# Personalised medicine of fluoropyrimidines using DPYD pharmacogenetics

Lunenburg, C.A.T.C.

# CHAPTER 12

## Genome-wide association study to discover novel genetic variants related to the onset of severe toxicity following fluoropyrimidine use

*Manuscript in preparation*

**Carin A.T.C. Lunenburg, Linda M. Henricks, Stefan Böhringer, Femke M. de Man, Geert-Jan Creemers, Arnold Baars, Vincent O. Dezentjé, Alexander L.T. Imholz, Frank J.F. Jeurissen, Johanna E.A. Portielje, Rob L.H. Jansen, Paul Hamberg, Albert J. ten Tije, Helga J. Droogendijk, Miriam Koopman, Peter Nieboer, Marlène H.W. van de Poel, Caroline M.P.W. Mandigers, Hans Gelderblom, Ron H.J. Mathijssen, Annemieke Cats, Henk-Jan Guchelaar, Jan H.M. Schellens, Jesse J. Swen**

**Abstract**

Fluoropyrimidines are widely used anticancer drugs, but may lead to severe toxicity in up to 30% of patients. Prospective *DPYD* genotyping is increasingly used in clinical practice to predict and prevent severe toxicity, by means of initial dose reductions in *DPYD* variant allele carriers. While this strategy successfully reduces the incidence of severe toxicity, substantial toxicity remains that is not attributable to genetic variation in *DPYD*. A genome-wide association study (GWAS) was initiated to discover novel genetic variants associated with the onset of severe fluoropyrimidine-induced toxicity.

We conducted a GWAS in 1,146 patients treated with fluoropyrimidines who participated in the Alpe DPD study. Patients were genotyped using the Illumina Global Screening Array and data was imputed using the 1000 Genomes reference panel. The primary outcome was severe (grade ≥3) fluoropyrimidine-induced toxicity, compared to grade 0 or 1 fluoropyrimidine-induced toxicity. Variants were tested for association with severe fluoropyrimidine-induced toxicity using logistic, Cox, and ordinal regressions. A Polygenic Risk Score (PRS) was constructed by extracting all variants with $p<0.01$ in the association test.

1,101 patients passed the quality control (QC) analyses and 599 patients were included in the primary analysis. After imputation, 4,650,899 variants were included in the analysis. None of the genetic variants showed genome-wide significance ($p<5x10^{-8}$). Six variants were suggestive ($p<5x10^{-6}$) for the onset of severe fluoropyrimidine-induced toxicity. A PRS was constructed including 5,055 variants and predicted 62% of severe toxicity by non-genetic covariates alone and 96% by the combined analysis including covariates.

While no genome-wide significant variants could be identified, six variants were suggestive for the onset of severe toxicity in merely Caucasian patients. These variants are located outside of known fluoropyrimidine-pathway genes. Using a PRS consisting of 5,055 variants combined with clinical variables explained 96% of toxicity in this discovery cohort. This GWAS is one of the first attempts to identify variants predictive for fluoropyrimidine-induced toxicity and identified variants and the PRS require replication in an independent cohort.

**Acknowledgements**

The authors thank the Human Genotyping Facility of the Erasmus Medical Center for the execution of the genotyping and all participating centres of the NCT02324452 study for their help in including patients in the study.

**Introduction**

Fluoropyrimidines, including 5-fluorouracil (5-FU) and capecitabine, represent the backbone of chemotherapeutic regimens used to treat solid tumours, such as colorectal and breast cancer. Severe (grade ≥3) fluoropyrimidine-induced toxicity can occur in up to 30% of the patients, depending on the treatment regimen and may even be lethal in up to 1% of the patients experiencing toxicity.[1,2] Common fluoropyrimidine-induced adverse events are diarrhoea, mucositis, hand-foot syndrome and myelosuppression.[1,3] Dihydropyrimidine dehydrogenase (DPD) plays a key role in the degradation of 5-FU into inactive metabolites[4] and is encoded by the gene *DPYD.* Both DPD and genetic variants in *DPYD* have been widely investigated to explain severe fluoropyrimidine-induced toxicity. Recently, we have shown that prospective genotyping and dose reduction based on four variants in *DPYD* (*DPYD**2A, c.2846A>T, c.1679T>G and c.1236G>A) reduces severe fluoropyrimidine-induced toxicity in these *DPYD* variant allele carriers.[5] These four variants were selected based on previous studies and meta-analyses in which the association with fluoropyrimidine-induced toxicity was reported.[6-13] Nonetheless, severe toxicity did still occur in 23% of patients wild-type for these four variants, showing that other genetic variants or non-genetic factors may play a role in the onset of severe toxicity.[5]

Variants in genes other than *DPYD* could also play a role in the onset of severe fluoropyrimidine-induced toxicity. Previously, research to identify genetic variants has been conducted based on the pharmacological background of fluoropyrimidines, for example in pathway analyses or candidate gene studies. Several variants in *CDA* (cytidine deaminase), *CES1* (carboxylesterase 1), *TYMS* (thymidylate synthase), *MTHFR* (methylenetetrahydrofolate reductase), *ENOSF1* (enolase superfamily, member 1), *SLC22A7* (solute carrier family 22, member 7), *UMPS* (uridine monophosphate synthase) and *TYMP* (thymidine phosphorylase) genes were previously identified and associated with severe fluoropyrimidine-induced toxicity.[1,14-21] However, genome-wide association studies (GWAS) have the potential to identify novel variants without making assumptions based on a pharmacological background. Previously, O'Donnell *et al*. executed a GWAS on 503 cell lines to identify novel single nucleotide polymorphisms (SNPs) associated with capecitabine sensitivity.[22] Five variants showed genome-wide significance in this cell-line based GWAS, but replication in 268 patients only showed an association with sensitivity for capecitabine for *ADCY2* rs4702484.[23] Fernandez-Rodzilla *et al.* analysed data of 221 colorectal cancer patients treated with 5-FU or FOLFOX (folinic acid, 5-FU and oxaliplatin).[24] Seven SNPs were associated with adverse drug reactions, yet none reached the genome-wide significance level. Low *et al.* executed a GWAS on 13,220 patients in total, of which 1,460 patients received 5-FU, focussing on neutropenic and leukopenic toxicities.[25] For 5-fluorouracil, they identified four SNPs associated to neutropenia and leukopenia, yet none reached the genome-wide significance level. We conducted a GWAS to discover novel genetic variants associated with the onset of severe fluoropyrimidine-induced toxicity.

## Materials and methods

### *Patients*

Patients were recruited for the Alpe DPD study (clinicaltrial.gov identifier NCT02324452)[5] between April 30, 2015 and December 21, 2017, and were newly treated with fluoropyrimidines and genotyped prospectively for four *DPYD* variants (*DPYD**2A, rs3918290, c.1905+1G>A, IVS14+1G>A; c.1679T>G, *DPYD**13, rs55886062, I560S; c.1236G>A/HapB3, rs56038477, E412E; and c.2846A>T, rs67376798, D949V). Upon identification of one of these variants, heterozygous variant allele carriers received an initial dose reduction (25 or 50%) based on pharmacogenetic guidelines to prevent severe fluoropyrimidine-induced toxicity. Wild-type patients for these four *DPYD* variants received standard fluoropyrimidine dosages. After the second cycle the dose could be titrated upwards or downwards according to the occurrence of toxicity. The study was reviewed and approved by the medical ethical committee of the Netherlands Cancer Institute, Amsterdam, the Netherlands, and approval of the board of directors of each individual hospital was obtained for all participating centres. All patients signed informed consent prior to inclusion in the study, which included approval for the use of clinical data and remaining DNA to perform the current GWAS. All patients of whom sufficient DNA was available were genotyped. *DPYD* variant allele carriers (*N*=85) received dose reductions based on the four variants mentioned and were therefore excluded in the GWAS analyses.

### *Clinical data*

Baseline characteristics, treatment type and toxicity data were collected for each patient. Ethnicity of the patients was self-reported, merely Caucasian patients participated in the Alpe DPD study. Toxicity was graded according to the National Cancer Institute common terminology criteria for adverse events (CTC-AE; version 4.03) and severe toxicity was defined as CTC-AE grade ≥3.[26] Relation to the study drugs 5-FU and capecitabine was recorded for each adverse event and only adverse events classified as possible, probable, or definite were taken into account.

### *Genotyping and quality control*

Patient DNA remaining from the Alpe DPD study was collected. For each patient 200 ng of DNA was required and genotyping was executed at the Human Genotyping Facility of the Erasmus Medical Center, using the Illumina Global Screening Array (GSA).[27] The array contains 692,842 SNPs and includes rare variants with allele frequencies <1%. 1000 Genomes reference phase 3 GRCh37.p13 was used to impute the data. Quality control (QC) checks were performed using software R version 3.5.0[28] and PLINK software, version 1.07.[29,30] Patients were excluded from analyses based on an individual genotype call rate <97%, gender mismatch between reported and estimated sex based on genotypes of the X-chromosome (using PLINK), or excess of heterozygous genotypes as measured by the inbreeding coefficient. An inbreeding statistic of F>0.1 was judged to be outlying and patients were removed from the analysis. Genetic markers were excluded based on a SNP call rate <97% and a p-value $\leq 10^{-7}$ for the Hardy-Weinberg equilibrium (HWE) goodness-of-fit test. After exclusion of patients and markers in these marginal QCs, the remaining set was used

for integrative QC assessment. In order to evaluate the possibility of population stratification or outliers, multidimensional scaling (MDS) analysis was performed in PLINK. In addition, pairwise identity by state (IBS)/identity by descent (IBD) statistics was calculated to assess duplicates. MDS, IBS and IBD were computed using PLINK. Patients who were identified as outliers based on IBS clustering were excluded from the analysis. MDS coordinates were extracted and used as covariates in the association analysis. SNP imputation was performed using the programs *shapeit* and *impute2* with default parameters in which the reference panel 1000Genomes build version 3 was used with total, 'cosmopolitan', set of individuals.[31] An MDS plot was created to compare self-reported ethnicity of patients.

***Statistical analyses***

Genetic variants were tested for an association with the onset of severe fluoropyrimidine-induced toxicity. The primary outcome was severe (grade ≥3) fluoropyrimidine-induced toxicity, compared to grade 0 or 1 fluoropyrimidine-induced toxicity. Grade 2 toxicity was excluded from this analysis to maximize the contrast between toxicities. Gender, age, baseline BSA and treatment type (grouped as previously published)[5] were used as pre-specified covariates. Statistical analyses were performed in R statistics version 2.3.2. Base packages *stats, survival* and *MASS* were used to evaluate logistic, Cox, and ordinal regressions, respectively. Associations with a p-value $\leq 5x10^{-8}$ were considered statistically genome-wide significant. Associations with a p-value between $5x10^{-8}$ and $\leq 5x10^{-6}$ were considered suggestive. Post association QC was performed by visual inspection of Quantile-Quantile (QQ) plots of p-values of association tests and computation of the inflation factor given as: $\lambda=(\text{median}(T_1,\ldots,T_n)/0.675)^2$, where $T_1,\ldots,T_n$ are square roots of $\chi^2$ quantiles.

A Polygenic Risk Score (PRS) was constructed by extracting all SNPs with a p-value <0.01 in the association test. To avoid problems due to collinearity, in the list sorted according to p-values, SNPs in a window of 100 kb were excluded after inclusion of a SNP. A penalized regression model was fitted using R-package *glmnet*. Included clinical parameters were gender, age, baseline BSA and treatment schedule.

12

**Results**

***Patients***

Sufficient DNA was available for 1,146 out of 1,181 recruited patients (97%). These patients entered the QC procedure prior to association analyses. The flowchart on patient inclusion is shown in Figure 1. The observed individual genotype call rates varied between 97% and 100% and therefore meet the quality criteria. Based on subsequent QC steps, 45 patients (3.9%) were excluded from the analyses. Of these 45 patients, 30 patients (2.6%) were excluded due to missing genotypes, four (0.3%) patients were excluded due to a gender mismatch with the clinical data, six patients (0.5%) were excluded based on outlier removal of IBS plots. The inbreeding coefficient was 0.01 (-0.03─0.004), therefore, five (0.4%) patients were excluded. Of the 1,101 remaining patients, screen failures (*N*=55), patients with missing BSA at baseline (*N*=24) and *DPYD* variant allele carriers who received initially reduced dosages (*N*=80) were excluded (Figure 1). In addition, we chose to exclude patients who experienced grade 2 toxicity (*N*=343) from the primary analysis to maximize contrast between severe and non-severe toxicity.

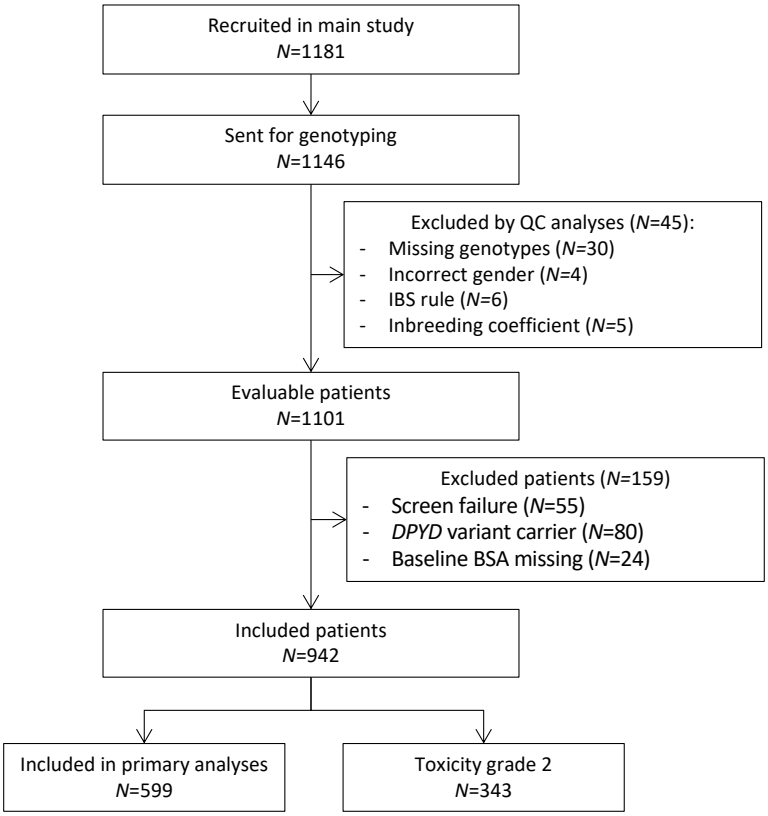


**Figure 1. Flowchart of patients in the study**

Patients who experienced toxicity grade 2 were excluded from the primary analyses to maximize contrast between severe and non-severe toxicity.

*Abbreviations:* QC: quality control; IBS rule: identity by state rule; *DPYD*: gene encoding dihydropyrimidine dehydrogenase; BSA: body surface area.

### *Association analysis*

In the primary analysis, severe (grade ≥3) toxicity was compared to grade 0 and 1 toxicity in 599 patients. Patient characteristics of patients included in the primary analysis of this study and patients from the Alpe DPD study are shown in Table 1. There were no statistical differences between the cohorts, except for the number of treatment cycles.

The number of patients varied per SNP due to genotype missingness, which was limited to up to 3% as per QC. For the primary analysis, a MAF of 2% was used on imputed data to ensure stable numeric evaluation of all logistic regression models. This resulted in a total of 4,650,899 markers for which an association test was performed.

The primary analysis association test for severe fluoropyrimidine-induced toxicity (grades 3—5) includes covariates gender, age, baseline BSA and treatment type. The corresponding

Manhattan plot is shown in Figure 2. The corresponding QQ-plot of p-values is shown in Figure 3. The inflation factor is 1.04. Table 2 shows the list of the top 30 identified markers. No variants were identified to be statistically significant associated with severe fluoropyrimidine-induced toxicity at the genome-wide level. However, six SNPs were found to be suggestive. None of these SNPs have previously been reported in publications or in the ClinVar database of the National Center for Biotechnology Information (NCBI).[32] The variants are reported in the SNP database of the NCBI.[33] Three variants are stated on the website as having 'no gene consequence', two were listed as an intron variant in RNA gene LOC101927414 (rs114105116) and protein coding gene COL6A3 (rs12622722), and one was listed as an 2KB upstream variant in LOC107984256 (rs10786179).

***Genotyping and quality control***

A set of 692,367 markers was genotyped. After several QC steps, 186,920 markers were excluded. Of these, 18,114 markers (2.6%) were excluded based on a deviation from Hardy-Weinberg equilibrium (HWE). Filtering for allele frequencies (threshold 0.5%) resulted in the exclusion of 147,607 markers (21.3%). The missingness cut-off was set at 10%, 23,835 markers (3.4%) were excluded based on the missing data analysis. Of the abovementioned excluded markers, 2,636 had multiple QC failures. In total, 505,447 markers met the QC for statistical analyses. These markers were imputed using the 1000 Genomes dataset as a reference panel. In total, 4,650,899 variants were available for statistical analyses. In the integrative QC, individuals and markers from the marginal QC steps were excluded. An MDS was executed in order to detect population stratification. No individuals were excluded. IBD/IBS clustering was executed to assess duplicates. No individuals were excluded.

***Polygenic risk score***

To calculate the PRS all SNPs with a p-value <0.01 for their association with severe fluoropyrimidine-induced toxicity were selected. To reduce linkage disequilibrium, SNPs were pruned for a minimum distance of $10^5$ bps. This resulted in a set of 5,055 SNPs. Finally, an elastic net regression (R package *glmnet*, $\alpha$=0.5) was performed and evaluated by cross-validation. The receiver operating characteristic (ROC) curve is shown in Figure 4, where it is compared to the model containing only clinical covariates. The two corresponding areas under the curve (AUCs) were 96% and 62%, respectively.

**Table 1. Patient characteristics**

Patient characteristics of patients included in the primary analysis of this GWAS (*N*=599) and patients included in the Alpe DPD study (*N*=1,103). Data are *N*(%) or median(IQR). P-values comparing patients from the primary analysis to the Alpe DPD study patients. We used a nonparametric test for independent samples to compare medians of age, BSA, and number of treatment cycles; and a $\chi^2$ test for gender, ethnic origin, tumour type, treatment regimen and WHO performance status.

| **Characteristic** | **GWAS cohort (*N*=599)** | **Alpe DPD study (*N*=1,103)** | **P-value** |
|---|---|---|---|
| Gender | | | 0.841 |
| *Male* | 319 (53.3%) | 593 (53.8%) | |
| *Female* | 280 (46.7%) | 510 (46.2%) | |
| Age in years [IQR] | 64 [57—71] | 64 [56—71] | 0.454 |
| Ethnic origin | | | 0.362 |
| *White* | 573 (95.7%) | 1048 (95%) | |
| *Black* | 14 (2.3%) | 19 (1.7%) | |
| *Asian* | 9 (1.5%) | 24 (2.2%) | |
| *Other*[a] | 3 (<1%) | 12 (1.1%) | |
| Tumour type | | | 0.991 |
| *Non-metastatic colorectal cancer* | 265 (44.2%) | 472 (42.8%) | |
| *Metastatic colorectal cancer* | 114 (19%) | 232 (21%) | |
| *Breast cancer* | 75 (12.5%) | 141 (12.8%) | |
| *Gastric cancer* | 32 (5.3%) | 63 (5.7%) | |
| *Other*[b] | 113 (18.9%) | 195 (17.7%) | |
| Type of treatment regimen | | | 0.234 |
| *Capecitabine monotherapy (±bevacizumab)* | 102 (17%) | 205 (18.6%) | |
| *Capecitabine + radiotherapy (±mitomycin)* | 172 (28.7%) | 264 (23.9%) | |
| *Capecitabine + oxaliplatin (±bevacizumab)* | 179 (29.9%) | 374 (33.9%) | |
| *Capecitabine + other anticancer drugs* | 41 (6.8%) | 72 (6.5%) | |
| *Fluorouracil monotherapy* | - | 2 (<1%) | |
| *Fluorouracil + radiotherapy (±mitomycin)* | 43 (7.2%) | 63 (5.7%) | |
| *Fluorouracil + oxaliplatin + folinic acid (±bevacizumab)* | 18 (3%) | 43 (3.9%) | |
| *Fluorouracil + other anticancer drugs* | 44 (7.3%) | 80 (7.3%) | |
| BSA [IQR] | 1.9 [1.8—2.1] | 1.9 [1.8—2.1] | 0.503 |
| WHO performance status | | | 0.257 |
| *0* | 317 (52.9%) | 554 (50.2%) | |
| *1* | 241 (40.2%) | 448 (40.6%) | |
| *2* | 21 (3.5%) | 42 (3.8%) | |
| *Not specified*[c] | - | 59 (5.3%) | |
| Number of treatment cycles [IQR] | 3 [1-7] | 3 [1-8] | 0.010 |
| *DPYD* status | | | NA |
| *Wild-type* | 599 (100%) | 1018 (92.3%) | |
| *c.1236G>A heterozygous* | | 51 (4.6%) | |
| *c.2846A>T heterozygous* | | 17 (1.5%) | |
| *DPYD*2A heterozygous* | | 16 (1.5%) | |
| *DPYD*13 heterozygous* | | 1 (<1%) | |

[a] Other ethnic origins included Hispanic descent, mixed racial parentage, and unknown ethnic origin;
[b] Other tumour types included anal cancer, oesophageal cancer, head and neck cancer, pancreatic cancer, bladder cancer, unknown primary tumour, vulva carcinoma, and several rare tumour types;
[c] WHO performance status was not specified for these patients, but was either 0, 1, or 2, as required by the study inclusion criteria.
*Abbreviations:* IQR: interquartile range; BSA: body surface area; DPD: dihydropyrimidine dehydrogenase; *DPYD*: gene encoding dihydropyrimidine dehydrogenase; WHO: world health organisation; NA: not applicable.

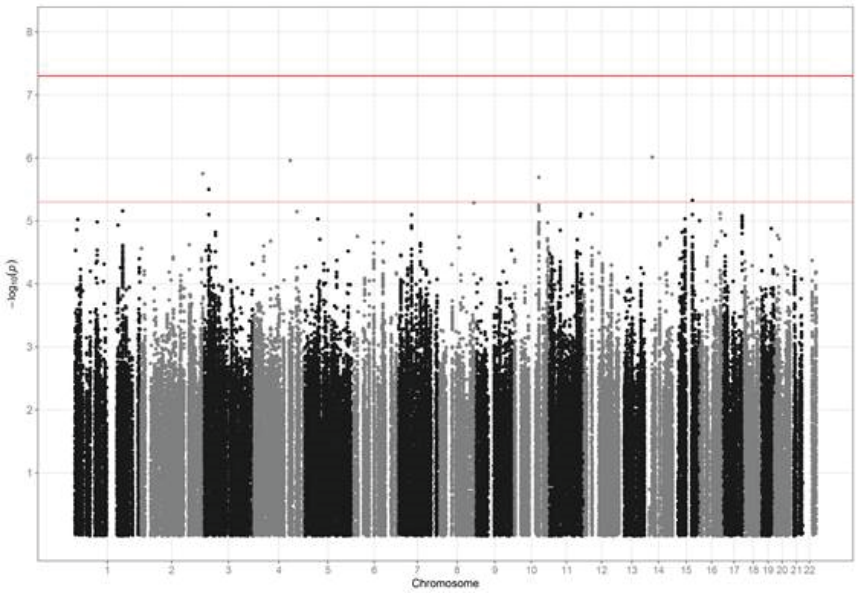


**Figure 2. Manhattan plot**
Manhattan plot for association with severe fluoropyrimidine-induced toxicity (grades 3—5), including de covariates gender, age, baseline BSA and treatment type. Genome-wide significance of the association with the onset of severe fluoropyrimidine-induced toxicity is indicated by the upper dark red line (≤p-value of $5x10^{-8}$). Suggestive association is indicated by the lower red line (p-value of ≤$5x10^{-6}$). No SNPs were found to be associated with severe fluoropyrimidine-induced toxicity. Six SNPs were found to be suggestive for association with severe fluoropyrimidine-induced toxicity, shown in Table 2.
*Abbreviations:* BSA: body surface area; SNPs: single nucleotide polymorphisms.

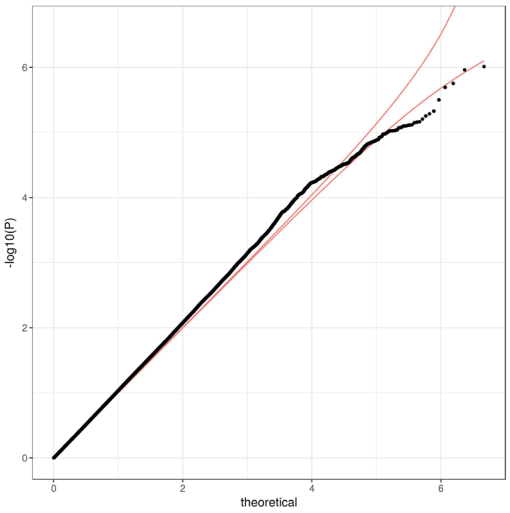


**Figure 3. QQ-plot of p-values**

The Quantile-Quantile (QQ)-plot shows the extent to which the observed distribution of the test statistic follows the theoretical null distribution. The inflation factor was λ=1.04.

**Table 2. Thirty SNPs with lowest p-values**

Variants are selected on allele frequency >0.01, β within -5 to 5, and are separated from another variant with more than 10 bps. Variants suggestive for the onset of severe toxicity are marked with an *.

| Nr. | Marker | Chr | Position | A0 | A1 | AF | β | P-value |
|---|---|---|---|---|---|---|---|---|
| 1 | rs17114875 * | 14 | 29999987 | G | A | 0,409 | 1,554562193 | 9,73E-07 |
| 2 | rs114105116 * | 4 | 138539880 | T | A | 0,02 | 1,213779485 | 1,1E-06 |
| 3 | rs12622722 * | 2 | 238269120 | G | A | 0,484 | 0,657789872 | 1,77E-06 |
| 4 | rs10786179 * | 10 | 96759531 | T | G | 0,889 | -0,345506543 | 2,03E-06 |
| 5 | rs367239 * | 3 | 21421935 | T | C | 0,546 | 1,719787858 | 3,16E-06 |
| 6 | rs11630087 * | 15 | 75261673 | G | T | 0,456 | 0,757448146 | 4,71E-06 |
| 7 | rs77579689 | 8 | 137130325 | G | A | 0,021 | 1,286849892 | 5,18E-06 |
| 8 | rs11187969 | 10 | 96231169 | G | A | 0,129 | 1,381495967 | 5,59E-06 |
| 9 | rs11187974 | 10 | 96239326 | G | A | 0,181 | 1,527446065 | 6,25E-06 |
| 10 | rs12414693 | 10 | 97228795 | C | T | 0,259 | 0,713658584 | 6,87E-06 |
| 11 | rs482061 | 1 | 182485749 | T | C | 0,847 | -0,737925221 | 6,96E-06 |
| 12 | chr4:164083322:D | 4 | 164083322 | TG | T | 0,051 | 1,218271821 | 7,11E-06 |
| 13 | chr16:78157332:I | 16 | 78157332 | G | GTT | 0,065 | 1,023072586 | 7,63E-06 |
| 14 | rs1838947 | 11 | 119691200 | T | C | 0,297 | 0,928423057 | 7,7E-06 |
| 15 | rs495426 | 12 | 31021833 | A | G | 0,689 | 0,073481654 | 7,77E-06 |

*table continues*

| Nr. | Marker | Chr | Position | A0 | A1 | AF | β | P-value |
|---|---|---|---|---|---|---|---|---|
| 16 | rs56338926 | 15 | 75259335 | C | A | 0,448 | 0,772401797 | 7,94E-06 |
| 17 | rs449973 | 3 | 21425977 | C | G | 0,548 | 1,69493935 | 7,97E-06 |
| 18 | rs1722291 | 7 | 56238936 | G | A | 0,198 | 1,553925389 | 8,02E-06 |
| 19 | rs2344989 | 17 | 70924851 | T | C | 0,04 | 1,54932572 | 8,33E-06 |
| 20 | rs2512155 | 11 | 117889448 | C | T | 0,179 | 1,449722565 | 8,45E-06 |
| 21 | rs8076418 | 17 | 70921917 | T | C | 0,042 | 1,608309687 | 8,5E-06 |
| 22 | rs2738545 | 16 | 78629320 | G | A | 0,673 | 0,420462271 | 9,13E-06 |
| 23 | rs10851447 | 15 | 47411086 | T | A | 0,059 | 1,464774341 | 9,21E-06 |
| 24 | rs722910 | 5 | 52781597 | A | T | 0,496 | 0,35577388 | 9,36E-06 |
| 25 | rs8070810 | 17 | 70921851 | G | A | 0,042 | 1,61395951 | 9,42E-06 |
| 26 | rs8067883 | 17 | 70921731 | C | T | 0,042 | 1,612564569 | 9,43E-06 |
| 27 | rs6501582 | 17 | 70921801 | T | C | 0,042 | 1,613613576 | 9,44E-06 |
| 28 | rs113309475 | 1 | 11430624 | A | T | 0,023 | 1,200617555 | 9,49E-06 |
| 29 | rs9911437 | 17 | 70922305 | T | C | 0,042 | 1,614540275 | 9,49E-06 |
| 30 | chr17:70923098:D | 17 | 70923098 | AC | A | 0,042 | 1,614669511 | 9,65E-06 |

*Abbreviations:* Nr: number; Chr: chromosome; A0: nucleotide on allele 0; A1: nucleotide on allele 1; AF: allele frequency.

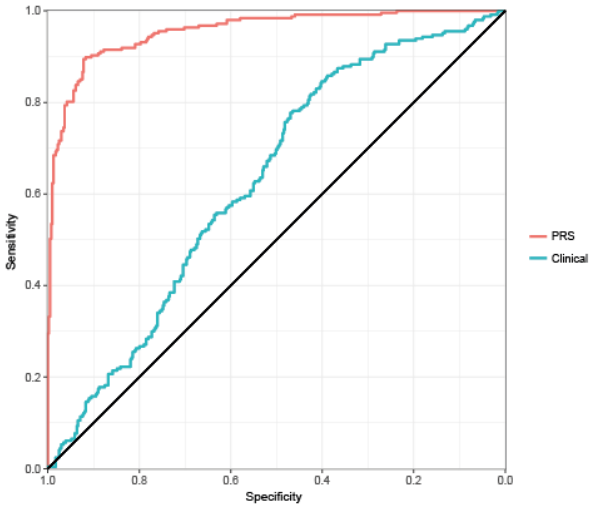


**Figure 4. PRS-plot**

The PRS-plot shows the predictive power of a covariate model (Clinical, blue line) and a covariate model plus 5,055 SNPs (PRS, red line). Included clinical parameters were gender, age, baseline BSA and treatment schedule. The AUC for the covariate model is 0.620 (95%CI: 0.577-0.663, $p=1.34^{-09}$), compared to the AUC for the combined model (prs) 0.956 (95%CI: 0.939-0.973, $p=3.98^{-140}$).

*Abbreviations:* PRS: polygenic risk score; AUC: area under the curve; 95%CI: 95% confidence interval.

**Discussion**

It is well recognized that *DPYD* genotyping is useful in preventing severe fluoropyrimidine-induced toxicity by applying initial dose reductions in patients who carry a specific *DPYD* variant.[5] However, not all toxicity can be predicted and prevented by the current four *DPYD* variants. Indeed, still ~20% of the patients experience toxicity, thus the search for genetic variants predictive for severe fluoropyrimidine-induced toxicity continues. We executed a GWAS in order to identify novel genetic variants possibly associated with the onset of severe fluoropyrimidine-induced toxicity.

To perform this GWAS on toxicity of fluoropyrimidines, over 1,100 patients were genotyped. Severe toxicity includes the National Cancer Institute CTC-AE grades 3—5. Scoring clinical toxicities can be difficult sometimes, as it can be open to interpretation. As severe toxicity has the most clinical impact, we chose to maximize contrast to the toxicity endpoint, and we excluded patients with grade 2 toxicity in the primary analysis. We were unable to identify genome-wide significant SNPs, yet we identified six SNPs suggestive (p-value of $5x10^{-6}$) of association with severe fluoropyrimidine-induced toxicity. Possibly the number of patients in our study is too small to reach genome-wide significance. However, we repeated the analysis including the patients with grade 2 toxicity, increasing the number of patients while reducing the contrast between toxicities. Yet, this did not result in a different outcome. The suggestive variants need to be re-tested in an independent cohort of patients who were treated with a fluoropyrimidine drug.

This GWAS was executed using DNA from patients participating in the Alpe DPD study. A formal comparison of GWAS analysed patients with the entire Alpe DPD cohort shows the cohorts were comparable. The range of number of treatment cycles was statistically different, with fewer cycles in the GWAS cohort. Possibly this is due to the exclusion of patients with grade 2 toxicity in the GWAS cohort, as grade 2 toxicity, if not developing into severe toxicity, may possibly arise from longer periods of fluoropyrimidine-treatment. We have no reason to believe that selection bias was introduced by leaving out patients with grade 2 toxicity. We believe the GWAS cohort is representative for patients in daily clinical care, as in the Alpe DPD study there were only limited restrictions on the inclusion criteria and the burden for patients to participate was very low.

With a large amount of genotyping data, we were able to compare ethnicity strings in the MDS plots to self-reported ethnicity from the Alpe DPD study. When adding two principal components, including ethnicity, to the statistical analysis, no differences were visible. Therefore, ethnicity was not of influence on the outcome of this GWAS and no patients were excluded based on self-reported ethnic origin.

Data on the functionality of the six SNPs suggestive of association with severe fluoropyrimidine-induced toxicity is limited. To the best of our knowledge, these six SNPs were not previously identified by other GWAS or other studies, or previously described in relation to the fluoropyrimidine pathway. Genome-wide significant SNPs (rs4702484, rs8101143, rs576523, rs361433) and suggestive SNPs (rs16857540, rs2465403, rs10876844, rs10784749, rs17626122, rs7325568, rs4243761, rs10488226, rs6740660, rs1567482 and rs6706693) identified in previously executed GWAS,[22,24,25] were not identified in this GWAS, possibly due to the differences in the design or endpoints of the study. For example, the

GWAS of Low *et al*. focused only on neutropenia and leukopenia as toxicity endpoint.[25] In the current GWAS we chose to include all types of fluoropyrimidine-induced toxicity, as we aim to improve fluoropyrimidine treatment by reducing all types of toxicity. Compared to the GWAS of O'Donnell *et al.* [22], we offer a cohort with clinical patient data representative of daily clinical care, in order to identify variants which could be clinically relevant. Compared to the GWAS of Fernandez-Rodzilla *et al.* our cohort is much larger.[24]

When applying prospective *DPYD* genotyping, still 23% of patients treated with fluoropyrimidines experience severe toxicity.[5] In order to further reduce this number, other genetic variants predictive for severe toxicity need to be identified. Options are to screen for rare variants in *DPYD*, investigate epigenetics, or look outside of the *DPYD* gene as was performed in this GWAS. The onset of severe toxicity might not only be linked to the start of fluoropyrimidines, but can be multifactorial and linked to patient characteristics or co-medication. In line of that thought, the onset of toxicity might be better predicted by multiple genetic variants. For this reason, we executed the PRS analysis. In addition, the future of genotyping is quickly evolving, with less single SNP-based assays and more assays with a panel of SNPs in genes or assays sequencing entire regions of genes, leading to future possibilities to apply a PRS in patients in clinical care. Our PRS analysis showed that a panel of 5,055 SNPs combined with clinical covariates outperforms clinical parameters alone, and can predict 96% of severe fluoropyrimidine-induced toxicity. Our PRS analysis possibly shows too optimistic results due to the pre-selection of significant SNPs into the score. Although we used cross-validation to verify the score, this step did not include the SNP selection as this would have been computationally prohibitive. We see this as an exploratory analysis that needs validation, but still suggests that low penetrance variants exist which are difficult to prove in a single-variant association test. The PRS analysis shows the possibility of future research with a multifactorial research approach. The panel of SNPs needs replication in a validation cohort and additional research is needed to be able to link the result to a dose adjustment advice in order to prevent toxicity.

In conclusion, while no genome-wide significant SNPs could be identified in our unique dataset of patients, six variants were suggestive for the onset of severe toxicity. These variants are located outside of known fluoropyrimidine-pathway genes. Using a PRS consisting of 5,055 SNPs combined with clinical variables explained 96% of toxicity in an optimistic analysis, suggesting highly polygenic nature of toxicity predisposition. This GWAS is one of the first attempts to identify variants predictive for fluoropyrimidine-induced toxicity. The identified variants and the PRS require replication in an independent cohort.

## References


1. Rosmarin D, Palles C, Pagnamenta A, et al. A candidate gene study of capecitabine-related toxicity in colorectal cancer identifies new toxicity variants at *DPYD* and a putative role for ENOSF1 rather than TYMS. *Gut.* 2015;64(1):111-120.
2. Saltz LB, Niedzwiecki D, Hollis D, et al. Irinotecan fluorouracil plus leucovorin is not superior to fluorouracil plus leucovorin alone as adjuvant treatment for stage III colon cancer: results of CALGB 89803. *J Clin Oncol.* 2007;25(23):3456-3461.
3. Saltz LB, Cox JV, Blanke C, et al. Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. Irinotecan Study Group. *N Engl J Med.* 2000;343(13):905-914.
4. Heggie GD, Sommadossi JP, Cross DS, Huster WJ, Diasio RB. Clinical pharmacokinetics of 5-fluorouracil and its metabolites in plasma, urine, and bile. *Cancer Res.* 1987;47(8):2203-2206.
5. Henricks LM, Lunenburg CATC, de Man FM, et al. *DPYD* genotype-guided dose individualisation of fluoropyrimidine therapy in patients with cancer: a prospective safety analysis. *Lancet Oncol.* 2018;19(11):1459-1467.
6. Rosmarin D, Palles C, Church D, 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(10):1031-1039.
7. Terrazzino S, Cargnin S, Del RM, Danesi R, Canonico PL, Genazzani AA. *DPYD* IVS14+1G>A and 2846A>T genotyping for the prediction of severe fluoropyrimidine-related toxicity: a meta-analysis. *Pharmacogenomics.* 2013;14(11):1255-1272.
8. Deenen MJ, Tol J, Burylo AM, et al. Relationship between single nucleotide polymorphisms and haplotypes in *DPYD* and toxicity and efficacy of capecitabine in advanced colorectal cancer. *Clin Cancer Res.* 2011;17(10):3455-3468.
9. Meulendijks D, Henricks LM, Sonke GS, et al. Clinical relevance of *DPYD* variants c.1679T>G, c.1236G>A/HapB3, and c.1601G>A as predictors of severe fluoropyrimidine-associated toxicity: a systematic review and meta-analysis of individual patient data. *Lancet Oncol.* 2015;16(16):1639-1650.
10. Morel A, Boisdron-Celle M, Fey L, et al. Clinical relevance of different dihydropyrimidine dehydrogenase gene single nucleotide polymorphisms on 5-fluorouracil tolerance. *Mol Cancer Ther.* 2006;5(11):2895-2904.
11. Schwab M, Zanger UM, Marx C, et al. Role of genetic and nongenetic factors for fluorouracil treatment-related severe toxicity: a prospective clinical trial by the German 5-FU Toxicity Study Group. *J Clin Oncol.* 2008;26(13):2131-2138.
12. Van Kuilenburg ABP, Meijer J, Mul ANPM, et al. Intragenic deletions and a deep intronic mutation affecting pre-mRNA splicing in the dihydropyrimidine dehydrogenase gene as novel mechanisms causing 5-fluorouracil toxicity. *Hum Genet.* 2010;128(5):529-538.
13. Amstutz U, Farese S, Aebi S, Largiader CR. Dihydropyrimidine dehydrogenase gene variation and severe 5-fluorouracil toxicity: a haplotype assessment. *Pharmacogenomics.* 2009;10(6):931-944.
14. Hamzic S, Kummer D, Milesi S, et al. Novel Genetic Variants in Carboxylesterase 1 Predict Severe Early-Onset Capecitabine-Related Toxicity. *Clin Pharmacol Ther.* 2017;102(5):796-804.
15. Pellicer M, Garcia-Gonzalez X, Garcia MI, et al. Identification of new SNPs associated with severe toxicity to capecitabine. *Pharmacol Res.* 2017;120:133-137.

16. Garcia-Gonzalez X, Cortejoso L, Garcia MI, et al. Variants in CDA and ABCB1 are predictors of capecitabine-related adverse reactions in colorectal cancer. *Oncotarget.* 2015;6(8):6422-6430.
17. Joerger M, Huitema AD, Boot H, et al. Germline TYMS genotype is highly predictive in patients with metastatic gastrointestinal malignancies receiving capecitabine-based chemotherapy. *Cancer Chemother Pharmacol.* 2015;75(4):763-772.
18. Loganayagam A, Arenas HM, Corrigan A, et al. Pharmacogenetic variants in the *DPYD*, TYMS, CDA and MTHFR genes are clinically significant predictors of fluoropyrimidine toxicity. *Br J Cancer.* 2013;108(12):2505-2515.
19. Roberto M, Romiti A, Botticelli A, et al. Evaluation of 5-fluorouracil degradation rate and Pharmacogenetic profiling to predict toxicity following adjuvant Capecitabine. *Eur J Clin Pharmacol.* 2017;73(2):157-164.
20. Thomas F, Motsinger-Reif AA, Hoskins JM, et al. Methylenetetrahydrofolate reductase genetic polymorphisms and toxicity to 5-FU-based chemoradiation in rectal cancer. *Br J Cancer.* 2011;105(11):1654-1662.
21. Jennings BA, Loke YK, Skinner J, et al. Evaluating predictive pharmacogenetic signatures of adverse events in colorectal cancer patients treated with fluoropyrimidines. *PloS one.* 2013;8(10):e78053.
22. O'Donnell PH, Stark AL, Gamazon ER, et al. Identification of novel germline polymorphisms governing capecitabine sensitivity. *Cancer.* 2012;118(16):4063-4073.
23. van Huis-Tanja LH, Ewing E, van der Straaten RJ, et al. Clinical validation study of genetic markers for capecitabine efficacy in metastatic colorectal cancer patients. *Pharmacogenet Genomics.* 2015;25(6):279-288.
24. Fernandez-Rozadilla C, Cazier JB, Moreno V, et al. Pharmacogenomics in colorectal cancer: a genome-wide association study to predict toxicity after 5-fluorouracil or FOLFOX administration. *Pharmacogenomics J.* 2013;13(3):209-217.
25. Low SK, Chung S, Takahashi A, et al. Genome-wide association study of chemotherapeutic agent-induced severe neutropenia/leucopenia for patients in Biobank Japan. *Cancer Sci.* 2013;104(8):1074-1082.
26. NCI. National Cancer Institute: Common Terminology Criteria for Adverse Events v4.03. https://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_8.5x11.pdf, 5 May 2017.
27. Illumina I. Infinium® Global Screening Array-24 v1.0. 2017; http://glimdna.org/assets/2017-infinium-global-screening-array-illumina-data-sheet.pdf. Accessed 15 Nov 2018.
28. R. Core Team. R: A Language and Environment for Statistical Computing. 2018; https://www.R-project.org.
29. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559-575.
30. S P. 2018; http://pngu.mgh.harvard.edu/purcell/plink/.
31. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet.* 2007;39(7):906-913.
32. NCBI. National Center for Biotechnology Information. ClinVar database https://www.ncbi.nlm.nih.gov/clinvar/. Accessed 08 January 2019.
33. NCBI. National Center for Biotechnology Information. SNP database (dbSNP). https://www.ncbi.nlm.nih.gov/snp/. Accessed 08 January 2019.

# GENERAL DISCUSSION, SUMMARIES AND APPENDIX