A candidate gene study of capecitabine-related toxicity in colorectal cancer identifies new toxicity variants at \textit{DPYD} and a putative role for \textit{ENOSF1} rather than \textit{TYMS}

Dan Rosmarin,1,2 Claire Palles,1 Alistair Pagnamenta,3 Kulvinder Kaur,3 Guillermo Pita,4 Miguel Martin,5 Enric Domingo,1,3 Angela Jones,1 Kimberley Howarth,1 Luke Freeman-Mills,1 Elaine Johnstone,2 Hai Tao Wang,2 Sharon Love,6 Claire Scudder,7 Patrick Julier,7 Ceres Fernández-Rozadilla,1 Clara Ruiz-Ponte,8 Angel Carracedo,8 Sergi Castellvi-Bel,9 Antoni Castells,10 Anna Gonzalez-Neira,4 Jenny Taylor,3 Rachel Kerr,2 David Kerr,2 Ian Tomlinson1,3

\textbf{ABSTRACT}

Objective Capcitabine is an oral 5-fluorouracil (5-FU) pro-drug commonly used to treat colorectal carcinoma and other tumours. About 35\% of patients experience dose-limiting toxicity. The few proven genetic biomarkers of 5-FU toxicity are rare variants and polymorphisms, respectively, at candidate loci dihydropyrimidine dehydrogenase (\textit{DPYD}) and thymidylate synthase (\textit{TYMS}).

Design We investigated 1456 polymorphisms and rare coding variants near 25 candidate 5-FU pathway genes in 968 UK patients from the QUASAR2 clinical trial.

Results We identified the first common \textit{DPYD} polymorphisms to be consistently associated with capcitabine toxicity, rs12132152 (toxicity allele frequency (TAF)=0.031, OR=3.83, \textit{p}=4.31\times10^{-7}) and rs12022243 (TAF=0.196, OR=1.69, \textit{p}=2.55\times10^{-5}). rs12132152 was particularly strongly associated with hand-foot syndrome (OR=6.1, \textit{p}=3.6\times10^{-6}). The rs12132152 and rs12022243 associations were independent of each other and of previously reported \textit{DPYD} toxicity variants. Next-generation sequencing additionally identified rare \textit{DPYD} variant p.Ala551Thr in one patient with severe toxicity. Using functional predictions and published data, we assigned p.Ala551Thr as causal for toxicity. We found that polymorphism rs2612091, which lies within an intron of \textit{ENOSF1}, was also associated with capcitabine toxicity (TAF=0.532, OR=1.59, \textit{p}=5.28\times10^{-5}). \textit{ENOSF1} is adjacent to \textit{TYMS} and there is a poorly characterised regulatory interaction between the two genes/proteins. Unexpectedly, rs2612091 fully explained the previously reported associations between capcitabine toxicity and the supposedly functional \textit{TYMS} variants, 5\’\textit{VNTR} 2R/3R and 3\’UTR 6 bp ins-del. rs2612091 genotypes were, moreover, consistently associated with \textit{ENOSF1} mRNA levels, but not with \textit{TYMS} expression.

Conclusions \textit{DPYD} harbours rare and common capcitabine toxicity variants. The toxicity polymorphism in the \textit{TYMS} region may actually act through \textit{ENOSF1}.

INTRODUCTION

Capecitabine (Xeloda, Roche) is an oral 5-fluorouracil (5-FU) pro-drug commonly given to patients with colorectal cancer (CRC). Capecitabine is activated to 5-FU, which then causes cytotoxicity by inhibiting production of thymidine and by being converted to metabolites that are incorporated into DNA and RNA.1 As with other 5-FU-based chemotherapy...
Significance of this study

How might it impact on clinical practice in the foreseeable future?

- The genetic architecture of 5-FU/capecitabine toxicity is complex and encompasses rare and common variants.
- Panels of markers in tests used to predict clinically actionable 5-FU/capecitabine toxicity should be updated to include the new DPYD and ENOSF1 alleles, while omitting the TYMS 5 VNTR and 3' UTR polymorphisms.
- Rare loss-of-function DPYD alleles remain the only genetic variants proven to have large positive predictive values for 5-FU/capecitabine toxicity, and these variants account for the majority of patients with life-threatening myelosuppression induced by capecitabine.

regimens, approximately one-third of capecitabine patients suffer dose-limiting levels of drug-induced adverse events. The rapid onset of toxicity results in mortality for 0.5–2% of patients in monotherapy and combination regimens of infusional and bolus 5-FU, and about half that number for capecitabine schedules.

The most common dose-limiting capecitabine toxicities are hand-foot syndrome (HFS) and diarrhoea. Additionally, an important proportion of patients develop neutropaenia and thrombocytopaenia, and others experience nausea, vomiting, mucositis and stomatitis. Some interpatient differences in toxicity can be explained by clinical factors, such as age, gender, local clinical practice and, possibly, diet. However, much variability in toxicity remains unexplained.

The biochemical pathway of capecitabine activation and subsequent 5-FU action and degradation is well established and provides 25 candidate genes in which variation might affect 5-FU toxicity (figure 1, and see online supplementary table 1). Upon absorption in the gut, capecitabine is partially converted to 5-FU in the liver, then preferentially converted to 5-FU at the CRC site. Much 5-FU is degraded in the liver by dihydropyrimidine dehydrogenase (DPYD) prior to activation. As part of the drug’s rationally designed activation, 5-FU is further activated in the tumour to cytotoxic compounds that inhibit DNA synthesis by competing with nucleotide precursors for binding with thymidylate synthase (TYMS). Various sources of toxicity may exist, including alternative activation pathways outside the tumour that result in direct DNA/RNA damage through incorporation, undesired transport of activated compounds, variable expression of drug targets, and reduced levels of drug degradation.

Over a decade of publications exists regarding inherited genetic biomarkers of 5-FU-related toxicity, but only a handful of polymorphisms and rare genetic variants associated with toxicity have been identified with high confidence. These include two common polymorphisms in TYMS (5 VNTR 2R/3R and 3' UTR 6 bp ins-del) and rare functionally deleterious DPYD variants, chiefly *2A and 2846T>A. In aggregate, these variants are potentially useful, but suboptimal, for the prediction of toxicity in clinical practice. Furthermore, there is only limited evidence that genetic variants are generalisable as predictors of toxicity across 5-FU regimens.

In this study, we have investigated the 25 candidate genes that act in the capecitabine/5-FU pathway for new common and rare genetic variants that are associated with capecitabine toxicity. Our patient set comprises about 1000 individuals from the QUASAR2 trial of capecitabine±bevacizumab (Avastin, Genentech/Roche). Our results show that comprehensive genetic studies in sufficiently large, homogeneous datasets can identify additional toxicity predisposition variants in known candidate genes.

METHODS

Patients

The QUASAR2 study (http://www.octo-oxford.org.uk/alltrials/ infollowup/q2.html; http://www.controlled-trials.com/ISRCTN 45133151/) is a phase III randomised controlled trial of adjuvant capecitabine (Xeloda) (1250 mg/m² twice daily for 14 days every 3 weeks, total of 8 cycles) ± bevacizumab (Avastin) (7.5 mg/kg every 3 weeks) following resection of stage II/III CRC. Patients were entered into the study between July 2005 and December 2011 at 123 UK and 81 non-UK sites. Survival analyses are scheduled for 2014. Of 1119 patients with blood collected as of July 2010, 1046 were selected for genetic study based on availability of clinical data and informed consent. Patient characteristics are shown in table 1. All work was performed with full UK Ethics Committee approval, according to the tenets of the Declaration of Helsinki.

Toxicity phenotypes

Toxicity phenotype data were collected as part of QUASAR2 according to the NCI Common Toxicity Criteria for Adverse Events (CTCAE) V3.0. Maximum toxicity (0–4) at any treatment cycle was derived for each of the following individual FU-related toxicities: diarrhoea, nausea and vomiting, mucositis/stomatitis, neutropaenia, thrombocytopaenia and HFS. We then derived a global measure of toxicity, defined as the maximum individual toxicity score measured for each patient. Global and individual toxicities were analysed using two approaches: (1) a binary classification into low toxicity (grade 0–2) versus high (dose-limiting) toxicity (grade 3–4) and (2) a quantitative measure of toxicity. In the latter, if <100 patients experienced a particular grade of toxicity, we combined them into a single bin with an adjacent grade. Specifically, we analysed grades 0–1 v 2 v 3–4 for global toxicity, diarrhoea and HFS and 0v1234 for the other, rarer toxicity phenotypes. Toxicity data by grade are shown in online supplementary table 2.

Genotype data

Exclusion of patients on the basis of availability of DNA and completeness of toxicity data, non-Caucasian ethnicity or genotyping quality control (see online supplementary methods) meant that data were available from 940 patients genotyped using Illumina tagging SNP arrays and a largely overlapping set of 968 QUASAR2 patients genotyped on Illumina HumanExome 12v1_A or -12v1-1_A arrays, which were designed to capture uncommon protein-coding variation. For each of the 25 capecitabine/5-FU pathway genes (see online supplementary table 1), we identified genetic variants that were present on the arrays and that lay within 25 kb either side of the coding region of the longest isoform of each gene. We used imputation to obtain missing genotypes arising from differences in array content. Further genotyping was performed using previously described methods.

For loci at which significant or borderline significant associations between genetic variants and toxicity were detected, we performed fine mapping studies by using the methods above to impute all SNPs in a 1.5 Mb flanking region, in order to refine the association signal.
Figure 1  Capcitabine activation and subsequent 5-FU activation, action, transport and catabolism. Capcitabine is an oral 5-FU prodrug that is rationally designed so that concentrations of the cytotoxic metabolites FdUMP, FdUTP and FUTP are higher within malignant cells than within normal cells. After absorption in the gut, most of the drug activation occurs via the common pro-drug activation route (a), starting in the liver and finishing in the tumour. Additionally, 5-FU can be converted to its active compounds via alternate activation routes (b) in colon tumour cells and cells from multiple other tissues. Toxicity may occur if non-target tissue is exposed to activated capcitabine/5-FU (i.e., FdUMP, FdUTP and FUTP), either when 5-FU is activated prior to arrival at the tumour or when it exits target tissue and is subsequently activated. 5-FU released in circulation may be quickly metabolised (c) in the liver by enzymes including DPYD. Gene names for candidate enzymes, provided in circles, are defined in online supplementary table S1. Drug catabolites are provided in rectangles. Primary (rationally designed) pathway are shown in solid lines, and alternative pathways shown in dashed lines.

Statistical analysis
For each of the 1456 SNP and exome array variants, our primary analysis was to test associations between global (any 5-FU-related) dose-limiting (grade 012v34) toxicity and genotype. Frequentist tests under a missing data linear or logistic regression model were implemented using SNPTESTv2. Samples were stratified by QUASAR2 treatment with age and gender as covariates. Meta-analysis of the two arms of QUASAR2 was performed using GWAMA, including tests of interarm heterogeneity. A stringent Bonferroni-corrected p value threshold of 3.43×10⁻⁵ (=0.05/1456) was used to indicate a significant association for the primary analysis of (binary) global dose-limiting toxicity.

For selected SNPs with association signals that reached or surpassed formal significance, we imputed additional SNPs within 1.5 Mb flanking regions (see online supplementary methods). We performed association tests for global and specific toxicities using the global grade 012v34 measure. Since the genotyped and imputed SNPs were non-independent, we declared associations with imputed SNPs significant using the same threshold of p=3.43×10⁻⁵. For any region within which one or more SNPs achieved significant associations with global toxicity, the underlying individual toxicities were investigated at the most strongly associated SNPs using quantitative measures and clinically actionable cut-offs for dose delay or reduction in QUASAR2 (generally, grade 012v34, except for grade 01v234 for diarrhoea).

To test for independent effects of variants within a region, we used logistic regression analysis in R, with age, sex and study arm as covariates. The best-fitting model was determined as that which minimised the Akaike information criterion (AIC) subject to a variant showing an association at p=0.05. Haplotype analyses at DPYD and TYMS/ENOSF1 were performed using the --hap-logistic and --independent-effect commands in PLINK. Tests to examine multiple genetic variants were performed in PLINK. Receiver operator characteristic (ROC) analysis was performed in Stata using a binary classification of patients to either grade 0/1/2 or grade 3/4 global toxicity, using a genetic score given for each individual by $\sum_{i} \beta_i$ (equation 2), where $\beta_i$ is the beta coefficient of the ith SNP significantly associated with global toxicity in a logistic regression model, and Ni is the number of harmful alleles carried by that individual at that locus.

Sequencing and validation of novel variants identified
Amplificon sequencing of the coding regions of DPYD and TYMS was performed by Roche/454 Titanium GS FLX technology in 100 patients with the highest levels of 5-FU-related toxicity (‘HiTox’), specifically grade 3 or grade 4 diarrhoea in the first 4 cycles of treatment and or other grade 3/4 toxicities in the first 4 cycles of treatment. We also sequenced the same amplicons in 100 patients with no adverse toxicity events during the entire duration of treatment (‘LoTox’). The missense DPYD variant p.Ala551Thr (A551T) was identified in the HiTox pool. The DNA for each individual comprising the pool was Sanger-sequenced to identify those carrying this variant. Only one heterozygous individual was found. For analysis of the whole sample set, we designed KASPar (http://cshprotocols.cshlp.org/content/2007/9/pdb.prot4841.abstract) allele-specific single-nucleotide variant primers to detect A551T and included three duplicates of the known variant sample in each run to facilitate genotype clustering (details available upon request). All samples
that did not cluster with the A allele homozygotes were subsequently examined by bidirectional Sanger sequencing.

RESULTS

A total of 1046 capecitabine-treated patients from the QUASAR2 trial were initially selected. Grade 3+ global toxicity was observed in 34% of these patients, with severe diarrhoea in 10% and HFS in 24% (see online supplementary table 2). Severe toxicity was less common for the other phenotypes, although five patients experienced grade 4 neutropaenia. After exclusions (see online supplementary methods), genotyping data were available for 940 patients genotyped on genome-wide tagSNPs arrays and 968 patients genotyped on the Exome array. Across the two types of arrays, we identified 1456 genetic variants that mapped to within 25 kb of the 25 capecitabine/5-FU candidate genes (see online supplementary table 1). In the primary analysis, we searched for associations between each genetic variant and global (grade 012v34) capecitabine toxicity. Rather than regarding the two arms of QUASAR2 as discovery and validation sets, with consequent reduction in statistical power, we performed a meta-analysis of data from the 439 patients in arm A (capecitabine) and the 501 patients in arm B (capecitabine+bevacizumab), and assessed evidence for between-arm heterogeneity. The results of this analysis on a per variants that did not cluster with the A allele homozygotes were subsequently examined by bidirectional Sanger sequencing.

New associations of common DPYD variants with capecitabine toxicity

We found that the A-allele (freq.=0.03) of SNP rs12132152 was associated with global capecitabine toxicity (ORglobalbinary=3.83, p=4.31×10⁻⁶; table 2). rs12132152 is an intragenic SNP 22 kb downstream of DPYD (chr1:97523004, b37). Upon imputation of variants in the region flanking this tag SNP (see online supplementary methods), we identified a few SNPs with marginally more significant associations, notably rs76387818 (chr1:97539400; ORbinary=4.05, p=2.11×10⁻⁶, r²tag=0.98; table 2; figure 2A). Results were similar when the quantitative measure of global toxicity was used (table 2). We investigated the individual phenotypes comprising the global toxicity measure. rs12132152 and rs76387818 were principally associated with HFS, under quantitative and binary models (for rs76387818, ORhfsquant=1.78, p=5.51×10⁻⁶, ORhfsbinary=6.44, p=1.75×10⁻⁶), albeit with some weaker evidence of associations with diarrhoea (table 2).

We then explored in silico possible functional mechanisms underlying the rs12132152/rs76387818 association. We examined ENCODE data (http://genome.ucsc.edu/ENCODE) for the region containing rs12132152 and seven strongly correlated SNPs (figure 2A: approximately chr1:97475000-97562000, b37). FAIRE and histone K4 methylation data suggested that this is a region of open chromatin and one correlated SNP in particular, rs12123160, lies at a methylated CpG. Although no suitable data from normal liver were available, we found that genotypes at rs12132152 and correlated SNPs were not associated with DPYD expression in adipose tissue, lymphoblastoid cells or skin (Genevar database, p>0.13) or in colon tissue (TCGA, p=0.72).

The second DPYD toxicity-associated variant (table 2) was identified following SNP imputation in the region of 1.5 Mb surrounding rs7548189, a tagSNP intronic to DPYD (chr1:97867713, b37). rs7548189 was borderline associated with the binary measure global toxicity (ORglobalbinary=1.67, p=3.79×10⁻⁵). Further investigation showed rs7548189 formally to be associated with the quantitative measure of global toxicity, and with diarrhoea under binary and quantitative models (ORglobalquant=1.23, p=6.82×10⁻⁶; ORdiarrhoeaquant=1.18, p=1.54×10⁻⁵; ORdiarrhoeabinary=1.76, p=1.72×10⁻⁵). As can be seen from table 2, HFS also contributed to the association with global toxicity. Following regional imputation, rs12022243 (r² with rs7548189=0.95) was found formally to be associated with global toxicity under a binary model (ORglobalbinary=1.69, p=2.55×10⁻⁵). We found that rs12022243 showed excellent imputation quality: of 190 independent individuals, only 4 (2%) genotypes were missing, and all the remaining genotypes were imputed correctly (see online supplementary methods).

We found little evidence from ENCODE data that rs7548189 is functional, but rs12022243 falls in a region of open chromatin that may have enhancer activity. rs7548189 was not associated with DPYD expression levels in lymphoblasts, fibroblasts, T-cells, adipose tissue, or skin on the Genevar database (p>0.07) or in colon tissue from TCGA (p=0.97, rs12022243 Was absent from these datasets).

Using logistic regression analysis, we found that the rs12132152 and rs7548189/rs12022243 signals were independent of each other and of the known DPYD toxicity variants rs5918290 and 2846T>A (rs67373796) (details not shown; linkage disequilibrium data in online supplementary figure 1). Further analysis of haplotypes based on 81 tagSNPs in
Table 2

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP</th>
<th>Minor allele</th>
<th>Genetyped</th>
<th>Imputed</th>
<th>OR (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPYD</td>
<td>rs12132132</td>
<td>A/G</td>
<td>0.031</td>
<td>0.993</td>
<td>4.31×10^-4</td>
<td>0.0096</td>
</tr>
<tr>
<td></td>
<td>rs7548189</td>
<td>A/G</td>
<td>0.031</td>
<td>0.993</td>
<td>4.31×10^-4</td>
<td>0.0096</td>
</tr>
<tr>
<td></td>
<td>rs778713</td>
<td>A/G</td>
<td>0.031</td>
<td>0.993</td>
<td>4.31×10^-4</td>
<td>0.0096</td>
</tr>
<tr>
<td></td>
<td>rs12023483</td>
<td>A/G</td>
<td>0.031</td>
<td>0.993</td>
<td>4.31×10^-4</td>
<td>0.0096</td>
</tr>
<tr>
<td></td>
<td>rs18637888</td>
<td>A/G</td>
<td>0.031</td>
<td>0.993</td>
<td>4.31×10^-4</td>
<td>0.0096</td>
</tr>
<tr>
<td></td>
<td>rs19735604</td>
<td>A/G</td>
<td>0.031</td>
<td>0.993</td>
<td>4.31×10^-4</td>
<td>0.0096</td>
</tr>
<tr>
<td></td>
<td>rs18883607</td>
<td>A/G</td>
<td>0.031</td>
<td>0.993</td>
<td>4.31×10^-4</td>
<td>0.0096</td>
</tr>
<tr>
<td>ENOSF1</td>
<td>rs830790</td>
<td>T/C</td>
<td>0.532</td>
<td>0.990</td>
<td>1.91×10^-3</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>rs18883607</td>
<td>T/C</td>
<td>0.532</td>
<td>0.990</td>
<td>1.91×10^-3</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>rs18883607</td>
<td>T/C</td>
<td>0.532</td>
<td>0.990</td>
<td>1.91×10^-3</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>rs18883607</td>
<td>T/C</td>
<td>0.532</td>
<td>0.990</td>
<td>1.91×10^-3</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>rs18883607</td>
<td>T/C</td>
<td>0.532</td>
<td>0.990</td>
<td>1.91×10^-3</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>rs18883607</td>
<td>T/C</td>
<td>0.532</td>
<td>0.990</td>
<td>1.91×10^-3</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>rs18883607</td>
<td>T/C</td>
<td>0.532</td>
<td>0.990</td>
<td>1.91×10^-3</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>rs18883607</td>
<td>T/C</td>
<td>0.532</td>
<td>0.990</td>
<td>1.91×10^-3</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>rs18883607</td>
<td>T/C</td>
<td>0.532</td>
<td>0.990</td>
<td>1.91×10^-3</td>
<td>0.061</td>
</tr>
</tbody>
</table>

The table shows the results of the meta-analysis of global and selected individual toxicities, measured as binary or continuous. ORs are expressed relative to this. Imputation quality Info Score is also shown, as are numbers of samples genotyped and directly genotyped. The final column shows results for toxicity phenotype classifications that could lead to treatment change or delay according to the QUASAR2 protocol. There was no evidence of heterogeneity between QUASAR2 arms in any of these analyses (P=0.1–0.75).
a 25 kb window either side of DPYD provided no further refinement of the rs12132152 and rs7548189/rs12022243 associations and showed no evidence of additional, independent association signals (details not shown).

Resolving the toxicity associations at the TYMS and ENOSF1 loci

We found that the G-allele (freq.=0.45) of SNP rs2612091 was associated with increased global toxicity (OR_{globalbinary}=1.59, p=5.28×10^{-6}; OR_{globalquant}=1.19, p=2.35×10^{-6}; table 2). rs2612091 lies 10 kb downstream of TYMS within an intron of enolase superfamily member 1 (ENOSF1, chr18:683 607). Fine mapping showed somewhat more significant associations for a SNP, rs2741171 (chr18:700 687), in linkage disequilibrium (r^2=0.73) with rs2612091, particularly for the quantitative measure of global toxicity and HFS (OR_{globalquant}=1.2, p=9.24×10^{-7}). The rs2741171 variant is further downstream of TYMS (27 kb) and again is intronic for ENOSF1, but both

Figure 2 Regional plots of DPYD and TYMS SNPs for associations with capecitabine-related toxicity. Regional associations between polymorphisms and (left) global grade 01v34 capecitabine toxicity or (right) selected specific toxicities were plotted using LocusZoom and are shown for SNPs 500 kb either side of (a) rs12132152 which maps 3' of DPYD, (b) rs12022243, mapping to an intron of DPYD and (c) rs2612091, which maps to an intron of ENOSF1. The x-axis shows chromosome position, while the y-axis shows the log_{10}P for the association. Circles represent genotyped SNPs included in the 1456-SNP test panel, with the most significantly associated SNP in purple. Squares represent imputed SNPs. Linkage disequilibrium (calculated using hg19/1000 genomes, March 2012 EUR population) with the most significantly associated SNP is shown by colour as indicated.

Colon

SNPs fall between recombination hotspots that flank the entirety of TYMS and ENOSF1 (figure 2B). The rs2612091/rs2741171 effect on toxicity was essentially driven by HFS, with an especially strong association being observed using the quantitative measure (rs2612091: OR_{haplant}=1.24, p=3.67×10^{-7}; rs2741171: OR_{haplant}=1.23, p=3.10×10^{-8}).

ENOSF1 is a largely uncharacterised gene that appears to encode a protein and RNAs antisense to TYMS. It has been proposed that ENOSF1 regulates TYMS mRNA and/or protein expression.\textsuperscript{21} We therefore analysed associations between rs2612091 genotype and TYMS and ENOSF1 expression using the Genevar and TCGA databases. Using Caucasian-matched twin data,\textsuperscript{26} the rs2612091 G-allele significantly decreased ENOSF1 expression in adipose tissue for both twin sets (p_{set1}=0.89, p_{set2}=0.0012). However, rs2612091 genotype was not associated with TYMS expression (p>0.30 for each of the same analyses).

We sequenced the coding regions of DPYD and TYMS in pools of Hi-Tox and Lo-Tox patients (see online supplementary figures S2 and S3). In the Hi-Tox pool, we identified a single missense variant that was not present on SNP and exome arrays (DPYD c. G1651A; p.Ala551Thr; chr1:97 981 371). We found no other occurrence of this variant in the full set of 968 patients. A551 T was predicted to be strongly damaging by SIFT, PolyPhen, PhyloP and MutationTaster, and the single patient with this allele had experienced grade 4 neutropaenia and thrombocytopenia. Database searches determined that this variant has been previously reported as causal for DPYD Deficiency Syndrome (OMIM 612779).\textsuperscript{28} We confirmed that A551 T was not in linkage disequilibrium with any of the other common or rare DPYD toxicity variants.

We then considered only the 19 patients with extreme (grade 4) toxicity at any cycle and determined which alleles they carried at the three toxicity SNPs and their complement of rare DPYD alleles from the literature, including 2846 A>T and “A2” that were shown to be associated with 5-FU toxicity in our previous meta-analysis\textsuperscript{7} (table 3). There was no good evidence that the risk alleles at the three new toxicity SNPs were overrepresented as a group in these 19 patients (table 3). Using the evaluation of Caudle et al\textsuperscript{8} as a guide, supplemented by data from this study, we then assessed the likely contributions of each rare DPYD variant to extreme toxicity. There was insufficient prior evidence\textsuperscript{8} to regard six DPYD alleles (*4, *5, *6, *9A, M166 V and K259E) as pathogenic (see online supplementary table S6).

DISCUSSION
Through the analysis of capcitabine-treated patients from the QUASAR2 trial for SNPs and rare genetic variants in 25 capcitabine/5-FU pathway genes and subsequent exon sequencing of DPYD and TYMS, we have identified new genetic predictors of capcitabine-induced toxicity and clarified the origins of previously reported signals. Two of the new toxicity variants (rs12022243 and rs12132152) map to DPYD and are independent of previously reported DPYD toxicity alleles. The toxicity-associated allele at rs12022243 is common (freq.=0.22)
Table 3  Genotypes of QUASAR2 individuals with grade 4 toxicity at selected DPYD variants

<table>
<thead>
<tr>
<th>Case</th>
<th>D</th>
<th>V</th>
<th>H</th>
<th>N</th>
<th>P</th>
<th>M</th>
<th>S</th>
<th>Possible explanation</th>
<th>2846</th>
<th>*2A</th>
<th>A551T</th>
<th>*13</th>
<th>*4</th>
<th>*5</th>
<th>*9A</th>
<th>M166V</th>
<th>K259E</th>
<th>rs12132152</th>
<th>rs12022243</th>
<th>rs2612091</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2846T&gt;A</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>2846T&gt;A</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>A551T</td>
<td></td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>*2A</td>
<td></td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>*13</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

The variants shown are (1) those identified by this study (rs12132152, rs12022243, rs2612091, DPYD A551 T), (2) DPYD alleles (2846 A>T and *2A) shown to be associated with 5-FU toxicity in the meta-analysis of Rosmarin et al and (3) potential DPYD toxicity alleles from Caudle et al. Genotypes shown are major allele homozygote (0), heterozygote (1) and minor or variant allele homozygote (2). Blank cells denote missing data. The allele that provides a plausible explanation for the severe toxicity is shown. Note that *4 and *5 DPYD alleles are in complete linkage disequilibrium (D’=1.0) with 2A or 2846T>A. Overall association statistics and genotype frequencies in cases and controls are shown for the analysis of binary global toxicity in online supplementary table S3.

For the toxicities, D=diarrhoea; H=HFS; M=mucositis; N=neutropaenia; P=thrombocytpaenia; S=stomatitis; V=vomiting.

The variants shown are (1) those identified by this study (rs12132152, rs12022243, rs2612091, DPYD A551 T), (2) DPYD alleles (2846 A>T and *2A) shown to be associated with 5-FU toxicity in the meta-analysis of Rosmarin et al and (3) potential DPYD toxicity alleles from Caudle et al. Genotypes shown are major allele homozygote (0), heterozygote (1) and minor or variant allele homozygote (2). Blank cells denote missing data. The allele that provides a plausible explanation for the severe toxicity is shown. Note that *4 and *5 DPYD alleles are in complete linkage disequilibrium (D’=1.0) with 2A or 2846T>A. Overall association statistics and genotype frequencies in cases and controls are shown for the analysis of binary global toxicity in online supplementary table S3.

For the toxicities, D=diarrhoea; H=HFS; M=mucositis; N=neutropaenia; P=thrombocytpaenia; S=stomatitis; V=vomiting.
and has a moderate effect (OR$_{\text{global binary}}$ = 1.8) on toxicity risk, influencing HFS and diarrhoea. The rs12132152 toxicity allele is less common, but not rare (freq.=0.03), and its effect is greater (OR$_{\text{global binary}}$ = 3.8). Although associations between other common DPYD polymorphisms and 5-FU toxicity have previously been proposed, we failed to confirm these in this analysis and in a previous meta-analysis of multiple datasets.

Through sequencing patients with high capecitabine toxicity, we identified a third new DPYD toxicity allele, p.Ala551Thr. Like the other rare DPYD variants (*2A and 2846T>A) that have established associations with 5-FU-related toxicity in the heterozygous state, A551T has been shown to cause the recessive DPYD deficiency syndrome when presented as the homozygote or compound heterozygote with another mutant allele. The identification of A551T in a patient experiencing grade 4 neutropenia and thrombocytopenia adds further support to the view that all rare DPYD variants that cause DPYD Deficiency Syndrome greatly increase the risk of 5-FU toxicity in heterozygotes. Furthermore, of the three other QUASAR2 patients who experienced grade 4 myelotoxicity—two of whom were the only toxicity-induced deaths in QUASAR2—one carried DPYD *2A, one 2846A>C and one *13. Of the 12 carriers of rare, functionally deleterious DPYD alleles who did not develop severe toxicity, seven suffered grade 3 toxicity and may therefore have been spared severe toxicity by capecitabine dose reduction. However, the reason for the remaining five individuals avoiding clinically important toxicity is unclear. Much interpatient difference in 5-FU toxicity remains unexplained, and it remains very plausible that some of this variability is heritable.

Our finding of another capecitabine toxicity SNP, rs2612091, produced some very unexpected results. This common variant is downstream of TYMS and intronic to ENOSF1. Our detailed analysis showed that the rs2612091 signal appears to explain downstream of the previously reported associations between 5-FU/capecitabine toxicity and TYMS polymorphisms (5′VNTR 2R/3R and 3′UTR 6 bp ins-del). This is especially surprising given published evidence that the two TYMS polymorphisms directly affect TYMS mRNA expression and protein levels. However, recent critical assessments have shown that the data linking TYMS expression to 5′VNTR 2R/3R and 3′UTR 6 bp ins-del genotypes are actually very mixed. In fact, our analysis of public mRNA expression data demonstrated rs2612091 to be associated with ENOSF1 expression and not with TYMS expression. Our data imply that the TYMS 5′VNTR and 3′UTR toxicity association signals result from LD between these polymorphisms and tagSNP rs2612091 or, more likely, fine-map SNP rs2741171 (see online supplementary figure 1A). We therefore conclude that ENOSF1 is most likely to be the target of the functional variation tagged by rs2612091. ENOSF1 and TYMS transcripts are overlapping, but toxicity SNP does not appear to act through antisense-mediated down-regulation of TYMS mRNA. However, ENOSF1 protein has been proposed as an influence on TYMS activity, and this remains a plausible mechanism of toxicity.

We found no evidence of heterogeneity between the two arms of QUASAR2 in terms of the effects of genetic variants on global toxicity or any specific toxicity. The toxicity profiles of capecitabine and bevacizumab principally overlap through the risk of HFS, although the effects of the former greatly outweigh those of the latter (estimated 18% vs 7% for QUASAR2). In principle, a polymorphism could predispose to HFS caused by either capecitabine and/or bevacizumab. However, such a variant is most likely to act at the level of the HFS target tissue, whereas, our variants were specifically chosen for potential effects on 5-FU metabolism. The most likely effect of the use of bevacizumab on our study was, therefore, a small decrease in statistical power owing to a higher ‘background’ level of HFS resulting from a non-capecitabine source in arm B.

In summary, we have identified four new variants associated with capecitabine toxicity. Further work is desirable in order to confirm and quantitate these associations in additional datasets, and to understand the mechanistic origins of capecitabine-related toxicity, especially for the TYMS and ENOSF1 loci. While not yet an ideal test for clinical use (see online supplementary figure 4), genetic testing can be used to highlight patients at increased risk of capecitabine toxicity. A two-tier test may be justifiable, comprising (1) a sensitive test for severe, life-threatening toxicity based principally on rare DPYD variants and (2) a test additionally incorporating SNPs to highlight the risk of clinically actionable toxicity. It is a moot point as to whether such a strategy would currently be cost effective, and it is not yet known how well our panel of variants predicts toxicity from 5-FU delivered by other means or in combination regimens. Although our dataset was relatively large, and power was good to detect toxicity variants with relatively large effects (typically >80% power for a common variant conferring an OR >1.5), power was very limited to detect variants with lower allele frequency and/or smaller effect size (see online supplementary table S3). Further efforts to identify additional polymorphisms and rare variants associated with 5-FU toxicity remain valid.

URLS

Primer3: http://sourceforge.net/projects/primer3/
LocusZoom: http://csg.sph.umich.edu/locuszoom/
SNPTEST2: https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html
IMPUTE2: http://mathgen.stats.ox.ac.uk/impute/impute_v2.html
GWAMA: http://www.well.ox.ac.uk/gwama/download.shtml
TCGA: http://tcga-data.nci.nih.gov/

Author affiliations

1Molecular and Population Genetics Laboratory, Oxford, UK
2Department of Oncology, University of Oxford, Old Road Campus Research Building, Oxford, UK
3Oxford NIHR Comprehensive Biomedical Research Centre, Wellcome Trust Centre for Human Genetics, Oxford, UK
4Human Genotyping Unit-CeGen, Human Cancer Genetics Programme, Spanish National Cancer Research Centre, Melchor Fernández Almagro 3, Madrid, Spain
5Department of Medical Oncology, Instituto de Investigacion Sanitaria Hospital General Universitario Gregorio Marañon, Universidad Complutense, Madrid, Spain
6Centre for Statistics in Medicine, University of Oxford, Botnar Research Centre, Oxford, UK
7OCTO, University of Oxford, Old Road Campus Research Building, Oxford, UK
8Galician Public Foundation of Genomic Medicine (FPGMX), CIBERer, Genomics Medicine Group, Hospital Clinico, University of Santiago de Compostela, Santiago de Compostela, Galicia, Spain
9Genetic Susceptibility to Colorectal Cancer Group, Gastrointestinal & Pancreatic Oncology Team, IDIBAPS/CIBERehd/Hospital Clinic, Centre Esther Koplowitz (CEK), Barcelona, Spain
10Institute of Digestive and Metabolic Diseases, Hospital Clinic, Barcelona, Spain
11Nuffield Department of Clinical Laboratory Sciences, University of Oxford, Oxford, UK

Collaborators QUASAR2 collaborators, The EPICOLON Consortium—Listed in online supplementary data.

Contributors IT, CP, AGN, GP, MM, CFR, CRP, AC, SCB, AC and DNR planned and designed the study. DNR, CP and IT wrote the manuscript. CP, AJ, AP, KK, KH, ED, EU, HW, and DNR performed sequencing, genotyping and/or prepared DNA from QUASAR2. Oversight of sequencing was provided by JT. SL, CS and PJ provided QUASAR2 clinical data; RM and DJK ran the QUASAR2 study. DNR, CP and IT analysed the QUASAR2 data and, with LFM, the public functional data.
REFERENCES

A candidate gene study of capecitabine-related toxicity in colorectal cancer identifies new toxicity variants at DPYD and a putative role for ENOSF1 rather than TYMS


Gut 2015 64: 111-120 originally published online March 19, 2014
doi: 10.1136/gutjnl-2013-306571