Genetic predictors of severe chemotherapy-induced neutropenia: a candidate-gene and GWAS study protocol

(**Version 1.0**)

Contents

[Abstract 3](#_Toc207194864)

[1. Introduction 3](#_Toc207194865)

[2. Methods 5](#_Toc207194866)

[2.1 Study Design and Population 5](#_Toc207194867)

[2.2 Genetic Data 6](#_Toc207194868)

[2.3 Candidate-Gene Study 6](#_Toc207194869)

[2.3.1 Candidate variants selection 6](#_Toc207194870)

[2.3.2 Statistical Analysis 9](#_Toc207194871)

[2.4 Genome-Wide Association Study 13](#_Toc207194872)

[2.4.1 Statistical Analysis 13](#_Toc207194873)

[Ethics approval and consent to participate 21](#_Toc207194874)

[References 21](#_Toc207194875)

# Abstract

**Background**: Chemotherapy-induced neutropenia (CIN) is a toxic effect of chemotherapeutic agents, often leading to dose reduction or treatment delays, which can compromise drug efficacy and increase cancer progression or death. CIN can also cause febrile neutropenia (FN), which is linked to mortality and financial costs in the healthcare system. The exact mechanism underlying CIN is not fully understood, but it is thought to be explained by pharmacokinetics (PK) and pharmacodynamics (PD) pathways. Therefore, a candidate-gene association and a genome-wide association study (GWAS) will be conducted to identify PK or PD genetic variants contributing to the development of severe neutropenia risk to help improve cancer treatment precision.

**Methods**: The study comprises a main cohort, which will include all the study participants, and the sub-cohorts, which will be selected based on the first regimen's drug group they received. The outcomes of interest for the candidate and GWAS are time-to-first severe neutropenia and first cycle neutropenia, respectively. Cox and logistic regressions will estimate hazard ratios (HRs) and odds ratios (ORs), respectively, with 95% confidence intervals and *p*-values. Significance will be determined using Bonferroni correction for the candidate-gene study and a genome-wide threshold of 5 × 10⁻⁸ for GWAS. To ensure we have enough power to detect the outcome for each approach, we applied sample size calculations in advance. The main cohort showed sufficient power to detect HRs of 2.0–2.3 and ORs of 2.5–3 across specific minor allele frequency (MAF) ranges. Subgroup power calculations also indicated adequate power for HRs of 2.0–2.3. For GWAS, only the platinum-treated and the antimetabolite-treated subgroups had sufficient power to detect an OR of 3 under the same conditions.

**Keywords**: Neutropenia; SACT; chemotherapy, pharmacogenomics; pharmacogenetics; outcomes; cancer.

# Introduction

Neutropenia is characterised by a decrease in the absolute neutrophil count, and it is classified by the “Common Terminology Criteria for Adverse Events (CTCAE) version 5” into Grade 1 (<2000–1500/mm³), Grade 2 (<1500–1000/mm³), Grade 3 (<1000–500/mm³), and Grade 4 (<500/mm³) (1). Chemotherapy-induced neutropenia (CIN) is a toxic effect that occurs when chemotherapy is administered for cancer management in adjuvant, neoadjuvant, or metastatic settings, and it is considered one of the most common toxicities associated with cytotoxic chemotherapies, often resulting in dose reductions or treatment delays. Reducing the chemotherapy dose intensity, which is the quantity of drugs given per unit of time, can compromise the drug's efficacy and increase disease progression or even death. In a study that included patients with early breast cancer receiving the cyclophosphamide, methotrexate, and fluorouracil (CMF) regimen and followed up for twenty years, they found those who received ≥ 85% of the planned dose had higher overall survival (OS) and relapse-free survival (RFS) compared to patients who received a lesser proportion (2).

Another serious complication of neutropenia is febrile neutropenia (FN), which is linked to mortality and an increase in the financial burden on the healthcare system (3). Although the granulocyte-colony stimulating factors (GCSF) can reduce the risk of FN when given as prophylaxis, there is no definitive method to determine high-risk patients. The current guidelines recommend primary GCSF prophylaxis with regimens that have ≥ 20% FN risk or in other situations like immunocompromised patients who are infected with the human immunodeficiency virus (4). However, not all patients on high-risk chemotherapy will develop FN, and some patients on chemotherapies deemed lower risk may develop it. In a real-world study, they found that 18.2% of FN cases didn’t exhibit apparent risk factors, while around 50% of cases occurred while patients were on low-risk regimens (5). This uncertainty can lead to unnecessary GCSF use while simultaneously limiting access for those in high therapeutic need. Taken together, these findings highlight the critical impact of neutropenia and its consequences on cancer care, leading to improper cancer management, either directly or indirectly, and depriving the healthcare system of resources.

The exact mechanism underlying CIN has not been fully elucidated, but it is thought to be explained by two pathways, pharmacokinetics (PK) or pharmacodynamics (PD). The PK pathway means neutropenia is a direct action of the concentrations of cytotoxic drugs in the blood; therefore, it can be predicted by pharmacokinetic determinants. Among these determinants are pharmacogenes that influence the absorption, distribution, metabolism, or excretion of the drug, which eventually result in increased toxicity or decreased efficacy. There are many approved pharmacogenes for pre-emptive testing that fall within this context. For example, the FDA requires testing for thiopurine S-methyl transferase (TPMT) and nucleotide diphosphatase (NUDT15) genotypes before azathioprine administration to avoid the risk of severe myelosuppression and even death when giving the standard dose (6). In oncology, there are also approved pharmacokinetic germline variants to be tested before drug administration. Such as testing for *UGT1A1\*28* polymorphism carriers before giving belinostat to avoid the risk of toxicity, as these patients have impaired UGT1A1 enzyme activity (7). Many other PK variants are recommended for testing, like *DPYD* (5-fluorouracil, capecitabine), *UGT1A1* (irinotecan), and *TPMT* (mercaptopurine) (8).

The PD pathway that could also explain neutropenia toxicity mainly concerns how the drug affects the body. Therefore, genetic variants that affect processes, like the mechanisms of action or drug targets, can lead to adverse drug reactions. One example is the antiviral drug abacavir, which can cause severe hypersensitivity reactions in patients carrying the HLA-B\*5701 allele; hence, this drug is contraindicated in these patients, and a prior test is required before administration (9). In the cancer context, genetic variants that increase bone marrow susceptibility to the cytotoxic drug can lead to neutropenia, such as genetic mutations in the DNA repair genes. The haematopoietic stem cells in the bone marrow are rapidly dividing cells, a characteristic shared with malignant cancer cells; thus, mutations in such genes increase the risk of developing myelosuppression. As found in a study, included breast cancer female patients received neoadjuvant chemotherapy; the carriers of *BRCA1* mutations had a higher risk for developing acute haematological toxicities, including grade 4 neutropenia, compared to *BRCA2* and non-carriers (10).

Identifying the factors that contribute to severe neutropenia, including predictive PK or PD genetic variants, can help improve oncology treatment precision through identifying high-risk patient groups before starting treatment and planning proper mitigation strategies in advance to prevent the previously mentioned complications. In our previous study, we identified many clinical factors that are related to severe neutropenia, such as the chemotherapy regimen, cancer type, presence of kidney disease, body surface area, and number of treatment cycles (unpublished); here we seek to uncover the role of germline genetic variation in severe neutropenia risk. Therefore, the aims of this study are to validate the already reported genetic predictors of severe neutropenia in the literature, validate the important genetic PK and PD variants of the common cancer drug groups that were reported in “The Pharmacogenetics and Pharmacogenomics Knowledge Base (PharmGKB)”, and discover new variants through a genome-wide association study (GWAS). Since allele frequencies can vary between ethnicities, and since we conducted a study of the non-genetic predictors of neutropenia in the White cancer population, our focus will be this population.

# Methods

## 2.1 Study Design and Population

This is a retrospective study of patients diagnosed with invasive cancers, including lung, breast, gastrointestinal (GIT), female genital organs, male genital organs, urinary tract and oral cancers, who received systemic anti-cancer therapy (SACT). The study will be composed of two parts, a cohort candidate-gene association study and a case-control GWAS. The inclusion criteria are age ≥18 years and availability of genetic data or blood samples available for genotyping. Since this study complements our previous work that investigated clinical determinants associated with neutropenia risk in SACT-treated cancer patients, the additional criteria for inclusion are the availability of haematology data to assess neutropenia and the initiation of SACT within one year of cancer diagnosis. Patients with multiple primary cancers within six months and patients who did not survive the first six weeks of treatment or until the administration of the next cycle, whichever was earlier, will be excluded.

## 2.2 Genetic Data

Genetic data is available for patients who were recruited in one of three cohorts: the “Genetics of Diabetes Audit and Research in Tayside Scotland GoDARTS” (11), the “Genetics of the Scottish Health Research Register GoSHARE” (12), and “Generation Scotland: Scottish Family Health Study GS:SFHS” (13). Additionally, we genotyped patients to whom the inclusion criteria apply, and blood samples were available in GoSHARE to increase the power of the study. The data is managed, anonymised and available through a Trusted Research Environment (TRE) by the Health Informatics Centre (HIC) at the University of Dundee, UK.

## 2.3 Candidate-Gene Study

### 2.3.1 Candidate variants selection

The genetic variants to be investigated are those relevant to the key drugs in the SACT regimens received by the study population. They are determined by a literature search of already published statistically significant associations and through knowledge of relevant genes that were involved in key SACTs' PK and PD pathways that were reported in the PharmGKB database (8).

#### **Literature Search**

A literature search was conducted in PubMed using a combination of keyword and Medical Subject Headings (MeSH) terms, with results limited to articles published in English. For the search terms details, refer to page 1 in the supplementary materials.

The inclusion criteria for the papers were an adult study population of European ancestry treated with SACT who had an invasive cancer, investigated the association/prediction of genetic determinants on neutropenia or leukopenia, found statistically significant associations, and included ≥ 30 individuals. Papers reporting findings on non-European patients, experimental cancer drugs, uncommon cancers, or paediatric patients, as well as reviews, editorials, letters, and meeting abstracts, were excluded.

The search was conducted in March 2025 and returned 284 papers. After title and abstract screening, 46 papers were considered for full-text review. Evaluation of the papers based on the inclusion criteria resulted in the selection of 38 studies, including one additional study identified through the reference list of a previously reviewed paper. Refer to figure 1 for more details. For each study, the treatment group was recorded as reported. If the group was not specified, the regimen was assigned to all relevant drug categories based on its individual components. Refer to Tables S1 and S2 in the supplementary materials for more details.

**Identification**

**Included**

Records screened

(n = 284)

Records excluded\*\*

(n = 238)

Reports sought for retrieval

(n = 46)

Reports not retrieved

(n = 0)

Reports assessed for eligibility

(n = 46)

**Screening**

Records identified from PubMed:

Databases (n = 284)

Reports excluded:

No significant association found (n = 3)

Study cohort < 30 (n = 2)

Mixed genotypes investigated (n = 1)

Variant and phenotype effect (n = 1)

Gene-gene interaction (n = 1)

Uncommon cancer (n = 1)

Studies included in review

(n = 38)

\*One study identified form the reference list

Figure 1: Study screening flow chart

#### **PharmGKB Search**

The candidate genetic variants are also searched in PharmGKB for the most common cancer groups, which included platinum agents, anthracyclines, alkylating agents, taxanes, vinca alkaloids, anti-metabolites and podophyllotoxin derivatives. The search was performed for each drug group individually, using the following search terms: “platinum compounds”, “anthracyclines”, “alkylating agents”, “taxanes”, “vinca alkaloids”, “antimetabolites” and “podophyllotoxin derivatives”.

The relevant variants were searched using two steps. Firstly, we searched for variants that were listed under clinical annotations with a strength of evidence level of 2 or higher regarding toxicity or efficacy phenotypes. Secondly, we searched for statistically significant variants that were associated with neutropenia or leukopenia in the population of European ancestry and that were listed under variant annotations. The data were downloaded in (.tsv) format with one filter applied before downloading (SIGNIFICANCE = yes). The downloaded data was further filtered in R using RStudio to be restricted to adults, populations with European ancestry and neutropenia/leukopenia toxicity using the below R functions, as demonstrated by the R code example below:

ptn\_pathway = PharmGKB\_file %>%

filter(Pediatric == "FALSE") %>%

filter(grepl("European|Caucasian", `Biogeographical Groups`, ignore.case = T))

filter(grepl("neutro|leuko", `Association`, ignore.case = T))

The search was conducted in June 2025 for each main drug group individually for relevant variants under the “Clinical Annotations” and “Variant Annotations” sections. In platinum and anthracycline groups, we found a strong level of clinically annotated variants; however, the one reported for anthracyclines was reported for cardiotoxicity in paediatric patients, so we decided not to include it. In addition, for the vinca alkaloid and podophyllotoxin derivative drug groups, no variants were found. Thirty-five SNVs were reported for anthracyclines, antimetabolites and alkylating agents in a one-pathway GWAS study for the fluorouracil, epirubicin, and cyclophosphamide (FEC) regimen in breast cancer patients (14). They reported that the alleles of these variants achieved a level of significance with a p-value < 1.0E-04, and they did not correct for multiple testing, as their goal was to identify variants for functional validation. PharmGKB-annotated variants that were found non-significant or irrelevant to the study were excluded. Refer to Tables S3 and S4 in the supplementary materials for more details.

We decided to exclude variants with a minor allele frequency (MAF) of less than 1% in European populations due to their rarity. In addition, variants reported for uncommon drugs that fall under other cytotoxic chemotherapy categories, like irinotecan and sunitinib, will be excluded from analysis. Although the *UGT1A1\*28* variant was frequently reported and linked to severe neutropenia in irinotecan-treated patients, the total number of genotyped patients who received this drug in our cohort is only 20, which would not allow a meaningful analysis. Furthermore, variants that resulted from copy number variation or gene deletion will be excluded due to the technical limitation of our genotyping array data. Refer to Table 1 for the final variants list for the candidate-gene study.

### 2.3.2 Statistical Analysis

The study uses a cohort design with a time-to-event outcome of neutropenia, classified as grade 3 or 4. The worst grade of neutropenia is selected for each treatment-time window that started from the cycle treatment day until administration of the next cycle or up to six weeks, whichever is earlier. With a Cox proportional hazards model, we will analyse the time from baseline to the first episode of severe neutropenia occurring during the initial treatment regimen. Assuming that the initial regimen is fixed at baseline, follow-up will be censored at the end of this regimen. Single nucleotide variant (SNV) genotypes will be classified as wild-type, heterozygous, or homozygous for the minor allele under a genotypic mode. All variables will be modelled as baseline variables. Hazard ratios (HR) with their corresponding 95% confidence intervals (95% CI) and the associated p-value will be reported.

The primary analysis will include all patients for each variant at a time, then a secondary analysis will be conducted by subgrouping patients according to the first drug group they received, with only SNVs for that drug group retained. Further subgrouping based on a cancer type, subtype, or chemotherapy regimen is considered if a sufficient number is available. Patients who were initiated on the other cytotoxic drug group will not be investigated as an individual group, as they are heterogeneous drugs, and further dividing this group for analysis won’t be feasible.

#### **Correction for multiple testing**

To control for a type I error of 5%, a Bonferroni adjustment will be performed for the primary analysis and each sub-analysis, and the significance threshold will be as follows:

*P* value ≤ , where is the number of tests.

#### **Power calculation**

The solid cancers cohort includes 1,073 patients with available genetic data who received a solid cancer diagnosis. There are 306 events, which results in an event proportion of 0.29. For the sample size calculation, we used the *ssizeEpiCont.default* function from the *powerSurvEpi* R package (15) with the following inputs: power = 0.80; HR = 2.0 or 2.3; variance derived from the MAFs; and α = 0.05. The formula also requires the squared multiple correlation between the covariate of interest, here, the SNV, and the remaining covariates; we set this to zero, assuming the genotype distribution is independent of other variables. Among the included studies, only one reported an HR (16), so we used that estimate to specify the detectable HR in our calculation.

The same sample size calculation was performed for all subgroups except the vinca alkaloid–treated cohort, as no SNV was identified for validation and its small sample size (n=38) was likely to yield low statistical power.

For the whole cohort, the estimated sample size needed to detect an HR of 2.0 ranges from 113 to 968 for MAFs between 0.50 and 0.03, while for an HR of 2.3, the required sample size is from 79 to 996 for MAFs between 0.50 and 0.02. For the subgroups, the study showed sufficient statistical power to detect an HR of 2 or 2.3 for reasonable MAF ranges. For each group, the number of tests corresponds to the number of unique SNVs to be evaluated. When a variant is listed as a haplotype, each SNV will be included individually. The solid cancer cohort will be tested for all variants, whereas the subgroups will be tested only for SNVs reported for regimens containing that group. In addition, the SNV list will be restricted to those with a minimum MAF that the study has sufficient power to detect for each group. Refer to Table 2 and Figures 2–7 for more details.

Table 2: Estimated sample size ranges needed for candidate-gene studies in the main study cohort and each subgroup

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Cancer cohort** | **N** | **Event proportion** | **HR** | **MAFs range** | **Required sample size range** | **Number of tests** | **Bonferroni-adjusted *p*-value** |
| **Solid cancer cohort** | 1073 | 0.29 | 2 | 0.50 – 0.03 | 113 –967 | 73 | 0.0007 |
| 2.3 | 0.50 – 0.02 | 79 – 996 |
| **Platinum cohort** | 525 | 0.31 | 2 | 0.50 – 0.06 | 106 – 468 | 46 | 0.001 |
| 2.3 | 0.50 – 0.04 | 73 – 476 |
| **Antimetabolites cohort** | 712 | 0.26 | 2 | 0.50 – 0.05 | 126 – 662 | 51 | 0.001 |
| 2.3 | 0.50 – 0.04 | 88 – 567 |
| **Anthracyclines cohort** | 302 | 0.39 | 2 | 0.50 – 0.08 | 84 – 285 | 7 | 0.007 |
| 2.3 | 0.50 – 0.06 | 59 – 258 |
| **Alkylating agent cohort** | 276 | 0.41 | 2 | 0.50 – 0.08 | 80 – 271 | 6 | 0.008 |
| 2.3 | 0.50 – 0.06 | 56 – 245 |
| **Taxanes cohort** | 130 | 0.27 | 2 | 0.50 – 0.37 | 122 – 130 | 4 | 0.01 |
| 2.3 | 0.50 – 0.21 | 84 – 127 |

***N:*** *Cohort sample size.* ***HR:*** *Hazard ratio.* ***MAF:*** *Mainor allele frequency.*

***Note:*** *The MAF range is presented from high to low to reflect the relationship between MAF and required sample size, as higher MAFs correspond to smaller sample sizes. The Bonferroni-adjusted p-values were rounded to the minimum number of decimal places needed to obtain the closest rounded value*

**A graph of a graph showing a number of different sizes of power

AI-generated content may be incorrect.**

Figure 2: Sample size for the solid cancer cohort

A graph of a graph showing a number of different sizes of power

AI-generated content may be incorrect.

Figure 3: Sample size for the platinum-treated cohort

A graph of a graph showing a number of different sizes of power

AI-generated content may be incorrect.

Figure 4: Sample size for the antimetabolites-treated cohort

A graph of a graph showing a number of different sizes of power

AI-generated content may be incorrect.

Figure 5: Sample size for the anthracyclines-treated cohort

**A graph of a graph showing a number of different sizes of power

AI-generated content may be incorrect.**

Figure 6: Sample size for the alkylating agent-treated cohort

**A graph of a line graph

AI-generated content may be incorrect.**

Figure7: Sample size for the taxanes-treated cohort

## 2.4 Genome-Wide Association Study

### 2.4.1 Statistical Analysis

In this approach, we will conduct an unmatched case-control design, assigning patients who experienced severe neutropenia during the first cycle as cases and those who had mild or no neutropenia as controls. The patient's main drug group will be chosen based on the first treatment group, with the worst neutropenia grade during this drug group's first treatment cycles as the outcome. Multiple logistic regression models for each SNV, adjusted for age, sex, and the principal component (PC), will be fitted to investigate their association with the outcomes. Under an additive genetic model, SNV genotypes will be coded as 0, 1, or 2 based on the number of minor alleles. Odd ratios (ORs), their 95% CI, and the associated *p*-value will be reported for each test.

In the primary analysis, all genetic variants will be investigated individually for all patients combined. In the secondary analysis, we will subgroup patients according to their initial drug groups and explore genetic variants relevant to that group. For a result to be significant, we will use the conventional genome-wide significance threshold of 5 × 10⁻⁸ (17). When the sample size allows, further subgrouping is considered based on cancer type, subtype, or regimen. Similarly to the candidate gene approach, we decided not to perform a GWAS for other cytotoxic drug groups as an individual group. Identified hits will be further investigated using the candidate-gene approach, as previously described.

#### **Power calculation**

The genotyped solid cancers cohort comprised 1,073 participants, of whom 193 are cases. Sample size calculations were performed using the *genpwr* package in R (18). Under an additive genetic mode, the analysis parameters were selected as the following: a statistical power of 0.80, a genome-wide significance threshold of α = 5 × 10⁻⁸, and a case rate of 0.18, as defined by the *genpwr* manual (i.e., number of cases divided by the total sample size). To estimate a detectable OR, we used the quartiles Q1 and Q3 of the ORs reported by studies identified in our previous literature survey and in PharmGKB, which applied a similar additive genotype model. We found the plausible ORs range from 0.63 to 3. Accordingly, we set the range to 0.5 to 3, in increments of 0.5. For ORs lower than 2, the required sample sizes exceeded those available for the study and were therefore excluded from further consideration. Refer to Table 3 and Figures 8 – 10 for more details.

Table 3: Estimated sample size ranges needed for GWAS in the main study cohort and each subgroup

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Cancer cohort** | **N** | **Case rate** | **OR** | **MAFs range** | **Required sample size range** |
| **Solid cancer cohort** | 1073 | 0.18 | 2.5 | 0.50 – 0.15 | 683 – 1053 |
| 3 | 0.50 – 0.09 | 489 – 1055 |
| **Platinum cohort** | 525 | 0.19 | 2.5 | - | - |
| 3 | 0.50 – 0.26 | 471 – 522 |
| **Antimetabolites cohort** | 712 | 0.17 | 2.5 | - | - |
| 3 | 0.50 – 0.17 | 510 – 696 |
| **Anthracyclines cohort** | 302 | 0.25 | 2.5 | - | - |
| 3 | - | - |
| **Alkylating agent cohort** | 276 | 0.27 | 2.5 | - | - |
| 3 | - | - |
| **Taxanes cohort** | 130 | 0.18 | 2.5 | - | - |
| 3 | - | - |
| **Vinca Alkaloids cohort** | 38 | 0.24 | 2.5 | - | - |
| 3 | - | - |

***N:*** *Cohort sample size.* ***OR:*** *Odds ratio.* ***MAF:*** *Mainor allele frequency.*

***Note:*** *The MAF range is presented from high to low to reflect the relationship between MAF and required sample size, as higher MAFs correspond to smaller sample sizes.*

For an OR of 2.5 to be detected in the main cohort, the analysis indicated adequate power to detect MAFs between 0.15 and 0.5, with the estimated total sample size ranging from 683 to 1,053 participants. While for an OR of 3, the analysis showed we are powered enough to detect MAFs between 0.09 and 0.5, with required sample sizes from 489 to 1055 participants. In addition, for the platinum-treated patients, the study has sufficient power to detect an OR of 3 for MAFs ranging from 0.26 to 0.5, with the required sample size estimated to be between 471 and 522. Furthermore, the required sample size was estimated to be between 510 and 696 for an OR of 3, with MAFs ranging from 0.17 to 0.50. Nonetheless, the analysis for the other subgroups lacks sufficient power to identify the other predefined ORs under the same parameters.

A graph of a line graph

AI-generated content may be incorrect.

Figure 8: Sample size for the solid cancer cohort

A graph with a red line

AI-generated content may be incorrect.

Figure 9: Sample size for the platinum-treated cohort

A graph showing a red line

AI-generated content may be incorrect.

Figure 10: Sample size for the antimetabolites-treated cohort

*Table 1: Final variants list for the candidate-gene study and GWAS*

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Gene** | **Variant** | **Allele frequency** | **Drug group** | **Source** | **Note** | **Patients No.** | **Genetic Model** | **Effect Size** | **Interpretation** | **Outcome** | **SACTs** |
|  | *XRCC1* | *Arg399Gln (rs25487)* | C=0.642131 | **Platinum** | PharmGKB | *rs25487* was listed twice under clinical and variant annotations | *104* | Dominant: Arg/Arg vs. Arg/Gln + Gln/Gln | OR = 3.02 (95% CI: 1.33–6.88; P=0.009*).* | Homozygous **wild** type has a 3-fold increase of severe neutropenia risk. Variant allele is protective. | G3/4 neutropenia | Cisplatin-based chemotherapy (cisplatin, cyclophosphamide) *(19)* |
|  | *ERCC2* | *Asp312Asn (rs1799793)* | T=0.35200 | - | Overdominant: Asp/Asn vs. Asp/Asp + Asn/Asn | OR = 2.33 (95% CI 1.05–5.33; P=0.045) | Heterozygous Asp/Asn genotype has a 2.33-fold increase of severe neutropenia risk. | G3/4 neutropenia |
|  | *TP53* | *Arg72Pro (rs1042522)* | C=0.73664 | Recessive: Pro/Pro vs. Arg/Arg or Arg/Pro | OR = 8.57 (95% CI 1.05–69.8; P=0.023) | Homozygous Pro/Pro genotype has an 8.57-fold increase of severe neutropenia risk. | G3/4 neutropenia |
|  | *MSH6* | *T>G*  *(rs3136228)* | G=0.364580 | *rs3136228* was also listed under antimetabolites (fluorouracil) | 154 | **Recessive: GG vs. TT + GG** | OR = 3.23 (95% CI: 1.38–7.57; P= 0.0071, q = **0.0937**) | Homozygous GG genotype has a 3.23-fold increase of severe neutropenia risk. | G3/4 neutropenia | FOLFOX4 (20) |
|  | *ERCC5* | *G>C*  *(rs17655)* | C=0.22121 | Thia variant was not annotated in PharmGKB but was reported in the same study | 104 | Codominant:  GG: reference  CG vs. GG  CC vs. GG | CG: OR = 0.174(0.070–0.431; P = 0.0001)  CC: OR = 0.051(0.005–0.490; P = 0.0043) | Homozygous GG genotype increases the risk | G3/4 neutropenia | Cisplatin-based chemotherapy (cisplatin, cyclophosphamide) (21) |
|  | *RAD52* | *T>C*  *(rs11226)* | A=0.45611 | - | Codominant:  CC: reference  CT vs. CC  TT vs. CC | CT: OR = 0.122 (0.045–0.328, P = 0.00002)  TT: OR: 0.140 (0.039–0.507; P = 0.0028) | Homozygous CC genotype increases the risk |
|  | *MUTYH* | *G>A*  *(rs3219484)* | T=0.070132 | Codominant:  GG: reference  AG vs. GG | OR = 0.013 (0.001–0.220; P = 0.00000004) | Homozygous GG genotype increases the risk |
|  | *LIG3* | *T>C*  *(rs1052536)* | T=0.462444 | Codominant:  TT: reference  CT vs. TT  CC vs. TT | CT: OR = 5.333 (1.845–15.416; P = 0.0019)  CC: OR = 73.5 (8.126–664.84; P = 0.0000002) | Homozygous TT genotype decreases the risk |
|  | *CDA* | [*Lys27Gln (A79C) (rs2072671)*](https://www.ncbi.nlm.nih.gov/snp/rs2072671#frequency_tab) | C=0.351089 | Literature search; Tibaldi-2008 (22) | *rs2072671* was reported in two studies.  Also, the variant was listed under anti-metabolites (gemcitabine) | 65 | Codominant: AA vs. AC vs. CC | No reported OR  48.1% vs. 10.3% vs. 22.2%; P = 0.006 |  | G3/4 neutropenia | All studies reported on platinum-based chemotherapies |
| Literature search; Joerger-2012 (23) | 137 | Dominant: CC + AC vs. AA | No reported OR  28% vs 11%; *P* = 0.02 |  | Severe leukocytopenia |
|  | *ERCC1* | [*Asn118Asn (rs11615)*](https://www.ncbi.nlm.nih.gov/snp/rs11615) | G=0.379441 | Literature search; Cortejoso-2013 (24) | - | 106 | Recessive: CC vs. CT + TT | OR = 0.205 (95 % CI: 0.061–0.690; P = 0.010) (24) | Homozygous CC genotype has a 79.5% decrease of severe neutropenia risk. | G3/4 neutropenia | All studies reported on platinum-based chemotherapies |
| Literature search; Grenda-2020 (25) | - | 113 | Codominant: AA vs. AG | No reported OR  (P = 0.0133; χ 2 = 6.12) | AA genotype had higher risk of neutropenia | Any grade neutropenia |
|  | *ERCC1* | ERCC1-118T/8092C *(rs11615/ rs3212986)* | G=0.376972/ A=0.246764 | Goekkurt-2009 (26) | Also reported under anti-metabolites (fluorouracil). | 121 | Haplotype | OR = 2.68; 95% CI: 1.03- 6.92; P = 0.042) | The haplotype ERCC1-118T/8092C has a 2.68-fold increase of severe neutropenia risk. | G3/4 neutropenia | Fluorouracil, leucovorin, and oxaliplatin vs. fluorouracil, leucovorin, and cisplatin |
|  | *RECQ1* | [*A159C (rs13035)*](https://www.ncbi.nlm.nih.gov/snp/rs13035) | G=0.43661 | Literature search; Joerger-2012 (23) | *rs13035* was also listed for anti-metabolites group | 146 | Dominant: CC + AC vs. AA | No reported OR  26% vs 10%; *P* = 0.03 | Carriers of the C allele has more increased risk compared to the AA genotype | Severe leukocytopenia | Platinum-gemcitabine chemotherapy |
|  | *HLA-G* | *C>G*  ([*rs1610696)*](https://www.ncbi.nlm.nih.gov/snp/rs1610696) | G=0.31090 | Literature search; Garziera-2017 (27) | Also, reported under anti-metabolites (fluorouracil). | 144 | **Recessive: CC vs. CG + GG** | OR = 3.76 (95 CI%: 1.29–10.96; P = 0.015) | Homozygous **wild** **CC** genotype has a 3.76 -fold increase of severe neutropenia risk. | G3/4 neutropenia | FOLFOX4 (fluorouracil, leucovorin, oxaliplatin) |
|  | *MTHFR* | *A2756G (rs1805087)* | G=0.190499 | Literature search; Goekkurt-2009 (26). | The two variants were also reported under anti-metabolites (fluorouracil). *rs1695* was also listed for alkylating agents (cyclophosphamide) in Yao-2010 study (28) | 121 | Codominant: AA: Reference vs. GG vs. AG | AG: OR= 0.07 (95% CI: 0.01 - 0.53; P = 0.011)  GG: OR = 0.08 (95% CI: 0.01 - 0.44; P = 0.004) | Homozygous **wild** AA genotype showed increased risk of severe neutropenia. | G3/4 neutropenia | Fluorouracil, leucovorin, and oxaliplatin (FOLFOX) vs. fluorouracil, leucovorin, and cisplatin (FLP) |
|  | *GSTP1* | [*Ile105Val (rs1695)*](https://www.ncbi.nlm.nih.gov/snp/rs1695#frequency_tab) | G=0.326272 | Recessive:  Ile/Ile vs. Ile/ Val + Val/Val | OR = 4.45 (95% CI: 1.25-15.78; P =0 .02) | Homozygous **wild** Ile/Ile genotype has a 4.45-fold increase of severe neutropenia risk. | G3/4 neutropenia |
|  | *OR4D6* | *G>C*  *(*[*rs1453542*](https://www.ncbi.nlm.nih.gov/snp/rs1453542)*)* | C=0.28128 | Literature search; Gréen-2016 (29) | Also, listed under anti-metabolites (gemcitabine) | 32 | **Recessive:** | OR = 5.2 (95% CI: 1.8–18; P = 0.0008)  *P* = 0.00087 nominal and *P* = 0.042 after Hochberg adjustment | The minor allele is the risk allele | ≥ G3 neutropenia | Carboplatin and gemcitabine |
|  | *XPD*  *(ERCC2)* | *Asp312Asn /Lys751Gln*  *rs1799793/ rs13181* | T=0.34727 / G=0.373251 | Literature search; Booton-2006 (30) | - | 108 | Haplotype | No reported OR  haplotype: \*A/\*A, 49.9%; \*B/\*B, 33.3%; *P* = 0.05 | XPD \*A allele had higher risk | G4 neutropenia | Platinum-based chemotherapy (docetaxel and carboplatin), MIC or MVP |
|  | *GTF2E1* | *rs447978* | A=0.65974 | Literature search; Lamba-2014 (31) | Also, listed under taxanes (docetaxel) | 90 |  | OR = 0.444 (-1.531 to - 0.093; P = 0.027) |  |  | Carboplatin or cisplatin + etoposide, gemcitabine or paclitaxel |
|  | ERCC4 | *rs744154* | C=0.26886 |  | OR = 2.176 (0.035- 1.52; P = 0.04) |  |  |
|  | TMEM63A | *rs10158985* | A=0.222785 |  | OR = 2.557 (0.087 to 1.79; P = 0.031) |  |  |
|  | *HFM1* | *G>C (rs17131429)* | C=0.28656 | Literature search; Svedberg-2020 (32) | Also, listed under antimetabolites (gemcitabine) | 212  (144 validation cohort) | **Additive (allele-dosage) model** | No reported OR  Beta = -0.629 |  | The nadir parameter is the lowest blood count recorded on days 8, 14, and 21. | Gemcitabine and carboplatin |
|  | *LRRC8C* | *A>G (rs12032393)* | G=0.141325 | No reported OR  Beta = 0.5911 |  |
|  | *ABCB11* | *T>C (rs497692)* | C=0.535827 | No reported OR  Beta = -0.42 |  |
|  | *ZNF512* | *C>A (rs11127071)* | A=0.499159, G=0.000000 | No reported OR  Beta = -0.4292 |  |
|  | *CGGBP1* | *C>T (rs7432838)* | T=0.857151 | No reported OR  Beta = -0.6313 |  |
|  | *FRAS1* | *A>G (rs34840208)* | G=0.25542 | No reported OR  Beta = -0.536 |  |
|  | *PCDHB17* | *C>T (rs246697)* | T=0.18669 | No reported OR  Beta = 0.555 |  |
|  | *GABRR2* | *T>C (rs282117)* | C=0.592977 | No reported OR  Beta = -0.389 |  |
|  | *DNAH11* | *C>T (rs12536928)* | G=0.000002, T=0.506236 | No reported OR  Beta = 0.448 |  |
|  | *IMPDH1* | *C>T (rs2288551)* | T=0.02941 | No reported OR  Beta = 1.632 |  |
|  | *WEE2* | *A>G (rs6967301)* | G=0.51282 | No reported OR  Beta = 0.384 |  |
|  | *PWWP2B* | *G>A (rs11817589)* | A=0.07263 | No reported OR  Beta = -0.756 |  |
|  | *ITGB1* | *G>T (rs2230396)* | C=0.000000, T=0.895080 | No reported OR  Beta = 0.708 |  |
|  | *MRE11A* | *C>T (rs535801)* | T=0.307306 | No reported OR  Beta = 0.466 |  |
|  | *MUCL1* | *G>T (rs1048371)* | T=0.462188 | No reported OR  Beta = -0.391 |  |
|  | *SSH1* | *G>GT (rs34849596)* | TT=0.29286 | No reported OR  Beta = 0.476 |  |
|  | *PSME1* | *G>A (rs11548692)* | A=0.029562 | No reported OR  Beta = 0.888 |  |
|  | *BBS2* | *A>G (rs191207351)* | G=0.01258 | No reported OR  Beta = -1.833 |  |
|  | *IL34* | *C>G (rs3813905)* | G=0.33373 | No reported OR  Beta = 0.504 |  |
|  | *MBP* | *C>G (rs138484926)* | A=0.00000, G=0.01399 | No reported OR  Beta = 1.833 |  |
|  | *MAST1* | *A>G (rs11085822)* | G=0.145659, T=0.000000 | No reported OR  Beta = 0.396 |  |
|  | *ATP8B3* | *C>T (rs2385387)* | A=0.000000, T=0.433866 | No reported OR  Beta = 0.473 |  |
|  | *PABPC1L* | *A>G (rs11780)* | G=0.371990 | No reported OR  Beta = -0.380 |  |
|  | *SPATA2* | *G>A (rs495337)* | A=0.415799, C=0.000000 | No reported OR  Beta = -0.4074 |  |
|  | *MCM5* | *C>G (rs2307340)* | G=0.07471, T=0.00000 | No reported OR  Beta = 0.688 |  |
|  | *MCM5* | *G>A (rs133427)* | A=0.07322 | No reported OR  Beta = 0.700 |  |
|  | *MYO18B* | *G>A (rs133885)* | A=0.447273, C=0.000000 | No reported OR  Beta = -0.420 |  |
|  | *PIK3IP1* | *A>G (rs4820044)* | G=0.93649 | No reported OR  Beta = 0.830 |  |
|  | CYP39A1 | rs7761731 | G=0.00000, T=0.25828 | Literature search; Melchardt-2015 (33) | Listed also under taxanes and antimetabolites | 78 | Recessive: AA vs. AT + TT | No reported OR  62.5 % vs 32.5 %; p = 0.01 | Homozygous **wild** AA genotype has more risk than AT + TT genotypes | G3/4 leukopenia | Docetaxel, cisplatin and 5-fluorouracil |
|  | *AQP1* | *G> C (rs1049305)* | C=0.36931 | Literature search; Senk-2019 (34) |  | 231 | **Additive & dominant genetic models** | Additive model (CC): OR = 3.03 (1.10–8.38; P = 0.033)  Dominant model (CC): OR = 2.09 (1.00–4.35; P = 0.049)  (Neutropenia results was not significant) |  | Leukopenia grade ≥ 2 | Cisplatin and gemcitabine or pemetrexed |
|  | *PERP* | *G>A (rs78428806)* | A=0.04808 | **Anthracyclines** | PharmGKB | Listed also for alkylating agents (cyclophosphamide) and antimetabolites (fluorouracil) | 3252 | **Additive genetic model of the minor allele** | OR = 0.63 (P =0.0000979) |  | G3/4 neutropenia/ leukopenia or febrile neutropenia | FEC (14) |
|  | *CBR3* | *G>T (rs74743371)* | T=0.01301 | OR = 0.49 (P = 0.0000868) |  |
|  | *CYP2C8* | *G>A (rs117458836)* | A=0.01393 | OR = 2.53 (P= 0.0000885) |  |
|  | *TP53AIP1* | *C>T (rs118088833)* | T=0.01316 | OR = 0.26 (P= 0.0000899) |  |
|  | *GNL3* | *C>T (rs112242273)* | T=0.03345 | OR = 0.52 (P= 0.0000958) |  |
|  | *PERP* | *G>T (rs117101815)* | T=0.02783 | OR = 0.63 (P= 0.0000717) |  |
|  | *PERP* | *G>T (rs9402944)* | T=0.05261 | OR = 0.63 (P= 0.000072) |  |
|  | *CBR3* | *C>T (rs112783657)* | T=0.02814 | OR = 0.49 P= 0.0000865 |  |
|  | *PIK3R2* | *C>T (rs56022120)* | T=0.011439 | OR = 3.12 (P= 0.0000229) |  |
|  | *PIK3R2* | *C>T (rs58695150)* | T=0.01364 | OR = 3.07 (P= 0.0000315) |  |
|  | *HMMR* | *C>T (rs299293)* | T=0.12878 | OR = 1.36 (P= 0.0000319) |  |
|  | *PERP* | *T>C (rs9389568)* | C=0.04934 | OR = 0.59 (P= 0.0000071) |  |
|  | *PIK3R2* | *G>A (rs8110364)* | A=0.01533 | OR = 3.09 (P= 0.0000223) |  |
|  | *PIK3R2* | *C>T (rs148013902)* | T=0.01519 | OR = 3.01(P= 0.0000546) |  |
|  | *PIK3R2* | *C>T (rs55633228)* | T=0.01220 | OR = 3 (P= 0.0000559) |  |
|  | *INSR* | *C>A (rs41412545)* | A=0.03334 | OR = 1.7 (P= 0.0000636) |  |
|  | *PPP2R5D* | *T>C (rs3805945)* | C=0.04254 | OR = 1.68 (P= 0.0000681) |  |
|  | *PIK3R2* | *C>T (rs79430272)* | T=0.011613 | OR = 2.94 (P= 0.000048) |  |
|  | *PIK3R2* | *G>A (rs118129530)* | A=0.01617 | OR = 3 (P= 0.0000482) |  |
|  | *CCNK* | *A>G (rs77769901)* | G=0.01218 | OR = 0.31(P= 0.0000499) |  |
|  | *PIK3R2* | *C>T (rs145623321)* | T=0.01156 | OR = 3.01 (P= 0.0000535) |  |
|  | *HMMR* | *G>A (rs299313)* | A=0.08407 | OR = 1.35 (P= 0.0000372) |  |
|  | *IRS1* | *G>A (rs115457081)* | A=0.013147 | OR = 5.04 (P = 0.0000389) |  |
|  | *PIK3R2* | *C>T (rs117341846)* | T=0.01568 | OR = 3.03 (P= 0.0000393) |  |
|  | *INSR* | *C>T (rs142244113)* | T=0.03388 | OR = 1.72 (P= 0.0000459) |  |
|  | *PIK3R2* | *G>A (rs138602176)* | A=0.01596 | OR = 2.97 (P= 0.0000342) |  |
|  | *PIK3R2* | *G>A (rs148235907)* | A=0.01589 | OR = 3.06 (P= 0.0000335) |  |
|  | *HMMR* | *T>C (rs299314)* | C=0.13400 | OR = 1.35 (P= 0.0000372) |  |
|  | *NLRC5* | *G>T (rs4784750)* | T=0.15302 | Literature search; Fasching-2022 (35) | GWAS  Also, reported under alkylating agents (cyclophosphamide) and antimetabolites (fluorouracil) | 3276 | **Additive genetic model of the minor allele** | OR = 1.38 (1.23–1.54, P =1.56E-8) |  | G3/4 neutropenia or leukopenia | FEC |
|  | *TNFSF13B* | *C>G (rs16972207)* | G=0.11770 | OR = 1.54 (1.32–1.79; P = 3.42E-8) |  |
|  | *SOD2* | *T>C (in the paper)*  *Val16Ala (rs4880)* | A=0.499972  G>A | Literature search; Yao-2010 (36) | Also, reported under alkylating agents (cyclophosphamide) and antimetabolites (fluorouracil).  In the same study another regimen was reported for alkylating agents and anti-metabolites (CMF) | 458 | Codominant:  TT (reference) vs. TC vs. CC | TC: OR = 0.81 (0.48–1.36; P = 0.42)  CC: OR = 0.52 (0.29–0.92; P= 0.03) | Homozygous **minor** CC genotype has more risk than TT genotypes | G3/4 neutropenia | CAF ± tamoxifen |
|  | *UGT2B7* | *C161T (rs7668258)* | C=0.478377 | Literature search; Sawyer-2016 (37) | Also, reported under alkylating agents (cyclophosphamide) and antimetabolites (fluorouracil) | 132 | Codominant:  CC (reference) vs. CT vs. TT | No reported OR  CT: 76% vs. 50.0%; P = .032  TT: 76% vs. 48.7%; P = .038 | Homozygous **wild** CC genotype has more risk than CT or TT genotypes | G3/4 leukopenia | FEC |
|  | *ABCC1/MRP1* | *G>T (rs4148350)* | T=0.058163 | Literature search; Vulsteke-2013 (38) | Also, reported under alkylating agents (cyclophosphamide) and antimetabolites (fluorouracil) | 1012 | Codominant:  GG (reference) vs. GT vs. TT | No reported OR  85.5% vs. 12.8% vs. 1.4%; P= 0.002 FDR 0.046 | Carriers of the minor T allele had more risk. | Prolonged G3/4 neutropenia or deep neutropenia (<100/µl) | FEC |
|  | CYP2C8-HapC | rs1113129 | A=0.00000, C=0.11746 | **Taxanes** | Literature search; Gréen-2011 (39) | - | 33 |  | No reported OR |  |  | Paclitaxel and carboplatin |
|  | CYP3A5 / CYP2C8-HapC | CYP3A5\*3  rs776746 /rs1113129 | C=0.930430 / A=0.00000, C=0.11746 |  | No reported OR |  |  |
|  | *MTHFR* | *677C>T*  *rs1801133* | A=0.343070 | **Anti-metabolites** | PharmGKB |  | 122 | **CC vs. CT and TT vs. CC** | OR = 4.86 (1.14-20.63; P =0.025) | TT genotypes increase the risk | G2-4 leukopenia | Methotrexate (40) |
|  | *ABCC5* | *rs2292997* | A=0.114556 | **Anti-metabolites** | PharmGKB |  | 167 | Haplotype | OR=5.93 (P=0.0002) | The co-occurrence of *ABCG1* rs225440T and *ABCC5* rs2292997A predicted the risk of severe neutropenia (OR=5.93; *P*=0.0002), | G3/4 neutropenia | Fluorouracil, irinotecan, leucovorin (FOLFIRI-based regimen) (41) |
|  | *ABCG1* | *rs225440* | T=0.413504 | **Anti-metabolites** | PharmGKB |  |
|  | *DPYD* | *c.2194G>A*  [*\*6 (rs1801160)*](https://www.ncbi.nlm.nih.gov/snp/rs1801160) | T=0.049261 | **Anti-metabolites** | PharmGKB |  | 132 | Dominant:  GA vs. GG  (No AA genotype) | No reported OR  P = 0.002 | The A allele increase the risk | G3/4 neutropenia | FOLFIRI, FOLFIRINOX, FOLFOX and others (42) |
|  | *TGFB1* | [*rs1800469*](https://www.ncbi.nlm.nih.gov/snp/rs1800469) | G=0.677299 | **Anti-metabolites** | Literature search; Korver-2023 (43) | - | 155 | Nominal categorical predictors of the number of the wild type alleles | 1 WT: -0.48 (-2.1 to 1.1, P = 0.555).  2 WT: 3.21 (1.2 to 5.2, P = 0.001) | Two copies of the wild type allele increase the risk | Neutropenia | Fluoropyrimidine-based chemotherapy (most common 5-FU+ (ECF, EOF, FEC, FOLFOX) regimens) |
|  | *BDNF* | [*rs6265*](https://www.ncbi.nlm.nih.gov/snp/rs6265#frequency_tab) | T=0.193746 | **Anti-metabolites** | 1 WT: 12.18 (-3513 to 3538, P = 0.995).  2 WT: 15.2 (-3510 to 3541, P = 0.993) |
|  | *DPYD* | *c.2194G>A*  [*\*6 (rs1801160)*](https://www.ncbi.nlm.nih.gov/snp/rs1801160) | T=0.051192 | **Anti-metabolites** | Literature search; Ruzzo-2017 (16) | - | 508 | **Codominant:**  GG (reference) vs. GA  GG (reference) vs. AA  **Dominant:**  GA + AA vs. GG (reference) | **Codominant:**  GA: HR = 2.30 (1.53–3.46; P <0.0001)  AA: HR = 2.00 (0.49–8.26; P = 0.3364)  **Dominant:**  HR = 2.28 (1.53–3.40; P<0.0001) | Carriers of the A minor allele had more risk. | G3/4 neutropenia  (time to neutropenia) | Fluoropyrimidine-based chemotherapy (FOLFOX-4 or XELOX) |
|  | *SOD2* | *T>C*  [*Val16Ala (rs4880)*](https://www.ncbi.nlm.nih.gov/snp/rs4880) | G=0.497902 | **Anti-metabolites** | Literature search; Yao-2010 (36) | Also, reported under alkylating agents (cyclophosphamide) and antimetabolites (fluorouracil).  In the same study another regimen was reported for anthracyclines, alkylating agents and anti-metabolites (CAF) | 458 | **Codominant:**  **TT (reference)**  **TC**  **CC** | TC: OR = 0.81 (0.48–1.36, P = 0.42)  CC: OR = 0.52 (0.29–0.92;P = 0.03) | TT wild type genotype increases the risk | G3/4 neutropenia | CMF ± tamoxifen |

**Variants** 51–78 were also reported for alkylating agents and antimetabolites in the variant annotation of the fluorouracil, epirubicin, and cyclophosphamide (FEC) regimen. **Variants** that were listed together were investigated as a haplotype. **Allele frequencies** were derived for the Allele Frequency Aggregator (ALFA) data in dpSNP of the European population (<https://www.ncbi.nlm.nih.gov/snp/>). For PharmGKB-listed variants, data on patient numbers, genetic models, effect sizes, outcomes, and SACT were obtained from the original publications. Patients number and treatment information for literature-reported variants were combined as appropriate. For all variants, if patients constituted a subgroup of the total study population, only that subgroup’s number is reported. **OR**: odds ratio. **IFL:** fluorouracil and irinotecan. **FOLFOX**: fluorouracil and oxaliplatin. **IROX**: irinotecan and oxaliplatin. **FOLFOX4**: fluorouracil, leucovorin, oxaliplatin. **MIC:** mitomycin, ifosfamide, and cisplatin. **MVP:** mitomycin, vinblastine, and cisplatin. **CMF**: cyclophosphamide, methotrexate, and fluorouracil. **CAF:** cyclophosphamide, Adriamycin (doxorubicin), and fluorouracil. **XELOX:** oxaliplatin and capecitabine.

## Ethics approval and consent to participate

The Tayside Medical Science Centre (TASC) approved the study protocol on April 26, 2022 (IRAS reference: 315039, REC Ref: 18/ES/0126). The Tayside Medical Ethics Committee approved the use of data linkage for electronic health record data (IRAS project ID: 31150, REC reference: 22/ES/0034). The Tayside Medical Ethics Committee granted approval for the GoSHARE, GoDARTS, and GS: SFHS studies, and all participants provided informed consent. The electronic health records are completely anonymous and made available to researchers through trusted research environments (TRE) under strict information governance protocols managed by the Health Informatics Centre (HIC).

## References

1. U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute. Common Terminology Criteria for Adverse Events (CTCAE) [Internet]. Version 5.0. Bethesda (MD): National Cancer Institute; 2017 Nov 27 [cited 2025 May 27]. Available from: <https://ctep.cancer.gov/protocoldevelopment/electronic_applications/docs/ctcae_v5_quick_reference_5x7.pdf>.

2. Bonadonna G, Valagussa P, Moliterni A, Zambetti M, Brambilla C. Adjuvant Cyclophosphamide, Methotrexate, and Fluorouracil in Node-Positive Breast Cancer — The Results of 20 Years of Follow-up. New England Journal of Medicine. 1995;332(14):901-6.

3. Kuderer NM, Dale DC, Crawford J, Cosler LE, Lyman GH. Mortality, morbidity, and cost associated with febrile neutropenia in adult cancer patients. Cancer. 2006;106(10):2258-66.

4. Crawford J, Caserta C, Roila F. Hematopoietic growth factors: ESMO Clinical Practice Guidelines for the applications. Annals of Oncology. 2010;21:v248-v51.

5. Borgeaud M, Perano S, Addeo A, Tsantoulis P. Rates of febrile neutropenia and its causes in the real world. Future Oncol. 2024;20(29):2203-12.

6. Sebela Pharmaceuticals Inc. IMURAN (azathioprine) 50-mg scored tablets [package insert] [Internet]. Roswell (GA): Sebela Pharmaceuticals Inc.; 2024 May 16 [revised 2024 Jul 9; cited 2025 May 27]. Available from: <https://www.accessdata.fda.gov/drugsatfda_docs/label/2024/016324s040lbl.pdf>.

7. Spectrum Pharmaceuticals, Inc. Beleodaq (belinostat) for injection [package insert] [Internet]. Henderson (NV): Spectrum Pharmaceuticals, Inc.; 2014 Jul [revised 2014 Jul 3; cited 2025 May 27]. Available from: <https://www.accessdata.fda.gov/drugsatfda_docs/label/2014/206256lbl.pdf>.

8. Whirl-Carrillo M, Huddart R, Gong L, Sangkuhl K, Thorn CF, Whaley R, et al. An Evidence-Based Framework for Evaluating Pharmacogenomics Knowledge for Personalized Medicine. Clin Pharmacol Ther. 2021;110(3):563-72.

9. ViiV Healthcare Company. ZIAGEN (abacavir) tablets and oral solution [package insert] [Internet]. Research Triangle Park (NC): ViiV Healthcare Company; 2020 Nov [revised 2024 Jul 24; cited 2025 May 27]. Available from: <https://www.accessdata.fda.gov/drugsatfda_docs/label/2020/020977s035,020978s038lbl.pdf>.

10. Friedlaender A, Vuilleumier A, Viassolo V, Ayme A, De Talhouet S, Combes J-D, et al. BRCA1/BRCA2 germline mutations and chemotherapy-related hematological toxicity in breast cancer patients. Breast Cancer Research and Treatment. 2019;174(3):775-83.

11. Hébert HL, Shepherd B, Milburn K, Veluchamy A, Meng W, Carr F, et al. Cohort Profile: Genetics of Diabetes Audit and Research in Tayside Scotland (GoDARTS). Int J Epidemiol. 2018;47(2):380-1j.

12. McKinstry B, Sullivan FM, Vasishta S, Armstrong R, Hanley J, Haughney J, et al. Cohort profile: the Scottish Research register SHARE. A register of people interested in research participation linked to NHS data sets. BMJ Open. 2017;7(2):e013351.

13. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, et al. Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. Int J Epidemiol. 2013;42(3):689-700.

14. Bidadi B, Liu D, Kalari KR, Rubner M, Hein A, Beckmann MW, et al. Pathway-Based Analysis of Genome-Wide Association Data Identified SNPs in HMMR as Biomarker for Chemotherapy- Induced Neutropenia in Breast Cancer Patients. Frontiers in Pharmacology. 2018;Volume 9 - 2018.

15. Qiu W, Chavarro J, Lazarus R, Rosner B, Ma J. powerSurvEpi: Power and Sample Size Calculation for Survival Analysis of Epidemiological Studies 2025 [R package version 0.1.5]. Available from: <https://CRAN.R-project.org/package=powerSurvEpi>

16. Ruzzo A, Graziano F, Galli F, Galli F, Rulli E, Lonardi S, 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(9):1269-77.

17. Panagiotou OA, Ioannidis JP. What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. Int J Epidemiol. 2012;41(1):273-86.

18. Moore C, Jacobson S. genpwr: Power Calculations Under Genetic Model Misspecification 2021 [R package version 1.0.4]. Available from: <https://CRAN.R-project.org/package=genpwr>

19. Khrunin AV, Moisseev A, Gorbunova V, Limborska S. Genetic polymorphisms and the efficacy and toxicity of cisplatin-based chemotherapy in ovarian cancer patients. The Pharmacogenomics Journal. 2010;10(1):54-61.

20. Cecchin E, D'Andrea M, Lonardi S, Zanusso C, Pella N, Errante D, et al. A prospective validation pharmacogenomic study in the adjuvant setting of colorectal cancer patients treated with the 5-fluorouracil/leucovorin/oxaliplatin (FOLFOX4) regimen. The Pharmacogenomics Journal. 2013;13(5):403-9.

21. Khrunin AV, Khokhrin DV, Moisseev AA, Gorbunova VA, Limborska SA. PHarmacogenomic Assessment of Cisplatin-Based Chemotherapy Outcomes in Ovarian Cancer. Pharmacogenomics. 2014;15(3):329-37.

22. Tibaldi C, Giovannetti E, Vasile E, Mey V, Laan AC, Nannizzi S, et al. Correlation of CDA, ERCC1, and XPD polymorphisms with response and survival in gemcitabine/cisplatin-treated advanced non-small cell lung cancer patients. Clin Cancer Res. 2008;14(6):1797-803.

23. Joerger M, Burgers SA, Baas P, Smit EF, Haitjema TJ, Bard MP, et al. Germline polymorphisms in patients with advanced nonsmall cell lung cancer receiving first-line platinum-gemcitabine chemotherapy: a prospective clinical study. Cancer. 2012;118(9):2466-75.

24. Cortejoso L, García MI, García-Alfonso P, González-Haba E, Escolar F, Sanjurjo M, et al. Differential toxicity biomarkers for irinotecan- and oxaliplatin-containing chemotherapy in colorectal cancer. Cancer Chemother Pharmacol. 2013;71(6):1463-72.

25. Grenda A, Błach J, Szczyrek M, Krawczyk P, Nicoś M, Kuźnar Kamińska B, et al. Promoter polymorphisms of TOP2A and ERCC1 genes as predictive factors for chemotherapy in non-small cell lung cancer patients. Cancer Med. 2020;9(2):605-14.

26. Goekkurt E, Al-Batran SE, Hartmann JT, Mogck U, Schuch G, Kramer M, et al. Pharmacogenetic analyses of a phase III trial in metastatic gastroesophageal adenocarcinoma with fluorouracil and leucovorin plus either oxaliplatin or cisplatin: a study of the arbeitsgemeinschaft internistische onkologie. J Clin Oncol. 2009;27(17):2863-73.

27. Garziera M, Virdone S, De Mattia E, Scarabel L, Cecchin E, Polesel J, et al. HLA-G 3'UTR Polymorphisms Predict Drug-Induced G3-4 Toxicity Related to Folinic Acid/5-Fluorouracil/Oxaliplatin (FOLFOX4) Chemotherapy in Non-Metastatic Colorectal Cancer. Int J Mol Sci. 2017;18(7).

28. Yao S, Barlow WE, Albain KS, Choi JY, Zhao H, Livingston RB, et al. Gene polymorphisms in cyclophosphamide metabolism pathway,treatment-related toxicity, and disease-free survival in SWOG 8897 clinical trial for breast cancer. Clin Cancer Res. 2010;16(24):6169-76.

29. Gréen H, Hasmats J, Kupershmidt I, Edsgärd D, de Petris L, Lewensohn R, et al. Using Whole-Exome Sequencing to Identify Genetic Markers for Carboplatin and Gemcitabine-Induced Toxicities. Clin Cancer Res. 2016;22(2):366-73.

30. Booton R, Ward T, Heighway J, Taylor P, Power F, Ashcroft L, et al. Xeroderma pigmentosum group D haplotype predicts for response, survival, and toxicity after platinum-based chemotherapy in advanced nonsmall cell lung cancer. Cancer. 2006;106(11):2421-7.

31. Lamba JK, Fridley BL, Ghosh TM, Yu Q, Mehta G, Gupta P. Genetic variation in platinating agent and taxane pathway genes as predictors of outcome and toxicity in advanced non-small-cell lung cancer. Pharmacogenomics. 2014;15(12):1565-74.

32. Svedberg A, Björn N, Sigurgeirsson B, Pradhananga S, Brandén E, Koyi H, et al. Genetic association of gemcitabine/carboplatin-induced leukopenia and neutropenia in non-small cell lung cancer patients using whole-exome sequencing. Lung Cancer. 2020;147:106-14.

33. Melchardt T, Hufnagl C, Magnes T, Weiss L, Hutarew G, Neureiter D, et al. CYP39A1 polymorphism is associated with toxicity during intensive induction chemotherapy in patients with advanced head and neck cancer. BMC Cancer. 2015;15(1):725.

34. Senk B, Goricar K, Kovac V, Dolzan V, Franko A. Genetic polymorphisms in aquaporin 1 as risk factors for malignant mesothelioma and biomarkers of response to cisplatin treatment. Radiol Oncol. 2019;53(1):96-104.

35. Fasching PA, Liu D, Scully S, Ingle JN, Lyra PC, Rack B, et al. Identification of Two Genetic Loci Associated with Leukopenia after Chemotherapy in Patients with Breast Cancer. Clin Cancer Res. 2022;28(15):3342-55.

36. Yao S, Barlow WE, Albain KS, Choi JY, Zhao H, Livingston RB, et al. Manganese superoxide dismutase polymorphism, treatment-related toxicity and disease-free survival in SWOG 8897 clinical trial for breast cancer. Breast Cancer Res Treat. 2010;124(2):433-9.

37. Sawyer MB, Pituskin E, Damaraju S, Bies RR, Vos LJ, Prado CM, et al. A Uridine Glucuronosyltransferase 2B7 Polymorphism Predicts Epirubicin Clearance and Outcomes in Early-Stage Breast Cancer. Clin Breast Cancer. 2016;16(2):139-44.e1-3.

38. Vulsteke C, Lambrechts D, Dieudonné A, Hatse S, Brouwers B, van Brussel T, et al. Genetic variability in the multidrug resistance associated protein-1 (ABCC1/MRP1) predicts hematological toxicity in breast cancer patients receiving (neo-)adjuvant chemotherapy with 5-fluorouracil, epirubicin and cyclophosphamide (FEC). Annals of Oncology. 2013;24(6):1513-25.

39. Gréen H, Khan MS, Jakobsen-Falk I, Åvall-Lundqvist E, Peterson C. Impact of CYP3A5\*3 and CYP2C8-HapC on Paclitaxel/Carboplatin-Induced Myelosuppression in Patients with Ovarian Cancer. Journal of Pharmaceutical Sciences. 2011;100(10):4205-9.

40. Alessia O, Monica De M, Matteo Giovanni Della P, GianMatteo R, Cristina A, Francesco Di R, et al. Gene polymorphisms in folate metabolizing enzymes in adult acute lymphoblastic leukemia: effects on methotrexate-related toxicity and survival. Haematologica. 2009;94(10):1391-8.

41. Chen S, Villeneuve L, Jonker D, Couture F, Laverdière I, Cecchin E, et al. ABCC5 and ABCG1 polymorphisms predict irinotecan-induced severe toxicity in metastatic colorectal cancer patients. Pharmacogenetics and Genomics. 2015;25(12):573-83.

42. Ardizzone A, Bulzomì M, De Luca F, Silvestris N, Esposito E, Capra AP. Dihydropyrimidine Dehydrogenase Polymorphism c.2194G>A Screening Is a Useful Tool for Decreasing Gastrointestinal and Hematological Adverse Drug Reaction Risk in Fluoropyrimidine-Treated Patients. Current Issues in Molecular Biology. 2024;46(9):9831-43.

43. Korver SK, Bowen JM, Gibson RJ, Ball IA, Secombe KR, Wain TJ, et al. Advanced statistics identification of participant and treatment predictors associated with severe adverse effects induced by fluoropyrimidine-based chemotherapy. Cancer Chemother Pharmacol. 2023;91(6):507-21.