–501 (2011).
101. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
102. Whirl-Carrillo, M. *et al.* An Evidence-Based Framework for Evaluating Pharmacogenomics Knowledge for Personalized Medicine. *Clin Pharmacol Ther* **110**, 563–572 (2021).
103. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature Methods* **7**, 248–249 (2010).
104. Davydov, E. V. *et al.* Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Computational Biology* **6**, (2010).
105. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics* **46**, 310–315 (2014).
106. Quang, D., Chen, Y. & Xie, X. DANN: A deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* **31**, 761–763 (2015).
107. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet* **81**, 559–575 (2007).
108. Zhan, X., Hu, Y., Li, B., Abecasis, G. R. & Liu, D. J. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* **32**, 1423–1426 (2016).
109. Lee, S. *et al.* Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *American Journal of Human Genetics* **91**, 224–237 (2012).
110. RStudio Team (2020). RStudio: Integrated Development for R. *RStudio, PBC, Boston, MA*.
111. Traylor, M. *et al.* Exome array analysis of adverse reactions to fluoropyrimidine-based therapy for gastrointestinal cancer. *PLoS ONE* (2018).
112. Hirschhorn, J. N., Lohmueller, K., Byrne, E. & Hirschhorn, K. A comprehensive review of genetic association studies. *Genetics in Medicine* **4**, 45–61 (2002).
113. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics* (2016).
114. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
115. Barbeira, A. N. *et al.* Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome Biology* **22**, 1–24 (2021).
116. Battle, A. *et al.* Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Research* **24**, 14–24 (2014).
117. Dampier, C. H. *et al.* Oncogenic features in histologically normal mucosa: Novel insights into field effect from a mega-analysis of colorectal transcriptomes. *Clinical and Translational Gastroenterology* (2020).
118. Theodoratou, E. *et al.* Dietary Flavonoids and the Risk of Colorectal Cancer. *Cancer Epidemiol Biomarkers Prev* **16**, 684–93 (2007).
119. Fernandez-Rozadilla, C. *et al.* Telomere length and genetics are independent colorectal tumour risk factors in an evaluation of biomarkers in normal bowel. *British Journal of Cancer* **118**, 727–732 (2018).
120. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nature Genetics 2015 47:9* **47**, 1091–1098 (2015).

121. Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nature Communications* **9**, 1–20 (2018).
122. Barbeira, A. N. *et al.* Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS Genetics* **15**, (2019).
123. Shrestha, S. *et al.* Gene-Specific Variant Classifier (DPYD-Varifier) to Identify Deleterious Alleles of Dihydropyrimidine Dehydrogenase. *Clinical Pharmacology and Therapeutics* **104**, 709–718 (2018).
124. Chen, S. *et al.* ABCC5 and ABCG1 polymorphisms predict irinotecan-induced severe toxicity in metastatic colorectal cancer patients. *Pharmacogenetics and Genomics* **25**, 573–583 (2015).
125. Di Martino, M. T. *et al.* Single nucleotide polymorphisms of ABCC5 and ABCG1 transporter genes correlate to irinotecan-associated gastrointestinal toxicity in colorectal cancer patients: A DMET microarray profiling study. *Cancer Biology and Therapy* **12**, 780–787 (2011).
126. Gonzalez-Haba, E. *et al.* ABCB1 gene polymorphisms are associated with adverse reactions in fluoropyrimidine-treated colorectal cancer patients. *Pharmacogenomics* **11**, 1715–1723 (2010).
127. De Jong, F. A. *et al.* Irinotecan-induced diarrhea: Functional significance of the polymorphic ABCC2 transporter protein. *Clinical Pharmacology and Therapeutics* **81**, 42–49 (2007).
128. Cortejoso, L. & Lopez-Fernandez, L. A. Pharmacogenetic markers of toxicity for chemotherapy in colorectal cancer patients. *Pharmacogenomics* **13**, 1173–1191 (2012).
129. Ab Mutalib, N. S., Md Yusof, N. F., Abdul, S. N. & Jamal, R. Pharmacogenomics DNA biomarkers in colorectal cancer: Current update. *Frontiers in Pharmacology* **8**, 1–8 (2017).
130. Offer, S. M. *et al.* A DPYD variant (Y186C) in individuals of african ancestry is associated with reduced DPD enzyme activity. *Clinical Pharmacology and Therapeutics* (2013) doi:10.1038/clpt.2013.69.
131. Pellicer, M. *et al.* Identification of new SNPs associated with severe toxicity to capecitabine. *Pharmacological Research* **120**, 133–137 (2017).
132. Gross, E. *et al.* Detailed analysis of five mutations in dihydropyrimidine dehydrogenase detected in cancer patients with 5-fluorouracil-related side effects. *Human Mutation* **22**, 498–498 (2003).
133. Caudle, K. E. *et al.* Clinical pharmacogenetics implementation consortium guidelines for dihydropyrimidine dehydrogenase genotype and fluoropyrimidine dosing. *Clinical Pharmacology and Therapeutics* **94**, 640–645 (2013).
134. Meulendijks, D. *et al.* Rs895819 in MIR27A improves the predictive value of DPYD variants to identify patients at risk of severe fluoropyrimidine-associated toxicity. *International Journal of Cancer* **138**, 2752–2761 (2016).
135. Lévesque, É. *et al.* Refining the UGT1A haplotype associated with irinotecan-induced hematological toxicity in metastatic colorectal cancer patients treated with 5-fluorouracil/irinotecan-based regimens. *Journal of Pharmacology and Experimental Therapeutics* **345**, 95–101 (2013).
136. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).

137. Samocha, K. E. *et al.* A framework for the interpretation of de novo mutation in human disease. *Nature Genetics* **46**, 944–950 (2014).
138. Statello, L., Guo, C.-J., Chen, L.-L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol* **22**, 96–118 (2021).
139. Richards, S. *et al.* Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine : official journal of the American College of Medical Genetics* **17**, (2015).
140. Pullarkat, S. T. & Lenz, H. J. Thymidylate synthase gene polymorphism determines response and toxicity of 5-FU chemotherapy. *Pharmacogenomics Journal* **1**, 65–70 (2001).
141. Lecomte, T. *et al.* Thymidylate synthase gene polymorphism predicts toxicity in colorectal cancer patients receiving 5-fluorouracil-based chemotherapy. *Clinical Cancer Research* **10**, 5880–5888 (2004).
142. Castro-Rojas, C. A. *et al.* Thymidylate synthase gene variants as predictors of clinical response and toxicity to fluoropyrimidine-based chemotherapy for colorectal cancer. *Drug Metabolism and Personalized Therapy* **32**, 209–218 (2017).
143. Hamzic, S. *et al.* Evaluating the role of ENOSF1 and TYMS variants as predictors in fluoropyrimidine-related toxicities: An IPD meta-analysis. *Pharmacological Research* **152**, 104594–104594 (2020).
144. Wu, Q. & Dolnick, B. J. Detection of thymidylate synthase modulators by a novel screening assay. *Mol Pharmacol* **63**, 167–173 (2003).
145. Vasiliou, V. & Pappa, A. Polymorphisms of Human Aldehyde Dehydrogenases. *Pharmacology* **61**, 192–198 (2000).
146. Zanoni, M., Bravaccini, S., Fabbri, F. & Arienti, C. Emerging Roles of Aldehyde Dehydrogenase Isoforms in Anti-cancer Therapy Resistance. *Frontiers in medicine* **9**, 1–10 (2022).
147. Mizukami, T., Izawa, N., Nakajima, T. E. & Sunakawa, Y. Targeting EGFR and RAS/RAF Signaling in the Treatment of Metastatic Colorectal Cancer: From Current Treatment Strategies to Future Perspectives. *Drugs* **79**, 633–645 (2019).
148. Chen, S. *et al.* FAM83A is amplified and promotes cancer stem cell-like traits and chemoresistance in pancreatic cancer. *Oncogenesis* **6**, 1–10 (2017).
149. Bartel, C. A., Parameswaran, N., Cipriano, R. & Jackson, M. W. FAM83 proteins: Fostering new interactions to drive oncogenic signaling and therapeutic resistance. *Oncotarget* **7**, 52597–52612 (2016).
150. Acharya, K. R. & Ackerman, S. J. Eosinophil granule proteins: Form and function. *Journal of Biological Chemistry* **289**, 17406–17415 (2014).
151. Romano, M., Baralle, F. E. & Patriarca, P. Expression and characterization of recombinant human eosinophil peroxidase. Impact of the R286H substitution on the biosynthesis and activity of the enzyme. *European Journal of Biochemistry* **267**, 3704–3711 (2000).
152. The Human Protein Atlas. https://www.proteinatlas.org/ENSG00000157680-DGKI/single+cell+type.
153. Sato, M. *et al.* Evaluations of the Selectivities of the Diacylglycerol Kinase Inhibitors R59022 and R59949 Among Diacylglycerol Kinase Isozymes Using a New Non-Radioactive Assay Method. *Pharmacology* **92**, 99–107 (2013).

154. Ding, L., Traer, E., McIntyre, T. M., Zimmerman, G. A. & Prescott, S. M. The Cloning and Characterization of a Novel Human Diacylglycerol Kinase, DGKι. *Journal of Biological Chemistry* **273**, 32746–32752 (1998).
155. The Human Protein Atlas. https://www.proteinatlas.org/ENSG00000183873-SCN5A/single+cell+type.
156. Boyette-Davis, J. A., Hou, S., Abdi, S. & Dougherty, P. M. An updated understanding of the mechanisms involved in chemotherapy-induced neuropathy. *Pain Manag* **8**, 363–375 (2018).
157. Adelsberger, H. *et al.* The chemotherapeutic oxaliplatin alters voltage-gated Na(+) channel kinetics on rat sensory neurons. *Eur J Pharmacol* **406**, 25–32 (2000).
158. Law, P. J. *et al.* Association analyses identify 31 new risk loci for colorectal cancer susceptibility. *Nat Commun* **10**, 2154 (2019).
159. Bao, D. *et al.* Integrative Analysis of Complement System to Prognosis and Immune Infiltrating in Colon Cancer and Gastric Cancer. *Front Oncol* **10**, 553297 (2021).
160. Liu, W., Zhuang, Z., Wang, W., Huang, T. & Liu, Z. An Improved Genome-Wide Polygenic Score Model for Predicting the Risk of Type 2 Diabetes. *Frontiers in Genetics* **12**, (2021).
161. Garcia-Etxebarria, K. *et al.* Local genetic variation of inflammatory bowel disease in Basque population and its effect in risk prediction. *Sci Rep* **12**, 3386 (2022).
162. Klarin, D. & Natarajan, P. Clinical utility of polygenic risk scores for coronary artery disease. *Nat Rev Cardiol* **19**, 291–301 (2022).
163. Adeyemo, A. *et al.* Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. *Nat Med* **27**, 1876–1884 (2021).
164. Zhao, B. *et al.* Transcriptome-wide association analysis of brain structures yields insights into pleiotropy with complex neuropsychiatric traits. *Nat Commun* **12**, 2878 (2021).
165. Marigorta, U. M. *et al.* Transcriptional risk scores link GWAS to eQTLs and predict complications in Crohn's disease. *Nat Genet* **49**, 1517–1521 (2017).
166. Liang, Y. Obtaining polygenic transcriptome risk scores (PTRS) directly from GWAS summary statistics. 2022.07.10.499509 Preprint at https://doi.org/10.1101/2022.07.10.499509 (2022).
167. Toffoli, G. *et al.* Clinical validity of a DPYD-based pharmacogenetic test to predict severe toxicity to fluoropyrimidines. *International Journal of Cancer* **137**, 2971–2980 (2015).
168. Meulendijks, D. *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. *The Lancet Oncology* **16**, 1639–1650 (2015).
169. Falvella, F. S. *et al.* DPD and UGT1A1 deficiency in colorectal cancer patients receiving triplet chemotherapy with fluoropyrimidines, oxaliplatin and irinotecan. *British Journal of Clinical Pharmacology* (2015) doi:10.1111/bcp.12631.
170. O'Donnell, P. H. *et al.* Clinical evaluation of germline polymorphisms associated with capecitabine toxicity in breast cancer: TBCRC-015. *Breast Cancer Research and Treatment* (2020) doi:10.1007/s10549-020-05603-8.
171. Madi, A. *et al.* Pharmacogenetic analyses of 2183 patients with advanced colorectal cancer; potential role for common dihydropyrimidine dehydrogenase variants in toxicity to chemotherapy. *European Journal of Cancer* (2018) doi:10.1016/j.ejca.2018.07.009.
172. Ruzzo, A. *et al.* Dihydropyrimidine dehydrogenase pharmacogenetics for predicting fluoropyrimidine-related toxicity in the randomised, phase III adjuvant TOSCA trial in

high-risk colon cancer patients. *British Journal of Cancer* 1–9 (2017) doi:10.1038/bjc.2017.289.
173. Zaanan, A., Dumont, L. M., Loriot, M. A., Taieb, J. & Narjoz, C. A case of 5-FU-related severe toxicity associated with the p.Y186C DPYD variant. *Clinical Pharmacology and Therapeutics* (2014) doi:10.1038/clpt.2013.183.
174. Nie, Q. *et al.* Quantitative Contribution of rs75017182 to Dihydropyrimidine Dehydrogenase mRNA Splicing and Enzyme Activity. *Clinical Pharmacology and Therapeutics* (2017) doi:10.1002/cpt.685.
175. Kouwaki, M. *et al.* Identification of novel mutations in the dihydropyrimidine dehydrogenase gene in a Japanese patient with 5-fluorouracil toxicity. *Clinical Cancer Research* (1998).
176. Vreken, P., Van Kuilenburg, A. B. P., Meinsma, R. & Van Gennip, A. H. Identification of novel point mutations in the dihydropyrimidine dehydrogenase gene. in *Journal of Inherited Metabolic Disease* (1997). doi:10.1023/A:1005357307122.
177. Meulendijks, D. *et al.* Increased risk of severe fluoropyrimidine-associated toxicity in patients carrying a G to C substitution in the first 28-bp tandem repeat of the thymidylate synthase 2R allele. *International Journal of Cancer* (2016) doi:10.1002/ijc.29694.
178. Kristensen, M. H., Pedersen, P. L., Melsen, G. V., Ellehauge, J. & Mejer, J. Variants in the dihydropyrimidine dehydrogenase, methylenetetrahydrofolate reductase and thymidylate synthase genes predict early toxicity of 5-fluorouracil in colorectal cancer patients. *Journal of International Medical Research* **38**, 870–883 (2010).
179. Rosmarin, D. *et al.* Genetic markers of toxicity from capecitabine and other fluorouracil-based regimens: Investigation in the QUASAR2 study, systematic review, and meta-analysis. *Journal of Clinical Oncology* **32**, 1031–1039 (2014).
180. Tsunoda, A. *et al.* Associations of various gene polymorphisms with toxicity in colorectal cancer patients receiving oral uracil and tegafur plus leucovorin: A prospective study. *Annals of Oncology* (2011) doi:10.1093/annonc/mdq358.
181. Loganayagam, A. *et al.* Pharmacogenetic variants in the DPYD, TYMS, CDA and MTHFR genes are clinically significant predictors of fluoropyrimidine toxicity. *British Journal of Cancer* **108**, 2505–2515 (2013).
182. Lee, K. H. *et al.* Pharmacogenetic analysis of adjuvant FOLFOX for Korean patients with colon cancer. *Cancer Chemotherapy and Pharmacology* **71**, 843–851 (2013).
183. Jennings, B. A. *et al.* Evaluating Predictive Pharmacogenetic Signatures of Adverse Events in Colorectal Cancer Patients Treated with Fluoropyrimidines. *PLoS ONE* **8**, 1–9 (2013).
184. Magdy, T. *et al.* ABCC11/MRP8 polymorphisms affect 5-fluorouracil-induced severe toxicity and hepatic expression. *Pharmacogenomics* **14**, 1433–1448 (2013).
185. García-González, X. *et al.* Variants in CDA and ABCB1 are predictors of capecitabine-related adverse reactions in colorectal cancer. *Oncotarget* **6**, 6422–6430 (2015).
186. Ribelles, N. *et al.* A Carboxylesterase 2 Gene Polymorphism as Predictor of Capecitabine on Response and Time to Progression. *Current Drug Metabolism* **9**, 336–343 (2008).
187. Hamzic, S. *et al.* Novel Genetic Variants in Carboxylesterase 1 Predict Severe Early-Onset Capecitabine-Related Toxicity. *Clinical Pharmacology and Therapeutics* (2017) doi:10.1002/cpt.641.
188. Wheeler, H. E. *et al.* Identification of genetic variants associated with capecitabine-induced hand-foot syndrome through integration of patient and cell line genomic analyses. *Pharmacogenetics and Genomics* **24**, 231–237 (2014).

189. Cecchin, E. *et al.* A prospective validation pharmacogenomic study in the adjuvant setting of colorectal cancer patients treated with the 5-fluorouracil/leucovorin/ oxaliplatin (FOLFOX4) regimen. *Pharmacogenomics Journal* (2013) doi:10.1038/tpj.2012.31.
190. Custodio, A. *et al.* Pharmacogenetic predictors of severe peripheral neuropathy in colon cancer patients treated with oxaliplatin-based adjuvant chemotherapy: A GEMCAD group study. *Annals of Oncology* **25**, 398–403 (2014).
191. Stoehlmacher, J. *et al.* A multivariate analysis of genomic polymorphisms: Prediction of clinical outcome to 5-FU/oxaliplatin combination chemotherapy in refractory colorectal cancer. *British Journal of Cancer* **91**, 344–354 (2004).
192. Ruzzo, A. *et al.* Genetic markers for toxicity of adjuvant oxaliplatin and fluoropyrimidines in the phase III TOSCA trial in high-risk colon cancer patients. *Scientific Reports* **4**, 1–7 (2014).
193. Boige, V. *et al.* Pharmacogenetic assessment of toxicity and outcome in patients with metastatic colorectal cancer treated with LV5FU2, FOLFOX, and FOLFIRI: FFCD 2000-05. *Journal of Clinical Oncology* **28**, 2556–2564 (2010).
194. Salimzadeh, H., Lindskog, E. B., Gustavsson, B., Wettergren, Y. & Ljungman, D. Association of DNA repair gene variants with colorectal cancer: Risk, toxicity, and survival. *BMC Cancer* **20**, (2020).
195. Ando, Y. *et al.* Polymorphisms of UDP-glucuronosyltransferase gene and irinotecan toxicity: A pharmacogenetic analysis. *Cancer Research* **60**, 6921–6926 (2000).
196. Onoue, M. *et al.* UGT1A1*6 polymorphism is most predictive of severe neutropenia induced by irinotecan in Japanese cancer patients. *International Journal of Clinical Oncology* (2009) doi:10.1007/s10147-008-0821-z.
197. Chen, S. *et al.* A novel UGT1 marker associated with better tolerance against irinotecan-induced severe neutropenia in metastatic colorectal cancer patients. *Pharmacogenomics Journal* **15**, 513–520 (2015).
198. Côté, J. F. *et al.* UGT1A1 polymorphism can predict hematologic toxicity in patients treated with irinotecan. *Clinical Cancer Research* **13**, 3269–3275 (2007).
199. Li, M. *et al.* ABC transporter polymorphisms are associated with irinotecan pharmacokinetics and neutropenia. *Pharmacogenomics Journal* **18**, 35–42 (2018).
200. De Mattia, E. *et al.* Pharmacogenetics of ABC and SLC transporters in metastatic colorectal cancer patients receiving first-line FOLFIRI treatment. *Pharmacogenetics and Genomics* **23**, 549–557 (2013).
201. Hoskins, J. M. *et al.* Irinotecan pharmacogenetics: Influence of pharmacodynamic genes. *Clinical Cancer Research* (2008) doi:10.1158/1078-0432.CCR-07-1472.
202. Li, J. *et al.* A novel genetic score model of UGT1A1 and TGFB pathway as predictor of severe irinotecan-related diarrhea in metastatic colorectal cancer patients. *Journal of Cancer Research and Clinical Oncology* (2016) doi:10.1007/s00432-016-2176-6.
203. Takahashi, H. *et al.* Application of a combination of a knowledge-based algorithm and 2-stage screening to hypothesis-free genomic data on irinotecan-treated patients for identification of a candidate single nucleotide polymorphism related to an adverse effect. *PLoS ONE* (2014) doi:10.1371/journal.pone.0105160.
204. Etienne-Grimaldi, M. C. *et al.* Multifactorial pharmacogenetic analysis in colorectal cancer patients receiving 5-fluorouracil-based therapy together with cetuximab-irinotecan. *British Journal of Clinical Pharmacology* **73**, 776–785 (2012).
205. Parmar, S. *et al.* Pharmacogenetic predictors for EGFR-inhibitor-associated skin toxicity. *Pharmacogenomics Journal* **13**, 181–188 (2013).

206. Dahan, L. *et al.* Pharmacogenetic profiling and cetuximab outcome in patients with advanced colorectal cancer. *BMC Cancer* **11**, (2011).
207. Sibertin-Blanc, C. *et al.* Vascular Endothelial Growth Factor A c.*237C>T polymorphism is associated with bevacizumab efficacy and related hypertension in metastatic colorectal cancer. *Digestive and Liver Disease* **47**, 331–337 (2015).
208. Schneider, B. P. *et al.* Association of vascular endothelial growth factor and vascular endothelial growth factor receptor-2 genetic polymorphisms with outcome in a trial of paclitaxel compared with paclitaxel plus bevacizumab in advanced breast cancer: ECOG 2100. *Journal of Clinical Oncology* **26**, 4672–4678 (2008).
209. Morita, S. *et al.* Association between bevacizumab-related hypertension and vascular endothelial growth factor (VEGF) gene polymorphisms in Japanese patients with metastatic colorectal cancer. *Cancer Chemotherapy and Pharmacology* **71**, 405–411 (2013).

Supplementary Tables

Supplementary Table S1. Toxicity grading criteria according to the CTCAE v3.0 (adapted from Trotti et al.[94]). Only the ADRs present in our patients are described.

Adverse Drug Reaction	1	2	3	4	5
BLOOD/BONE MARROW					
Lymphopenia	<LLN – 800/mm^3 <LLN x0.8 – 10^9/L	<800 – 500/mm^3 <0.8 – 0.5x10^9/L	<500 – 200 mm^3 <0.5 – 0.2x10^9/L	<200/mm^3 <0.2x10^9/L	Death
Leucopenia	<LLN – 3000/mm^3 <LLN – 3.0x10^9/L	<3000 – 2000/mm^3 <3.0 – 2.0x10^9/L	<2000 – 1000/mm^3 <2.0 – 1.0x10^9/L	<1000/mm^3 <1.0x10^9/L	Death
Neutropenia	<LLN – 1500/mm^3 <LLN – 1.5x10^9/L	<1500 – 1000/mm^3 <1.5 – 1.0x10^9/L	<1000 – 500/mm^3 <1.0 – 0.5x10^9/L	<500/mm^3 <0.5x10^9/L	Death
GASTROINTESTINAL					
Diarrhoea	Increase of <4 stools/day over baseline; mild increase in ostomy output compared to baseline	Increase of 4-6 stools/day over baseline; IV fluids indicated <24h; moderate increase in ostomy output compared to baseline; not interfering with ADL	Increase of ≥7 stools/ day over baseline; incontinence; IV fluids ≥24h; hospitalisation; severe increase in ostomy output compared to baseline; interfering with ADL	Life-threatening consequences (e.g., hemodynamic collapse)	Death
Enteritis (inflammation of the small bowel)	Asymptomatic, pathologic, or radiographic findings only	Abdominal pain; mucus or blood in stool	Abdominal pain, fever, change in bowel habits with ileus; peritoneal signs	Life-threatening consequences (e.g., perforation, bleeding, ischemia, necrosis)	Death
Mucositis (clinical exam)	Erythema of the mucosa	Patchy ulcerations or pseudo membranes	Confluent ulcerations or pseudomembranes; bleeding with minor trauma	Tissue necrosis; significant spontaneous bleeding; life-threatening consequences	

Mucositis (functional/ symptomatic)	Upper aerodigestive tract sites: Minimal symptoms, normal diet; minimal respiratory symptoms but not interfering with function Lower GI sites: Minimal discomfort, intervention not indicated or other symptoms interfering with ADL	Upper aerodigestive tract sites: Symptomatic but can eat and swallow modified diet; respiratory symptoms interfering with function but not interfering with ADL Lower GI sites: Symptomatic, medical intervention indicated but not interfering with ADL	Upper aerodigestive tract sites: Symptomatic and unable to adequately aliment or hydrate orally; respiratory symptoms interfering with ADL Lower GI sites: Stool incontinence	Symptoms associated with life-threatening consequences	Death
Nausea	Loss of appetite without alteration in eating habits	Oral intake decreased without significant weight loss, dehydration or malnutrition; IV fluids indicated <24h	Inadequate oral caloric or fluid intake; IV fluids, tube feedings, or TPN indicated ≥24h	Life-threatening consequences	Death
ALLERGY/IMMUNOLOGY					
Allergic reaction/ hypersensitivity (including drug fever)	Transient flushing or rash; drug fever <38°C (<100.4°F)	Rash; flushing; urticaria; dyspnea; drug fever ≥38°C (≥100.4°F)	Symptomatic bronchospasm, with or without urticaria; parenteral medication(s) indicated; allergy-related edema/angioedema; hypotension	Anaphylaxis	Death
HEPATOBILIARY/PANCREAS					
Hepatobiliary/Pancreas	Mild	Moderate	Severe	Life-threatening	Death
Rash/desquamation	Macular or papular eruption or erythema without associated symptoms	Macular or popular eruption or erythema with pruritus or other associated symptoms; localized desquamation or other lesions covering <50% of body surface area (BSA)	Severe, generalized erythroderma or macular, papular or vesicular eruption; desquamation covering ≥50% BSA	Generalized exfoliative, ulcerative, or bullous dermatitis	Death

Rash: acne	Intervention not indicated	Intervention indicated	Associated with pain, disfigurement, ulceration, or desquamation	—	Death
Rash: erythema multiforme	—	Scattered, but not generalized eruption	Severe (e.g., generalized rash or painful stomatitis); IV fluids, tube feedings, or TPN indicated	Life-threatening; disabling	Death
Burn	Minimal symptoms; intervention not indicated	Medical intervention; minimal debridement indicated	Moderate to major debridement or reconstruction indicated	Life-threatening consequences	Death
Dry skin	Asymptomatic	Symptomatic, not interfering with ADL	Interfering with ADL	—	—
Induration/fibrosis (skin and subcutaneous tissue)	Increased density on palpation	Moderate impairment of function not interfering with ADL; marked increase in density and firmness on palpation with or without minimal retraction	Dysfunction interfering with ADL; very marked density, retraction or fixation	—	—
Nail changes	Discoloration; ridging (koilonychias); pitting	Partial or complete loss of nail(s); pain in nailbed(s)	Interfering with ADL	—	—
Photosensitivity	Painless erythema	Painful erythema	Erythema with desquamation	Life-threatening; disabling	Death
Pruritus/itching	Mild or localized	Intense or widespread	Intense or widespread and interfering with ADL	—	—
Rash/desquamation	Macular or popular eruption or erythema without associated symptoms	Macular or popular eruption or erythema with pruritus or other associated symptoms; localized desquamation or other lesions covering <50% of body surface area (BSA)	Severe, generalized erythroderma or macular, papular or vesicular eruption; desquamation covering ≥50% BSA	Generalized exfoliative, ulcerative, or bullous dermatitis	Death

Skin breakdown/ decubitus ulcer	—	Local wound care; medical intervention indicated	Operative debridement or other invasive intervention indicated (e.g., hyperbaric oxygen)	Life-threatening consequences; major invasive intervention indicated (e.g., tissue reconstruction, flap, or grafting)	Death
Telangiectasia	Few	Moderate	Many and confluent	—	—
Ulceration	—	Superficial ulceration <2 cm; local wound care; medical intervention indicated	Ulceration ≥2 cm; operative debridement, primary closure or other invasive intervention indicated (e.g., hyperbaric oxygen)	Life-threatening consequences; major invasive intervention indicated (e.g., complete resection, tissue reconstruction, flap, or grafting)	Death
Urticaria (hives, welts, wheals)	Intervention not indicated	Intervention indicated for <24h	Intervention indicated for ≥24h	—	—
Wound complication, non-infectious	Incisional separation of ≤25% of wound, no deeper than superficial fascia	Incisional separation >25% of wound with local care; asymptomatic hernia	Symptomatic hernia without evidence of strangulation; fascial disruption/dehiscence without evisceration; primary wound closure or revision by operative intervention indicated; hospitalization or hyperbaric oxygen indicated	Symptomatic hernia with evidence of strangulation; fascial disruption with evisceration; major reconstruction flap, grafting, resection, or amputation indicated	Death
Dermatology/Skin – Other	Mild	Moderate	Severe	Life-threatening; disabling	Death
Rash: hand-foot skin reaction	Minimal skin changes or dermatitis (e.g., erythema) without pain	Skin changes (e.g., peeling, blisters, edema) or pain, not interfering with function	Ulcerative dermatitis or skin changes with pain interfering with function	—	—

CARDIAC GENERAL					
Cardiac ischemia/infarction	Asymptomatic arterial narrowing without ischemia	Asymptomatic and testing suggesting ischemia; stable angina	Symptomatic and testing consistent with ischemia; unstable angina; intervention indicated	Acute myocardial infarction	Death
Cardiac troponin I (cTnI)			Levels consistent with unstable angina as defined by the manufacturer	Levels consistent with myocardial infarction as defined by the manufacturer	Death
Cardiac troponin I (cTnI)	0.03 – <0.05 ng/mL	0.05 – <0.1 ng/mL	0.1 – <0.2 ng/mL	0.2 ng/mL	Death
Cardiopulmonary arrest, cause unknown	—	—	—	Life-threatening	—
Hypertension	Asymptomatic, transient >20 mmHg (diastolic) or to >150/100 if previously WNL; intervention not indicated	Recurrent or persistent (≥24h) or symptomatic increase by >20 mmHg (diastolic) or to >150/100 if previously WNL; monotherapy may be indicated	Requiring more than one drug or more intensive therapy than previously	Life-threatening consequences (e.g., hypertensive crisis)	Death
Hypotension	Changes, intervention not indicated	Brief (<24h) fluid replacement or other therapy; no physiologic consequences	Sustained (≥24h) therapy, resolves without persisting physiologic consequences	Shock (e.g., acidemia; impairment of vital organ function)	Death
Left ventricular diastolic dysfunction	Asymptomatic diagnostic finding; intervention not indicated	Asymptomatic, intervention indicated	Symptomatic CHF responsive to intervention	Refractory CHF, poorly controlled; intervention such as ventricular assist device or heart transplant indicated	Death
Left ventricular systolic dysfunction	Asymptomatic, resting ejection fraction (EF) <60 – 50%; shortening fraction (SF) <30 – 24%	Asymptomatic, resting EF <50 – 40%; SF <24 – 15%	Symptomatic CHF responsive to intervention; EF <40 – 20% SF <15%	Refractory CHF or poorly controlled; EF <20%; intervention such as ventricular assist device,	Death

				ventricular reduction surgery, or heart transplant indicated	
Myocarditis	—	—	CHF responsive to intervention	Severe or refractory CHF	Death
Pericardial effusion (non-malignant)	Asymptomatic effusion	—	Effusion with physiologic consequences	Life-threatening consequences (e.g., tamponade); emergency intervention indicated	Death
Pericarditis	Asymptomatic, ECG or physical exam (rub) changes consistent with pericarditis	Symptomatic pericarditis (e.g., chest pain)	Pericarditis with physiologic consequences (e.g., pericardial constriction)	Pericarditis Life-threatening consequences; emergency intervention indicated	Death
Pulmonary hypertension	Asymptomatic without therapy	Asymptomatic, therapy indicated	Symptomatic hypertension, responsive to therapy	Symptomatic hypertension poorly controlled	Death
Restrictive cardiomyopathy	Asymptomatic, therapy not indicated	Asymptomatic, therapy indicated	Symptomatic CHF responsive to intervention	Refractory CHF, poorly controlled; intervention such as ventricular assist device, or heart transplant indicated	Death
Right ventricular dysfunction (cor pulmonale)	Asymptomatic without therapy	Asymptomatic, therapy indicated	Symptomatic cor pulmonale, responsive to intervention	Symptomatic cor pulmonale poorly controlled; intervention such as ventricular assist device, or heart transplant indicated	Death

Valvular heart disease	Asymptomatic valvular thickening with or without mild valvular regurgitation or stenosis; treatment other than endocarditis prophylaxis not indicated	Asymptomatic; moderate regurgitation or stenosis by imaging	Symptomatic; severe regurgitation or stenosis; symptoms controlled with medical therapy	Life-threatening; disabling; intervention (e.g., valve replacement, valvuloplasty) indicated	Death
Cardiac General	Mild	Moderate	Severe	Life-threatening; disabling	Death
MUSCULOSKELETAL/SOFT TISSUE					
Neuropathy-cranial	Asymptomatic, detected on exam/testing only	Symptomatic, not interfering with ADL	Symptomatic, interfering with ADL	Life-threatening; disabling	Death
Neuropathy-motor	Asymptomatic, weakness on exam/testing only	Symptomatic weakness interfering with function, but not interfering with ADL	Weakness interfering with ADL; bracing or assistance to walk (e.g., cane or walker) indicated	Life-threatening; disabling (e.g., paralysis)	Death
Neuropathy-sensory	Asymptomatic; loss of eep tendon reflexes or paresthesia (including tingling) but not interfering with function	Sensory alteration or paresthesia (including tingling), interfering with function, but not interfering with ADL	Sensory alteration or paresthesia interfering with ADL	Disabling	Death
CONSTITUTIONAL SYMPTOMS					
Fatigue (asthenia, lethargy, malaise)	Mild fatigue over baseline	Moderate or causing difficulty performing ADL	Severe fatigue interfering with ADL	Disabling	—
OCULAR/VISUAL					
Ocular surface disease (including conjunctivitis)	Asymptomatic or minimally symptomatic but not interfering with function	Symptomatic, interfering with function but not interfering with ADL; topical antibiotics or other topical intervention indicated	Symptomatic, interfering with ADL; operative intervention indicated	—	—

IV: intravenous; LLN: local laboratory value; WNL: within normal limits; ADL: activities of daily living; CHF: cardiac heart failure.

Supplementary Table S2. Summary of CRC pharmacogenomics, including all variants reported so far related to pharmacogenomics of toxicity to CRC chemotherapy (adapted from Simoes et al.[47]).

Drug	Gene	SNP (rsID)	Change	Alternative	Risk allele frequency[a]	Associated ADR	OR (95% CI)	Evidence level[b]
5-FU or capecitabine	*DPYD*	rs55886062	NM_000110.3:c.1679T>G; NP_000101.2:p.(I560S)	*DPYD**13	$3x10^{-4}$ (C)	Global[167]	6.0 (0.6-61)	1A
		rs3918290	NM_000110.4:c.1905+1G>A	*DPYD**2A	$7x10^{-3}$ (T)	Global[55]	8.5 (1.8-40.9)	1A
		rs67376798	NM_000110.3:c.2846A>T; NP_000101.2:p.(D949V)		$3x10^{-3}$ (A)	Global[53]	7.8 (1.6-39.2)	1A
		rs56038477	NM_000110.3:c.1236G>A; NP_000101.2:p.(E412=)		0.01 (T)	Digestive; Haematological[168]	2.0 (1.5-2.8) 2.8 (1.2-3.7)	1A
		rs2297595	NM_000110.3:c.496A>G; NP_000101.2:p.(M166V)		0.09 (C)	Global[169]	5.9 (1.3- 27.2)	1A
		rs1801159	NM_000110.3:c.1627A>G; NP_000101.2:p.(I543V)	*DPYD**5	0.19 (C)	Diarrhoea[170]	4.9 (-)	1A
		rs1801158	NM_000110.3:c.1601G>A; NP_000101.2:p.(S534N)	*DPYD**4	0.02 (T)	Global[171]	1.7 (1.1-2.6)	1A
		rs1801160	NM_000110.3:c.2194G>A; NP_000101.2:p.(V732I)	*DPYD**6	0.05 (T)	Global[172]	2.1 (1.5-3.0)	1A
		rs17376848	NM_000110.3:c.1896T>C; NP_000101.2:p.(P632=)		0.05 (G)	Global[169]	14.5 (1.4-155.2)	1A
		rs115232898	NM_000110.3:c.557A>G; NP_000101.2:p.(Y186C)		$2x10^{-3}$	Neutropenia, mucositis, alopecia[173]	-	1A
		rs75017182	NM_000110.4:c.1129-5923C>G		0.01 (C)	Global[174]	6.8 (2.0-23)	1A
		rs72549303[c]	NM_000110.4:c.1898del; NP_000101.2:p.(P633fs)	*DPYD**3	NA	NA[53]	NA	1A
		rs72549309[c]	NM_000110.4:c.299_302del; NP_000101.2:p.(P100fs)	*DPYD**7	$6x10^{-5}$ delATGA	NA[53]	NA	1A
		rs1801266[c]	NM_000110.4:c.703C>T; NP_000101.2:p.(R235W)	*DPYD**8	$3x10^{-5}$ (A)	NA[53]	NA	1A
		rs1801268[c]	NM_000110.4:c.2983G>T; NP_000101.2:p.(V995P)	*DPYD**10	NA	NA[53]	NA	1A

		rs78060119	NM_000110.3:c.1156G>T; NP_000101.2:p.(E386*)	*DPYD**12	$8x10^{-6}$ (A)	Leucopenia, thrombopenia, mucositis[175]	NA	1A
		rs1801265	NM_000110.3:c.85T>C; NP_000101.2:p.(C29R)	*DPYD**9A	0.22 (G)	Diarrhoea[171]	0.8 (0.7-1)	1A
		rs1801267[c]	NM_000110.4:c.2657G>A; NP_000101.2:p.(R886H)	*DPYD**9B	$1x10^{-4}$ (T)	NA[176]	NA	NA
		rs17376848	NM_000110.3:c.1896T>C; NP_000101.2:p.(P632=)		0.05 (G)	Global[169]	14.5 (1.4-155.2)	1A
		rs12022243	NM_000110.4:c.1906-14763G>A		0.18 (T)	Global[67]	1.7 (1.5- 1.9)	3
		rs12119882	NM_000110.4:c.680+2545T>C		0.08 (G)	Hyperbilirubinemia[131]	4.9 (1.2-20.8)	3
	Intergenic	rs76387818	Intergenic		0.02 (A)	Global[67]	4.1 (3.5-4.6)	3
	Intergenic	rs12132152	Intergenic		0.02 (A)	HFS; Global[67]	6.1 (5.5-6.8); 1.6 (1.4-1.8)	3
	TYMS	rs183205964	NM_001071.4:c.-86= (5' UTR)		$3x10^{-5}$(C)	Global[177]	3.0 (1.1- 8.4)	3
		rs2853741	NC_000018.9:g.657352T>C		0.32 (C)	Diarrhoea[131]	0.3 (0.1-0.7)	3
		rs699517	NG_028255.1:g.20413C>T		0.38 (T)	Nausea/vomiting; Asthenia[131]	7.9 (1.5-41.6); 0.3 (0.1-0.8)	3
		rs45445694	NG_028255.1:g.5054_5109del (5' UTR)		$7x10^{-3}$ (2R2R[d])	Global[178]	1.7 (-)	3
		rs2853542	NG_028255.1:g.5082G>C (5' UTR)		0.26 (C)	Global; HFS[179]	1.5 (1.2-1.8); 1.4 (1.2-1.8)	NA
		rs11280056	NG_028255.1:g.20844_20849del (3' UTR)		0.30	Global[179]	1.7 (1.2-2.2)	NA
	ENOSF1	rs2612091	NM_017512.7:c.742-227G>C		0.37 (C)	Global[67]	1.6 (1.4-1.8)	3
	UMPS	rs2279199	NG_017037.1:g.5040T>C		0.56 (T)	Nausea/vomiting[131]	0.2 (0.1-1.0)	3
		rs4678145	NM_000373.4:c.156+607G>C		0.09 (C)	Asthenia[131]	4.5 (1.6-13.2)	3
		rs1801019[e]	NM_000373.4:c.638G>C; NP_000364.1:p.(G213A)		0.17 (C)	Global[180]	17.6 (1.6-195.9)	3
	MTHFR	rs1801131	NM_001330358.1:c.1409A>C; NP_001317287.1:p.(E470A)		0.29 (G)	HFS[181]	10.0 (3.8-27.8)	3
		rs1801133	NM_001330358.1:c.788C>T; NP_001317287.1:p.(A263V)		0.32 (A)	Neutropenia[182]	2.3 (1.2-4.6)	3

	TYMP	rs11479	NM_001113755.3:c.1412C>T; NP_001244917.1:p.(S471L)		0.09 (A)	Global[183]	2.7 (1.2-5.9)	3
	MIR27A	rs895819	NR_029501.1:n.40A>G		0.34 (C)	Global[134]	1.6 (1.1-2.2)	3
	ABCC1	rs7194667	NM_032583.4:c.1609-491A>C		0.06 (G)	Leukopenia[184]	3.31 (1.3-8.7)	3
	ABCB1	rs1045642	NM_001348945.1:c.3645T>C; NP_001335874.1:p.(I1215=)	*ABCB1**6	0.50 (G)	HFS[126]	NA	3
		rs2032582	NM_001348945.1:c.2887T>G; NP_001335874.1:p.(S963A)	*ABCB1**7	0.64 (C)	HFS[126]	NA	3
		rs1128503	NM_001348945.1:c.1446T>C; NP_001335874.1:p.(G482=)	*ABCB1**8	0.61 (G)	Neutropenia[126]	NA	3
	SLC22A7	rs2270860	NM_006672.3:c.1269C>T; NP_006663.2:p.(S423=)		0.37 (T)	Global[131]	17.1 (1.7-170.3)	3
		rs4149178	NM_006672.3:c.1586+206A>G		0.79 (A)	Diarrhoea[131]	0.3 (0.1-0.9)	3
	CDA	rs2072671	NM_001785.3:c.79A>C; NP_001776.1:p.(K27Q)		0.28 (C)	Global[185]	1.8 (1.1-3.0)	3
		rs1048977	NM_001785.3:c.435C>T; NP_001776.1:p.(T145=)		0.31 (T)	Hyperbilirubinemia[131]	8.6 (1.1-70.3)	3
		rs602950	NC_000001.10:g.20915531A>G		0.22 (G)	Diarrhoea[181]	2.3 (1.3-4.2)	3
		rs3215400	NC_000001.10:g.20915592del (5' UTR)		0.56 (delC)	HFS[186]	0.5 (0.3-1.0)	3
		rs532545	NC_000001.10:g.20915172C>T		0.22 (T)	Diarrhoea[181]	2.3 (1.3-4.2)	NA
	CES1	rs3217164	NM_001025195.2:c.693+129del		0.61 (G)	Global[187]	4.1 (1.8-9.0)	3
		rs2244614	NM_001025195.2:c.1171-41C>T		0.48 (G)	Global[187]	4.7 (1.9-12.0)	3
		rs2244613	NM_001025195.2:c.1171-33C>T		0.23 (G)	Global[187]	6.4 (1.5-27.7)	3
	CES1P1	rs7187684	NC_000016.9:g.55794951T>C		0.28 (T)	Global[187]	6.5 (1.5-28.0)	3
		rs11861118	NC_000016.9:g.55793279A>G		0.16 (G)	Global[187]	6.5 (1.5-28.0)	3
	Intergenic	rs9936750	Intergenic		0.16 (C)	Global[188]	4.6 (1.5-13.9)	3
	Intergenic	rs10876844	Intergenic		0.44 (A)	Diarrhoea[68]	6.5 (1.6-27.2)	NA
Oxaliplatin	*ABCC2*	rs717620	NG_011798.2:g.5224C>T (5' UTR)		0.17 (T)	Neuropathy[189]	14.4 (1.6-127.0)	3
		rs3740066	NM_000392.5:c.3972C>T NP_000383.2:p.(I1324=)		0.37	Neuropathy[189]	3.0 (1.2-7.7)	NA
		rs1885301	NG_011798.2:g.3699A>G		0.41 (G)	Neuropathy[189]	3.1 (1.4-6.9)	NA
		rs4148396	NM_000392.5:c.3258+56T>C		0.35 (T)	Neuropathy[189]	4.7 (1.6-13.7)	NA
	ABCG2	rs3114018	NM_004827.3:c.-19-3415T>G		0.52 (A)	Neuropathy[190]	2.7 (1.0-4.4)	NA
	GSTP1	rs1695	NM_000852.3:c.313A>G; NP_000843.1:p.(I105V)	*GSTP1**B	0.34 (G)	Death[191]	3.0 (1.2-7.6)	3

	GSTM1	Null genotype	Biallelic deletion of gene	*GSTM1**0		Neutropenia[192]	2.0 (1.1-3.7)	NA
	GSTT1	Null genotype	Biallelic deletion of gene			Neutropenia[192]	2.0 (1.1-3.7)	NA
	ERCC1	rs11615	NM_202001.3:c.354T>C; NP_001356337.1:p.(N118=)		0.49 (A)	Neutropenia[182]	4.6 (1.2-17.4)	3
	ERCC2	rs13181	NM_000400.3:c.2251A>C; NP_000391.1:p.(K751Q)		0.32 (G)	Haematological[193]	2.2 (1.3-3.8)	3
		rs238406	NM_000400.4:c.468A>C NP_000391.1:p.(R156=)		0.65 (C)	Thrombopenia[194]	NA	NA
	PARD3B	rs17626122	NM_001302769.2:c.3261-6168T>C		0.55 (T)	Global[68]	3.4 (1.9-6.8)	3
	Intergenic	rs7325568	Intergenic		0.41 (T)	Haematological[68]	1.8 (1.3-2.4)	3
Irinotecan	*UGT1A1*	rs3064744	NG_002601.2:g.175492dupTA	*UGT1A1**28	0.35 (dupTA)	Global[195]	7.2 (2.5-22.3)	1A
		rs4148323[c]	NM_000463.2:c.211G>A; NP_000454.1:p.(G71R)	*UGT1A1**6	0.01 (A)	NA[196]	NA	1B
		rs11563250	NG_002601.2:g.189961A>G		0.89 (A)	Neutropenia[197]	0.3 (0.2-0.6)	3
		rs4124874	NM_001072.3:c.862-10021T>G	*UGT1A1**60	0.45 (T)	Neutropenia[58]	NA	3
		rs10929302	NM_019075.2:c.856-9898G>A	*UGT1A1**93	0.29 (A)	Global[198]	8.4 (1.9-37.2)	3
	UGT1A9	rs11692021	NM_021027.3:c.855+9770T>C		0.35 (C)	Global[135]	2.0 (1.1-3.6)	3
	UGT1A6	rs2070959	NM_001072.4:c.541A>G		0.69 (A)	Global[135]	2.1 (1.1-3.9)	3
	ABCG1	rs225440	NM_016818.3:c.286+7029C>T		0.43 (T)	Neutropenia[124]	3.1 (1.1-8.6)	3
		rs425215	NM_016818.3:c.974-898C>G		0.62 (G)	Digestive[125]	11.4 (1.7- 78.4)	NA
	ABCB1	rs12720066	NM_001348945.1:c.2529+971T>G		0.04 (C)	Neutropenia[199]	NA	3
	ABCC1	rs17501331	NM_004996.4:c.49-12232A>G		0.93 (A)	Neutropenia[199]	NA	3
		rs3743527	NG_028268.2:g.197248C>T (3' UTR)		0.77 (C)	Neutropenia[199]	NA	3
	ABCC5	rs2292997	NM_005688.4:c.129+7980C>T		0.13 (A)	Neutropenia[124]	3.2 (1.3-7.9)	3
		rs10937158	NM_005688.4:c.130-1268A>T		0.61 (C)	Diarrhoea[124]	0.4 (0.2-0.8)	3
		rs3749438	NM_005688.4:c.591+374C>T		0.32 (A)	Diarrhoea[124]	5.9 (1.3-26.3)	3
		rs562	NG_047115.1:g.102954A>G (3' UTR)		0.46 (C)	Digestive[125]	32.0 (2.8-370.8	NA
	ABCG2	rs7699188	NC_000004.11:g.89096061G>A		0.23 (A)	Global; Non-haematological[200]	7.3 (1.5-34.5); 15.2 (2.5-78.2)	3
	SLCO1B1	rs2306283	NM_006446.5:c.388A>G NP_006437.3:p.(N130D)	*SLCO1B1**1b	0.53 (G)	Digestive[125]	2.3 (0.4-15.1)	4
	TOP1	rs6072262	NM_003286.4:c.279+61G>A		0.14 (A)	Diarrhoea[201]	NA	3
	TGFBR2	rs3087465	NG_007490.1:g.4167A>G		0.66 (G)	Diarrhoea[202]	3.7 (1.0-13.3)	3
	TGFB1	rs1800469	NG_013364.1:g.4536T>C		0.70 (G)	Diarrhoea[202]	4.4 (1.0-18.9)	3
	KCNQ5	rs9351963	NM_019842.4:c.490-1798A>C		0.18 (C)	Neuropathy[203]	3.3 (1.8-5.6)	3

	Intergenic	rs10486003	Intergenic		0.91 (C)	Neuropathy[69]	0.3 (0.2-0.5)	NA
	Intergenic	rs2338	Intergenic		0.28 (A)	Neuropathy[69]	2.3 (1.6-3.3)	NA
	Intergenic	rs830884	Intergenic		0.92 (T)	Neuropathy[69]	0.3 (0.2-0.5)	NA
	ACYP2	rs843748	NM_001320586.2:c.405-28913G>A		0.38 (A)	Neuropathy[69]	2.4 (1.6-3.7)	NA
	DLEU7	rs797519	NC_000013.11:g.50656996G>C		0.55 (G)	Neuropathy[69]	0.5 (045-0.7)	NA
	FARS2	rs17140129	NM_001318872.2:c.-22+36771A>G		0.16 (G)	Neuropathy[69]	3.3 (1.8-6.4)	NA
Cetuximab	*EGFR*	rs712830	NG_007726.3:g.5056A>C (5' UTR)		0.89 (C)	Global[204]	6.1 (1.6-23.8)	3
		rs2227983	NM_005228.5:c.1562G>A NP_005219.2:p.(R521K)		0.77 (G)	Skin toxicity[205]	3.2 (1.3-8.3)	3
		rs11568315	NM_005228.5:c.88+1195dup		$4x10^{-4}$ (CA>35)	Skin toxicity[206]	2.9 (1.0-8.9)	NA
	RPS7	rs10203413	NG_011744.1:g.11326G>A		0.78 (G)	Skin toxicity[70]	0.1 (0.1-0.4)	NA
	ZNF827	rs12646351	NC_000002.12:g.3581588G>A		0.82 (G)	Skin toxicity[70]	0.04 (0.01-0.3)	NA
		rs17806780	NM_001306215.2:c.2383+11920A>T		0.82 (T)	Skin toxicity[70]	0.04 (0.01-0.4)	NA
	EPHA5	rs7692430	NM_004439.8:c.2237-1876A>G		0.16 (G)	Skin toxicity[70]	4.6 (2.5-8.5)	NA
Bevacizumab	*VEGF*	rs3025039	NG_008732.1:g.19584C>T (3' UTR)		0.13 (T)	Hypertension[207]	0.2 (0.03-0.8)	NA
		rs2010963	NG_008732.1:g.5398C>G (5' UTR)		0.69 (G)	Hypertension[208]	NA	NA
		rs833061	NG_008732.1:g.4534C>T		0.45 (T)	Hypertension[209]	0.2 (0.03-0.8)	NA
		rs699947	NG_008732.1:g.3437A>C		0.41 (C)	Hypertension[209]	0.1 (0.01-0.6)	NA

a: Risk alleles frequencies, consulted on gnomAD v2.1.1; b: Evidence level, measure of confidence in the association, according to PharmGKB[24,102]; *DPYD* variants in bold are those included in the pharmacogenetic guidelines, therefore clinically actionable currently; c: Associated with changes in enzymatic activity, but with a particular ADR; d: 2R2R: double tandem repeat of 28bp; e: Described for tegafur, a 5-FU prodrug; NA: not available.

Supplementary Table S3. Datasets and the gene expression models used in the TWAS.

Dataset and tissue	N samples
DGN whole blood	922
BarcUVa colonic mucosa	191
SOCCS colonic mucosa	221
INTERMPHEN rectum, distal colon, and proximal colon	327
GTEx Adipose - Subcutaneous	581
GTEx Adipose - Visceral (Omentum)	469
GTEx Adrenal Gland	233
GTEx Artery - Aorta	387
GTEx Artery - Coronary	213
GTEx Artery - Tibial	584
GTEx Brain - Amygdala	129
GTEx Brain - Anterior cingulate cortex (BA24)	147
GTEx Brain - Caudate (basal ganglia)	194
GTEx Brain - Cerebellar Hemisphere	175
GTEx Brain - Cerebellum	209
GTEx Brain - Cortex	205
GTEx Brain - Frontal Cortex (BA9)	175
GTEx Brain - Hippocampus	165
GTEx Brain - Hypothalamus	170
GTEx Brain - Nucleus accumbens (basal ganglia)	202
GTEx Brain - Putamen (basal ganglia)	170
GTEx Brain - Spinal cord (cervical c-1)	126
GTEx Brain - Substantia nigra	114
GTEx Breast - Mammary Tissue	396
GTEx Cells - Cultured fibroblasts	483
GTEx Cells - EBV-transformed lymphocytes	147
GTEx Colon - Sigmoid	318
GTEx Colon - Transverse	368
GTEx Esophagus - Gastroesophageal Junction	330
GTEx Esophagus - Mucosa	497
GTEx Esophagus - Muscularis	465
GTEx Heart - Atrial Appendage	372
GTEx Heart - Left Ventricle	386
GTEx Kidney - Cortex	73
GTEx Liver	208
GTEx Lung	515
GTEx Minor Salivary Gland	144
GTEx Muscle - Skeletal	706
GTEx Nerve - Tibial	532
GTEx Ovary	167
GTEx Pancreas	305
GTEx Pituitary	237
GTEx Prostate	221
GTEx Skin - Not Sun Exposed (Suprapubic)	517
GTEx Skin - Sun Exposed (Lower leg)	605
GTEx Small Intestine - Terminal Ileum	174
GTEx Spleen	227
GTEx Stomach	324
GTEx Testis	322
GTEx Thyroid	574
GTEx Uterus	129
GTEx Vagina	141
GTEx Whole Blood	670

Supplementary Table S4. Primers used for vector construction.

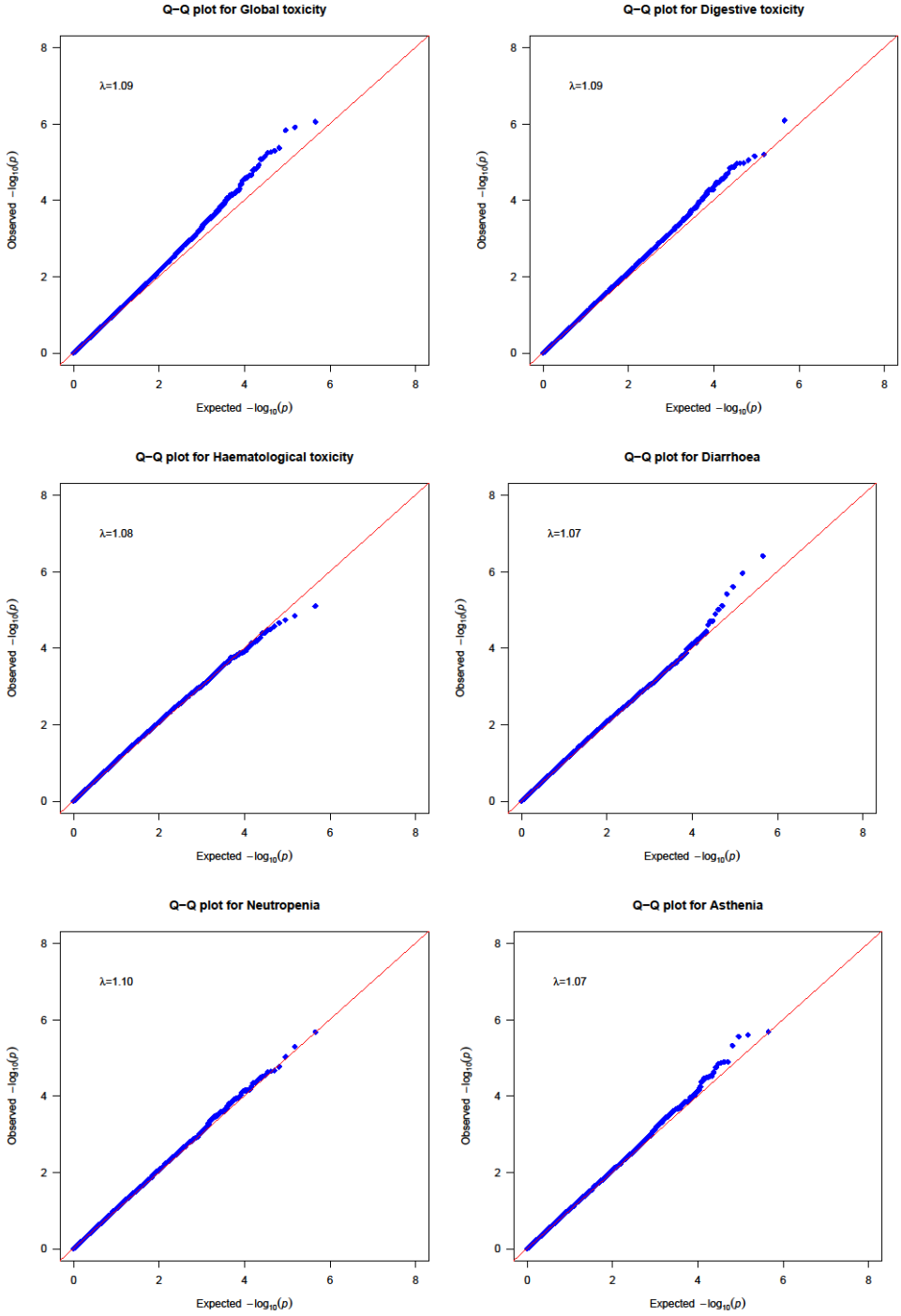
Variant	Primer sequence
c.2071	5'-CTGCGGAACATCTGCCGCTGG-3'; 5'-CAGCTCTGGATCCTGCCCAC-3'
c.1905+1G>A	5'-gcctccctcgcgccatcagTTTAAAATCATGATCCTTGTGTTCAGC-3'; 5'-gccttgccagcccgctcagTCTTCATTTGTACTTTATGGAGCTAACTAC-3'
c.2846A>T	5'- TTGAAGAAATGTGTATCAACTGTGGTAAATGC-3'; 5'- CAATCATAGCCACAACTTGCTCTACG-3'

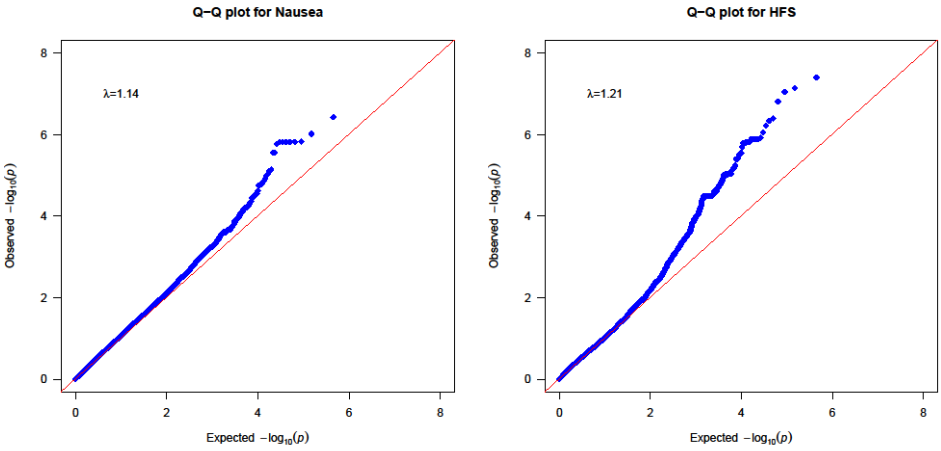
Supplementary Table S5. Composition of the SDS-polyacrylamide gels.

Reagent	Running gel	Stacking gel
Gel stock solution (acrylamide 30%, bisacrylamide 0.8%)	2.4 mL	0.4 mL
4x Running gel buffer	2 mL	-
4x Stacking gel buffer	-	1 mL
H_2O distilled	3.6 mL	2.6 mL
10% ammonium persulphate (add only at the end)	50 µL	25 µL
TEMED (add only at the end)	5 µL	2.5 µL

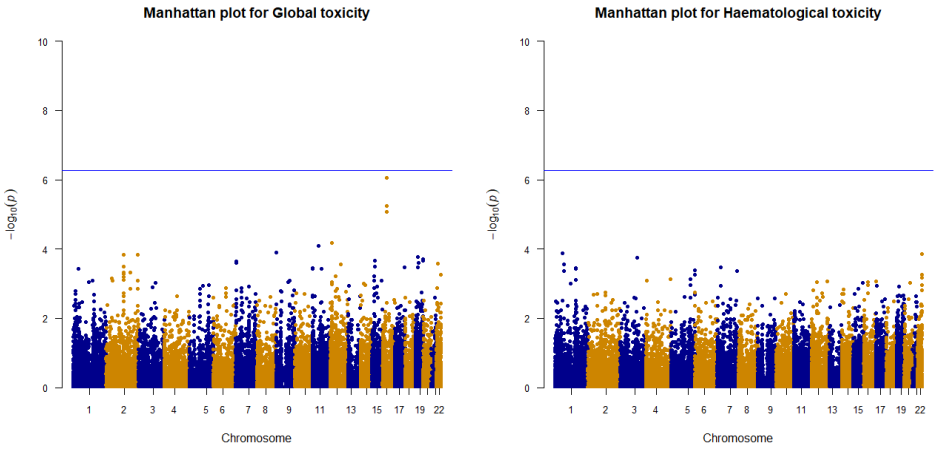
Supplementary Figures

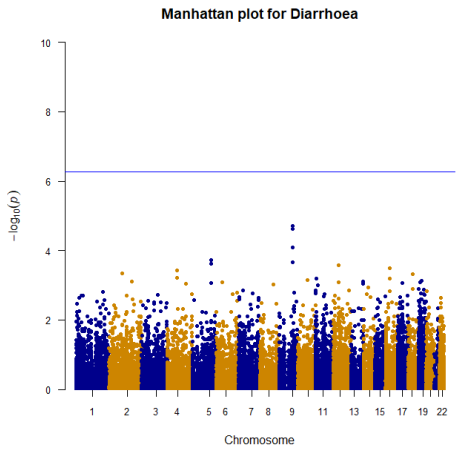
Supplementary Figure 1. Quantile-Quantile (Q-Q) plots of the observed and expected CHIS-Q values from the association analysis between genotypes and ADRs.

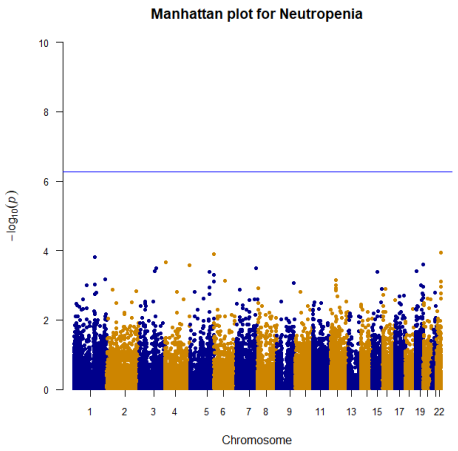


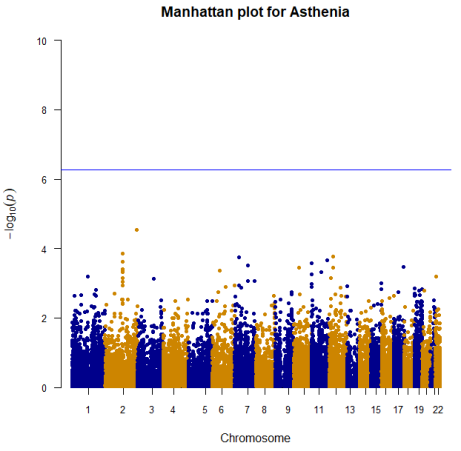


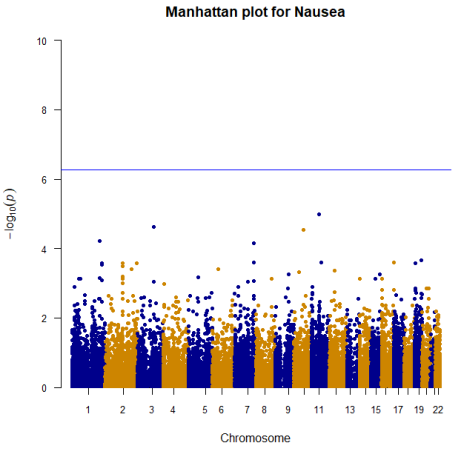
Supplementary Figure 2. Manhattan plots for all ADRs analysed in the toxicity GWASs. All SNPs were above the threshold for exome-wide significance (suggestive line at 5.5×10^{-7}).

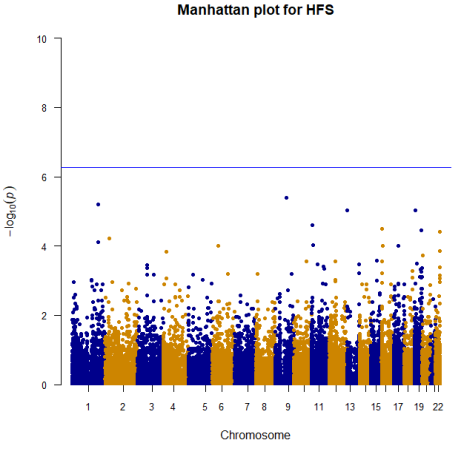












Annex 1. Informed consent

INFORMATIVE DOCUMENT
PARTICIPATION IN BIOMEDICAL RESEARCH

LINE OF RESEARCH TITLE: PHARMACOGENETICS AND PHARMACOGENOMICS
LIABLE ENTITY: GALICIAN PUBLIC FOUNDATION OF GENOMIC MEDICINE
RESEARCH GROUP*: GENOMIC MEDICINE GROUP*

PURPOSE OF PARTICIPATION.-

Every biomedical research study requires collection of data and biological samples from people affected by the pathology to be studied and from people that are not affected by the disease, and analysis of the collected data to obtain conclusions for a better understanding and advance in the diagnostic and/or treatment of the diseases targeted by the study. In fact, many of the scientific advances obtained in medicine over the last years are a result of this type of studies, in which human biological samples are used that are either remnants of samples obtained for the diagnostic/control of the diseases, or are requested *ex professo* for a specific study. For this reason, **we invite you to collaborate in the line of research of PHARMACOGENETICS AND PHARMACOGENOMICS.**

The main objective of this **line of research** is to search genetic markers of response to treatment, both efficacy and side effects, and common diseases stratification, as base to a personalized medicine.

You must know that the execution of the research projects in which human samples are used will be previously authorised by a competent Ethics Committee.

All the information that we provide in this document and the activity of the research group and the Biobank are regulated by the Organic Law 15/1999, on the 13th of December, of Personal Data Privacy and Protection, the Constitutional Law 14/2007, of 3 July, on Biomedical Research, the Royal Decree 1716/2011, of 11 November.

Your participation is completely voluntary. If you sign this consent form, you will confirm your participation. You can refuse to participate or withdraw your consent at any time after the time of signature without having to explain your reasons. **Your non-participation or posterior withdrawal of your consent will not affect in any way your present or future medical attention. You can ask any third parties or us about any of your doubts before signing the consent form.**

BIOLOGICAL SAMPLES AND ASSOCIATED INFORMATION. RISKS.-

BIOLOGICAL SAMPLES Blood, other fluids, cells or tissues obtained during your medical attention that are no longer needed for the purpose that they were obtained for, and are requested from you to be included in this line of research.

The person in charge will store and have access of these samples to carry out the biomedical research studies. The samples and their associated information will be stored in the established areas for said purpose, within the facilities of the Galician Public Foundation of Genomic Medicine under the responsibility of Dr. Ángel Carracedo Álvarez.

Donation of these samples will not stop your or your family to use them when needed for medical reasons, as long as they are available and have not been anonymized.

CLINICAL INFORMATION that, together with the results of the studies carried out on the samples will allow us to draw useful conclusions for the study of the diseases. For the optimal development of the research studies, it is necessary to obtain information relative to the donor, and so, we need access to your clinical history and/or genomic data (if existent) to archive said information with your sample. *(More information in the Confidentiality section.)*

In the case that any information or additional sample is needed, and as long as you authorise it in the consent form, the medical institution could contact you to request your collaboration once again.

CONDITIONS OF THE DONATION.-

The donation and use of human biological samples will be free, and you will not receive remuneration at the time of the donation or in the future. You relinquish any right of economical nature, patrimonial or facultative over the results or potential economical profits that may derive from the research carried out with the sample that you cede.

You will not receive any other direct benefit to your health condition as a result of your donation. However, the knowledge obtained from this study using your sample and many others may help medical progress, and by extension, other people.

Ownership of the results of the research will correspond to the researcher and the institution where the research takes place.

CONFIDENTIALITY.-

All your information will be considered confidential and treated accordingly with respect to the established regulations. Such information will be only at the disposal of the authorised personnel, who have the **legal obligation to keep secrecy**. Your samples and associated clinical data will be part of the "Biomedical Research and Biobanks" database, responsibility of the Ministry of Health and the Galician Health Service.

To guarantee the confidentiality of your identity (assure that the information of your sample cannot be traced to your identity), your samples will only be identified, from the moment they enter the research, using a code. Only this code, and never your identity, will appear in the material used to work. The link between your code and your identity will be kept and guarded by the authorised personnel in a database that fulfils all legal requirements. This way, we can assure you that any information obtained using your samples will remain confidential, but may later be associated with your clinical data if needed for clinical purposes.

Cession of the samples and associated data to the collaborating researchers of this line of research will take place in an encoded manner (the researcher in charge will keep the link that relates your samples to your identity). They will be demanded to guarantee that they will work with the same level of data protection required by Spanish law.

On the other hand, it is possible that the results of this research study will be published in scientific literature, but interpreting these results as those obtained from all of the samples of the study, and not individual results. If this were the case, your identity will remain completely confidential and will never be part of any publication.

With your acceptance of participation in this line of research of pharmacogenetics and pharmacogenomics, you acknowledge that this information may be transferred in the conditions mentioned earlier. You may or may not authorise us to carry out the indicated cessions, stating your decision in the corresponding section in the consent form.

You may exercise the rights of **ACCESS, RECTIFICATION, CANCELLATION and/or OBJECTION (ARCO RIGHTS)** recognised by Spanish law, directing your written application together with a copy of your national identification document (with the purpose of guaranteeing that the exercise of these rights is carried out by the authorised individual) to:

- **DELIVERY BY HAND**, to the Galician Public Foundation of Genomic Medicine directed to Ángel Carracedo Álvarez.
- **POST MAIL**: Fundación Pública Galega de Medicina Genómica, CHUS, Edif Consultas, planta -2 Trav Choupana s/n, 15706 Santiago de Compostela
- **E-MAIL**: angel.carracedo@usc.es

If you have any doubts regarding your participation in this research, you can contact the researchers of this line of research in cancer genetics on the **telephone number: 981951491.** All necessary mediums will be employed to provide you with any information at all times.

If you decide to **REVOKE YOUR CONSENT**, you must direct the filled and signed revocation form (at the end of this document) to Dr. Ángel Carracedo Álvarez, Executive Director of the Galician Public Foundation of Genomic Medicine.

You must know that the revocation will come into effect from the moment of formalisation and will not affect any data resulting from research that has taken place before this moment.

RESTRICTIONS OF USE OF THE SAMPLE.-

You may indicate if you want to establish some type of restriction on your samples and data, in relation to their possible uses in certain research projects or about certain cessions. For this, you have at your disposal a specific section in the consent form.

INFORMATION REGARDING RESULTS OF THE RESEARCH.-

Evaluation of the results will only be done by set groups (for example, men and women, age groups, diagnostics, etc.), and not in an individual manner. You must understand that valuable results that may be obtained will derive from the study of multiple samples, and never exclusively from your single samples or data.

The medical implications of the results of the many different tests, if there are any will only be known once the research is complete.

The one responsible for the research will have the information about the research projects in which samples and data are used at the disposal of the participants.

In certain circumstances, the competent Ethics Committee will be able to decide if it is necessary to contact a specific participant to provide him or her with individual information.

The different studies in which your samples and data are used may need to carry out **cellular and genetic assays** on them, in order to obtain relevant information for you or your family's health. In **genetic studies**, information that was not originally being actively looking for may be found, known as unexpected findings. If or when this happens, the obtained results will be validated and analysed by professionals to determine if they have any clinical applications or treatments that may have to be informed to any affected people, according to the present medical knowledge.

You must know that you have the right to decide to be informed or not be informed of the information obtained from the analysis of your samples.

In the case that you decide not to be informed, law dictates that when the information obtained is necessary to avoid serious detriment to the health of your family members, a committee of experts will study the case and must decide between the convenience of informing or not the affected people or their legal representatives.

FINAL DESTINATION OF THE SAMPLE

The remains of the samples will be stored in a Sample Collection, created in the Galician Public Foundation of Genomic Medicine by Dr. Ángel Carracedo with register number C.0001387 for the national registry of ISCIII, specific to the line of research of pharmacogenetics and pharmacogenomics with the purpose of using these samples and associated data in different research projects in which Dr. Ángel Carracedo and other collaborators of the involved research groups participate; as long as such studies fulfil the purpose of this line of research. You must know that all of the research projects in which these samples are used will have previously been approved by the competent Ethics Committee.

In a second phase, when deemed right by the person in charge of the research with whom you accepted to collaborate with your samples and data, and as long as you authorise it in the signing page on the consent form, the remains of the samples will be deposited in the CHUS Biobank.

Biobanks are storage banks of samples of human origin for their used in national or international research studies within the field of biomedicine. Their operation mainly revolves around the management of the reception, processing, storage and posterior cession of samples, while fulfilling the necessary criteria of safety, quality and efficiency, to researchers that may request to use them in their research projects; as long as these projects fulfil all ethical and legal requisites.

The Biobank will keep the remains of these samples to carry out these biomedical research studies. The samples and associated information will be stored in the established areas for such purpose, within the facilities of the University Hospital Complex of Santiago de Compostela, under direct responsibility of the University Hospital Complex of Santiago de Compostela Biobank Division.

Cession of the samples and associated data to researchers that request them, intra-communitary (countries part of the European Union) or extra-communitary (not part of the EU), will be carried out in an anonymous or dissociated manner, and so, only the sample and general associated information will be given, without the possibility of tracing this information to your identity. In research projects in which it may be deemed necessary for the development of the study by the researcher in charge, and after

the approval from the competent Ethics Committee, your samples may be transferred (the Biobank will keep the link that related your sample to your identity). They will be demanded to guarantee that they will work with the same level of data protection required by Spanish law.

The CHUS Biobank agrees with everything described above in the sections relating to confidentiality, possible findings, etc.

The Biobank will only be able to ask for costs of obtainment, processing and postage of the samples to the researchers/institutions that may request the samples.

CHUS BIOBANK.

SCIENTIFIC DIVISION HOLDER: Máximo Fraga.
TELEPHONE CONTACT: 981 955 148 E-MAIL: Maximo.Francisco.Fraga.Rodriguez@sergas.es

CLOSURE OF THE BIOBANK.-

In the case of the closure of the Biobank or revocation of the authorisation for its constitution and operation, the information regarding the destiny of your samples will be at your disposal in the Biobank and or Hospital Centre website www http://www.idisantiago.es/biobanco and in the National Registry of Biobanks for Biomedical Research of the Healthcare Institute Carlos III (ISC III), with website www.isciii.es, with the purpose of manifesting your accordance or disagreement of the programmed destiny for your samples.

Many thanks for your collaboration.

INFORMED CONSENT OF THE ADULT PARTICIPANT

(Space provided to place the tag with the donor's information)

I,…… (*Handwritten name and surnames of the participant*), **have read the information brochure and could ask all the questions that I deemed necessary, and so accept to participate in the line of research of cancer genetics with my samples and data so that they may be used in the conditions I have been notified of.**

- ☐ **Tick in the case that it is a CONTROL sample**

RESTRICTIONS OF USE (you may indicate any restrictions that you may consider necessary by ticking the corresponding boxes. If you don't tick any boxes, we will understand that you DO NOT manifest any conditions of use).

- ☐ **I DO NOT** authorise the used of my samples and data in research studies that……
- ☐ **I DO NOT** authorise access to my clinical history.
- ☐ **I DO NOT** accept to be contacted when necessary.
- ☐ **I DO NOT** want to be informed of the results of the research that may be of interest for my health.
- ☐ **I DO NOT** authorise the cession of my samples and data in a coded (dissociated) manner.
- ☐ **I DO NOT** authorise the cession of my samples and data outside of the European Union.
- ☐ **I DO NOT** authorise my samples and data to be deposited in the **CHUS Biobank** in the conditions mentioned in information brochure.

If you accept to be contacted when necessary, please indicate your telephone number/e-mail address:

……

Signature of participant
Identity of participant: ……..

Signature of informant
Identity of informant: ……..

Signature of witness
I confirm the verbal consent of the participant, who authorises me to sign in his name
Identity of witness: …… ID: ……..
Only to be signed if the participant cannot read/write and delegates in the witness to sign.

Signature of legal representative
Identity of legal representative: …… ID: ……..
Only to be signed if the participant is legally incapacitated.

In ……., on the ……. of …….. of ……..

We thank your selfless collaboration with the progress of science and medicine.

REVOCATION OF CONSENT

I, Mr./Mrs./Miss./Ms.: ..... with ID .....
Revoke / nullify the consent presented with date: .....
And I do not wish to continue with the voluntary donation done to .....*(name of healthcare centre)*, that I finalise on this date.
□ I REQUEST ONLY THE ELIMMINATION OF THE SAMPLES.
□ I REQUEST ONLY THE ELIMMINATION OF MY PERSONAL DATA.
The sample will remain anonymous irreversibly and may be used in research projects.
□ I REQUEST THE TOTAL ELIMMINATION OF MY DATA AND SAMPLES.

Signed:

In....., on the ..... of ..... of 20.....

I, Mr./Mrs./Miss./Ms.: ..... with ID .....
As: *(tick the option applicable to your case)*
- □ Witness that signs in the name
- □ Legal representative

Of Mr./Mrs./Miss./Ms.: ..... with ID .....
Revoke / nullify the consent presented with date: .....
And I do not wish to continue with the voluntary donation done to .....*(name of healthcare centre)*, that I finalise on this date.
□ I REQUEST ONLY THE ELIMMINATION OF THE SAMPLES.
□ I REQUEST ONLY THE ELIMMINATION OF MY PERSONAL DATA.
The sample will remain anonymous irreversibly and may be used in research projects.
□ I REQUEST THE TOTAL ELIMMINATION OF MY DATA AND SAMPLES.

Signed:

In....., on the ..... of ..... of 20.....

Annex 2. Project's approval by the local ethics committee

XUNTA DE GALICIA
CONSELLERÍA DE SANIDADE
Secretaría Xeral Técnica

Secretaria Técnica
Comité Autonómico de Ética da Investigación de Galicia
Secretaria Xeral. Consellería de Sanidade
Edificio Administrativo San Lázaro
15703 SANTIAGO DE COMPOSTELA
Tel: 881546425. Correo-e: ceic@sergas.es

galicia

DICTAMEN DEL COMITÉ DE ÉTICA DE LA INVESTIGACIÓN DE SANTIAGO-LUGO

Guillermo José Prada Ramallal, Secretario del Comité de Ética de la Investigación de Santiago-Lugo,

CERTIFICA:

Que este Comité evaluó en su reunión del día 26 de octubre de 2017 el estudio**:**

Título: Descubrimiento y validación de marcadores farmacogenéticos de toxicidad para quimioterapia con fluoropirimidinas, oxaliplatino e irinotecan en cáncer colorrectal
Promotor*:* Ángel Carracedo Álvarez
Tipo de estudio: EPA-AS
Versión:
Código del Promotor: ANG-QUI-2017-01
Código de Registro: 2017/354

Y, tomando en consideración las siguientes cuestiones:

- La pertinencia del estudio, teniendo en cuenta el conocimiento disponible, así como los requisitos legales aplicables, y en particular la Ley 14/2007, de investigación biomédica, el Real Decreto 1716/2011, de 18 de noviembre, por el que se establecen los requisitos básicos de autorización y funcionamiento de los biobancos con fines de investigación biomédica y del tratamiento de las muestras biológicas de origen humana, y se regula el funcionamiento y organización del Registro Nacional de Biobancos para investigación biomédica, la ORDEN SAS/3470/2009, de 16 de diciembre, por la que se publican las Directrices sobre estudios Postautorización de Tipo Observacional para medicamentos de uso humano, y la Circular nº 07/2004, de investigaciones clínicas con productos sanitarios.
- La idoneidad del protocolo en relación con los objetivos del estudio, justificación de los riesgos y molestias previsibles para el sujeto, así como los beneficios esperados.
- Los principios éticos da Declaración de Helsinki vigente.
- Los Procedimientos Normalizados de Trabajo del Comité.

Emite un dictamen **FAVORABLE** para la realización del estudio **por el/la investigador/a del centro**:

Centros	**Investigadores Principales**
Fundación Pública Galega de Medicina Xenómica	Ángel Carracedo Álvarez

En Santiago de Compostela, a 3 de noviembre 2017.

El Secretario del Comité Territorial de Ética de la Investigación de Santiago Lugo,

XUNTA DE GALICIA SECRETARÍA XERAL TÉCNICA SECRETARÍA DO COMITÉS DE ÉTICA DA INVESTIGACIÓN CONSELLERÍA DE SANIDADE

PRADA RAMALLAL GUILLERMO JOSE - 44841744R
2017.11.04 19:49:31 +02'00'

Guillermo José Prada Ramallal

XUNTA DE GALICIA
CONSELLERÍA DE SANIDADE
Secretaría Xeral Técnica

Secretaria Técnica
Comité Autonómico de Ética da Investigación de Galicia
Secretaria Xeral. Consellería de Sanidade
Edificio Administrativo San Lázaro
15703 SANTIAGO DE COMPOSTELA
Tel: 881546425. Correo-e: ceic@sergas.es

galicia

Se emite un informe **FAVORABLE CONDICIONADO** a que se tengan en cuenta lo siguiente:

Debe responderse al punto *1. Aspectos formales: Se debe aportar el CV del Investigador Principal (IP)*, de las primeras aclaraciones solicitadas.

Modelos de documentos:
- Para sucesivas solicitudes de evaluación, se sugiere utilizar el modelo de protocolo recomendado por el Comité. Puede encontrar y descargar el formulario para presentación de protocolos de investigación elaborado y aprobado por el CEI-SL en nuestra página web: https://acis.sergas.es/DXerais/118/20170709_Formulario%20Modelo%20Protocolo_ES.pdf
- Para sucesivas solicitudes de evaluación, se sugiere utilizar el modelo de consentimiento informado recomendado por el Comité, que puede encontrar en nuestra página web: http://acis.sergas.es/Paxinas/web.aspx?tipo=paxtab&idLista=3&idContido=83&migtab=83&idTax=15534&idioma=es

NOTA: Se recuerda que para la valoración de las condiciones por parte del Comité, se debe aportar, junto a la nueva versión del protocolo que incluya las oportunas modificaciones resaltadas e identificadas, un documento dónde se expliciten cada una de las aclaraciones aportadas y la ubicación exacta en el nuevo protocolo.

En Santiago de Compostela, a 3 de noviembre 2017.

El Secretario del Comité Territorial de Ética de la Investigación de Santiago Lugo,

PRADA RAMALLAL GUILLERMO JOSE - 44841744R
2017.11.04 19:49:34 +02'00'

Guillermo José Prada Ramallal

XUNTA DE GALICIA
CONSELLERÍA DE SANIDADE
Secretaría Xeral Técnica

Secretaria Técnica
Comité Autonómico de Ética da Investigación de Galicia
Secretaria Xeral. Consellería de Sanidade
Edificio Administrativo San Lázaro
15703 SANTIAGO DE COMPOSTELA
Tel: 881546425. Correo-e: ceic@sergas.es

galicia

Guillermo José Prada Ramallal, Secretario del Comité de Ética de la Investigación de Santiago-Lugo,

HACE CONSTAR QUE:

1.- El Comité Territorial de Ética de la Investigación de Santiago-Lugo cumple tanto en su composición como en sus PNTs los requisitos legales vigentes (RD 1090/2015 de ensayos clínicos, y la Ley 14/2007 de Investigación Biomédica).

2.- La composición actual del Comité Territorial de Ética de la Investigación de Santiago-Lugo es:

- **Juan Manuel Vázquez Lago (Presidente).** Médico especialista en Medicina Preventiva y Salud Pública. Área de Gestión Integrada de Santiago**.**
- **Pilar Rodríguez Ledo (Vicepresidenta).** Médico especialista en Medicina Familiar y Comunitaria. Área de Gestión Integrada de Lugo.
- **Guillermo José Prada Ramallal (Secretario).** Médico especialista en Farmacología Clínica. Área de Gestión Integrada de Santiago. Fundación Ramón Domínguez.
- **Lorenzo Armenteros del Olmo (Vicesecretario).** Médico especialista en Medicina Familiar y Comunitaria. Área de Gestión Integrada de Lugo.
- **Francisco Campos Pérez.** Biólogo. Instituto de Investigación Sanitaria de Santiago de Compostela.
- **Rosana Castelo Domínguez.** Farmacéutica de Atención Primaria. Área de Gestión Integrada de Santiago.
- **Ricardo García Martínez.** Licenciado en Derecho. Área de Gestión Integrada de Lugo.
- **Jaime Gulín Dávila**. Farmacéutico especialista en Farmacia Hospitalaria. Área de Gestión Integrada de Lugo.
- **Victor Herrán Carreira.** Paciente. ADIL-Asociación de Diabéticos Lucense.
- **María Jesús Lamas Díaz.** Farmacéutica especialista en Farmacia Hospitalaria. Área de Gestión Integrada de Santiago.
- **Carlos Rodríguez Moreno.** Médico especialista en Farmacología Clínica. Área de Gestión Integrada de Santiago.
- **Rafael Carlos Vidal Pérez.** Médico especialista en Cardiología. Área de Gestión Integrada de Lugo.
- **María Jesús Wandosell Picatoste.** Enfermera. Área de Gestión Integrada de Santiago.

Para que conste donde proceda, y a petición del promotor/investigador, en Santiago de Compostela, a 3 de noviembre de 2017.

El Secretario del Comité Territorial de Ética de la Investigación de Santiago Lugo,

XUNTA DE GALICIA SECRETARÍA XERAL TÉCNICA CONSELLERÍA DE SANIDADE

PRADA RAMALLAL GUILLERMO JOSE - 44841744R
2017.11.04 19:49:37 +02'00'

Guillermo José Prada Ramallal

Annex 3. Case Report Form

PATIENT CHN STICKER	Date:
	Project code for patient:
	Hospital:

Informed consent: (name of the doctor)
Date of IC:
Date of sampling:
Sample collected by:

☐ EDTA

Epidemiological Information

Sex:_____ Age:_____ Weight (kg):_____ Height (cm):_____

Personal History of Cancer:

Type of cancer1:	Age at diagnosis1:
Treatment1:	Adverse effects1:
Type of cancer2:	Age at diagnosis2:
Treatment2:	Adverse effects2:

Family History of CRC: ☐ YES ☐ NO

1st degree Age at diagnosis	☐ Father Age(____)	☐ Mother Age(____)	☐ Sibling Age(____)	☐ Son/daughter Age(____)
	☐ Sibling Age (____)	☐ Sibling Age (____)	☐ Sibling Age (____)	☐ Sibling Age (____)
	☐ Son/daughter Age (____)	☐ Son/daughter Age (____)	☐ Son/daughter Age (____)	☐ Son/daughter Age(____)

2nd degree	☐ Grandparent	☐ Uncle/Aunt	☐ Nephew/Niece
	☐ Grandparent	☐ Uncle/Aunt	☐ Nephew/Niece
	☐ Grandparent	☐ Uncle/Aunt	☐ Nephew/Niece
	☐ Grandparent	☐ Uncle/Aunt	☐ Nephew/Niece

Family history of other related cancers:

Degree of kinship and age at diagnosis:

Clinical Data

Type of cancer: colon ☐ rectal ☐ Location: right ☐ left ☐
Date of diagnosis:_____

ECOG: ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Stage:

A1	IA2	IA3	IB	IIA	IIB	IIIA	IIIB	IIIC	IVA	IVB
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Treatment Guideline:

Surgery: NO ☐ YES ☐

QT: NO ☐ YES ☐: Neoadyuvant ☐ Adyuvant ☐ Palliative ☐

TT start date:_____ TT end date:_____ End reason:_____

TT plan:_____

Dose_oxali:_____ Dose_fu:_____ Dose_irino:_____

Dose redution_2cycle:_____ Dose redution_ncycles:_____

Dose delay_2cycle:_____ Dose delay_ncycle:_____

Progression: NO ☐ YES ☐: Date of progression:_____

DPYD Status:_____

Exitus: NO ☐ YES ☐: Date of exitus:_____

RECIST: CR ☐ **PR** ☐ **PD** ☐ **SD** ☐

Adverse Effects (CTCAE)

1st cycle (Symptom)	(Count)	(Grade)		2nd cycle (Symptom)	(Count)	(Grade)		3rd cycle (Symptom)	(Count)	(Grade)	
Hemoglobin				Hemoglobin				Hemoglobin			
Leukocytes				Leukocytes				Leukocytes			
Neutrophils				Neutrophils				Neutrophils			
Thrombocytes				Thrombocytes				Thrombocytes			
Sickness/Vomiting				Sickness/Vomiting				Sickness/Vomiting			
Mucositis				Mucositis				Mucositis			
Diarrhea				Diarrhea				Diarrhea			
Neuropathy				Neuropathy				Neuropathy			
Neutropenia				Neutropenia				Neutropenia			
Hand-Foot Syndrome				Hand-Foot Syndrome				Hand-Foot Syndrome			
Alopecia				Alopecia				Alopecia			
Cardiotoxicity				Cardiotoxicity				Cardiotoxicity			
Other				Other				Other			

Annex 4. Protocols

Protocol 1 - Preparation of buffer $A^{+/+}$

1. Buffer $A^{-/-}$, (A1 + A2) pH=7.4 (store at 4°C, reproduce when cloudy sediments visible)
   a. Prepare Buffer A1 in 500 mL volumetric flask (pKa=12.4) → dissolve K2HPO4 before adding MgCl2
      - 35 mM K2HPO4 – 3.05 g K2HPO4 (anhydrous, MW=174.2 g/mol)
      - 2.5 mM MgCl2 – 1.25 mL 1M MgCl2 (40 mL MQ-H2O + 3.81g MgCl2, anhydrous MW=95.21 g/mol) (reaction heats up)
      - Bring to 500 mL with MQ-H2O, stir until clear
   b. Prepare Buffer A2 in 500 mL volumetric flask (pKa=7.2) → dissolve KH2PO4 before adding MgCl2
      - 35 mM KH2PO4 – 2.38 g KH2PO4 (anhydrous MW=136.1 g/mol)
      - 2.5 mM MgCl2 – 1.25 mL 1M MgCl2
      - Bring to 500 mL with MQ-H2O, stir until clear
   c. Add approx. 200 mL Buffer A1 in clean, rinsed, and dried glass beaker. Add stir bar and place on a stir plate (gently stirring).
   d. Add Buffer A2 until pH reaches 7.4. This is now buffer $A^{-/-}$.
2. Buffer $A^{+/+}$ (can be stored at 4°C for 2 weeks, check pH)
   a. 50 mL Buffer A-/-, check pH to be 7.4
   b. 17.5 µL β-Mercaptoethanol (=0.035%) (Handle with extreme care in chemical fume hood, double latex-gloves).
   c. 1 PIC Protease Inhibitor Tablet (cOmplete mini EDTA-free)

Protocol 2 - Total protein determination.

Determine the total protein concentration of each lysate using DC Protein Assay Kit (Bio-Rad Laboratories, Inc, California, USA).

1. Dilute cell lysate 1:10 before measuring (3 µL sample in 27 µL H2O).
2. Reserve the first 3 columns of a 96-well plate for the standards loaded in triplicates. The remaining wells will be for samples loaded in triplicates as well (8 samples at once possible).
3. Add 5 µL of sample and standards to each well.
4. Mix 3 mL Reagent A and 60 µL Reagent S, add 25 µL of this mixture to each well.
5. Add 200 µL Reagent B to each well, shake plate for 15s, incubate on bench for 15 minutes, and scan plate. Make sure that the correct wavelength is selected for the dye being used (method: Protein 690 nm).
6. Dilute all lysates to 400ng/µL with buffer $A^{+/+}$.
   - Prepare aliquots containing 4 µg of total protein for western blotting.
   - Prepare 320 µL aliquots for LC-MS/MS; store at -80°C.

Annex 5. Published article related to this thesis

Journal of *Personalized Medicine*

MDPI

Review

The Road so Far in Colorectal Cancer Pharmacogenomics: Are We Closer to Individualised Treatment?

Ana Rita Simões [1,2,†], Ceres Fernández-Rozadilla [1,2,*,†], Olalla Maroñas [1] and Ángel Carracedo [1,2,3,4]

[1] Grupo de Medicina Xenómica, Universidade de Santiago de Compostela (USC), 15706 Santiago de Compostela, Spain; anarita.santos.simoes@rai.usc.es (A.R.S.); olalla.maronas@usc.es (O.M.); angel.carracedo@usc.es (Á.C.)
[2] Instituto de Investigación Sanitaria de Santiago (IDIS), 15706 Santiago de Compostela, Spain
[3] Fundación Pública Galega de Medicina Xenómica; SERGAS, 15706 Santiago de Compostela, Spain
[4] Consorcio Centro de Investigación Biomédica en Red de Enfermedades Raras—CIBERER, 28029 Madrid, Spain
* Correspondence: ceres.fernandez.rozadilla@gmail.com
† These authors contributed equally to this work.

Received: 13 October 2020; Accepted: 17 November 2020; Published: 19 November 2020

Abstract: In recent decades, survival rates in colorectal cancer have improved greatly due to pharmacological treatment. However, many patients end up developing adverse drug reactions that can be severe or even life threatening, and that affect their quality of life. These remain a limitation, as they may force dose reduction or treatment discontinuation, diminishing treatment efficacy. From candidate gene approaches to genome-wide analysis, pharmacogenomic knowledge has advanced greatly, yet there is still huge and unexploited potential in the use of novel technologies such as next-generation sequencing strategies. This review summarises the road of colorectal cancer pharmacogenomics so far, presents considerations and directions to be taken for further works and discusses the path towards implementation into clinical practice.

Keywords: colorectal cancer; adverse drug reactions; pharmacogenomics; personalised medicine; toxicity

1. Introduction

Colorectal cancer (CRC) is the second leading cause of cancer-related death and the third most commonly diagnosed cancer [1]. Surgical resection is the preferable treatment independently of stage, but chemotherapy is widely used too across stages. There are different chemotherapeutic schemes for CRC treatment (Table 1).

Table 1. Guidelines for colorectal cancer (CRC) treatment.

CRC Stage	Treatment	
	Surgery	Pharmacological Treatment
I	Wide surgical resection and anastomosis	No adjuvant chemotherapy recommended
II	Wide surgical resection and anastomosis	Adjuvant chemotherapy for high-risk could be considered

Table 1. *Cont.*

CRC Stage	Treatment	
	Surgery	**Pharmacological Treatment**
III	Wide surgical resection and anastomosis	Adjuvant administration of oxaliplatin plus 5-FU or capecitabine
IV	The majority of patients have metastases that initially are not suitable for potentially curative resection. Revaluate after chemotherapy	Cytotoxic agents: 1st line: 5-FU or capecitabine alone or in combination either with oxaliplatin or irinotecan 2nd line: if refractory to irinotecan-based treatment, FOLFOX is recommended; and if refractory to oxaliplatin-based treatment, FOLFIRI is recommended
		Biological targeted agents: 1st line: monoclonal antibodies against VEGF (bevacizumab, aflibercept) and/or EGFR (cetuximab, panitumumab), if *RAS* mutation excluded Multi-kinase inhibitor: regorafenib

FOLFOX: folinic acid (leucovorin-LV) + fluorouracil + oxaliplatin; FOLFIRI: leucovorin + fluorouracil + irinotecan; VEGF: vascular endothelial growth factor; EGFR: epidermal growth factor receptor.

Usually, the first line of treatment is based on fluoropyrimidines: 5-fluorouracil (5-FU) or its oral prodrug capecitabine, either alone or in different combinations with other agents, the most common being leucovorin, oxaliplatin (named FOLFOX or XELOX -if capecitabine is used instead of 5FU) or irinotecan (FOLFIRI) [2–5]. Besides these cytotoxic agents, metastatic CRC (mCRC) treatment may in addition include biological targeted agents to improve patient outcome, such as monoclonal antibodies against vascular endothelial growth factor (VEGF) (bevacizumab), or against epidermal growth factor receptor (EGFR) (cetuximab and panitumumab) (Table 1) [4].

There are two essential factors to be taken into account when considering efficacy and appropriateness of a treatment: response and toxicity. Response is often evaluated based on overall survival, progression-free survival or response evaluation criteria in solid tumors (RECIST), in the case of unresectable CRC [6]. On the other hand, patients subject to chemotherapy are prone to develop adverse drug reactions (ADRs) that might be severe or even fatal, and have a considerable impact on healthcare and burden. These ADRs can affect the patients' quality of life (even in the long term) and may hinder treatment, due to necessary delays or dose reduction. A study with more than four thousand mCRC patients receiving FOLFOX, FOLFIRI or XELOX saw that 90% of patients had one ADR, and 66% of patients had >1 ADR during the first line of treatment [7]. These toxic events also come with an increased economic burden to resolve them, with haematological toxicities being the most costly to resolve, followed by respiratory, endocrine/metabolic, central nervous system and cardiovascular ones.

Since both response and toxicity events have heterogeneous distributions amongst patients, it has been hypothesised that these ADRs may be caused by underlying genetic variants. Moreover, because chemotherapy agents have only been used since the 1950s, any genomic variants having large effects on toxicity responses have not had time to be washed away by negative selection [8,9]. Moreover, because cancer is usually related to later stages of life and does not affect fitness, purifying selection against these variants is not in place. Therefore, it is feasible that genetic variants having moderate-to-large effects (detectable by classical association studies) could be responsible for the observed variability.

Pharmacogenetics is a science that aims to learn about the inherited inter-variability in response and ADRs after drug exposure. First-generation studies were focused on the analysis of genes with an a priori relationship to drug effect, i.e., those involved mainly in the adsorption, distribution, metabolism and excretion (ADME) of chemotherapeutic agents. Later, these studies started to apply global approaches without a previous functional hypothesis, like genome-wide association studies (GWAS). The Pharmacogenomics Knowledgebase (PharmGKB [10]) is a free database that aggregates,

curates, integrates and disseminates the knowledge obtained from these studies regarding the impact of human genetic variation on drug response and toxicity. Other important sources of pharmacogenomic information have also launched from the efforts of The Clinical Pharmacogenetics Implementation Consortium (CPIC), which aims to create, curate and post free detailed gene/drug clinical practice guidelines (https://cpicpgx.org/ (accessed on 29 October 2020)).

In this review, we summarise the available data on CRC pharmacogenomics to date and go beyond the typically discussed candidate gene approaches, to cover genome-wide studies and next-generation sequencing. We also reflect on the necessity of comprehensive works including molecular studies to assess variant functionality, and discuss the limitations towards clinical implementation in the light of cost-effectiveness to health systems. Last but not least, we discuss considerations for further studies towards a routine implementation of personalised medicine strategies in clinical practice.

2. Chemotherapeutic Agents in CRC Treatment

Chemotherapy based on fluoropyrimidines, specifically 5-FU, has been used for over thirty years now, and is still the backbone of CRC treatment (Figure 1) [11]. However, there have been reports that show that up to 94% of patients treated with this drug end up developing ADRs, some of which may be severe or life threatening (Table 2) [12]. For instance, some studies have shown that around 40–56% of patients treated with 5-FU develop severe neutropenia, and 10–15% present grade 3–4 diarrhoea [13]. Patients receiving capecitabine have a similar incidence of ADRs, although with less severe neutropenia, but present hand–foot syndrome (HFS) at a high incidence (54%) instead (Table 2) [14].

Table 2. The most common toxicity profile of CRC treatments.

Treatment	Significant ADRs (According to FDA Labels) *	ADR Incidence (% Patients)	Ref.
5-Fluororacil	Diarrhoea, neutropenia, mucositis, nausea/vomiting, stomatitis, asthenia, leukopenia, anaemia.	94%	[12]
Capecitabine	Hand-and-foot syndrome, diarrhoea, nausea/vomiting, abdominal pain, fatigue, hyperbilirubinemia.	96%	[12]
Oxaliplatin	Peripheral sensory neuropathy, neutropenia, thrombocytopenia, anaemia, nausea/vomiting, increase in transaminases and alkaline phosphatase, diarrhoea, fatigue, stomatitis.	>92%	[15]
Irinotecan	Nausea/vomiting, diarrhoea, neutropenia, alopecia, abdominal pain, constipation, anorexia, leukopenia, anaemia, asthenia, fever, body weight decreasing.	100%	[16]
Cetuximab	Cutaneous adverse reactions, headache, diarrhoea, infection.	>87%	[17]
Panitumumab	Skin rash, paronychia, fatigue, nausea, diarrhoea.	>90%	[18]
Bevacizumab	Haemorrhage, hypertension, headache, rhinitis, proteinuria, taste alteration, dry skin, lacrimation disorder, back pain, exfoliative dermatitis.	>60%	[19]

* According to Food and Drug Administration (FDA) label section: Warnings and Precautions, Contraindications, and Boxed Warning Sections of Labelling for Human Prescription Drug and Biological Products.

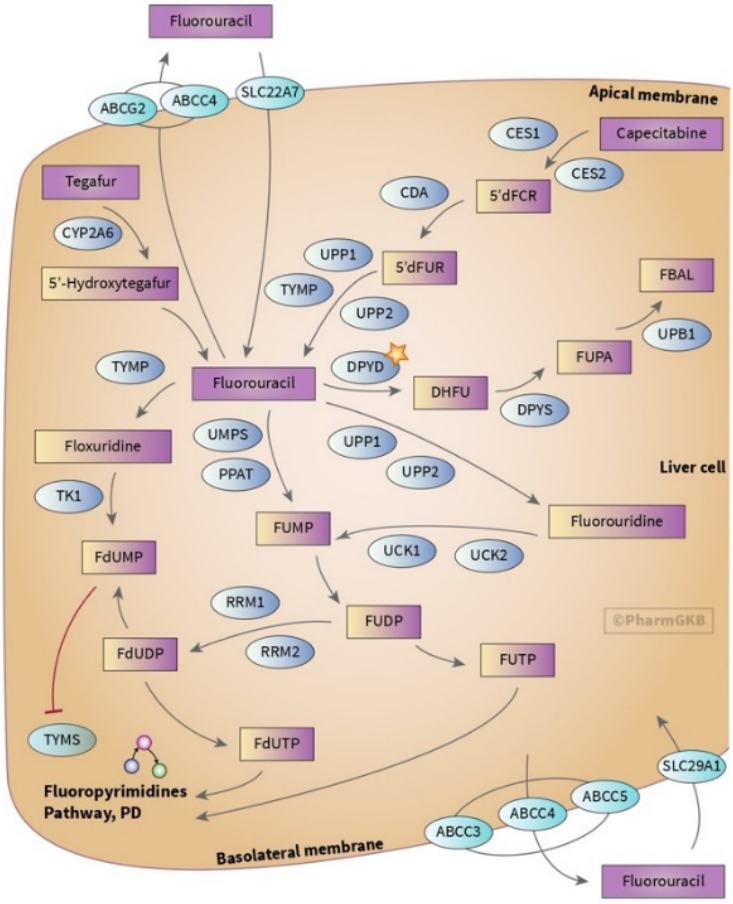


Figure 1. Graphic scheme of the genes involved in the adsorption, distribution, metabolism and excretion (ADME) of fluoropyrimidines [20]. Capecitabine passes through the gut wall and is metabolised into 5-deoxyfluorocytidine (5′dFCR) and 5′-deoxy-5-fluorouridine (5′dFUR) by carboxyl esterases (CES) and cytidine deaminase (CDA), respectively, and activated into 5-FU by thymidine phosphorylase (TP). - 5-FU is metabolised mostly in the liver by dihydropyrimidine dehydrogenase (DPD) (<80%) into dihydrofluorouracil (DHFU). The secondary elimination pathway is through urinary excretion or catabolism in extrahepatic tissues [21]. Its mechanism of action involves the methylenetetrahydrofolate reductase (MTHFR)—converting 5,10-methylentetrahydrofolate (5,10-MTHF) into 5-MTHF, which is required for purine and thymidine synthesis, and thymidylate synthase (TS) enzymes—forming a complex with 5,10-MTHF and deoxyuridine monophosphate (dUMP), which in the end disrupts DNA replication and repair. Used with PharmGKB and Stanford University permission (available at https://www.pharmgkb.org/pathway/PA150653776 (accessed on 24 September 2020)).

Platinum-based drugs, mainly oxaliplatin, are cytotoxic agents that prevent neoplastic proliferation, by forming DNA–platinum adducts, which block replication and transcription and induce apoptosis (Figure 2) (Table 1) [22]. The main oxaliplatin dose-limiting toxicity is neuropathy, occurring in about >90% of treated patients (Table 2) [15].

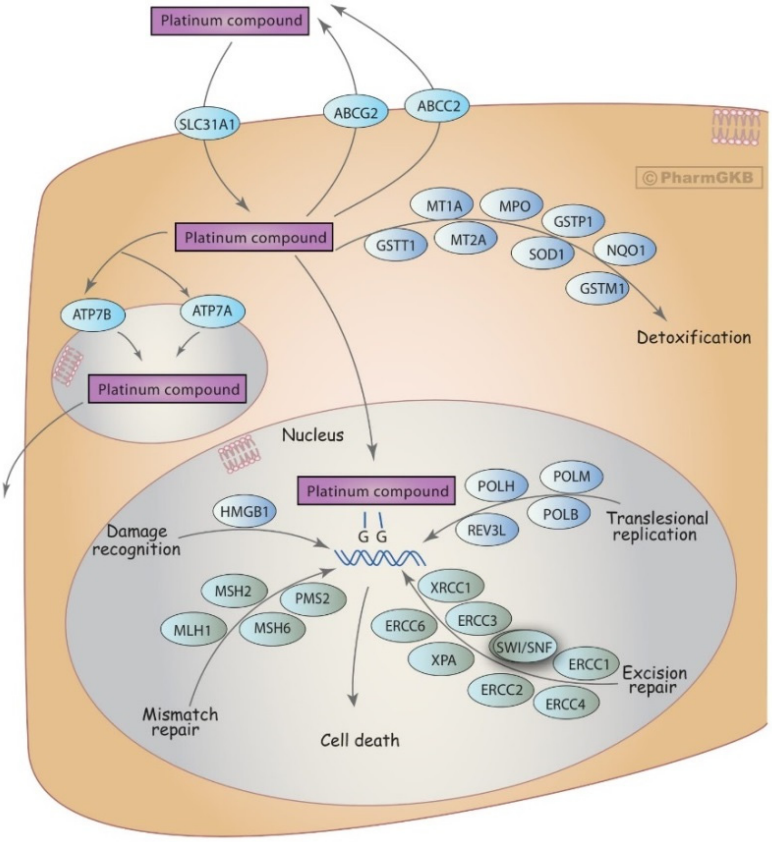


Figure 2. Graphic scheme of the genes involved in the ADME of platinum compounds, including oxaliplatin [23]. The glutathione S-transferases (GSTs), a multigene family of enzymes, undertake oxaliplatin detoxification. The solute carriers (SLCs) and adenosine-triphosphate binding cassette (ABC) transporters are responsible for oxaliplatin uptake and efflux in the liver, respectively, and so impact on drug bioavailability and toxicity profile. Further, the nucleotide excision repair (NER) and base excision repair (BER) pathways, which include the ERCC1 and ERCC2, and XRCC1 proteins, respectively, repair the damages cause by this drug. Used with PharmGKB and Stanford University permission (available at https://www.pharmgkb.org/pathway/PA150642262 (accessed on 24 September 2020)).

Irinotecan (CPT-11) is another cytotoxic agent used in the treatment of CRC in combination with 5-FU (FOLFIRI) (Table 1). FOLFIRI treatments result in better response rates and longer progression-free survival and overall survival of mCRC patients (Figure 3) [2,24]. CPT-11 is a semi-synthetic soluble analogue of the natural alkaloid camptothecin [25,26]. Some clinical trials report an ADR incidence for this drug of up to 100% of patients, where common ADRs include diarrhoea, neutropenia and alopecia (Table 2) [16].

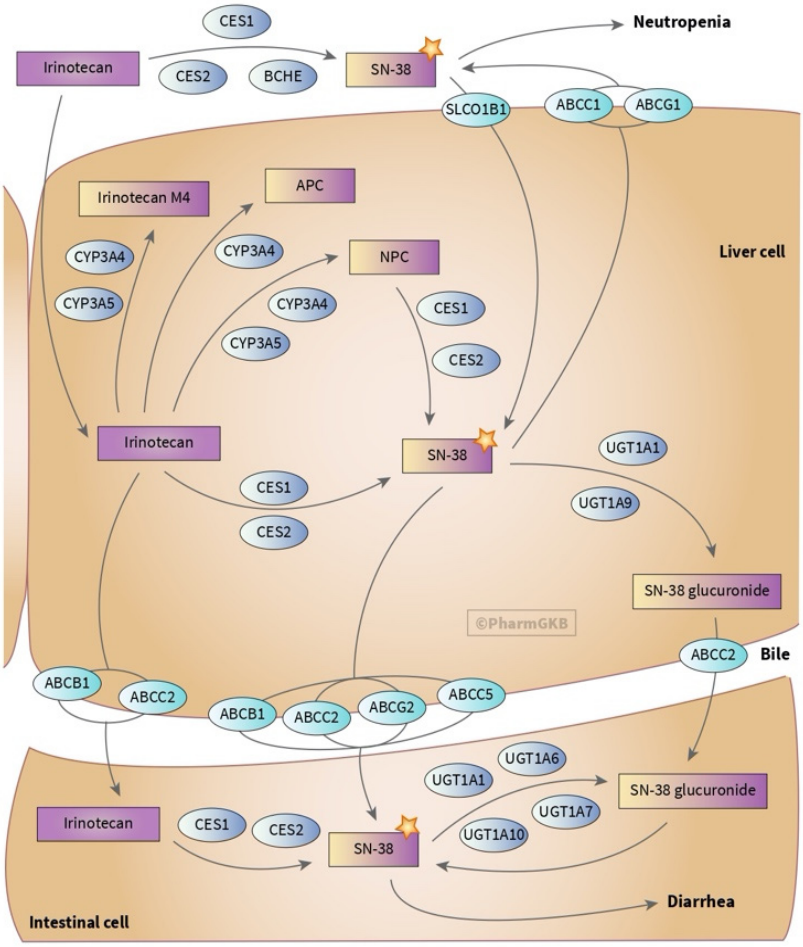


Figure 3. Graphic scheme of the genes involved in the ADME of irinotecan [10]. Irinotecan is converted into SN-38 by CES, which inhibits topoisomerase I, an enzyme essential for DNA replication and then into inactive SN-38G by UGTs. Further, it can suffer oxidation into 7-ethyl-10-[4-N-(5-aminopentanoic acid)-1-piperidino] carbonyloxycamptothecin (APC), M4 and 7-ethyl-10-[4-(1-piperidino)-1-amino] carbonyloxycamptothecin (NPC) by CYP3A4 and CYP3A5. NPC can be reactivated by CES into SN-38. Irinotecan and its metabolites' uptake and efflux are conducted by SLCs and ABC transporters, respectively. Used with PharmGKB and Stanford University permission (available at https://www.pharmgkb.org/pathway/PA2001 (accessed on 24 September 2020)).

In case of unresectable CRC, patients may also be given biological targeted agents. Cetuximab and panitumumab bind specifically to the human EGFR protein, which is constitutively expressed in normal epithelial tissues and overexpressed in some cancers like CRC. Some of the pioneer pharmacogenetics studies on treatment efficacy found, however, that because *RAS* mutations can constitutively activate the response pathway downstream from EGFR, anti-EGFR therapy efficacy is limited to patients' wild type for *KRAS* and *NRAS* [4]. These belong to signalling pathways downstream of EGFR, and mutations in these genes may cause EGFR-independent pathway activation, leading to resistance to anti-EGFR treatments [27]. More than 87% of patients receiving cetuximab develop an ADR and are commonly (>25%) prone to develop cutaneous reactions, headache, diarrhoea and infection, whereas patients receiving panitumumab (>20%) will probably have cutaneous reactions, fatigue, nausea and diarrhoea [17,18,25,26]. On the other hand, bevacizumab binding to VEGF blocks the interactions with its receptors on the endothelial cell surface. This interaction allows cell proliferation and angiogenesis, and thus bevacizumab reduces microvascular growth and inhibits metastatic progression. Over 60% of patients receiving bevacizumab develop ADRs, where the most common are hypertension, proteinuria, mucosal bleeding and wound healing problems [4,19].

3. Pharmacogenetics: Candidate Gene Studies

As we mentioned before, pharmacogenetic studies arose in the context of studying the genetic factors that contribute to ADRs. Initial efforts utilised candidate gene approaches to inspect mainly genetic variation in genes that might have a great influence on the drug pharmacokinetics and pharmacodynamics, and that can alter drug concentration levels, leading to toxicity.

3.1. Dihydropyrimidine Dehydrogenase (DPYD)

DPD, encoded by the *DPYD* gene, is responsible for the vast majority of 5-FU hepatic metabolism and is responsible for the first step and rate-limiting factor in the 5-FU catabolic pathway (Figure 1). Several single nucleotide polymorphisms (SNPs) have so far been identified in this gene in association with different toxicities [28]. The most studied *DPYD* variant is rs3918290 (*DPYD**2A, IVS14+1G>A), which causes exon 14 skipping and results in a truncated and catalytically inactive protein [29,30]. A study by Toffoli et al. on 603 patients treated with 5-FU-based chemotherapy reported the association of rs3918290 (OR = 8.5, $p = 0.008$), rs67376798 (OR = 7.8, $p = 0.012$) and rs55886062 (OR = 6.0, $p = 0.131$) with general toxicity (Table 3) [28].

Table 3. Summary of CRC pharmacogenomics.

Drug	Gene	SNP (rsID)	Change	Alternative Nomenclature	Frequency of Risk Allele [a]	Associated ADR	OR (95% CI)	Evidence Level [b]	Ref.
Fluoropyrimidines	*DPYD*	rs55886062	NM_000110.3:c.1679T>G; NP_000101.2:p.Ile560Ser	*DPYD*13	3×10^{-4} (C)	Global toxicity	6.0 (0.6–61)	1A	[28]
		rs3918290	NM_000110.4:c.1905+1G>A (Splice donor)	*DPYD*2A	0.007 (T)	Global toxicity	8.5 (1.8–40.9)	1A	[29]
		rs67376798	NM_000110.3:c.2846A>T; NP_000101.2:p.Asp949Val		0.003 (A)	Global toxicity	7.8 (1.6–39.2)	1A	[31]
		rs115232898	NM_000110.3:c.557A>G; NP_000101.2:p.Tyr186Cys		0.002 (Afr: 0.023) (C)	Neutropenia, mucositis, alopecia	-	1A	[32]
		rs75017182	NM_000110.4:c.1129-5923C>G (Intronic)		0.013 (C)	Global toxicity	6.8 (2.0–23)	1A	[33]
		rs56038477	NM_000110.3:c.1236G>A; NP_000101.2:p.Glu412=		0.014 (T)	Gastrointestinal; haematological	2.0 (1.5–2.8) 2.8 (1.2–3.7)	3	[34]
		rs72549303 c	NM_000110.4:c.1898del; NP_000101.2:p.Pro633fs	*DPYD*3	NA	NA	NA	1A	[31]
		rs72549309 c	NM_000110.4:c.295_298TCAT [1]; NP_000101.2:p.Phe100fs	*DPYD*7	6×10^{-5} (delATGA)	NA	NA	1A	[31]
		rs1801266 c	NM_000110.4:c.703C>T; NP_000101.2:p.Arg235Trp	*DPYD*8	3×10^{-5} (A)	NA	NA	1A	[31]
		rs1801268 c	NM_000110.4:c.2983G>T; NP_000101.2:p.Val995Phe	*DPYD*10	NA	NA	NA	1A	[31]
		rs78060119	NM_000110.3:c.1156G>T; NP_000101.2:p.Glu386Ter	*DPYD*12	8×10^{-6} (A)	Leucopenia, thrombocytopenia, mucositis	NA	1A	[35]
		rs2297595	NM_000110.3:c.496A>G; NP_000101.2:p.Met166Val		0.085 (C)	Global toxicity	5.9 (1.3–27.2)	3	[36]
		rs1801265	NM_000110.3:c.85T>C; NP_000101.2:p.Cys29Arg	*DPYD*9A	0.228 (G)	Diarrhoea	0.8 (0.7–1)	3	[37]
		rs1801267 c	NM_000110.4:c.2657G>A; NP_000101.2:p.Arg886His	*DPYD*9B	1×10^{-4} (T)	NA	NA	NA	[38]

Table 3. *Cont.*

Drug	Gene	SNP (rsID)	Change	Alternative Nomenclature	Frequency of Risk Allele [a]	Associated ADR	OR (95% CI)	Evidence Level [b]	Ref.
		rs1801159	NM_000110.3:c.1627A>G; NP_000101.2:p.Ile543Val	*DPYD*5	0.198 (C)	Diarrhoea	4.9 (-)	3	[39]
		rs1801158	NM_000110.3:c.1601G>A; NP_000101.2:p.Ser534Asn	*DPYD*4	0.015 (T)	Global toxicity	1.7 (1.1–2.6)	3	[37]
		rs17376848	NM_000110.3:c.1896T>C; NP_000101.2:p.Phe632=		0.051 (G)	Global toxicity	14.5 (1.4–155.2)	3	[36]
		rs1801160	NM_000110.3:c.2194G>A; NP_000101.2:p.Val732Ile	*DPYD*6	0.048 (T)	Global toxicity	2.1 (1.5–3.0)	3	[40]
		rs12022243	NM_000110.4:c.1906-14763G>A (Intronic)		0.181 (T)	Global toxicity	1.7 (1.5- 1.9)	3	[41]
		rs12119882	NM_000110.4:c.680+2545T>C (Intronic)		0.075 (G)	Hyperbilirubinemia	4.9 (1.2–20.8)	3	[42]
		rs76387818	Intergenic		0.019 (A)	Global toxicity	4.1 (3.5–4.6)	3	[41]
		rs12132152	Intergenic		0.020 (A)	HFS;global toxicity	6.1 (5.5–6.8); 1.6 (1.4–1.8)	3	[41]
	TYMS	rs183205964	NM_001071.4:c.-86= (5′ UTR)		3×10^{-5} (C)	Global toxicity	3.0 (1.1- 8.4)	3	[43]
		rs2853741	NM_001071.4:c. (Upstream transcript)		0.322 (T)	Diarrhoea	0.3 (0.1–0.7)	3	[42]
		rs699517	NM_017512.7:c.*1289= (3′ UTR)		0.379 (T)	Nausea/ vomiting;asthenia	7.9 (1.5–41.6); 0.3 (0.1–0.8)	3	[42]
		rs45445694	NM_001071.4:c. (5′ UTR)		0.007 (2R2R)	Global toxicity	1.7 (-)	3	[44]
		rs2853542	NM_001071.4:c.-58= (5′ UTR)			Global toxicity; HFS	1.5 (1.2–1.8); 1.4 (1.2–1.8)	NA	[45]
		rs11280056	NM_017512.7:c.*853_*861= (3′ UTR)			Global toxicity	1.7 (1.2–2.2)	NA	[45]
	ENOSF1	rs2612091	NM_017512.7:c.742-227G>C (Intronic)		0.373 (C)	Global toxicity	1.6 (1.4–1.8)	3	[41]

Table 3. *Cont.*

Drug	Gene	SNP (rsID)	Change	Alternative Nomenclature	Frequency of Risk Allele [a]	Associated ADR	OR (95% CI)	Evidence Level [b]	Ref.
	UMPS	rs2279199	NM_000373.4:c. (Genic upstream transcript)		0.556 (T)	Nausea/vomiting	0.2 (0.1–1.0)	3	[42]
		rs4678145	NM_000373.4:c.156+607G>C (Intronic)		0.096 (C)	Asthenia	4.5 (1.6–13.2)	3	[42]
		rs1801019 d	NM_000373.4:c.638G>C; NP_000364.1:p.Gly213Ala		0.169 (C)	Global toxicity	17.6 (1.6–195.9)	3	[46]
	MTHFR	rs1801131	NM_001330358.1:c.1409A>C; NP_001317287.1:p.Glu470Ala		0.289 (G)	HFS	10.0 (3.8–27.8)	3	[47]
		rs1801133	NM_001330358.1:c.788C>T; NP_001317287.1:p.Ala263Val		0.315 (A)	Neutropenia	2.3 (1.2–4.6)	3	[48]
	TYMP	rs11479	NM_001113755.3:c.1412C>T; NP_001244917.1:p.Ser471Leu		0.094 (A)	Global toxicity	2.7 (1.2–5.9)	3	[49]
	MIR27A	rs895819	NR_029501.1:n.40A>G (Non-coding transcript)		0.335 (C)	Global toxicity	1.6 (1.1–2.2)	3	[50]
	ABCC1	rs7194667	NM_032583.4:c.1609-491A>C (Intronic)		0.063 (G)	Leucopenia	3.31 (1.3–8.7)	3	[51]
	ABCB1	rs1045642	NM_001348945.1:c.3645T>C; NP_001335874.1:p.Ile1215=	*ABCB1**6	0.504 (G)	HFS	NA	3	[52]
		rs2032582	NM_001348945.1:c.2887T>G; NP_001335874.1:p.Ser963Ala	*ABCB1**7	0.637 (C)	HFS	NA	3	[52]
		rs1128503	NM_001348945.1:c.1446T>C; NP_001335874.1:p.Gly482=	*ABCB1**8	0.614 (G)	Neutropenia	NA	3	[52]
	SLC22A7	rs2270860	NM_006672.3:c.1269C>T; NP_006663.2:p.Ser423=		0.368 (T)	Global toxicity	17.1 (1.7–170.3)	3	[42]
		rs4149178	NM_006672.3:c.1586+206A>G (Intronic)		0.795 (A)	Diarrhoea	0.3 (0.1–0.9)	3	[42]

Table 3. *Cont.*

Drug	Gene	SNP (rsID)	Change	Alternative Nomenclature	Frequency of Risk Allele [a]	Associated ADR	OR (95% CI)	Evidence Level [b]	Ref.
	CDA	rs2072671	NM_001785.3:c.79A>C; NP_001776.1:p.Lys27Gln		0.279 (C)	Global toxicity	1.8 (1.1–3.0)	3	[53]
		rs1048977	NM_001785.3: c.435C>T; NP_001776.1:p.Thr145=		0.307 (T)	Hyperbilirubinemia	8.6 (1.1–70.3)	3	[42]
		rs602950	NM_001785.3:c. (Upstream transcript)		0.224 (G)	Diarrhoea	2.3 (1.3–4.2)	3	[47]
		rs3215400	NM_001785.3:c.-33_-31= (5′ UTR)		0.555 (delC)	HFS	0.5 (0.3–1.0)	3	[54]
		rs532545	NM_001785.3:c. (Upstream transcript)		0.220 (T)	Diarrhoea	2.3 (1.3–4.2)	NA	[47]
	CES1	rs3217164	NM_001025195.2:c.693+129del (Intronic)		0.607 (G)	Global toxicity	4.1 (1.8–9.0)	3	[55]
		rs2244614	NM_001025195.2:c.1171-41C>T (Intronic)		0.482 (G)	Global toxicity	4.7 (1.9–12.0)	3	[55]
		rs2244613	NM_001025195.2:c.1171-33C>T (Intronic)		0.232 (G)	Global toxicity	6.4 (1.5–27.7)	3	[55]
	CES1P1	rs7187684	NR_003276.2:n. (Intronic)		0.278 (T)	Global toxicity	6.5 (1.5–28.0)	3	[55]
		rs11861118	NR_003276.2:n. (Upstream transcript)		0.161 (G)	Global toxicity	6.5 (1.5–28.0)	3	[55]
	Intergenic	rs9936750	Intergenic		0.161 C	Global toxicity	4.6 (1.5–13.9)	3	[56]
	Intergenic	rs10876844	Intergenic		0.439 (A)	Diarrhoea	6.5 (1.6–27.2)	NA	[57]

Table 3. *Cont.*

Drug	Gene	SNP (rsID)	Change	Alternative Nomenclature	Frequency of Risk Allele [a]	Associated ADR	OR (95% CI)	Evidence Level [b]	Ref.
Oxaliplatin	*ABCC2*	rs717620	NM_000392.5:c.-24= (5′ UTR)		0.171 (T)	Neuropathy	14.4 (1.6–127.0)	3	[58]
		rs3740066	NM_000392.5:c.3972C> TNP_000383.2:p.Ile1324=			Neuropathy	3.0 (1.2–7.7)	NA	[58]
		rs1885301	NM_000392.5:c. (Upstream Transcript)		0.413 (A)	Neuropathy	3.1 (1.4–6.9)	NA	[58]
		rs4148396	NM_000392.5:c.3258+56T>C (Intronic)		0.347 (T)	Neuropathy	4.7 (1.6–13.7)	NA	[58]
	ABCG2	rs3114018	NM_004827.3:c.-19-3415T>G (Intronic)		0.516 (A)	Neuropathy	2.7 (1.0–4.4)	NA	[59]
	GSTP1	rs1695	NM_000852.3:c.313A>G; NP_000843.1:p.Ile105Val	*GSTP1**B	0.339 (G)	Dying	3.0 (1.2–7.6)	3	[60]
	GSTM1	Null genotype	-	*GSTM1**0		Neutropenia	2.0 (1.1–3.7)	NA	[61]
	GSTT1	Null genotype	-			Neutropenia	2.0 (1.1–3.7)	NA	[61]
	ERCC1	rs11615	NM_202001.3:c.354T>C; NP_001356337.1:p.Asn118=		0.498 (A)	Neutropenia	4.6 (1.2–17.4)	3	[48]
	ERCC2	rs13181	NM_000400.3:c.2251A>C; NP_000391.1:p.Lys751Gln		0.323 (G)	Haematological	2.2 (1.3–3.8)	3	[62]
		rs238406	NM_000400.4:c.468A>C NP_000391.1:p.Arg156=		0.645 (C)	Thrombocytopenia	NA	NA	[63]
	PARD3B	rs17626122	NM_001302769.2:c.3261-6168T>C (Intronic)		0.550 (T)	Global toxicity	3.4 (1.9–6.8)	3	[57]
	Intergenic	rs7325568	Intergenic		0.409 (T)	Haematological	1.8 (1.3–2.4)	3	[57]

Table 3. *Cont.*

Drug	Gene	SNP (rsID)	Change	Alternative Nomenclature	Frequency of Risk Allele [a]	Associated ADR	OR (95% CI)	Evidence Level [b]	Ref.
Irinotecan	*UGT1A1*	rs3064744	NM_000463.3:c. (Upstream transcript)	*UGT1A1*28*	0.347 (dupTA) (EAS:0.122)	Global toxicity	7.2 (2.5–22.3)	2A	[64]
		rs4148323 c	NM_000463.2:c.211G>A; NP_000454.1:p.Gly71Arg	*UGT1A1*6*	0.014 (EAS: 0.144) (A)	NA	NA	2A	[65]
		rs11563250	NM_001367507.1:c. (Genic upstream transcript)		0.893 (A)	Neutropenia	0.3 (0.2–0.6)	3	[66]
		rs4124874	NM_001072.3:c.862-10021T>G (Intronic)	*UGT1A1*60*	0.452 (T)	Neutropenia	NA	3	[67]
		rs10929302	NM_019075.2:c.856-9898G>A (Intronic)	*UGT1A1*93*	0.299 (A)	Global toxicity	8.4 (1.9–37.2)	3	[68]
	UGT1A9	rs11692021	NM_021027.3:c.855+9770T>C (Intronic)		0.349 (C)	Global toxicity	2.0 (1.1–3.6)	3	[69]
		rs3832043 e	NM_021027.3:c. (Upstream Transcript)		0.609 (delT)	Diarrhoea	6.3 (1.3–31.7)	3	[70]
	UGT1A6	rs2070959	NM_001072.4:c.541A>G (Intronic)		0.689 (A)	Global toxicity	2.1 (1.1–3.9)	3	[69]
	ABCG1	rs225440	NM_016818.3:c.286+7029C>T (Intronic)		0.428 (T)	Neutropenia	3.1 (1.1–8.6)	3	[71]
		rs425215	NM_016818.3:c.974-898C>G (Intronic)		0.623 (G)	Gastrointestinal	11.4 (1.7–78.4)	NA	[72]
	ABCB1	rs12720066	NM_001348945.1:c.2529+971T>G (Intronic)		0.035 (C)	Neutropenia	NA	3	[73]
	ABCC1	rs17501331	NM_004996.4:c.49-12232A>G (Intronic)		0.928 (A)	Neutropenia	NA	3	[73]
		rs3743527	NM_004996.4:c.*543= (3′ UTR)		0.774 (C)	Neutropenia	NA	3	[73]

Table 3. *Cont.*

Drug	Gene	SNP (rsID)	Change	Alternative Nomenclature	Frequency of Risk Allele [a]	Associated ADR	OR (95% CI)	Evidence Level [b]	Ref.
	ABCC5	rs2292997	NM_005688.4:c.129+7980C>T (Intronic)		0.126 (A)	Neutropenia	3.2 (1.3–7.9)	3	[71]
		rs10937158	NM_005688.4:c.130-1268A>T (Intronic)		0.612 (C)	Diarrhoea	0.4 (0.2–0.8)	3	[71]
		rs3749438	NM_005688.4:c.591+374C>T (Intronic)		0.324 (A)	Diarrhoea	5.9 (1.3–26.3)	3	[71]
		rs562	NM_005688.4:c.*1243= (3′ UTR)		0.515 (C)	Gastrointestinal	32.0 (2.8–370.8	NA	[72]
	ABCG2	rs7699188	NM_004827.3:c. (Genic upstream transcript)		0.227 (A)	Global toxicity; non-haematological	7.3 (1.5–34.5); 15.2 (2.5–78.2)	3	[74]
	SLCO1B1	rs2306283	NM_006446.5:c.388A>G NP_006437.3:p.Asn130Asp	*SLCO1B1*1b	0.529 (G)	Gastrointestinal	2.3 (0.4–15.1)	NA	[72]
	TOP1	rs6072262	NM_003286.4:c.279+61G>A (Intronic)		0.144 (A)	Neutropenia	NA	3	[75]
	TGFBR2	rs3087465	NM_001024847.2:c. (2KB upstream)		0.659 (G)	Diarrhoea	3.7 (1.0–13.3)	3	[76]
	TGFB1	rs1800469	NM_000660.7:c. (Upstream transcript)		0.701 (G)	Diarrhoea	4.4 (1.0–18.9)	3	[76]
	KCNQ5	rs9351963	NM_019842.4:c.490-1798A>C (Intronic)		0.178 (C)	Diarrhoea	3.3 (1.8–5.6)	3	[77]
	Intergenic	rs10486003	Intergenic		0.913 (C)	Neuropathy	0.3 (0.2–0.5)	NA	[78]
	Intergenic	rs2338	Intergenic		0.275 (A)	Neuropathy	2.3 (1.6–3.3)	NA	[78]
	Intergenic	rs830884	Intergenic		0.92 (T)	Neuropathy	0.3 (0.2–0.5)	NA	[78]
	ACYP2	rs843748	NM_001320586.2:c.405-28913G>A (Intronic)		0.379 (A)	Neuropathy	2.4 (1.6–3.7)	NA	[78]
	DLEU7	rs797519	NC_000013.11:g.50656996G>C (Intronic)		0.548 (G)	Neuropathy	0.5 (045–0.7)	NA	[78]
	FARS2	rs17140129	NM_001318872.2:c.-22+36771A>G (Intronic)		0.158 (G)	Neuropathy	3.3 (1.8–6.4)	NA	[78]

Table 3. *Cont.*

Drug	Gene	SNP (rsID)	Change	Alternative Nomenclature	Frequency of Risk Allele [a]	Associated ADR	OR (95% CI)	Evidence Level [b]	Ref.
Cetuximab	*EGFR*	rs712830	NM_005228.5:c.-191= (5′ UTR)		0.894 (C)	Global toxicity	6.1 (1.6–23.8)	3	[79]
		rs2227983	NM_005228.5:c.1562G>A NP_005219.2:p.Arg521Lys		0.768 (G)	Skin toxicity	3.2 (1.3–8.3)	3	[80]
		rs11568315	NM_005228.5:c.88+1195AC [10] (Intronic)		3.9 × 10−4 (CA > 35)	Skin toxicity	2.9 (1.0–8.9)	NA	[81]
	RPS7	rs10203413	NC_000002.12:g.3581588G>A (Regulatory region)		0.776 (G)	Skin toxicity	0.1 (0.1–0.4)	NA	[82]
	ZNF827	rs12646351	NC_000002.12:g.3581588G>A (Intronic)		0.815 (G)	Skin toxicity	0.04 (0.01–0.3)	NA	[82]
		rs17806780	NM_001306215.2:c.2383+11920A>T (Intronic)		0.818 (T)	Skin toxicity	0.04 (0.01–0.4)	NA	[82]
	EPHA5	rs7692430	NM_004439.8:c.2237-1876A>G (Intronic)		0.156 (G)	Skin toxicity	4.6 (2.5–8.5)	NA	[82]
Bevacizumab	*VEGF*	rs3025039	NM_001171623.1:c.*237= (3′ UTR)		0.134 (T)	Hypertension	0.2 (0.03–0.8)	NA	[83]
		rs2010963	NM_001171623.1:c.-634= (5′ UTR)		0.698 (G)	Hypertension	NA	NA	[84]
		rs833061	NM_001025366.3:c. (Upstream transcript)		0.452 (C)	Hypertension	0.2 (0.03–0.8)	NA	[85]
		rs699947	NM_001025366.3:c. (Upstream transcript)		0.414 (A)	Hypertension	0.1 (0.01–0.6)	NA	[85]

a: The risk alleles frequencies were consulted on gnomAD. b: Measure of confidence in the association, according to PharmGKB [10]. c: Associated with changes in enzymatic activity, but with a particular adverse drug reaction (ADR). d: Described for tegafur, a prodrug of 5-FU. e: Described for non-small-cell lung carcinoma. NA: not available. Note: In case of multiple studies, we have chosen a publication used by PharmGKB to support the level of evidence of the referred variant, and the corresponding OR and *p*-value.

A further meta-analysis including 7365 patients from eight different studies confirmed the association between *DPYD* rs55886062 (*DDYD**13) and *DPYD* rs56038477 with gastrointestinal (OR = 5.72, $p = 0.015$; 2.04, $p < 0.0001$, respectively) and haematological toxicities (OR = 9.76, $p = 0.00014$; and 2.07, $p = 0.013$, respectively), and also between *DPYD* rs3918290 and rs67376798 with overall toxicity (OR = 20.5, $p < 0.0001$; and 3.02, $p < 0.0001$, respectively) [34].

3.2. Thymidylate Synthetase (TYMS)

TS, encoded by the *TYMS* gene, is the main target of fluoropyrimidines and low levels of expression may influence toxicity [86,87]. The two most studied SNPs in *TYMS* are rs2853542 (5'VNTR 2R/3R) and rs11280056 (3'UTR 6bp ins-del). This gene has been widely studied, but with no conclusive results so far. Some studies have reported a correlation between rs2853542 and 5-FU/capecitabine toxicity, where the haplotype 2R/ins 6-bp was found to be significantly associated with severe toxicity [45,87], but other works could not replicate this association [61]. This might be explained by a work of Rosmarin et al. in 2015, which reported an association of an intronic variant located in the overlapping *ENOSF1* gene capable of explaining the toxicity attributed to the two previous *TYMS* polymorphisms. They discovered that SNP rs2612091 and *TYMS* 5'VNTR and 3'UTR are in moderate linkage disequilibrium (LD) ($r^2 = 0.40$ and 0.32, respectively), but after testing for dependency, they concluded that it was the rs2612091 G allele alone that increased the risk of toxicity ($p = 0.0021$). Although it has been proposed that the ENOSF1 protein could influence TYMS activity, the interaction between these two genes is not yet well understood [41]. Interestingly, genetic variation in *TYMS* has also been related to response to pyrimidine treatments, with higher levels of TS implicating worse response and poorer overall survival [88,89].

3.3. Methylenetetrahydrofolate Reductase (MTHFR)

MTHFR is the other major enzyme involved in 5-FU metabolism. Polymorphisms in this gene (namely rs1801133 and rs1801131) might impact enzyme activity, causing an accumulation of 5,10-MTHF, which increases toxicity [90]. Indeed, a study involving 292 stage II/III colon cancer patients found that the rs1801133 TT genotype was associated with neutropenia (OR = 2.32, $p = 0.014$) [48]. Another study involving 118 mCRC patients found that the same genotype was associated with diarrhoea ($p = 0.02$) [91]. However, other studies have not been able to find any association between polymorphisms in this gene and toxicity events [61,62,92,93].

3.4. Carboxyl Esterases (CES) and Cytidine Deaminase (CDA)

CES2 is the first enzyme in the conversion of capecitabine to 5-FU, followed by a second step catalysed by CDA (Figure 1). There have been some attempts to prove the association of polymorphisms on these two genes with ADRs, but there are still no concrete positive results. Ribelles et al. studied 136 patients and showed a trend ($p = 0.07$) between HFS and *CDA* SNP rs3215400 [54]. A study including 239 patients found an association of *CDA* rs2072671 with a high risk of overall toxicity (OR= 1.84, $p = 0.029$) [53]. Another work including 430 patients linked the *CDA* rs602950 and *CDA* rs532545 variants with diarrhoea (OR = 2.3, $p = 0.0055$, and 2.3, $p = 0.0082$, respectively) [47]. There have also been some smaller studies on *CES* polymorphisms and their association with capecitabine toxicity [45,54,94]. CES proteins are also important in the catabolic pathway of irinotecan (Figure 3) [95]. *CES1* rs2244613 was found to be associated with diarrhoea and patients with low *CES2* expression are more prone to develop neutropenia or diarrhoea [95–98].

3.5. DNA Repair Genes

DNA repair pathways have been extensively studied in pharmacogenomic studies [99]. A meta-analysis of more than 1000 CRC patients receiving oxaliplatin found a single significant association of the *ERCC1* rs11615 C allele with a higher risk of having haematological toxicity in Asian populations (HR = 1.97, $p < 0.05$) [100]. Boige et al. could not, however, replicate this association, perhaps due

to population differences, but did associate the *ERCC2* rs13181 C allele with a higher risk of severe haematological toxicity caused by FOLFOX (OR = 2.16, $p = 0.01$) [62]. A recent study on 596 CRC patients found that *ERCC1* rs11615 was significantly associated with stomatitis ($p = 0.03$) and nausea ($p = 0.04$), and that *ERCC2* rs13181 and rs238406 were associated with thrombocytopenia ($p = 0.004$ and $p = 0.03$, respectively) [63]. On the other hand, a study of 517 patients with stage II/III colon cancer concluded that polymorphisms in *ERCC1* and *XRCC1* did not have a clinically significant association with adverse effects [61]. Further smaller studies could neither confirm the relationship between these variants and toxicity [91,101].

3.6. Glutathione S-Transferases (GSTs)

GST enzymes are proteins from a multigene family, and specifically, GSTP1, GSTM1 and GSTT1 are involved in oxaliplatin detoxification (Figure 2). The most studied variations are *GSTP1* rs1695 and the complete deletion of the *GSTT1* and *GSTM1* genes. McLeod et al. tested these on 300 patients receiving FOLFOX in an advanced CRC setting. Patients bearing the *GSTM1* null genotype had a 1.7-fold increased risk of having severe neutropenia ($p = 0.016$), whereas homozygous patients for the rs1695 T allele had higher probability of discontinuing FOLFOX treatment due to neurotoxicity ($p = 0.01$) [102]. In contrast to these findings, Boige et al. did not find any significant association between these same SNPs and severe neurotoxicity on a study enrolling 349 patients [62]. Ruzzo et al. studied 517 patients and suggested a weak association between the *GST-T1/M1* null/null genotype and severe neutropenia (OR = 1.99, $p = 0.032$) [61], whereas Cecchin et al. analysed 154 patients receiving FOLFOX but could not replicate any markers of neurotoxicity. Interestingly, they suggested that variants other than genetics, such as the biological state of patients or disease stage, may also influence the detoxification pathway, and could therefore be responsible for the FOLFOX-related neurotoxicity [58].

3.7. Adenosine-Triphosphate Binding Cassette (ABC) Transporters

Genes within the ABC transporter family are responsible for the efflux of a variety of drugs and their metabolites, including oxaliplatin and irinotecan. However, there is a lot of controversy on the relationship of polymorphisms on *ABC* genes and chemotherapy-related toxicity. For 206 patients receiving FOLFOX, Custodio et al. reported that the *ABCG2* rs3114018 AA genotype had a significantly higher risk of neuropathy (OR = 2.67, $p = 0.059$) [59]. In a study including 144 patients, Cecchin et al. reported positive associations with neurotoxicity for SNPs in *ABCC2*: rs3740066 (OR = 2.99, $p = 0.0231$), rs1885301 (OR = 3.06, $p = 0.0072$), rs4148396 (OR = 4.69, $p = 0.0048$) and rs717620 (OR = 14.39, $p = 0.0164$), which are in high LD with one another. Others studies have been less successful in linking genetic variants in this gene with neurotoxicity or other toxicities [58,61,103].

In relation to irinotecan-based regimens, Salvador-Martín et al. showed that SNPs rs1128503, rs2032582 and rs1045642 in *ABCB1*, which are in LD, were associated with haematological and overall toxicity [92]. Others proposed the association of solely *ABCB1* rs1128503 (OR = 2.02, $p = 0.401$) with global toxicity, or of *ABCB1* rs1045642 with early toxicity (OR = 3.79, $p = 0.098$) (not strictly significant), while others did not find any association at all [74,93,95,96]. There have also been some reports on other *ABC transporter* genes, with conflicting results. For instance, a study on 26 mCRC patients showed that patients with the CC genotype in *ABCC5* rs562 or the GG genotype in *ABCG1* rs425215 presented higher gastrointestinal toxicity ($p < 0.02$) [72]. A study including 250 patients with mCRC linked the *ABCG2* rs7699188 variant with severe global toxicity (OR = 7.26, $p = 0.013$) [74].

3.8. Uridine Disphosphate Glucuronosyltransferases (UGTs)

UGT1A1 is the main enzyme responsible for SN-38 inactivation, followed by UGT1A7 and UGT1A9. Several groups have studied the influence of *UGT* polymorphisms on toxicity development. One of the most studied polymorphisms in *UGT1A1* is a change in the number of TA repeats $(TA)_n$TAA in the promoter region. The wild-type allele for this polymorphism is $(TA)_6$TAA, with $(TA)_7$TAA

(rs3064744, *UGT1A1*28) being frequent in Caucasians, but not in Asian populations (≈30% and ≈10%, respectively). However, rs4148323 (*UGT1A1*6) is more frequent in Asian populations comparing with Caucasians (≈14% and ≈1%, respectively). Ando and colleagues reported that patients carrying the *UGT1A1*28 genotype were at significantly higher risk of having irinotecan-related severe toxicity (OR = 7.23, $p < 0.001$) [64]. Innocenti et al. also stated that patients with *UGT1A1*28 had more events of severe neutropenia (OR = 9.3, $p = 0.001$) [67]. Others have also showed a correlation between *UGT1A1*28 and neutropenia, diarrhoea and vomiting ($p < 0.01$) [104–107]. Additionally, as for *TYMS*, it has been proven that the *UGT1A1* genotype also affects maximum tolerated dose and therefore response [108,109].

3.9. Solute Carriers (SLCs)

Reduction or elimination of the function of SLC genes due to genetic variation can lead to a decrease in SN-38 uptake, with further accumulation in plasma, ultimately leading to toxicity [97]. rs2306283 (*SLCO1B1*1b) has been shown to cause severe gastrointestinal toxicity, particularly diarrhoea and neutropenia [72,110,111]. A discovery study on 167 mCRC patients receiving irinotecan also revealed a protective effect of the *SLCO1B1* rs2291076 T allele against neutropenia but associated the rs2306283 GG genotype with significantly higher neutropenia events. These results were, however, not replicated in a posterior study of 250 mCRC patients [71].

3.10. Cytochrome p Gene Family (CYP)

CYP3A4 and CYP3A5 are responsible for the oxidation of irinotecan into the inactive metabolites APC, M4 and NPC. Some researchers have studied the possible association of polymorphisms on these genes and chemo-related toxicity but have not found any positive correlation [68,96,112], probably because over 80% of variants in *CYP* genes coding regions are very rare and the sample sizes of these studies were not large enough [113]. It has also been suggested that their enzymatic function might be altered by non-genetic factors such as diet, concomitant medications, altered liver function or patient's performance status [114].

3.11. Epidermal Growth Factor Receptor (EGFR)

Skin toxicity is the major ADR related to anti-EGFR agents. Parmar et al. studied 109 cancer patients and concluded that skin toxicity was linked to the *EGFR* rs2227983 GG genotype (OR = 3.24, $p = 0.014$) [80]. Dahan et al. studied 58 patients treated with third-line cetuximab and irinotecan, and reported a trend between the presence of rs11568315 (CA repeats ≤ 35) and skin toxicity (OR = 2.91, $p = 0.058$) [81]. Sunakawa et al. studied 77 patients treated with cetuximab in combination with oxaliplatin and also correlated rs11568315 (CA repeats ≤ 19) with skin toxicity [115]. A study on 52 patients treated with cetuximab and FOLFIRI found that *EGFR* rs712830 was significantly associated with severe global toxicity (OR = 6.13, $p = 0.010$), but not specifically with skin toxicity. rs712829, rs11568315 (CA repeats cut-off = 17) and rs4444903 were, however, not associated with any toxicity [79]. Another study on 46 mCRC patients receiving XELOX-bevacizumab with or without cetuximab also found no evidence for the association of either rs4444903 or rs11568315 (CA repeats cut-off = 20) with skin toxicity [116].

3.12. Vascular Endothelial Growth Factor (VEGF)

Hypertension is the major toxicity derived from anti-VEGF agent treatment. Studies on the relationship of *VEGF* polymorphisms and bevacizumab-related toxicity have also been controversial. For instance, a study on 89 patients reported a positive link between rs3025039 and hypertension (OR = 0.15, $p = 0.022$), but a meta-analysis of over 1000 cancer patients did not validate this finding [83,117]. Moreover, some researchers have reported that patients with the rs833061 TT, rs2010963 CC or rs699947 CC genotypes were less prone to hypertension caused by bevacizumab

($p < 0.03$) [84,85], but Etienne-Grimaldi et al. saw that patients harbouring the rs2010963 CC genotype alone had more toxicity than patients with other genotypes ($p = 0.01$) [118].

3.13. Immunotherapy and Toxicity

Immunotherapy has arisen in the past few years as a promising therapeutic option in many cancers, and has particular relevance in the case of tumours with microsatellite instability (MSI) [119]. Hence, the FDA approved, in 2018, the use of ipilimumab and nivolumab (anti-CTLA-4 and anti-PD1 monoclonal antibodies, respectively) for the treatment of metastatic CRC patients previously treated with standard chemotherapy [120]. In 2020, pembrolizumab (anti–PD-1) was also approved as a first-line treatment of patients with unresectable, MSI-high or mismatch repair-deficient metastatic CRC [121]. Although there have been some studies suggesting the influence of genetic variants on the development of toxicity due to these treatments in other cancer types, to date there is no sufficient data on CRC [122–124]. Surely novel data on this will shortly become available for pharmacogenomic studies as more patients undergo immunotherapy treatment.

4. Pharmacogenomic Approaches

4.1. Genome-Wide Association Studies (GWAS)

Despite the large effect sizes for toxicity variants discovered by candidate gene approaches, chemotherapy-related toxicity is likely complex and multigenic. Therefore, other discovery strategies may be more suitable to inspect genomic variation in a more comprehensive manner. This has been made possible by the increasing availability of higher-throughput technologies at increasingly affordable prices, which has allowed pharmacogenetics to go genomic. In these upcoming sections, we will describe the more recent approaches that have further expanded the knowledge on pharmacogenomics in recent years (Table 4).

Table 4. Advantages and disadvantages of different pharmacogenomics approaches.

Approach	Advantages	Disadvantages
Candidate genes	• offers biological plausibility • associates variants with known functional consequences and direct clinical implication	• bias toward certain genes/pathways (usually, ADME genes) based on prior information of relevance to phenotype, which may be incomplete • unable to discover novel genes/pathways • the selected SNPs may not represent the full variation of the studied genes • limited to protein-coding regions
SNP arrays (GWAS)	• unbiased by a priori functional knowledge • potential discovery of other relevant genes/pathways • potential to identify variation in regulatory regions such as promoters or enhancers • high-throughput	• need to be adequately powered to detect moderate-effect variants • require large sample sizes • multiple testing correction needs to be applied • variants might be intergenic; harder to interpret • inspects common populational variation (potential loss of rarer variants) • not suitable for CNV studies
SNP arrays (targeted fine-mapping approaches)	• denser coverage • cheaper • may be population-specific	• design bias • may require a priori knowledge of region to study (i.e., as defined by GWAS, for example).

Table 4. *Cont.*

Approach	Advantages	Disadvantages
NGS (targeted panels, WES, WGS)	• possibility of densely resequencing an entire gene (targeted genes) • allows a more comprehensive and unbiased identification of novel genetic biomarkers • allows the identification of relevant rare variants and CNV • rapid evolution of NGS technologies	• large number of false positives and VUS • need for validation by Sanger or other genotyping methods • higher turnaround time and costs (although decreasing) • need for high data storage capacity • need for deeper bioinformatic knowledge
Functional assays	• give mechanistic perspective on how variants exert their effect • validate the findings at the molecular level, giving further validity to the statistical association results • potentially applicable to a specific desired tissue	• assay design may be difficult, particularly in the case of intergenic variants • results must be replicated in clinical studies

GWAS make use of LD inheritance patterns to inspect common genetic variation across the entire genome. The main two advantages of GWAS over candidate gene studies are that they are unbiased by a priori functional knowledge on the variants (which may help in the discovery of other toxicity relevant pathways) and also have the potential to identify variation in regulatory regions such as promoters or enhancers, which have been largely unexplored by candidate gene approaches.

Several GWAS have been performed to inspect chemotherapy-related toxicity in CRC. In the QUASAR2 trial, Rosmarin et al. analysed over 1000 stage II/III CRC patients receiving capecitabine with or without bevacizumab to identify 1456 variants on 25 candidate genes (Table 3) [41]. Fernandez-Rozadilla et al. used 1012 patients in a two-stage study in patients treated with 5-FU and FOLFOX [57] to find a moderate association for the rs10876844 variant and diarrhoea in patients treated with 5-FU. Won et al. also completed a GWAS on 343 Korean patients receiving oxaliplatin-based regimens to identify possible genetic markers associated with chronic oxaliplatin-induced peripheral neurotoxicity (OXCPN) [78]. They found some evidence for an association that was intronic or within 100 Kb of genes related to various neuronal activities. Two subsequent and independent studies by Oguri et al. and Terrazzino et al. tried to validate these findings, but a single association between the *FARS2* rs17140129 G allele and OXCPN (OR = 6.5, $p = 0.034$) was found [125,126]. Lastly, the CAIRO2 trial included 282 advanced or metastatic CRC individuals treated with XELOX plus bevacizumab and cetuximab. They found some novel SNPs to be moderately associated with toxicity (Table 3) [82].

In general, although GWAS present several advantages over candidate gene strategies, there are also some important limitations, some of which could be overcome post hoc. Firstly, there is a lack of replication due to discrepancies in variant frequencies amongst the different populations used between studies, as seen when comparing the works from Won et al. and Terrazzino et al. mentioned above (Asian vs. Caucasian populations, respectively). Further, most of the associated variants are intergenic, which makes it harder to interpret the results directly and design appropriate validatory functional assays. Moreover, because we are evaluating thousands to millions of variants at a time, statistical power is a concern, and adequate study sample sizes are needed [127]. As an illustration, for a GWAS with a sample size of 200 patients, assessing variants with minor allele frequency (MAF) $\geq 5\%$, and a statistical threshold of 80% power, the OR that we would be able to discover is OR ≈ 2, which reflects a moderate effect.

GWAS are limited to inspecting common variation (i.e., generally over 5% MAF), but it is likely that toxicity variants may be of rarer prevalence [128,129]. Some approaches have been developed to overcome this limitation. For instance, targeted SNP panels can be designed to fine-map regions of interest spanning a large section of the gene or specific to a desired population. As an example, a commercially available array has been designed to include both common and low-frequency variation as well as Mendelian and functional alleles specific to Spanish genomes, which allows for better genotyping of the Spanish population when comparing with the generic global arrays [130]. Moreover, albeit possible, GWAS strategies are not usually suitable for CNV studies, because they demand that the CNV be in high LD with a genotyped SNP [57].

Despite these limitations, GWAS still hold great potential for discovery, given appropriate study conditions. Surely, there are still pathways contributing to toxicity development to be discovered, as proven by the contribution of *RPS7* to cetuximab-related toxicity. This gene is normally overexpressed in dermal papilla cells, which makes it reasonable that genetic variants could be associated with skin toxicity [82].

4.2. Next-Generation Sequencing (NGS)

NGS, either whole-exome (WES) or whole-genome sequencing (WGS), allows for a more comprehensive identification of novel genetic biomarkers in this regard, and several studies have reported the added value of NGS to identify relevant rare pharmacogenetic variants that would not be detected by other conventional methods (Table 4) [131–136].

In 2014, Mizzi et al. compared the data from 482 healthy individuals (data from Genomes Data and the Wellderly Study) obtained either with WGS or SNP array genotyping that included 1936 known pharmacogenomic variants within 231 ADMET genes (Table 5) [131]. Focusing on these genes, the WGS revealed an average of 17,733 variants vs. 249.5 found with the SNP array. In silico analysis with the PROVEAN and SIFT algorithms, which are in silico functional predictors, showed some missense variants likely to be deleterious. Specifically, they found that 254 of the 332 variants in *UGT1A1* were novel, of which 31 were functional and 26 had a frequency of <1%. In general, the WGS approach allowed the identification of a significantly higher number of variants compared to the SNP array, which might impact the pharmacological processes.

Table 5. Summary of relevant next-generation sequencing (NGS) results.

N	Cohort	Method	Genes	Results	Ref.
482	Genomes Data, Wellderly Study	WGS or SNP array genotyping	231 pharmacogenes	≈17,733 (WGS) vs. 249.5 (SNP array) *UGT1A1* (WGS): 254 of 332 variants were novel, 31 functional and 26 with frequency < 1%.	[131]
>6500	1KG phase 3; ESP	WES and WGS	146 pharmacogenes	19,328 SNV, 62.9% exonic 6225 and 6258 variants in *ABC* transporter (22 genes) and SLC genes (49), respectively, 253 variants in *UGTs* (16) and *GTSs* (14) 92.9% rare, 82.7% very rare 56.2% missenses ≈30–40% of the functional variability in pharmacogenes	[133]
141,456	gnomAD v2.1 [a]	WES and WGS	*SLC* genes	204,287 SNVs and indels, 56.9% missenses, 2.5% frameshifts, 1.7% stop-gains and 1.5% variations in canonical splice sites Each individual had ≈29.7 putatively functional *SLC* variants, 18% of functional variability due to rare variants	[136]

Table 5. *Cont.*

N	Cohort	Method	Genes	Results	Ref.
100	QUASAR	Amplicon sequencing	*DPYD* and *TYMS* coding regions	Novel rare independent *DPYD* variant (c.1651G>A; p.Ala551Thr)—classified as strongly damaging	[41]
62,402	1 KG phase 3; ExAC [b]	WES and WGS	208 pharmacogenes	201 (97%) genes had 5589 novel CNVs, 47% deletions and 54% duplications Novel deletions responsible for >5% of loss-of-function alleles in 87, 25, 49, 48, 59 and 51 genes in non-Finnish Europeans, Finnish, East Asians, South Asians, Africans and admixed Americans, respectively	[134]

1 KG: 1000 Genomes Project; ESP: Exome Sequencing Project; a: non-Finnish Europeans, Finns, Africans, East Asians, South Asians, Latinos, Ashkenazi Jews and other populations; ExAC: Exome Aggregation Consortium; b: included six major populations: non-Finnish Europeans, Finns, Africans, South Asians, East Asians and admixed Americans.

Another study analysed sequencing data for 146 genes related to pharmacological traits from over 6500 individuals (data from the 1000 Genomes Project (1KGP) and Exome Sequencing Project (ESP)) (Table 5) [133]. They detected 19,328 single nucleotide variants (SNVs), 62.9% of which were exonic; for example, 6225 and 6258 variants in *ABC transporter* (22 genes) and *SLC* genes (49) respectively, and 253 variants in *UGTs* (16) and *GTSs* (14). Most of these variants were indeed rare (MAF < 1%; 92.9%) or very rare (MAF < 0.1%; 82.7%)—meaning that they would not be detected by conventional methods—and the majority were missenses (56.2%). The functional impact from rare variants was different across the genes, yet they concluded that rare variants contribute on average 30–40% of the functional variability in the studied pharmacogenes.

Schaller et al. analysed WES and WGS data from 141,456 individuals (data from gnomAD v2.1) and assessed the genetic variability of *SLC* genes (Table 5) [136]. They detected 204,287 SNVs and indels, of which 56.9% were missenses, and several were loss-of-function variants, such as 2.5% frameshifts, 1.7% stop-gains and 1.5% variations in canonical splice sites. They concluded that each individual presents, on average, 29.7 putatively functional *SLC* variants, with rare variants contributing 18% of this functional variability.

Following on from the results obtained from their initial GWAS, Rosmarin et al. sequenced the complete *DPYD* and *TYMS* coding regions and identified a further novel rare independent *DPYD* variant (c.1651G>A; p.Ala551Thr). This change was present in a single patient that had presented with grade 4 neutropenia and thrombocytopenia, and was predicted to be "strongly damaging" by in silico predictors (Table 4) [41].

NGS approaches can not only be useful to identify rarer variants but can be an important asset to reveal copy number variations (CNVs). The case in point is the work by Santos et al. that included CNV available data from 2504 whole genomes and 59,898 exomes (data from 1KGP and Exome Aggregation Consortium (ExAC)) and focused on 208 ADME genes (Table 5) [134]. Within these, 201 (97%) genes had a total of 5589 novel CNVs, where 47% were deletions and 54% were duplications. These novel deletions were responsible for >5% of loss-of-function alleles in a considerable number of genes (87, 25, 49, 48, 59 and 51 genes in non-Finnish Europeans, Finnish, East Asians, South Asians, Africans and admixed Americans, respectively). This demonstrates the impact that CNV might have on ADME genes, and hence the development of ADRs.

As the conventional screening methods only include common variants, a high number of variants are missed, thus explaining the need for unbiased and more comprehensive approaches. These interesting works emphasise the potential of NGS to detect novel rarer variants or CNV, not only in ADME genes, but in other pathways, which might help to explain the pharmacogenetic variability possibly associated with toxicity caused by chemotherapy.

5. Functional Assays

Functional assays on candidate variants are essential to ultimately clarify the mechanisms by which the genetic variants exert their effect on ADR development. Pharmacokinetic (PK) studies have been the most used approach to assess the functional impact of toxicity SNPs (Table 4). They have been used for many years now to evaluate enzymatic activity in patients carrying the desired variants, as they measure the level of drug and its metabolites that influence drug bioavailability and could hence lead to the toxicity profile.

By far, the most studied gene in PK studies has been *DPYD*, and there is an agreement that the DPD protein plays a crucial role in 5-FU metabolism. There are several methods to determine DPD deficiency [30,137]: testing for DPD activity in peripheral blood mononuclear cells, the uracil breath test, the uracil test dose and endogenous DHU/U ratio, or high-performance liquid chromatography (HPLC).

A study including 30 patients heterozygous for the *DPYD* rs3918290 variant analysed 5-FU plasma concentrations by HPLC and found that the mean maximum enzymatic 5-FU conversion capacity value was 40% lower in these patients (Table 6) [138].

Table 6. Pharmacokinetic studies on fluoropyrimidines and irinotecan.

Genes	Significant Variants	N	Pharmacokinetic Results	Ref.
		Fluoropyrimidines		
DPYD	rs3918290	1 case (heterozygous for IVS14+1G>A) vs. 6 controls (CRC)	inactivation of one *DPYD* allele: strong $\downarrow CL_{5\text{-FU}}$: severe toxicity	[139]
DPYD	rs1801265 rs115232898 rs55886062	175 CRC patients	rs55886062: lowest activity ($p = 0.014$) rs115232898: 46% ↓activity ($p = 0.026$) rs1801265: 27% ↑activity ($p = 0.013$)	[140]
DPYD	rs3918290 rs67376798 rs55886062	487 advanced carcinoma patients	rs3918290, rs67376798, or rs55886062: $\downarrow CL_{5\text{-FU}}$ ($p < 0.001$)	[141]
DPYD	rs3918290	30 patients (heterozygous for IVS14+1G>A) and 18 controls	rs3918290: 40% ↓Vmax ($p < 0.001$)	[142]
DPYD	rs1801159	112 gastric or colon cancer patients	rs1801159: ↓k ($p = 0.022$) and nausea/vomiting ($p = 0.005$)	[143]
DPYD	rs55886062 rs1801265 rs1801158	Expression vector	rs1801158: 36% ↑activity ($p = 3.4 \times 10^{-7}$) rs1801265: 13% ↑activity ($p = 0.0013$) rs55886062: 75% ↓activity ($p = 5.2 \times 10^{-9}$)	[29]
DPYD	rs141044036 rs72549308 rs1801268 rs145773863 rs55674432 rs72547601 rs137999090 rs59086055 rs1801266 rs111858276 rs183385770 rs72549307 rs138616379 rs72549304 rs112766203 rs183105782 rs143986398 rs115232898 rs2297595	Expression vector	rs141044036, rs72549308, rs1801268, rs145773863, rs55674432, rs137999090, rs72547601, rs59086055: <12.5% activity ($p < 3.5 \times 10^{-4}$) rs1801266, rs72549307, rs111858276, rs138616379, rs183385770, rs72549304: 12.5–25% activity ($p < 0.0021$) rs112766203, rs143986398, rs183105782, rs115232898: >25% ↓activity ($p < 0.05$) rs2297595: 120% ↑activity ($p = 0.025$)	[31]
ABC	rs2271862	48 CRC patients	*ABCA2* rs2271862: $\uparrow CL_{5\text{-FU}}$	[144]

Table 6. *Cont.*

Genes	Significant Variants	N	Pharmacokinetic Results	Ref.
ABCB1 *ABCC1* *ABCG2* *UGT1A1*	rs12720066 rs6498588 rs10929302	85 advanced cancer patients	*ABCB1* rs12720066 ($p = 6.24 \times 10^{-4}$) and rs6498588 ($p = 9.50 \times 10^{-4}$), and *UGT1A1* rs10929302 ($p = 9.00 \times 10^{-5}$): $\uparrow AUC_{SN-38}$ $\uparrow AUC_{SN-38}$: G ≥ 3 neutropenia ($p = 0.0001$)	[73]
ABCG2 *SLCO1B1* *ABCB1* *ABCC1* *ABCC2* *UGT1A1* *UGT1A9* *UGT1A7* *CES* *CYP3A4* *CYP3A5* *HNF1A*	rs717620 rs1169288 rs4149056 rs35605 rs1092302 rs3740066	85 advanced cancer patients	*ABCC2* rs717620 ($p = 0.002$), *HNF1A* rs1169288 ($p = 0.007$), *SLCO1B1* rs4149056 ($p = 0.015$): $\uparrow AUC_{CPT-11}$ *ABCC1* rs35605 ($p = 0.031$), UGT1A1 rs1092302 ($p = 0.007$): $\uparrow AUC_{SN-38}$ *ABCC2* rs3740066: $\uparrow AUC_{SN-38G}$ and $\uparrow AUC_{APC}$ ($p = 0.012$) *ABCC1* rs35605 ($p = 0.023$), rs3064744 ($p < 0.0001$): ↓GR	[111]
UGT1A1	rs3064744	250 mCRC patients	↓GR ($p = 0.01$) and ↑BI ($p = 0.003$): G ≥ 3 toxicity rs3064744: ↓GR ($p = 0.01$) and ↑BI ($p = 0.007$)	[145]
UGT1A1 *UGT1A7* *UGT1A9*	rs4124874 rs10929302 *UGT1A7**3	Subset of 71 patients	*UGT1A1* rs4124874 and rs10929302: ↑BI ($p = 0.03$ and $p = 0.04$, respectively) *UGT1A7**3: ↓GR ($p = 0.02$) and ↑BI ($p = 0.007$)	[146]
HNF1A	rs2244608	Subset of 49 patients	rs2244608: $\uparrow AUC_{SN-38}$ ($p = 0.032$), ↑BI ($p = 0.021$) and ↓GR ($p = 0.035$)	[147]
ABCC2	rs2273697 rs17216114 rs1885301 rs2804402 rs717620 rs3740066	31 mCRC patients	rs2273697: $\downarrow AUC_{CPT-11}$ ($p = 0.011$) rs17216114: $\downarrow AUC_{SN-38}$ rs1885301, rs2804402, rs717620 and rs3740066: $\uparrow AUN_{SN-38}$ ($p < 0.03$)	[148]
ABCB1 *ABCC1* *ABCC2* *ABCG2* *CES1* *CES2* *CYP3A4* *CYP3A5* *UGT1A* *XRCC1*	rs1128503	65 solid tumour patients	*ABCB1* rs1128503: $\uparrow AUC_{CPT-11}$ ($p = 0.038$), AUC_{SN-38} ($p = 0.031$) and $\downarrow CL_{SN-38}$ ($p = 0.015$)	[96]
UGT1A1	rs3064744	20 solid tumour patients	rs3064744 ↓GR ($p = 0.001$) and ↑BI ($p = 0.001$) AUC_{SN-38}: neutropenia ($p < 0.0001$)	[105]
UGT1A1 *UGT1A9*	rs3064744	94 solid tumour patients	rs3064744: ↓GR ($p = 0.022$)	[149]
UGT1A1	rs4148323 rs4124874 rs3064744	85 solid tumour patients	rs4148323: ↓GR ($p = 0.0372$) rs4124874: ↑BI ($p = 0.0048$) rs3064744: ↑BI ($p = 0.0007$)	[150]
ABCC2 *UGT1A1*	*ABCC2**2 rs306474	167 solid tumour patients	*ABCC2**2: $\downarrow CL_{CPT-11}$ ($p = 0.020$) rs3064744: $\downarrow CL_{SN-38}$ ($p < 0.001$), GR and BI ($p = 0.014$)	[151]
UGT1A1	rs3064744	62 solid tumour patients	rs3064744: $\downarrow CL_{SN-38}$ ($p < 0.01$) ↑SN-38 exposure: G2–3 diarrhoea ($p = 0.03$)	[152]
UGT1A1	rs3064744	65 solid tumour or lymphoma patients	rs3064744: ↑BI ($p = 0.0003$) and ↓GR ($p = 0.03$) ↑BI: G4 neutropenia ($p = 0.001$)	[67]
UGT1A1 *UGT1A7* *UGT1A9* *UGT1A10*	rs3064744 rs4148323	176 cancer patients	rs3064744 or rs4148323: ↓GR ($p < 0.0001$)	[153]
UGT1A1 *ABCG2*	rs4148323	45 cancer patients	rs4148323: $\uparrow AUC_{SN-38}$ ($p = 0.018$), ↓GR ($p = 0.006$) and 61% ↑BI ($p = 0.003$)	[154]
ABCB1	*ABCB1**2	49 cancer patients	*ABCB1**2: $\downarrow CL_{CPT-11, SN-38, APC}$ ($p = 0.0154, 0.0043, 0.0169$, respectively)	[155]

Enzymatic activities were measured by high performance liquid chromatography (HPLC).

Another study reported the effect of *DPYD* rs75017182 on DPD expression and activity and showed that heterozygous carriers presented a 35% activity reduction that was caused by alternative splicing [33].

By these means, at least four SNPs in *DPYD* have been proven deleterious: rs55886062, rs3918290, rs67376798 and rs56038477/HapB3 [30,34,156]. Studies on other variations have so far led to inconclusive or contradicting results [157].

Of late, other approaches have also been used to assess the functionality of pharmacogenetic variants. For instance, Offer et al. proposed the construction of a vector for rapid phenotypic assessment of *DPYD* variants and their relation with 5-FU sensitivity (Table 6) [29,31]. *DPYD* constructs were expressed in mammalian cells and the enzymatic activity of the expressed proteins was measured by HPLC and compared to the wild type. By these means, they could confirm that 30 of the variants caused a significant reduction in enzymatic activity. Interestingly, 19 of the variants tested displayed <25% activity. In turn, *DPYD* rs1801158, rs1801265, rs2297595, rs200687447, rs60139309 and rs114096998 had higher enzymatic activity, and therefore cells expressing these variants were more resistant to 5-FU, which may not confer susceptibility to toxicity development, but may in turn influence response rates.

In 2015, Henricks et al. proposed to assign an activity value (AV) to *DPYD* alleles, to adjust the initial dose of 5-FU. In this context, fully functional alleles had an AV = 1, reduced activity alleles had an AV = 0.5 and non-functional alleles had an AV = 0 (wild-type AV = 1; rs67376798 and rs56038477 AV = 0.5; and rs3918290 and rs55886062 AV = 0). Based on the AV of both alleles, the gene activity score (AS) is calculated, thus representing the enzymatic phenotype of the patient [30].

For genes other than *DPYD*, there is much less functional evidence (Table 6). Some research has been conducted on the relation of irinotecan PK variants. These studies were able to significantly associate polymorphisms in *ABCC1* and *ABCB1* with SN-38 exposure and the glucuronidation ratio (GR)—measured as AUC SN-38G/AUC SN-38 [73,111]. Demattia et al. investigated the possible association between *ABCG2* rs7699188 and *ABCB1* rs2032582 with irinotecan PK parameters on patients with advanced CRC by measuring plasma concentrations of irinotecan, SN38 and SN38G, but did not find any significant correlation [74]. Toffoli et al. evaluated irinotecan PK in 71 patients with metastatic CRC. They associated severe toxicity with a significantly lower GR ($p = 0.01$) and an increased biliary index (BI) ($p = 0.003$), which indicates SN-38 accumulation. Further, they reported a significant correlation between *UGT1A1**28 and lower GR ($p = 0.01$), and higher BI ($p = 0.007$) [145]. Other works showed that patients with the wild-type genotype had a significantly higher clearance of SN-38 compared to *UGT1A1**28 ($p < 0.001$), and that the homozygous genotype was significantly associated with GR ($p = 0.005$) and BI ($p = 0.014$) [151]. Iyer et al. also reported significantly lower SN-38 glucuronidation in patients with *UGT1A1**28 ($p = 0.001$) [105]. Other *UGT1A* polymorphisms, such as *UGT1A1**60 ($p = 0.005$), *UGT1A1**93 ($p < 0.0001$), *UGT1A1**6 ($p = 0.037$) and *UGT1A7**3 ($p < 0.02$), were also associated with GR and BI [73,146,150].

6. Cost-Effectiveness Analysis

Besides the need for clear evidence on the functional relevance of a pharmacogenetic biomarker, a proof of cost-effectiveness—that the pharmacogenetic strategy is more effective with an acceptable additional cost or even a cost saving—is crucial to facilitate its introduction into clinical practice and acceptance from healthcare professionals and institutions.

In 2015, Deenen et al. evaluated the safety and costs of upfront *DPYD**2A genotyping with individualised dose adjustment treatment for fluoropyrimidines [158]. They showed that genotype-guided dosing represented a reduction in severe toxicity from 73% to 28%. Moreover, dose adjustment based on genotype produced shorter and easier to control toxicities, and a significant reduction in drug-induced death from 10% to 0%. Therefore, they demonstrated that screening for *DPYD**2A before treatment could be lifesaving and potentially cost-efficient. Cortejoso et al. complementarily evaluated the costs of genotyping three *DPYD* variants (rs3918290, rs67376798 and rs55886062) and the management of severe neutropenia caused by fluoropyrimidines. Considering an average cost

of management of EUR 3044.40 vs. EUR 6.40 per patient for *DPYD* testing, they concluded that genotyping is cost-effective if severe neutropenia is prevented in at least 2.1 cases per 1000 treated patients [159]. Given that the combined frequency of these three markers is about 1%, this provides evidence that *DPYD* testing should be considered by healthcare systems. Murphy et al. further compared the reactive vs. prospective *DPYD* genotyping of variants rs3918290, rs67376798, rs1801158 and rs55886062. Of the 134 included patients, five carried a *DPYD* variant and the costs of their hospitalisation were EUR 232,061, whereas the total cost of genotyping prior to treatment for all patients would have been only EUR 23,718. Even if patients still had to endure some ADRs, the cost would have been considerably smaller, making pharmacogenetic analysis again cost-efficient [160]. In 2019, Henricks et al. also compared the costs from prospective *DPYD* screening (rs3918290, rs67376798, rs55886062 and rs56038477) with no screening on 1103 patients receiving fluoropyrimidine-based therapy. Patients with variants rs67376798 or rs56038477 had a 25% dose reduction, while patients with rs3918290 or rs55886062 had a 50% dose reduction. They concluded that the expected costs of the screening approach were EUR 2599 vs. EUR 2650 for the non-screening approach, representing a cost saving of EUR 51 per patient. These results strongly suggested that upfront *DPYD*-guided dose individualisation does not result in extra costs, and therefore solidly supports *DPYD* screening implementation prior to fluoropyrimidine treatment as a standard of care [161]. It also constituted the basis for pharmGKB EMA guideline changes from actionable to recommended.

Gold et al. assessed the safety and costs of testing for *UGT1A1*28* before irinotecan treatment [162]. Assuming no treatment efficacy reduction, the average cost saving per patient was EUR 250. Obradovic et al. compared the standard irinotecan dose with dose reduction based on *UGT1A1* genotyping, and evaluated the cases of severe neutropenia, the number of life-years gained and the associated costs. They concluded that genotyping with dose reduction in homozygotes was cost-saving in African and Caucasian populations, but not in Asians, given the population frequency of this variant [163]. Another study by Butzke et al. compared severe neutropenia and grade 4 diarrhoea in a similar setting, to find that dosage calculations based on *UGT1A1*28* genotypes save about EUR 600 per patient [164]. More recently, Roncato et al. calculated that the costs for toxicity management per patient increased 1.4-fold for heterozygotes and 6-fold for homozygotes compared to wild-type individuals, and they were superior to the costs related to genotyping all patients before treatment [165].

7. Pharmacogenomic Testing Guidelines

Although, as we have described so far, there is a considerable amount of evidence on the effect of genetic variants on CRC chemotoxicity, translation into clinical practice is yet far from routine implementation. For now, guidelines from leading authorities, including the European Medicines Agency (EMA), the Food and Drug Administration (FDA), the private pharmacogenetic consortia, the CPIC, the Royal Dutch Association for the Advancement of Pharmacy-Pharmacogenetics Working Group (DPWG) and the Spanish Agency for Medicines and Health Products (Agencia Española de Medicamentos y Productos Sanitarios, AEMPS) have only produced a very limited list of recommendations (Table 7) [166–170].

Table 7. Current CRC pharmacogenetic guidelines for treatment administration.

Drug	Gene	Annotation by Drug Regulatory Agencies and Guidelines Recommendations				
		FDA	CPIC	AEMPS	EMA	DPWG
Fluoropyrimidines	*DPYD*	Actionable PGx [a] Withhold or permanently discontinue treatment in patients with evidence of acute early-onset or severe toxicity, which may indicate near complete or absence of DPD activity. No dose has been proven safe for patients with no DPD activity. There is insufficient data to recommend a specific dose in patients with partial DPD activity [170]	Actionable PGx [a] Intermediate metaboliser (individual with one normal function allele plus one no function allele or one decreased function allele, or with two decreased function alleles)—decreased DPD activity and increased risk for severe/fatal ADR. Reduce starting dose based on AS followed by titration of dose based on toxicity or therapeutic drug monitoring (if available). AS 1: Reduce dose by 50% AS 1.5: Reduce dose by 25–50% Poor metaboliser (individual with two no function alleles, or with one no function plus one decreased function allele)—complete DPD deficiency and increased risk for severe/fatal ADR. AS 0.5: Avoid treatment, and in case alternative agents are not suitable, strongly reduce starting dose with early therapeutic drug monitoring. AS 0: Avoid treatment [167]	Testing required [b] Test for *DPYD* genotype (c.1905+1G>A, c.1679T>G, c.2846A>T and c.1236G>A/HapB3) and/or DPD deficiency (measure blood uracil level) before treatment. Treatment is contraindicated in patients with complete DPD deficiency. In case of partial DPD deficiency with no suitable alternative agents, reduce initial dose and monitor levels. No concrete reduction has been established [166]	Testing required [b] Test for the lack of DPD activity before treatment (measure blood uracil level, or check for *DPYD* variants—c.1905+1G>A, c.1679T>G, c.2846A>T y c.1236G>A/HapB3). Treatment is contraindicated in patients with complete DPD deficiency. A reduced starting dose should be considered in patients with partial DPD deficiency [169]	-
Irinotecan	*UGT1A1*	Actionable PGx [a] Consider reduction in starting dose for patients homozygous for the *UGT1A1*28* allele. The precise dose reduction is not known and subsequent dose modifications should be considered based on individual patient tolerance to treatment [170]	-	-	-	Actionable PGx [a] Start with 70% of standard dose. If the patient tolerates it, the dose can be increased, guided by the neutrophil count [168]
Cetuximab/panitumumab	*EGFR* *KRAS* *NRAS*	Testing required [b] Determine *EGFR* expression status and confirm the absence of an *RAS* mutation before treatment [170]	-	-	Testing required [b] Test *RAS* status (*KRAS* and *NRAS* exons 2, 3 and 4) before treatment [171]	-

a: Actionable PGx—it may inform about changes in efficacy, dosage, metabolism or toxicity due to gene/protein/chromosomal variants or phenotypes, or contraindicate a drug in a subset of patients with particular variants/genotypes/phenotypes, without requiring prior testing. b: Testing required—it states that testing should be conducted before using a drug. This requirement may only be for a subset of patients.

Pharmacogenetic guidelines from CPIC for the administration of fluoropyrimidines recommend that the *DPYD* metaboliser status (based on variants rs3918290, rs67376798, rs55886062 and rs56038477) is characterised prior to treatment administration. Poor metabolisers (AS 0–0.5) either: (a) receive an alternative drug; or (b) if 5-FU/capecitabine is still considered the better suited option of treatment, it is recommended to strongly reduce the given dose and accompany with close monitoring. For intermediate metabolisers (AS 1–1.5), a 50% dose reduction is recommended [156,167]. On the other hand, the FDA only contraindicates the administration of 5-FU/capecitabine in patients with DPD deficiency, but does not directly recommend screening for DPD deficiency before treatment, neither does it distinguish heterozygous nor homozygous DPD-deficient patients (www.pharmgkb.org/gene/PA145/labelAnnotation (accessed on 07 October 2020)) [170].

As for irinotecan treatments, pharmacogenomic testing criteria are merely based on the *UGT1A1* genotype (rs3064744). DPWG recommends a 30% reduction in the standard dose if patients are *UGT1A1*28/28 [168], whereas the FDA vaguely recommends dose reduction (www.pharmgkb.org/chemical/PA450085/labelAnnotation (accessed on 07 October 2020)) [170]).

With the growing knowledge on CRC pharmacogenomics, more guidelines including other genes/variants will most likely be available in the next coming years. For instance, the *ABC* transporter genes, like *ABCC1* and *ABCB1*, have been quite studied so far and there is good evidence of their relation to the development of ADRs, both by association studies and functional assays.

8. Limitations in Pharmacogenomic Studies

In this appraisal, we have presented a comprehensive review of the field of CRC pharmacogenomics, since its early inception to the latest trends. Although remarkable findings have been produced, the road towards widespread clinical implementation is still far from over, and is inherently hindered by some of the limitations that pharmacogenetic analysis encounters. One of the main problems in pharmacogenomic studies is the extensive phenotype heterogeneity. This could be attributable to at least three different factors: (a) heterogeneity in clinical inclusion, i.e., differences in tumour staging and treatment strategies and lines (i.e., the genetic contribution to toxicity may be different in patients that have received FOLFOX as first-line treatment compared to those who have received it as second line); (b) pharmacogenomic data are not kept in a standardised manner, and it is usually hard to find in the patient's clinical record case report forms, including the appropriate scaling, timing and line of treatment, and Eastern Cooperative Oncology Group (ECOG) performance status of the patient (amongst others) should therefore be used to produce robust study designs; and (c) some ADRs, like haematological counts, can be measured quantitatively, whereas others, like diarrhoea, are subject to clinician interpretation. To overcome this, toxicity grading scales such as the Common Terminology Criteria for Adverse Events (CTCAE) should be used across studies [172].

Secondly, the influence of each therapeutical agent alone is hard, if not impossible, to assess, as the great majority of patients undergo combination therapies, and many of the ADRs are shared amongst treatments. This could be due, for instance, to the backbone presence of 5-FU in most settings but could also result from a pleiotropic effect of different drugs.

Thirdly, there has been, in general, a lack of unambiguous association findings. This could be due to the abovementioned phenotypic heterogeneity, but also other factors such as study sample sizes or population stratification issues. For instance, the overwhelming majority of studies reported in this review have been performed exclusively on Caucasian populations, and there are few published works in non-Europeans. Moreover, those that have been published in Asians show considerable differences in the allelic frequencies of the variants. Therefore, validation of findings in cohorts with appropriate statistical power is essential. On this topic, an outstanding example is the Radiogenomics Consortium, which advocates for the standardisation of toxicity data collection derived from radiation treatments. They have published guidelines for STrengthening the Reporting Of Genetic Association studies in Radiogenomics (STROGAR), which allow for multi-institutional approaches towards large-scale

radiotherapy patient biorepositories and databanks. Indeed, this consortium has already successfully completed several GWAS of radiotherapy toxicity [173–177].

Fourthly, there are no implemented guidelines for the reporting of pharmacogenetic studies. There have been recent efforts to overcome this, including a publication on the STrengthening the Reporting Of Pharmacogenetic Studies (STROPS). This work produces guidelines to standardise pharmacogenetic reporting. This could be essential for the homogenisation of pharmacogenetic data leading to improved systematic reviewing and meta-analyses, hence improving the power and applicability for pharmacogenetic associations [178].

Overall, the evidence gathered so far has brilliantly supported the relevance of pharmacogenomic testing in personalised medicine approaches. Novel genomic technologies such as GWAS and NGS offer unprecedented and affordable access to genomic information that can be assessed to discover novel pharmacogenomic variants related to toxic ADRs [179,180]. Pharmacokinetic profiling has proven to be useful for the identification of patients that might benefit from modified treatment strategies and might help improve the prediction value of genetic testing. Cost-effective analyses produced so far have validated the thought that the treatment design should be designed based on pharmacogenomic data, and that these strategies are always cost-effective vs. having to palliate toxicity issues.

Nevertheless, widespread testing is still anecdotic including in regulatory guideline recommendations. Researchers must hence make additional efforts to produce sound and relevant data that can be presented to the regulatory agencies to support pre-treatment testing. Surely, we must continue working in this direction towards a more meaningful implementation of pharmacogenomics in the routine clinical practice.

Author Contributions: Conceptualisation, C.F.-R.; writing—original draft preparation, A.R.S. and C.F.-R.; writing—review and editing, A.R.S., O.M., Á.C. and C.F.-R.; supervision, C.F.-R. and Á.C.; funding acquisition, Á.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by grant FIS PI 16/01057- ISCIII (to Á.C.), ISCIII PFIS grant FI17/00215 (to A.R.S.) and Fundación Olga Torres (to C.F.-R.).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424. [CrossRef] [PubMed]
2. Douillard, J.Y.; Cunningham, D.; Roth, A.D.; Navarro, M.; James, R.D.; Karasek, P.; Jandik, P.; Iveson, T.; Carmichael, J.; Alakl, M.; et al. Irinotecan combined with fluorouracil compared with fluorouracil alone. as first-line treatment for metastatic colorectal cancer: A multicentre randomised trial. *Lancet* **2000**, *355*, 1041–1047. [CrossRef]
3. de Gramont, A.; Figer, A.; Seymour, M.; Homerin, M.; Hmissi, A.; Cassidy, J.; Boni, C.; Cortes-Funes, H.; Cervantes, A.; Freyer, G.; et al. Leucovorin and fluorouracil with or without oxaliplatin as first-line treatment in advanced colorectal cancer. *J. Clin. Oncol.* **2000**, *18*, 2938–2947. [CrossRef] [PubMed]
4. Van Cutsem, E.; Cervantes, A.; Nordlinger, B.; Arnold, D. The ESMO Guidelines Working Group Metastatic colorectal cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **2014**, *25*, iii1–iii9. [CrossRef]
5. Iacovelli, R.; Pietrantonio, F.; Palazzo, A.; Maggi, C.; Ricchini, F.; De Braud, F.; Di Bartolomeo, M. Incidence and relative risk of grade 3 and 4 diarrhoea in patients treated with capecitabine or 5-fluorouracil: A meta-analysis of published trials. *Br. J. Clin. Pharmacol.* **2014**, *78*, 1228–1237. [CrossRef] [PubMed]
6. Schwartz, L.H.; Litière, S.; De Vries, E.; Ford, R.; Gwyther, S.; Mandrekar, S.; Shankar, L.; Bogaerts, J.; Chen, A.; Dancey, J.; et al. RECIST 1.1-Update and clarification: From the RECIST committee. *Eur. J. Cancer* **2016**, *62*, 132–137. [CrossRef]

7. Latremouille-Viau, D.; Chang, J.; Guerin, A.; Shi, S.; Wang, E.; Yu, J.; Ngai, C. The economic burden of common adverse events associated with metastatic colorectal cancer treatment in the United States. *J. Med. Econ.* **2016**, *20*, 54–62. [CrossRef]
8. Peters, E.J.; Motsinger-Reif, A.; Havener, T.M.; Everitt, L.; Hardison, N.E.; Watson, V.G.; Wagner, M.; Richards, K.L.; Province, M.A.; McLeod, H.L. Pharmacogenomic characterization of US FDA-approved cytotoxic drugs. *Pharmacogenomics* **2011**, *12*, 1407–1415. [CrossRef]
9. Watters, J.W.; Kraja, A.; Meucci, M.A.; Province, M.A.; McLeod, H.L. Genome-wide discovery of loci influencing chemotherapy cytotoxicity. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 11809–11814. [CrossRef]
10. Whirl-Carrillo, M.; McDonagh, E.M.; Hebert, J.M.; Gong, L.; Sangkuhl, K.; Thorn, C.F.; Altman, R.B.; Klein, T.E. Pharmacogenomics knowledge for personalized medicine. *Clin. Pharmacol. Ther.* **2012**, *92*, 414–417. [CrossRef]
11. Longley, D.B.; Harkin, D.P.; Johnston, P.G. 5-Fluorouracil: Mechanisms of action and clinical strategies. *Nat. Rev. Cancer* **2003**, *3*, 330–338. [CrossRef] [PubMed]
12. Food and Drug Administration. Highlights of prescribing information. *Capecitabine* **2015**. Available online: https://www.accessdata.fda.gov/drugsatfda_docs/label/2015/020896s037lbl.pdf (accessed on 13 November 2020).
13. De Matti, E.; Roncato, R.; Dalle Fratte, C.; Ecca, F.; Toffoli, G.; Cecchin, E. The use of pharmacogenetics to increase the safety of colorectal cancer patients treatedwith fluoropyrimidines. *Cancer Drug Resist.* **2019**, *2*, 116–130. [CrossRef]
14. Van Cutsen, E.; Twelves, C.; Cassidy, J.; Allman, D.; Bajetta, E.; Boyer, M.; Bugat, R.; Findlay, M.; Frings, S.; Jahn, M.; et al. Oral capecitabine compared with intravenous fluorouracil plus leucovorin in patients with metastatic colorectal cancer: Results of a large phase III study. *J. Clin. Oncol.* **2001**, *19*, 4097–4106. [CrossRef] [PubMed]
15. Food and Drug Administration. Highlights of prescribing information. *Oxaliplatin* **2002**. Available online: https://www.accessdata.fda.gov/drugsatfda_docs/label/2020/021492s016lbl.pdf (accessed on 13 November 2020).
16. Food and Drug Administration. Highlights of prescribing information. *Irinotecan* **2014**. Available online: https://www.accessdata.fda.gov/drugsatfda_docs/label/2014/020571s048lbl.pdf (accessed on 13 November 2020).
17. Food and Drug Administration. Highlights of prescribing information. *Cetuximab* **2009**. Available online: https://www.accessdata.fda.gov/drugsatfda_docs/label/2019/125084s273lbl.pdf (accessed on 13 November 2020).
18. Food and Drug Administration. Highlights of prescribing information. *Panitumumab* **2015**. Available online: http://www.accessdata.fda.gov/drugsatfda_docs/label/2015/125147s200lbl.pdf (accessed on 13 November 2020).
19. Food and Drug Administration. Highlights of prescribing information. *Bevacizumab* **2017**. Available online: https://www.accessdata.fda.gov/drugsatfda_docs/label/2011/125085s225lbl.pdf (accessed on 13 November 2020).
20. Thorn, C.F.; Marsh, S.; Carrillo, M.W.; McLeod, H.L.; Klein, T.E.; Altman, R.B. Pharm GKB summary: Fluoropyrimidine pathways. *Pharm. Genom.* **2011**, *21*, 237–242. [CrossRef]
21. Diasio, R.B.; Harris, B.E. *Clinical Pharmacokinetics*; Lippincott Williams and Wilkins: Philadelphia, PA, USA, 1989.
22. Raymond, E.; Chaney, S.G.; Taamma, A.; Cvitkovic, E. Oxaliplatin: A review of preclinical and clinical studies. *Ann. Oncol.* **1998**, *9*, 1053–1071. [CrossRef]
23. Marsh, S.; McLeod, H.; Dolan, E.; Shukla, S.J.; Rabik, C.A.; Gong, L.; Hernandez-Boussard, T.; Lou, X.J.; Klein, T.E.; Altman, R.B. Platinum pathway. *Pharm. Genom.* **2009**, *19*, 563–564. [CrossRef]
24. Saltz, L.B.; Cox, J.V.; Blanke, C.; Rosen, L.S.; Fehrenbacher, L.; Moore, M.J.; Maroun, J.A.; Ackland, S.P.; Locker, P.K.; Pirotta, N.; et al. Irinotecan plus Fluorouracil and Leucovorin for Metastatic Colorectal Cancer. *N. Engl. J. Med.* **2000**, *343*, 905–914. [CrossRef]
25. Wall, M.E.; Wani, M.C. Discovery to Clinic. *Ann. N. Y. Acad. Sci.* **1996**, *803*, 1–12. [CrossRef]
26. Rothenberg, M.L. Topoisomerase I inhibitors: Review and update. *Ann. Oncol.* **1997**, *8*, 837–855. [CrossRef]
27. Zhao, B.; Wang, L.; Qiu, H.; Zhang, M.; Sun, L.; Peng, P.; Yu, Q.; Yuan, X. Mechanisms of resistance to anti-EGFR therapy in colorectal cancer. *Oncotarget* **2016**, *8*, 3980–4000. [CrossRef]
28. Toffoli, G.; Giodini, L.; Buonadonna, A.; Berretta, M.; De Paoli, A.; Scalone, S.; Miolo, G.; Mini, E.; Nobili, S.; Lonardi, S.; et al. Clinical validity of a DPYD-based pharmacogenetic test to predict severe toxicity to fluoropyrimidines. *Int. J. Cancer* **2015**, *137*, 2971–2980. [CrossRef] [PubMed]
29. Offer, S.M.; Wegner, N.J.; Fossum, C.; Wang, K.; Diasio, R.B. Phenotypic profiling of DPYD variations relevant to 5-fluorouracil sensitivity using real-time cellular analysis and in vitro measurement of enzyme activity. *Cancer Res.* **2013**, *73*, 1958–1968. [CrossRef] [PubMed]

30. Henricks, L.M.; Lunenburg†, C.A.; Meulendijks, D.; Gelderblom, H.; Cats, A.; Swen, J.J.; Schellens, J.H.; Guchelaar, H. TranslatingDPYDgenotype into DPD phenotype: Using theDPYDgene activity score. *Pharmacogenomics* **2015**, *16*, 1275–1284. [CrossRef] [PubMed]
31. Offer, S.M.; Fossum, C.C.; Wegner, N.J.; Stuflesser, A.J.; Butterfield, G.L.; Diasio, R.B. Comparative Functional Analysis of DPYD Variants of Potential Clinical Relevance to Dihydropyrimidine Dehydrogenase Activity. *Cancer Res.* **2014**, *74*, 2545–2554. [CrossRef]
32. Zaanan, A.; Dumont, L.M.; Loriot, M.A.; Taieb, J.; Narjoz, C. *Clinical Pharmacology and Therapeutics*; CRC Press: Boca Raton, FL, USA, 2014.
33. Nie, Q.; Shrestha, S.; Tapper, E.E.; Trogstad-Isaacson, C.; Bouchonville, K.J.; Lee, A.M.; Wu, R.; Jerde, C.R.; Wang, Z.; Kubica, P.; et al. Quantitative Contribution of rs75017182 to Dihydropyrimidine Dehydrogenase mRNA Splicing and Enzyme Activity. *Clin. Pharmacol. Ther.* **2017**, *102*, 662–670. [CrossRef] [PubMed]
34. Meulendijks, D.; Henricks, L.M.; Sonke, G.S.; Deenen, M.J.; Froehlich, T.K.; Amstutz, U.; Largiadèr, C.R.; Jennings, B.A.; Marinaki, A.M.; Sanderson, J.D.; 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.* **2015**, *16*, 1639–1650. [CrossRef]
35. Kouwaki, M.; Hamajima, N.; Sumi, S.; Nonaka, M.; Sasaki, M.; Dobashi, K.; Kidouchi, K.; Togari, H.; Wada, Y. Identification of novel mutations in the dihydropyrimidine dehydrogenase gene in a Japanese patient with 5-fluorouracil toxicity. *Clin. Cancer Res.* **1998**, *4*, 2999–3004.
36. Falvella, F.S.; Cheli, S.; Martinetti, A.; Mazzali, C.; Iacovelli, R.; Maggi, C.; Gariboldi, M.; Pierotti, M.A.; Di Bartolomeo, M.; Sottotetti, E.; et al. DPD and UGT1A1 deficiency in colorectal cancer patients receiving triplet chemotherapy with fluoropyrimidines, oxaliplatin and irinotecan. *Br. J. Clin. Pharmacol.* **2015**, *80*, 581–588. [CrossRef]
37. Madi, A.; Fisher, D.; Maughan, T.S.; Colley, J.P.; Meade, A.M.; Maynard, J.; Humphreys, V.; Wasan, H.; Adams, R.A.; Idziaszczyk, S.; 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* **2018**, *102*, 31–39. [CrossRef]
38. Vreken, P.; Van Kuilenburg, A.B.P.; Meinsma, R.; Van Gennip, A.H. Identification of novel point mutations in the dihydropyrimidine dehydrogenase gene. *J. Inherit. Metab. Dis.* **1997**, *20*, 335–338. [CrossRef]
39. O'Donnell, P.H.; Trubetskoy, V.; Nurhussein-Patterson, A.; Hall, J.P.; Nath, A.; Huo, D.; Fleming, G.F.; Ingle, J.N.; Abramson, V.G.; Morrow, P.K.; et al. Clinical evaluation of germline polymorphisms associated with capecitabine toxicity in breast cancer: TBCRC-015. *Breast Cancer Res. Treat.* **2020**, *181*, 623–633. [CrossRef] [PubMed]
40. Ruzzo, A.; Graziano, F.; Galli, F.; Galli, F.; Rulli, E.; Lonardi, S.; Ronzoni, M.; Massidda, B.; Zagonel, V.; Pella, N.; 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* **2017**, *117*, 1269–1277. [CrossRef] [PubMed]
41. Rosmarin, D.; Palles, C.; Pagnamenta, A.; Kaur, K.; Pita, G.; Martin, M.; Domingo, E.; Jones, A.; Howarth, K.; Freeman-Mills, L.; 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* **2015**, *64*, 111–120. [CrossRef] [PubMed]
42. Pellicer, M.; García-González, X.; García, M.I.; Robles, L.; Grávalos, C.; García-Alfonso, P.; Pachón, V.; Longo, F.; Martínez, V.; Blanco, C.; et al. Identification of new SNPs associated with severe toxicity to capecitabine. *Pharmacol. Res.* **2017**, *120*, 133–137. [CrossRef] [PubMed]
43. Meulendijks, D.; Jacobs, B.A.; Aliev, A.; Pluim, D.; Van Werkhoven, E.; Deenen, M.J.; Beijnen, J.H.; Cats, A.; Schellens, J.H. Increased risk of severe fluoropyrimidine-associated toxicity in patients carrying a G to C substitution in the first 28-bp tandem repeat of the thymidylate synthase 2R allele. *Int. J. Cancer* **2015**, *138*, 245–253. [CrossRef]
44. Kristensen, M.H.; Pedersen, P.L.; Melsen, G.V.; Ellehauge, J.; Mejer, J. Variants in the dihydropyrimidine dehydrogenase, methylenetetrahydrofolate reductase and thymidylate synthase genes predict early toxicity of 5-fluorouracil in colorectal cancer patients. *J. Int. Med. Res.* **2010**, *38*, 870–883. [CrossRef]

45. Rosmarin, D.; Palles, C.; Church, D.; Domingo, E.; Jones, A.; Johnstone, E.; Wang, H.; Love, S.; Julier, P.; Scudder, C.; et al. Genetic markers of toxicity from capecitabine and other fluorouracil-based regimens: Investigation in the QUASAR2 study, systematic review, and meta-analysis. *J. Clin. Oncol.* **2014**, *32*, 1031–1039. [CrossRef]
46. Tsunoda, A.; Nakao, K.; Watanabe, M.; Matsui, N.; Ooyama, A.; Kusano, M. Associations of various gene polymorphisms with toxicity in colorectal cancer patients receiving oral uracil and tegafur plus leucovorin: A prospective study. *Ann. Oncol.* **2011**, *22*, 355–361. [CrossRef]
47. Loganayagam, A.; Arenas Hernandez, M.; Corrigan, A.; Fairbanks, L.; Lewis, C.M.; Harper, P.; Maisey, N.; Ross, P.; Sanderson, J.D.; Marinaki, A.M. Pharmacogenetic variants in the DPYD, TYMS, CDA and MTHFR genes are clinically significant predictors of fluoropyrimidine toxicity. *Br. J. Cancer* **2013**, *108*, 2505–2515. [CrossRef]
48. Lee, K.H.; Chang, H.J.; Han, S.W.; Oh, D.Y.; Im, S.A.; Bang, Y.J.; Kim, S.Y.; Lee, K.W.; Kim, J.H.; Hong, Y.S.; et al. Pharmacogenetic analysis of adjuvant FOLFOX for Korean patients with colon cancer. *Cancer Chemother. Pharm.* **2013**, *71*, 843–851. [CrossRef]
49. Jennings, B.A.; Loke, Y.K.; Skinner, J.; Keane, M.; Chu, G.S.; Turner, R.; Epurescu, D.; Barrett, A.; Willis, G. Evaluating Predictive Pharmacogenetic Signatures of Adverse Events in Colorectal Cancer Patients Treated with Fluoropyrimidines. *PLoS ONE* **2013**, *8*, 1–9. [CrossRef] [PubMed]
50. Meulendijks, D.; Henricks, L.M.; Amstutz, U.; Froehlich, T.K.; Largiadèr, C.R.; Beijnen, J.H.; de Boer, A.; Deenen, M.J.; Cats, A.; Schellens, J.H.M. Rs895819 in MIR27A improves the predictive value of DPYD variants to identify patients at risk of severe fluoropyrimidine-associated toxicity. *Int. J. Cancer* **2016**, *138*, 2752–2761. [CrossRef] [PubMed]
51. Magdy, T.; Arlanov, R.; Winter, S.; Lang, T.; Klein, K.; Toyoda, Y.; Ishikawa, T.; Schwab, M.; Zanger, U.M. ABCC11/MRP8 polymorphisms affect 5-fluorouracil-induced severe toxicity and hepatic expression. *Pharmacogenomics* **2013**, *14*, 1433–1448. [CrossRef] [PubMed]
52. Gonzalez-Haba, E.; García, M.I.; Cortejoso, L.; López-Lillo, C.; Barrueco, N.; García-Alfonso, P.; Alvarez, S.; Jiménez, J.L.; Martín, M.L.; Muñóz-Fernández, M.A.; et al. ABCB1 gene polymorphisms are associated with adverse reactions in fluoropyrimidine-treated colorectal cancer patients. *Pharmacogenomics* **2010**, *11*, 1715–1723. [CrossRef] [PubMed]
53. García-González, X.; Cortejoso, L.; García, M.I.; García-Alfonso, P.; Robles, L.; Grávalos, C.; González-Haba, E.; Marta, P.; Sanjurjo, M.; López-Fernández, L.A. Variants in CDA and ABCB1 are predictors of capecitabinerelated adverse reactions in colorectal cancer. *Oncotarget* **2015**, *6*, 6422–6430. [CrossRef]
54. Ribelles, N.; Lopez-Siles, J.; Sanchez, A.; Gonzalez, E.; Sanchez, M.; Carabantes, F.; Sanchez-Rovira, P.; Marquez, A.; Duenas, R.; Sevilla, I.; et al. A Carboxylesterase 2 Gene Polymorphism as Predictor of Capecitabine on Response and Time to Progression. *Curr. Drug Metab.* **2008**, *9*, 336–343. [CrossRef]
55. Hamzic, S.; Kummer, D.; Milesi, S.; Mueller, D.; Joerger, M.; Aebi, S.; Amstutz, U.; Largiadèr, C.R. Novel Genetic Variants in Carboxylesterase 1 Predict Severe Early-Onset Capecitabine-Related Toxicity. *Clin. Pharm. Ther.* **2017**, *102*, 796–804. [CrossRef]
56. Wheeler, H.E.; González-Neira, A.; Pita, G.; De La Torre-Montero, J.C.; Alonso, R.; Lopez-Fernandez, L.A.; Alba, E.; Martín, M.; Dolan, M.E. Identification of genetic variants associated with capecitabine-induced hand-foot syndrome through integration of patient and cell line genomic analyses. *Pharm. Genom.* **2014**, *24*, 231–237. [CrossRef]
57. Fernandez-Rozadilla, C.; Cazier, J.B.; Moreno, V.; Crous-Bou, M.; Guinó, E.; Durán, G.; Lamas, M.J.; López, R.; Candamio, S.; Gallardo, E.; et al. Pharmacogenomics in colorectal cancer: A genome-wide association study to predict toxicity after 5-fluorouracil or FOLFOX administration. *Pharm. J.* **2013**, *13*, 209–217. [CrossRef]
58. Cecchin, E.; D'Andrea, M.; Lonardi, S.; Zanusso, C.; Pella, N.; Errante, D.; De Mattia, E.; Polesel, J.; Innocenti, F.; Toffoli, G. A prospective validation pharmacogenomic study in the adjuvant setting of colorectal cancer patients treated with the 5-fluorouracil/leucovorin/oxaliplatin (FOLFOX4) regimen. *Pharm. J.* **2012**, *13*, 403–409. [CrossRef]
59. Custodio, A.; Moreno-Rubio, J.; Aparicio, J.; Gallego-Plazas, J.; Yaya, R.; Maurel, J.; Higuera, O.; Burgos, E.; Ramos, D.; Calatrava, A.; et al. Pharmacogenetic predictors of severe peripheral neuropathy in colon cancer patients treated with oxaliplatin-based adjuvant chemotherapy: A GEMCAD group study. *Ann. Oncol.* **2014**, *25*, 398–403. [CrossRef] [PubMed]

60. Stoehlmacher, J.; Park, D.J.; Zhang, W.; Yang, D.; Groshen, S.; Zahedy, S.; Lenz, H.J. A multivariate analysis of genomic polymorphisms: Prediction of clinical outcome to 5-FU/oxaliplatin combination chemotherapy in refractory colorectal cancer. *Br. J. Cancer* **2004**, *91*, 344–354. [CrossRef] [PubMed]
61. Ruzzo, A.; Graziano, F.; Galli, F.; Giacomini, E.; Floriani, I.; Galli, F.; Rulli, E.; Lonardi, S.; Ronzoni, M.; Massidda, B.; et al. Genetic markers for toxicity of adjuvant oxaliplatin and fluoropyrimidines in the phase III TOSCA trial in high-risk colon cancer patients. *Sci. Rep.* **2014**, *4*, 1–7. [CrossRef]
62. Boige, V.; Mendiboure, J.; Pignon, J.P.; Loriot, M.A.; Castaing, M.; Barrois, M.; Malka, D.; Trégouët, D.A.; Bouché, O.; Le Corre, D.; et al. Pharmacogenetic assessment of toxicity and outcome in patients with metastatic colorectal cancer treated with LV5FU2, FOLFOX, and FOLFIRI: FFCD 2000-05. *J. Clin. Oncol.* **2010**, *28*, 2556–2564. [CrossRef] [PubMed]
63. Salimzadeh, H.; Lindskog, E.B.; Gustavsson, B.; Wettergren, Y.; Ljungman, D. Association of DNA repair gene variants with colorectal cancer: Risk, toxicity, and survival. *BMC Cancer* **2020**, *20*, 1–10. [CrossRef] [PubMed]
64. Ando, Y.; Saka, H.; Ando, M.; Sawa, T.; Muro, K.; Ueoka, H.; Yokoyama, A.; Saitoh, S.; Shimokata, K.; Hasegawa, Y. Polymorphisms of UDP-glucuronosyltransferase gene and irinotecan toxicity: A pharmacogenetic analysis. *Cancer Res.* **2000**, *60*, 6921–6926. [PubMed]
65. Onoue, M.; Terada, T.; Kobayashi, M.; Katsura, T.; Matsumoto, S.; Yanagihara, K.; Nishimura, T.; Kanai, M.; Teramukai, S.; Shimizu, A.; et al. UGT1A1*6 polymorphism is most predictive of severe neutropenia induced by irinotecan in Japanese cancer patients. *Int. J. Clin. Oncol.* **2009**, *14*, 136–142. [CrossRef] [PubMed]
66. Chen, S.; Laverdiere, I.; Tourancheau, A.; Jonker, D.; Couture, F.; Cecchin, E.; Villeneuve, L.; Harvey, M.; Court, M.H.; Innocenti, F.; et al. A novel UGT1 marker associated with better tolerance against irinotecan-induced severe neutropenia in metastatic colorectal cancer patients. *Pharm. J.* **2015**, *15*, 513–520. [CrossRef]
67. Innocenti, F.; Undevia, S.D.; Iyer, L.; Chen, P.X.; Das, S.; Kocherginsky, M.; Karrison, T.; Janisch, L.; Ramírez, J.; Rudin, C.M.; et al. Genetic variants in the UDP-glucuronosyltransferase 1A1 gene predict the risk of severe neutropenia of irinotecan. *J. Clin. Oncol.* **2004**, *22*, 1382–1388. [CrossRef]
68. Côté, J.F.; Kirzin, S.; Kramar, A.; Mosnier, J.F.; Diebold, M.D.; Soubeyran, I.; Thirouard, A.S.; Selves, J.; Laurent-Puig, P.; Ychou, M. UGT1A1 polymorphism can predict hematologic toxicity in patients treated with irinotecan. *Clin. Cancer Res.* **2007**, *13*, 3269–3275. [CrossRef]
69. Lévesque, É.; Bélanger, A.S.; Harvey, M.; Couture, F.; Jonker, D.; Innocenti, F.; Cecchin, E.; Toffoli, G.; Guillemette, C. Refining the UGT1A haplotype associated with irinotecan-induced hematological toxicity in metastatic colorectal cancer patients treated with 5-fluorouracil/irinotecan-based regimens. *J. Pharm. Exp. Ther.* **2013**, *345*, 95–101. [CrossRef]
70. Han, J.-Y.; Lim, H.-S.; Park, Y.H.; Lee, S.Y.; Lee, J.S. Integrated pharmacogenetic prediction of irinotecan pharmacokinetics and toxicity in patients with advanced non-small cell lung cancer. *Lung Cancer* **2009**, *63*, 115–120. [CrossRef]
71. Chen, S.; Villeneuve, L.; Jonker, D.; Couture, F.; Laverdière, I.; Cecchin, E.; Innocenti, F.; Toffoli, G.; Lévesque, E.; Guillemette, C. ABCC5 and ABCG1 polymorphisms predict irinotecan-induced severe toxicity in metastatic colorectal cancer patients. *Pharm. Genom.* **2015**, *25*, 573–583. [CrossRef]
72. Di Martino, M.T.; Arbitrio, M.; Leone, E.; Guzzi, P.H.; Rotundo, M.S.; Ciliberto, D.; Tomaino, V.; Fabiani, F.; Talarico, D.; Sperlongano, P.; et al. Single nucleotide polymorphisms of ABCC5 and ABCG1 transporter genes correlate to irinotecan-associated gastrointestinal toxicity in colorectal cancer patients: A DMET microarray profiling study. *Cancer Biol. Ther.* **2011**, *12*, 780–787. [CrossRef]
73. Li, M.; Seiser, E.L.; Baldwin, R.M.; Ramirez, J.; Ratain, M.J.; Innocenti, F.; Kroetz, D.L. ABC transporter polymorphisms are associated with irinotecan pharmacokinetics and neutropenia. *Pharm. J.* **2018**, *18*, 35–42. [CrossRef]
74. De Mattia, E.; Toffoli, G.; Polesel, J.; D'Andrea, M.; Corona, G.; Zagonel, V.; Buonadonna, A.; Dreussi, E.; Cecchin, E. Pharmacogenetics of ABC and SLC transporters in metastatic colorectal cancer patients receiving first-line FOLFIRI treatment. *Pharm. Genom.* **2013**, *23*, 549–557. [CrossRef]
75. Hoskins, J.M.; Marcuello, E.; Altes, A.; Marsh, S.; Maxwell, T.; Van Booven, D.J.; Paré, L.; Culverhouse, R.; McLeod, H.L.; Baiget, M. Irinotecan Pharmacogenetics: Influence of Pharmacodynamic Genes. *Clin. Cancer Res.* **2008**, *14*, 1788–1796. [CrossRef]

76. Li, J.; Yu, Q.; Fu, S.; Xu, M.; Zhang, T.; Xie, C.; Feng, J.; Chen, J.; Zang, A.; Cai, Y.; et al. A novel genetic score model of UGT1A1 and TGFB pathway as predictor of severe irinotecan-related diarrhea in metastatic colorectal cancer patients. *J. Cancer Res. Clin. Oncol.* **2016**, *142*, 1621–1628. [CrossRef]
77. Takahashi, H.; Sai, K.; Saito, Y.; Kaniwa, N.; Matsumura, Y.; Hamaguchi, T.; Shimada, Y.; Ohtsu, A.; Yoshino, T.; Doi, T.; et al. Application of a Combination of a Knowledge-Based Algorithm and 2-Stage Screening to Hypothesis-Free Genomic Data on Irinotecan-Treated Patients for Identification of a Candidate Single Nucleotide Polymorphism Related to an Adverse Effect. *PLoS ONE* **2014**, *9*, e105160. [CrossRef]
78. Won, H.H.; Lee, J.; Park, J.O.; Park, Y.S.; Lim, H.Y.; Kang, W.K.; Kim, J.W.; Lee, S.Y.; Park, S.H. Polymorphic markers associated with severe oxaliplatin-induced, chronic peripheral neuropathy in colon cancer patients. *Cancer* **2012**, *118*, 2828–2836. [CrossRef]
79. Etienne-Grimaldi, M.C.; Bennouna, J.; Formento, J.L.; Douillard, J.Y.; Francoual, M.; Hennebelle, I.; Chatelut, E.; Francois, E.; Faroux, R.; El Hannani, C.; et al. Multifactorial pharmacogenetic analysis in colorectal cancer patients receiving 5-fluorouracil-based therapy together with cetuximab-irinotecan. *Br. J. Clin. Pharm.* **2012**, *73*, 776–785. [CrossRef]
80. Parmar, S.; Schumann, C.; Rüdiger, S.; Boeck, S.; Heinemann, V.; Kächele, V.; Seeringer, A.; Paul, T.; Seufferlein, T.; Stingl, J.C. Pharmacogenetic predictors for EGFR-inhibitor-associated skin toxicity. *Pharm. J.* **2013**, *13*, 181–188. [CrossRef] [PubMed]
81. Dahan, L.; Norguet, E.; Etienne-Grimaldi, M.-C.; Formento, J.-L.; Gasmi, M.; Nanni, I.; Gaudart, J.; Garcia, S.; Ouafik, L.; Seitz, J.-F.; et al. Pharmacogenetic profiling and cetuximab outcome in patients with advanced colorectal cancer. *BMC Cancer* **2011**, *11*, 496. [CrossRef]
82. Baas, J.; Krens, L.; Bohringer, S.; Mol, L.; Punt, C.; Guchelaar, H.; Gelderblom, H. Genome wide association study to identify predictors for severe skin toxicity in colorectal cancer patients treated with cetuximab. *PLoS ONE* **2018**, *13*, e0208080. [CrossRef]
83. Sibertin-Blanc, C.; Mancini, J.; Fabre, A.; Lagarde, A.; Del Grande, J.; Levy, N.; Seitz, J.F.; Olschwang, S.; Dahan, L. Vascular Endothelial Growth Factor A c.*237C>T polymorphism is associated with bevacizumab efficacy and related hypertension in metastatic colorectal cancer. *Dig. Liver Dis.* **2015**, *47*, 331–337. [CrossRef]
84. Schneider, B.P.; Wang, M.; Radovich, M.; Sledge, G.W.; Badve, S.; Thor, A.; Flockhart, D.A.; Hancock, B.; Davidson, N.; Gralow, J.; et al. Association of vascular endothelial growth factor and vascular endothelial growth factor receptor-2 genetic polymorphisms with outcome in a trial of paclitaxel compared with paclitaxel plus bevacizumab in advanced breast cancer: ECOG 2100. *J. Clin. Oncol.* **2008**, *26*, 4672–4678. [CrossRef]
85. Morita, S.; Uehara, K.; Nakayama, G.; Shibata, T.; Oguri, T.; Inada-Inoue, M.; Shimokata, T.; Sugishita, M.; Mitsuma, A.; Ando, Y. Association between bevacizumab-related hypertension and vascular endothelial growth factor (VEGF) gene polymorphisms in Japanese patients with metastatic colorectal cancer. *Cancer Chemother. Pharm.* **2013**, *71*, 405–411. [CrossRef]
86. Ma, T.; Zhu, Z.G.; Ji, Y.B.; Zhang, Y.; Yu, Y.Y.; Liu, B.Y.; Yin, H.R.; Lin, Y.Z. Correlation of thymidylate synthase, thymidine phosphorylase and dihydropyrimidine dehydrogenase with sensitivity of gastrointestinal cancer cells to 5-fluorouracil and 5-fluoro-2′-deoxyuridine. *World J. Gastroenterol.* **2004**, *10*, 172–176. [CrossRef]
87. Lecomte, T.; Ferraz, J.M.; Zinzindohoué, F.; Loriot, M.A.; Tregouet, D.A.; Landi, B.; Berger, A.; Cugnenc, P.H.; Jian, R.; Beaune, P.; et al. Thymidylate synthase gene polymorphism predicts toxicity in colorectal cancer patients receiving 5-fluorouracil-based chemotherapy. *Clin. Cancer Res.* **2004**, *10*, 5880–5888. [CrossRef]
88. Popat, S.; Matakidou, A.; Houlston, R.S. Thymidylate-synthase expression and prognosis in colorectal cancer: A systematic review and meta-analysis. *J. Clin. Oncol.* **2004**, *22*, 529–536. [CrossRef]
89. Danenberg, P.V.; Leichman, L.; Lenz, H.J.; Leichman, C.G.; Danenberg, K.D. Thymidylate Synthase Gene and Protein Expression Correlate and Are Associated with Response to 5-Fluorouracil in Human Colorectal and Gastric Tumors. *Cancer Res.* **1995**, *55*, 1407–1412.
90. De Mattia, E.; Toffoli, G. C677T and A1298C MTHFR polymorphisms, a challenge for antifolate and fluoropyrimidine-based therapy personalisation. *Eur. J. Cancer* **2009**, *45*, 1333–1351. [CrossRef] [PubMed]
91. Chua, W.; Goldstein, D.; Lee, C.K.; Dhillon, H.; Michael, M.; Mitchell, P.; Clarke, S.J.; Iacopetta, B. Molecular markers of response and toxicity to FOLFOX chemotherapy in metastatic colorectal cancer. *Br. J. Cancer* **2009**, *101*, 998–1004. [CrossRef] [PubMed]
92. Salvador-Martín, S.; García-González, X.; García, M.I.; Blanco, C.; García-Alfonso, P.; Robles, L.; Grávalos, C.; Pachón, V.; Longo, F.; Martínez, V.; et al. Clinical utility of ABCB1 genotyping for preventing toxicity in treatment with irinotecan. *Pharm. Res.* **2018**, *136*, 133–139. [CrossRef] [PubMed]

93. Glimelius, B.; Garmo, H.; Berglund, Å.; Fredriksson, L.A.; Berglund, M.; Kohnke, H.; Byström, P.; Sørbye, H.; Wadelius, M. Prediction of irinotecan and 5-fluorouracil toxicity and response in patients with advanced colorectal cancer. *Pharm. J.* **2011**, *11*, 61–71. [CrossRef] [PubMed]
94. Martin, M.; Martinez, N.; Ramos, M.; Calvo, L.; Lluch, A.; Zamora, P.; Munoz, M.; Carrasco, E.; Caballero, R.; Garcia-Saenz, J.A.; et al. Standard Versus Continuous Administration of Capecitabine in Metastatic Breast Cancer (GEICAM/2009-05): A Randomized, Noninferiority Phase II Trial With a Pharmacogenetic Analysis. *Oncologist* **2015**, *20*, 111–112. [CrossRef]
95. Rhodes, K.E.; Zhang, W.; Yang, D.; Press, O.A.; Gordon, M.; Vallböhmer, D.; Schultheis, A.M.; Lurje, G.; Ladner, R.D.; Fazzone, W.; et al. ABCB1, SLCO1B1 and UGT1A1 gene polymorphisms are associated with toxicity in metastatic colorectal cancer patients treated with first-line irinotecan. *Drug Metab. Lett.* **2007**, *1*, 23–30. [CrossRef]
96. Mathijssen, R.H.J.; Marsh, S.; Karlsson, M.O.; Xie, R.; Baker, S.D.; Verweij, J.; Sparreboom, A.; McLeod, H.L. Irinotecan pathway genotype analysis to predict pharmacokinetics. *Clin. Cancer Res.* **2003**, *9*, 3246–3253.
97. Teft, W.A.; Welch, S.; Lenehan, J.; Parfitt, J.; Choi, Y.H.; Winquist, E.; Kim, R.B. OATP1B1 and tumour OATP1B3 modulate exposure, toxicity, and survival after irinotecan-based chemotherapy. *Br. J. Cancer* **2015**, *112*, 857–865. [CrossRef]
98. Cecchin, E.; Corona, G.; Masier, S.; Biason, P.; Cattarossi, G.; Frustaci, S.; Buonadonna, A.; Colussi, A.; Toffoli, G. Carboxylesterase isoform 2 mRNA expression in peripheral blood mononuclear cells is a predictive marker of the irinotecan to SN38 activation step in colorectal cancer patients. *Clin. Cancer Res.* **2005**, *11*, 6901–6907. [CrossRef]
99. Cortejoso, L.; López-Fernández, L.A. Pharmacogenetic markers of toxicity for chemotherapy in colorectal cancer patients. *Pharmacogenomics* **2012**, *13*, 1173–1191. [CrossRef] [PubMed]
100. Qian, Y.-Y.; Liu, X.-Y.; Wu, Q.; Song, X.; Chen, X.-F.; Liu, Y.-Q.; Pei, D.; Shen, L.-Z.; Shu, Y.-Q. The ERCC1 C118T Polymorphism Predicts Clinical Outcomes of Colorectal Cancer Patients Receiving Oxaliplatin-Based Chemotherapy: A Meta-analysis Based on 22 Studies. *Asian Pac. J. Cancer Prev.* **2014**, *15*, 8383–8390. [CrossRef] [PubMed]
101. Etienne-Grimaldi, M.C.; Milano, G.; Maindrault-Gœbel, F.; Chibaudel, B.; Formento, J.L.; Francoual, M.; Lledo, G.; André, T.; Mabro, M.; Mineur, L.; et al. Methylenetetrahydrofolate reductase (MTHFR) gene polymorphisms and FOLFOX response in colorectal cancer patients. *Br. J. Clin. Pharm.* **2010**, *69*, 58–66. [CrossRef] [PubMed]
102. McLeod, H.L.; Sargent, D.J.; Marsh, S.; Green, E.M.; King, C.R.; Fuchs, C.S.; Ramanathan, R.K.; Williamson, S.K.; Findlay, B.P.; Thibodeau, S.N.; et al. Pharmacogenetic predictors of adverse events and response to chemotherapy in metastatic colorectal cancer: Results from North American Gastrointestinal Intergroup Trial N9741. *J. Clin. Oncol.* **2010**, *28*, 3227–3233. [CrossRef]
103. Gamelin, L.; Capitain, O.; Morel, A.; Dumont, A.; Traore, S.; Anne, L.B.; Gilles, S.; Boisdron-Celle, M.; Gamelin, E. Predictive factors of oxaliplatin neurotoxicity: The involvement of the oxalate outcome pathway. *Clin. Cancer Res.* **2007**, *13*, 6359–6368. [CrossRef]
104. Massacesi, C.; Terrazzino, S.; Marcucci, F.; Rocchi, M.B.; Lippe, P.; Bisonni, R.; Lombardo, M.; Pilone, A.; Mattioli, R.; Leon, A. Uridine diphosphate glucuronosyl transferase 1A1 promoter polymorphism predicts the risk of gastrointestinal toxicity and fatigue induced by irinotecan-based chemotherapy. *Cancer* **2006**, *106*, 1007–1016. [CrossRef]
105. Iyer, L.; Das, S.; Janisch, L.; Wen, M.; Ramírez, J.; Karrison, T.; Fleming, G.F.; Vokes, E.E.; Schilsky, R.L.; Ratain, M.J. UGT1A1*28 polymorphism as a determinant of irinotecan disposition and toxicity. *Pharm. J.* **2002**, *2*, 43–47. [CrossRef]
106. Liu, C.Y.; Chen, P.M.; Chiou, T.J.; Liu, J.H.; Lin, J.K.; Lin, T.C.; Chen, W.S.; Jiang, J.K.; Wang, H.S.; Wang, W.S. UGT1A1*28 polymorphism predicts irinotecan-induced severe toxicities without affecting treatment outcome and survival in patients with metastatic colorectal carcinoma. *Cancer* **2008**, *112*, 1932–1940. [CrossRef]
107. Liu, X.; Cheng, D.; Kuang, Q.; Liu, G.; Xu, W. Association of UGT1A1*28 polymorphisms with irinotecan-induced toxicities in colorectal cancer: A meta-analysis in Caucasians. *Pharm. J.* **2014**, *14*, 120–129. [CrossRef]

108. Toffoli, G.; Cecchin, E.; Gasparini, G.; D'Andrea, M.; Azzarello, G.; Basso, U.; Mini, E.; Pessa, S.; De Mattia, E.; Re, G.L.; et al. Genotype-driven phase I study of irinotecan administered in combination with fluorouracil/leucovorin in patients with metastatic colorectal cancer. *J. Clin. Oncol.* **2010**, *28*, 866–871. [CrossRef]
109. Marcuello, E.; Páez, D.; Paré, L.; Salazar, J.; Sebio, A.; Del Rio, E.; Baiget, M. A genotype-directed phase I-IV dose-finding study of irinotecan in combination with fluorouracil/leucovorin as first-line treatment in advanced colorectal cancer. *Br. J. Cancer* **2011**, *105*, 53–57. [CrossRef] [PubMed]
110. Treenert, A.; Areepium, N.; Tanasanvimon, S. Effects of ABCC2 and SLCO1B1 Polymorphisms on treatment responses in Thai metastatic colorectal cancer patients treated with Irinotecan-based chemotherapy. *Asian Pac. J. Cancer Prev.* **2018**, *19*, 2757–2764. [CrossRef] [PubMed]
111. Innocenti, F.; Kroetz, D.L.; Schuetz, E.; Dolan, M.E.; Ramírez, J.; Relling, M.; Chen, P.; Das, S.; Rosner, G.L.; Ratain, M.J. Comprehensive pharmacogenetic analysis of irinotecan neutropenia and pharmacokinetics. *J. Clin. Oncol.* **2009**, *27*, 2604–2614. [CrossRef] [PubMed]
112. Riera, P.; Salazar, J.; Virgili, A.C.; Tobeña, M.; Sebio, A.; Gallano, P.; Barnadas, A.; Páez, D. Relevance of CYP3A4*20, UGT1A1*37 and UGT1A1*28 variants in irinotecan-induced severe toxicity. *Br. J. Clin. Pharm.* **2018**, *84*, 1389–1392. [CrossRef] [PubMed]
113. Fujikura, K.; Ingelman-Sundberg, M.; Lauschke, V.M. Genetic variation in the human cytochrome P450 supergene family. *Pharm. Genom.* **2015**, *25*, 584–594. [CrossRef] [PubMed]
114. Paulík, A.; Grim, J.; Filip, S. Predictors of irinotecan toxicity and efficacy in treatment of metastatic colorectal cancer. *Acta Medica* **2012**, *55*, 153–159. [CrossRef]
115. Sunakawa, Y.; Yang, D.; Moran, M.; Astrow, S.H.; Tsuji, A.; Stephens, C.; Zhang, W.; Cao, S.; Takahashi, T.; Denda, T.; et al. Combined assessment of EGFR-related molecules to predict outcome of 1st-line cetuximab-containing chemotherapy for metastatic colorectal cancer. *Cancer Biol. Ther.* **2016**, *17*, 751–759. [CrossRef]
116. Pander, J.; Gelderblom, H.; Antonini, N.F.; Tol, J.; van Krieken, J.H.J.M.; van der Straaten, T.; Punt, C.J.A.; Guchelaar, H.J. Correlation of FCGR3A and EGFR germline polymorphisms with the efficacy of cetuximab in KRAS wild-type metastatic colorectal cancer. *Eur. J. Cancer* **2010**, *46*, 1829–1834. [CrossRef]
117. Lambrechts, D.; Moisse, M.; Delmar, P.; Miles, D.W.; Leighl, N.; Escudier, B.; Van Cutsem, E.; Bansal, A.T.; Carmeliet, P.; Scherer, S.J.; et al. Genetic markers of bevacizumab-induced hypertension. *Angiogenesis* **2014**, *17*, 685–694. [CrossRef]
118. Etienne-Grimaldi, M.C.; Formento, P.; Degeorges, A.; Pierga, J.Y.; Delva, R.; Pivot, X.; Dalenc, F.; Espié, M.; Veyret, C.; Formento, J.L.; et al. Prospective analysis of the impact of VEGF-A gene polymorphisms on the pharmacodynamics of bevacizumab-based therapy in metastatic breast cancer patients. *Br. J. Clin. Pharm.* **2011**, *71*, 921–928. [CrossRef]
119. Le, D.T.; Durham, J.N.; Smith, K.N.; Wang, H.; Bartlett, B.R.; Aulakh, L.K.; Lu, S.; Kemberling, H.; Wilt, C.; Luber, B.S.; et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* **2017**, *357*, 409–413. [CrossRef] [PubMed]
120. Food and Drug Administration. FDA Grants Accelerated Approval to Ipilimumab for MSI-H or dMMR Metastatic Colorectal Cancer. Available online: https://www.fda.gov/drugs/resources-information-approved-drugs/fda-grants-accelerated-approval-ipilimumab-msi-h-or-dmmr-metastatic-colorectal-cancer (accessed on 30 October 2020).
121. Food and Drug Administration. FDA Approves Pembrolizumab for First-Line Treatment of MSI-H/dMMR Colorectal Cancer. Available online: https://www.fda.gov/drugs/drug-approvals-and-databases/fda-approves-pembrolizumab-first-line-treatment-msi-hdmmr-colorectal-cancer (accessed on 30 October 2020).
122. Bins, S.; Basak, E.A.; El Bouazzaoui, S.; Koolen, S.L.W.; De Hoop, E.O.; Van Der Leest, C.H.; Van Der Veldt, A.A.M.; Sleijfer, S.; Debets, R.; Van Schaik, R.H.N.; et al. Association between single-nucleotide polymorphisms and adverse events in nivolumab-treated non-small cell lung cancer patients. *Br. J. Cancer* **2018**, *118*, 1296–1301. [CrossRef] [PubMed]
123. Al-Samkari, H.; Snyder, G.D.; Nikiforow, S.; Tolaney, S.M.; A Freedman, R.; Losman, J.-A. Haemophagocytic lymphohistiocytosis complicating pembrolizumab treatment for metastatic breast cancer in a patient with the PRF1A91V gene polymorphism. *J. Med. Genet.* **2018**, *56*, 39–42. [CrossRef] [PubMed]

124. Queirolo, P.; Dozin, B.; Morabito, A.; Banelli, B.; Carosio, R.; Fontana, V.; Ferrucci, P.F.; Martinoli, C.; Cocorocchio, E.; Ascierto, P.A.; et al. CTLA-4 gene variant-1661A>G may predict the onset of endocrine adverse events in metastatic melanoma patients treated with ipilimumab. *Eur. J. Cancer* **2018**, *97*, 59–61. [CrossRef]
125. Oguri, T.; Mitsuma, A.; Inada-Inoue, M.; Morita, S.; Shibata, T.; Shimokata, T.; Sugishita, M.; Nakayama, G.; Uehara, K.; Hasegawa, Y.; et al. Genetic polymorphisms associated with oxaliplatin-induced peripheral neurotoxicity in Japanese patients with colorectal cancer. *Int. J. Clin. Pharm. Ther.* **2013**, *51*, 475–481. [CrossRef]
126. Terrazzino, S.; Argyriou, A.A.; Cargnin, S.; Antonacopoulou, A.G.; Briani, C.; Bruna, J.; Velasco, R.; Alberti, P.; Campagnolo, M.; Lonardi, S.; et al. Genetic determinants of chronic oxaliplatin-induced peripheral neurotoxicity: A genome-wide study replication and meta-analysis. *J. Peripher. Nerv. Syst.* **2015**, *20*, 15–23. [CrossRef]
127. Motsinger-Reif, A.A.; Jorgenson, E.; Relling, M.V.; Kroetz, D.L.; Weinshilboum, R.; Cox, N.J.; Roden, D.M. Genome-wide association studies in pharmacogenomics:successes and lessons. *Pharm. Genom.* **2013**, *23*, 1744–6872. [CrossRef]
128. Singleton, A.B.; Hardy, J.; Traynor, B.J.; Houlden, H. Towards a complete resolution of the genetic architecture of disease. *Trends Genet.* **2010**, *26*, 438–442. [CrossRef]
129. Manolio, T.A.; Collins, F.S.; Cox, N.J.; Goldstein, D.B.; Hindorff, L.A.; Hunter, D.J.; McCarthy, M.I.; Ramos, E.M.; Cardon, L.R.; Chakravarti, A.; et al. Finding the missing heritability of complex diseases. *Nature* **2009**, *461*, 747–753. [CrossRef]
130. Scientific, T.F. Axiom Spain Biobank Array. Available online: http://www.usc.es/cegen/wp-content/uploads/2019/08/COL32017-1217-Axiom-Spain-EN_FLR_FINAL.pdf. (accessed on 7 October 2020).
131. Mizzi, C.; Peters, B.; Mitropoulou, C.; Mitropoulos, K.; Katsila, T.; Agarwal, M.R.; Van Schaik, R.H.N.; Drmanac, R.; Borg, J.; Patrinos, G.P. Personalized pharmacogenomics profiling using whole-genome sequencing. *Pharmacogenomics* **2014**, *15*, 1223–1234. [CrossRef]
132. Apellániz-Ruiz, M.; Lee, M.Y.; Sánchez-Barroso, L.; Gutiérrez-Gutiérrez, G.; Calvo, I.; García-Estévez, L.; Sereno, M.; García-Donás, J.; Castelo, B.; Guerra, E.; et al. Whole-exome sequencing reveals defective CYP3A4 variants predictive of paclitaxel dose-limiting neuropathy. *Proc. Clin. Cancer Res.* **2015**, *21*, 322–328. [CrossRef] [PubMed]
133. Kozyra, M.; Ingelman-Sundberg, M.; Lauschke, V.M. Rare genetic variants in cellular transporters, metabolic enzymes, and nuclear receptors can be important determinants of interindividual differences in drug response. *Genet. Med.* **2017**, *19*, 20–29. [CrossRef] [PubMed]
134. Santos, M.; Niemi, M.; Hiratsuka, M.; Kumondai, M.; Ingelman-Sundberg, M.; Lauschke, V.M.; Rodríguez-Antona, C. Novel copy-number variations in pharmacogenes contribute to interindividual differences in drug pharmacokinetics. *Genet. Med.* **2018**, *20*, 622–629. [CrossRef] [PubMed]
135. Ingelman-Sundberg, M.; Mkrtchian, S.; Zhou, Y.; Lauschke, V.M. Integrating rare genetic variants into pharmacogenetic drug response predictions. *Hum. Genom.* **2018**, *12*, 1–12. [CrossRef] [PubMed]
136. Schaller, L.; Lauschke, V.M. The genetic landscape of the human solute carrier (SLC) transporter superfamily. *Hum. Genet.* **2019**, *138*, 1359–1377. [CrossRef] [PubMed]
137. Van Staveren, M.C.; Jan Guchelaar, H.; Van Kuilenburg, A.B.P.; Gelderblom, H.; Maring, J.G. Evaluation of predictive tests for screening for dihydropyrimidine dehydrogenase deficiency. *Pharm. J.* **2013**, *13*, 389–395. [CrossRef] [PubMed]
138. Van Kuilenburg, A.B.P.; Häusler, P.; Schalhorn, A.; Tanck, M.W.T.; Proost, J.H.; Terborg, C.; Behnke, D.; Schwabe, W.; Jabschinsky, K.; Maring, J.G. Evaluation of 5-fluorouracil pharmacokinetics in cancer patients with a c.19051G>A Mutation in DPYD by means of a bayesian limited sampling strategy. *Clin. Pharm.* **2012**, *51*, 163–174. [CrossRef]
139. Maring, J.G.; Van Kuilenburg, A.B.P.; Haasjes, J.; Piersma, H.; Groen, H.J.M.; Uges, D.R.A.; Van Gennip, A.H.; De Vries, E.G.E. Reduced 5-FU clearance in a patient with low DPD activity due to heterozygosity for a mutant allele of the DPYD gene. *Br. J. Cancer* **2002**, *86*, 1028–1033. [CrossRef]
140. Offer, S.M.; Lee, A.M.; Mattison, L.K.; Fossum, C.; Wegner, N.J.; Diasio, R.B. A DPYD Variant (Y186C) in Individuals of African Ancestry Is Associated With Reduced DPD Enzyme Activity. *Clin. Pharm. Ther.* **2013**, *94*, 158–166. [CrossRef]

141. Morel, A.; Boisdron-Celle, M.; Fey, L.; Soulie, P.; Craipeau, M.C.; Traore, S.; Gamelin, E. Clinical relevance of different dihydropyrimidine dehydrogenase gene single nucleotide polymorphisms on 5-fluorouracil tolerance. *Mol. Cancer Ther.* **2006**, *5*, 2895–2904. [CrossRef]
142. Van Kuilenburg, A.; Hausler, P.; Schalhorn, A.; Tanck, M.; Proost, J.H.; Terborg, C.; Behnke, D.; Schwabe, W.; Jabschinsky, K.; Maring, J.G. Evaluation of 5-FU pharmacokinetics in cancer patients with DPD deficiency using a Bayesian limited sampling strategy. *Ther. Drug Monit.* **2011**, *33*, 478.
143. Zhang, H.; Li, Y.-M.; Zhang, H.; Jin, X. DPYD*5 gene mutation contributes to the reduced DPYD enzyme activity and chemotherapeutic toxicity of 5-FU. *Med. Oncol.* **2007**, *24*, 251–258. [CrossRef] [PubMed]
144. Sáez-Belló, M.; Mangas-Sanjuán, V.; Martínez-Gómez, M.A.; López-Montenegro Soria, M.Á.; Climente-Martí, M.; Merino-Sanjuán, M. Evaluation of ABC gene polymorphisms on the pharmacokinetics and pharmacodynamics of capecitabine in colorectal patients: Implications for dosing recommendations. *Br. J. Clin. Pharm.* **2020**, online. [CrossRef] [PubMed]
145. Toffoli, G.; Cecchin, E.; Corona, G.; Russo, A.; Buonadonna, A.; D'Andrea, M.; Pasetto, L.M.; Pessa, S.; Errante, D.; De Pangher, V.; et al. The role of UGT1A1*28 polymorphism in the pharmacodynamics and pharmacokinetics of irinotecan in patients with metastatic colorectal cancer. *J. Clin. Oncol.* **2006**, *24*, 3061–3068. [CrossRef] [PubMed]
146. Cecchin, E.; Innocenti, F.; D'Andrea, M.; Corona, G.; De Mattia, E.; Biason, P.; Buonadonna, A.; Toffoli, G. Predictive role of the UGT1A1, UGT1A7, and UGT1A9 genetic variants and their haplotypes on the outcome of metastatic colorectal cancer patients treated with fluorouracil, leucovorin, and irinotecan. *J. Clin. Oncol.* **2009**, *27*, 2457–2465. [CrossRef]
147. Labriet, A.; De Mattia, E.; Cecchin, E.; Lévesque, É.; Jonker, D.; Couture, F.; Buonadonna, A.; D'Andrea, M.; Villeneuve, L.; Toffoli, G.; et al. Improved progression-free survival in irinotecan-treated metastatic colorectal cancer patients carrying the HNF1A coding variant p.I27L. *Front. Pharm.* **2017**, *8*, 712. [CrossRef]
148. Fujita, K.I.; Nagashima, F.; Yamamoto, W.; Endo, H.; Sunakawa, Y.; Yamashita, K.; Ishida, H.; Mizuno, K.; Matsunaga, M.; Araki, K.; et al. Association of ATP-binding cassette, sub-family C, number 2 (ABCC2) genotype with pharmacokinetics of irinotecan in Japanese patients with metastatic colorectal cancer treated with irinotecan plus infusional 5-fluorouracil/leucovorin (FOLFIRI). *Biol. Pharm. Bull.* **2008**, *31*, 2137–2142. [CrossRef]
149. Paoluzzi, L.; Singh, A.S.; Price, D.K.; Danesi, R.; Mathijssen, R.H.J.; Verweij, J.; Figg, W.D.; Sparreboom, A. Influence of genetic variants in UGT1A1 and UGT1A9 on the in vivo glucuronidation of SN-38. *J. Clin. Pharm.* **2004**, *44*, 854–860. [CrossRef]
150. Sai, K.; Saeki, M.; Saito, Y.; Ozawa, S.; Katori, N.; Jinno, H.; Hasegawa, R.; Kaniwa, N.; Sawada, J.; Komamura, K.; et al. UGT1A1 haplotypes associated with reduced glucuronidation and increased serum bilirubin in irinotecan-administered Japanese patients with cancer*1. *Clin. Pharm. Ther.* **2004**, *75*, 501–515. [CrossRef]
151. De Jong, F.A.; Scott-Horton, T.J.; Kroetz, D.L.; McLeod, H.L.; Friberg, L.E.; Mathijssen, R.H.; Verweij, J.; Marsh, S.; Sparreboom, A. Irinotecan-induced diarrhea: Functional significance of the polymorphic ABCC2 transporter protein. *Clin. Pharm. Ther.* **2007**, *81*, 42–49. [CrossRef]
152. de Jong, F.A.; Kehrer, D.F.S.; Mathijssen, R.H.J.; Creemers, G.; de Bruijn, P.; van Schaik, R.H.N.; Planting, A.S.T.; van der Gaast, A.; Eskens, F.A.L.M.; Janssen, J.T.P.; et al. Prophylaxis of Irinotecan-Induced Diarrhea with Neomycin and Potential Role for UGT1A1*28 Genotype Screening: A Double-Blind, Randomized, Placebo-Controlled Study. *Oncologist* **2006**, *11*, 944–954. [CrossRef] [PubMed]
153. Minami, H.; Sai, K.; Saeki, M.; Saito, Y.; Ozawa, S.; Suzuki, K.; Kaniwa, N.; Sawada, J.I.; Hamaguchi, T.; Yamamoto, N.; et al. Irinotecan pharmacokinetics/pharmacodynamics and UGT1A genetic polymorphisms in Japanese: Roles of UGT1A1*6 and *28. *Pharm. Genom.* **2007**, *17*, 497–504. [CrossRef] [PubMed]
154. Jada, S.R.; Lim, R.; Wong, C.I.; Shu, X.; Lee, S.C.; Zhou, Q.; Goh, B.C.; Chowbay, B. Role of UGT1A1*6, UGT1A1*28 and ABCG2 c.421C>A polymorphisms in irinotecan-induced neutropenia in Asian cancer patients. *Cancer Sci.* **2007**, *98*, 1461–1467. [CrossRef] [PubMed]
155. Sai, K.; Kaniwa, N.; Itoda, M.; Saito, Y.; Hasegawa, R.; Komamura, K.; Ueno, K.; Kamakura, S.; Kitakaze, M.; Shirao, K.; et al. Haplotype analysis of ABCB1/MDR1 blocks in a Japanese population reveals genotype-dependent renal clearance of irinotecan. *Pharmacogenetics* **2003**, *13*, 741–757. [CrossRef]

156. Henricks, L.M.; Lunenburg, C.A.T.C.; de Man, F.M.; Meulendijks, D.; Frederix, G.W.J.; Kienhuis, E.; Creemers, G.-J.; Baars, A.; Dezentjé, V.O.; Imholz, A.L.T.; et al. DPYD genotype-guided dose individualisation of fluoropyrimidine therapy in patients with cancer: A prospective safety analysis. *Lancet Oncol.* **2018**, *19*, 1459–1467. [CrossRef]
157. Kleibl, Z.; Fidlerova, J.; Kleiblova, P.; Kormunda, S.; Bilek, M.; Bouskova, K.; Sevcik, J.; Novotny, J. Influence of dihydropyrimidine dehydrogenase gene (DPYD) coding sequence variants on the development of fluoropyrimidine-related toxicity in patients with high-grade toxicity and patients with excellent tolerance of fluoropyrimidine-based chemotherapy. *Neoplasma* **2009**, *56*, 303–316. [CrossRef]
158. Deenen, M.J.; Meulendijks, D.; Cats, A.; Sechterberger, M.K.; Severens, J.L.; Boot, H.; Smits, P.H.; Rosing, H.; Mandigers, C.M.; Soesan, M.; et al. Upfront Genotyping of DPYD*2A to Individualize Fluoropyrimidine Therapy: A Safety and Cost Analysis. *J. Clin. Oncol.* **2016**, *34*, 227–234. [CrossRef]
159. Cortejoso, L.; García-González, X.; García, M.I.; García-Alfonso, P.; Sanjurjo, M.; López-Fernández, L.A. Cost-effectiveness of screening for DPYD polymorphisms to prevent neutropenia in cancer patients treated with fluoropyrimidines. *Pharmacogenomics* **2016**, *17*, 979–984. [CrossRef]
160. Murphy, C.; Byrne, S.; Ahmed, G.; Kenny, A.; Gallagher, J.; Harvey, H.; O'Farrell, E.; Bird, B. Cost Implications of Reactive Versus Prospective Testing for Dihydropyrimidine Dehydrogenase Deficiency in Patients with Colorectal Cancer: A Single-Institution Experience. *Dose-Response* **2018**, *16*, 16. [CrossRef]
161. Henricks, L.M.; Lunenburg, C.A.; De Man, F.M.; Meulendijks, D.; Frederix, G.W.; Kienhuis, E.; Creemers, G.-J.; Baars, A.; Dezentjé, V.O.; Imholz, A.L.; et al. A cost analysis of upfront DPYD genotype–guided dose individualisation in fluoropyrimidine-based anticancer therapy. *Eur. J. Cancer* **2019**, *107*, 60–67. [CrossRef]
162. Gold, H.T.; Hall, M.J.; Blinder, V.; Schackman, B.R. Cost effectiveness of pharmacogenetic testing for uridine diphosphate glucuronosyltransferase 1A1 before irinotecan administration for metastatic colorectal cancer. *Cancer* **2009**, *115*, 3858–3867. [CrossRef] [PubMed]
163. Obradovic, M.; Mrhar, A.; Kos, M. Cost-effectiveness of UGT1A1 genotyping in second-line, high-dose, once every 3 weeks irinotecan monotherapy treatment of colorectal cancer. *Pharmacogenomics* **2008**, *9*, 539–549. [CrossRef] [PubMed]
164. Butzke, B.; Oduncu, F.S.; Severin, F.; Pfeufer, A.; Heinemann, V.; Giesen-Jung, C.; Stollenwerk, B.; Rogowski, W.H. The cost-effectiveness of UGT1A1 genotyping before colorectal cancer treatment with irinotecan from the perspective of the German statutory health insurance. *Acta Oncol.* **2016**, *55*, 318–328. [CrossRef] [PubMed]
165. Roncato, R.; Cecchin, E.; Montico, M.; De Mattia, E.; Giodini, L.; Buonadonna, A.; Solfrini, V.; Innocenti, F.; Toffoli, G. Cost Evaluation of Irinotecan-Related Toxicities Associated With the UGT1A1*28 Patient Genotype. *Clin. Pharm. Ther.* **2017**, *102*, 123–130. [CrossRef]
166. Agencia Española de Medicamentosy Produtos Sanitarios Fluorouracilo, Capecitabina, Tegafury Flucitosina en Pacientes con Déficit de Dihidropirimidina Deshidrogenasa. Available online: https://www.aemps.gob.es/informa/notasinformativas/medicamentosusohumano-3/seguridad-1/2020-seguridad-1/fluorouracilo-capecitabina-tegafur-y-flucitosina-en-pacientes-con-deficit-de-dihidropirimidina-deshidrogenasa/ (accessed on 24 September 2020).
167. Amstutz, U.; Henricks, L.M.; Offer, S.M.; Barbarino, J.; Schellens, J.H.; Swen, J.J.; Klein, T.E.; McLeod, H.L.; Caudle, K.E.; Diasio, R.B.; et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Dihydropyrimidine Dehydrogenase Genotype and Fluoropyrimidine Dosing: 2017 Update. *Clin. Pharm. Ther.* **2018**, *103*, 210–216. [CrossRef]
168. The Pharmacogenetics Working Group. UGT1A1: Irinotecan. Available online: https://www.g-standaard.nl/risicoanalyse/B0001694.PDF (accessed on 24 September 2020).
169. European Medicines Agency. Recommendations on DPD Testing Prior to Treatment with Fluorouracil, Capecitabine, Tegafur and Flucytosine. Available online: https://www.ema.europa.eu/en/news/ema-recommendations-dpd-testing-prior-treatment-fluorouracil-capecitabine-tegafur-flucytosine (accessed on 24 September 2020).
170. Food and Drug Administration. Table of Pharmacogenomic Biomarkers in Drug Labeling. Available online: https://www.fda.gov/media/124784/download (accessed on 24 September 2020).
171. European Medicines Agency. Erbitux (cetuximab). *Am. J. Neuroradiol.* **2010**, *31*, 626–627.

172. Cancer Therapy Evaluation Program. Common Terminology Criteria for Adverse Events (CTCAE).v.5.0. Available online: https://ctep.cancer.gov/protocoldevelopment/electronic_applications/docs/ctcae_v5_quick_reference_5x7.pdf (accessed on 23 July 2020).
173. Kerns, S.L.; Fachal, L.; Dorling, L.; Barnett, G.C.; Baran, A.; Peterson, D.R.; Hollenberg, M.; Hao, K.; Narzo, A.D.; Ahsen, M.E.; et al. Radiogenomics Consortium Genome-Wide Association Study Meta-Analysis of Late Toxicity After Prostate Cancer Radiotherapy. *J. Natl. Cancer Inst.* **2020**, *112*, 179–190. [CrossRef]
174. Barnett, G.C.; Thompson, D.; Fachal, L.; Kerns, S.; Talbot, C.; Elliott, R.M.; Dorling, L.; Coles, C.E.; Dearnaley, D.P.; Rosenstein, B.S.; et al. A genome wide association study (GWAS) providing evidence of an association between common genetic variants and late radiotherapy toxicity. *Radiother. Oncol.* **2014**, *111*, 178–185. [CrossRef]
175. Fachal, L.; Gómez-Caamaño, A.; Barnett, G.C.; Peleteiro, P.; Carballo, A.M.; Calvo-Crespo, P.; Kerns, S.L.; Sánchez-García, M.; Lobato-Busto, R.; Dorling, L.; et al. A three-stage genome-wide association study identifies a susceptibility locus for late radiotherapy toxicity at 2q24.1. *Nat. Genet.* **2014**, *46*, 891–894. [CrossRef]
176. Kerns, S.L.; Stock, R.G.; Stone, N.N.; Blacksburg, S.R.; Rath, L.; Vega, A.; Fachal, L.; Gómez-Caamaño, A.; De Ruysscher, D.; Lammering, G.; et al. Genome-wide association study identifies a region on chromosome 11q14.3 associated with late rectal bleeding following radiation therapy for prostate cancer. *Radiother. Oncol.* **2013**, *107*, 372–376. [CrossRef]
177. Kerns, S.L.; Dorling, L.; Fachal, L.; Bentzen, S.; Pharoah, P.D.; Barnes, D.R.; Gómez-Caamaño, A.; Carballo, A.M.; Dearnaley, D.P.; Peleteiro, P.; et al. Meta-analysis of Genome Wide Association Studies Identifies Genetic Markers of Late Toxicity Following Radiotherapy for Prostate Cancer. *EBioMedicine* **2016**, *10*, 150–163. [CrossRef] [PubMed]
178. Richardson, M.; Kirkham, J.; Dwan, K.; Sloan, D.J.; Davies, G.R.; Jorgensen, A.L. STrengthening the Reporting of Pharmacogenetic Studies: Development of the STROPS guideline. *PLoS Med.* **2020**, *17*, e1003344. [CrossRef]
179. Hegde, M.; Santani, A.; Mao, R.; Ferreira-Gonzalez, A.; Weck, K.E.; Voelkerding, K.V. Development and validation of clinical whole-exome and whole-genome sequencing for detection of germline variants in inherited disease. *Arch. Pathol. Lab. Med.* **2017**, *141*, 798–805. [CrossRef] [PubMed]
180. Zhou, Y.; Fujikura, K.; Mkrtchian, S.; Lauschke, V.M. Computational methods for the pharmacogenetic interpretation of next generation sequencing data. *Front. Pharm.* **2018**, *9*, 1–17. [CrossRef]

Colorectal cancer (CRC) is often linked to adverse drug reactions (ADRs), which can be severe or even fatal. To address this issue, pharmacogenetic studies have emerged to help personalise treatments based on molecular information. However, these have been limited by a lack of standardized data, the complexity of ADR phenotyping and the need for novel strategies and functional assays. Therefore, the aim of this project was to find and validate new CRC chemotoxicity biomarkers, using state-of-the-art approaches. We have conducted different omic analyses to identify low to high-penetrance variants and performed in vitro assays on a variant of interest. Our work successfully identified several candidate toxicity variants/genes that might be influencing the development of ADRs in CRC patients.