Using whole exome sequencing to identify genetic markers for carboplatin and gemcitabine-induced toxicities.

Henrik Gréen 1,2,3*, Johanna Hasmats 1*, Ilya Kupershmidt 1,4, Daniel Edsgärd 1, Luigi de Petris 5, Rolf Lewensohn 5, Fiona Blackhall 6,7, Svante Vikingsson 3, Benjamin Besse 8, Andrea Lindgren 9,10,11, Eva Brandén 5,12, Hirsh Koyi 5,12, Curt Peterson 3, Joakim Lundeberg 1

1 Science for Life Laboratory, School of Biotechnology, Division of Gene Technology, Royal Institute of Technology, Solna, Sweden
2 Department of Forensic Genetics and Forensic Toxicology, National Board of Forensic Medicine, Linköping, Sweden
3 Division of Drug Research, Clinical Pharmacology, Department of Medical and Health Sciences, Linköping University, Linköping, Sweden
4 NextBio, 3 Results Way, Cupertino, CA, USA
5 Department of Oncology and Pathology, Karolinska Institutet and Oncology Clinic, Karolinska University Hospital, Stockholm Sweden
6 Department of Medical Oncology, Christie Hospital, Manchester, U.K.
7 Institute of Cancer Sciences, University of Manchester, Manchester, U.K.
8 Unité INSERM U 981, Université Paris Sud, Département de Médecine, Institut Gustave Roussy, Villejuif, France
9 Department of Clinical Physiology, University Hospital, Linköping, Sweden
10 Department of Medical and Health Sciences, Linköping University, Linköping, Sweden
11 Department of Pulmonary Medicine, University Hospital, Linköping, Sweden
12 Present affiliation: Department of Respiratory Medicine, Gävle Hospital, Gävle, Sweden

* Henrik Green and Johanna Hasmats contributed equally to this publication.

Acknowledgements: The authors thank Ellen Sherwood, Magnus Bjursell, Henrik Stranneheim, Nathaniel Street, and Mikael Huss for technical assistance during the genotyping and bioinformatics analysis. This work was financially supported by grants from the European Commission (CHEMORES LSHC-CT-2007-037665), the Swedish Cancer Society, the Swedish Research Council, Fondkistan, Stiftelsen Sigurd och Elsa Goljes Minne, and Marcus Borgströms stiftelse. We would like to acknowledge Science for Life Laboratory (SciLifeLab Stockholm), National Genomics Infrastructure (NGI – Sweden), and Uppmax for providing massive parallel sequencing and computational infrastructure. The support by BILS (Bioinformatics Infrastructure for Life Sciences) is gratefully acknowledged.

Short title: Exome sequencing & gemcitabine/carboplatin toxicity

Keywords: Exome sequencing, Gemcitabine, Carboplatin, Thrombocytopenia, Neutropenia, Toxicity

Corresponding author: Henrik Gréen, Clinical Pharmacology, Division of Drug Research, Department of Medical and Health Sciences, Faculty of Health Sciences, Linköpings Universitet, SE-581 85 Linköping, Sweden. E-mail: henrik.green@liu.se Phone: +46-(0)13-252157

Disclaimer: Joakim Lundeberg has a patent/intellectual property together with Spatial Transcriptomics AB. NextBio was involved in funding part of the development of the informatics approach. Otherwise we do not have any conflict of interest.
Abstract

Purpose: Chemotherapies are associated with significant inter-individual variability in therapeutic effect and adverse drug reactions. In lung cancer, the use of gemcitabine and carboplatin induces grade 3 or 4 myelosuppression in about a quarter of the patients, while an equal fraction of patients are basically unaffected in terms of myelosuppressive side effects. We therefore set out to identify genetic markers for gemcitabine/carboplatin-induced myelosuppression.

Experimental design: We exome sequenced 32 patients that suffered extremely high neutropenia and thrombocytopenia (grade 3 or 4 after first chemotherapy cycle) or were virtually unaffected (grade 0 or 1). The genetic differences/polymorphism between the groups were compared using six different bioinformatics strategies: 1) whole exome non-synonymous single nucleotide variants association analysis, 2) deviation from Hardy–Weinberg equilibrium, 3) analysis of genes selected by a priori biological knowledge, 4) analysis of genes selected from gene expression meta-analysis of toxicity datasets, 5) Ingenuity Pathway Analysis, and 6) FunCoup network enrichment analysis.

Results: A total of 53 genetic variants that differed among these groups were validated in an additional 291 patients and were correlated to the patients’ myelosuppression. In the validation, we identified rs1453542 in OR4D6 (p = 0.0008, OR = 5.2, 95% CI = 1.8–18) as a marker for gemcitabine/carboplatin-induced neutropenia and rs5925720 in DDX53 (p = 0.0015, OR = 0.36, 95% CI = 0.17–0.71) as a marker for thrombocytopenia. Patients homozygous for the minor allele of rs1453542 had a higher risk of neutropenia, and for rs5925720 the minor allele was associated with a lower risk for thrombocytopenia.
**Conclusions:** We have identified two new genetic markers with the potential to predict myelosuppression induced by gemcitabine/carboplatin chemotherapy.
Statement of translational relevance

The success of personalized chemotherapy based on genetic information is built on two cornerstones. First, the selection of the most effective drug(s) must take into account mutations in the tumor, and second, selecting the optimal dose and predicting toxicity for that specific drug must be based on the patient’s constitutional genotype. In this study we have focused on the latter – to identify patients at risk of adverse drug reactions – and we set out to identify genetic markers for gemcitabine/carboplatin-induced myelosuppression in lung cancer patients using exome sequencing and bioinformatics. This is the first study that uses exome sequencing to compare the genotypes of patients suffering from severe neutropenia and thrombocytopenia with those of patients who are virtually unaffected by gemcitabine/carboplatin treatment. We have identified two new potential genetic markers for gemcitabine/carboplatin-induced myelosuppression using six different bioinformatics approaches, and these markers were validated in an additional set of patients.
**Introduction**

There is a continuously growing number of identified sequence variations in genes encoding drug targets, drug-metabolizing enzymes, and drug transporters that influence the clinical effects after drug treatment (1-3). Gemcitabine and carboplatin are two chemotherapeutic agents used for the treatment of patients with various solid tumors, including non-small cell lung cancer. Dose-limiting adverse drug reactions (ADRs) are mainly hematological toxicities, including neutropenia, leukopenia, and thrombocytopenia (4). There is also large inter-individual variability when it comes to the severity and frequency of these ADRs. Our data show that 15%–17% of lung cancer patients suffer grade 4 (National Cancer Institute’s Common Toxicity Criteria, CTC scale) neutropenia and/or thrombocytopenia (life threatening toxicity, grade 5 = death) during the first chemotherapy cycle, while 20%–37% are basically unaffected (grade 0) when treated with the same drug regimen. The severity of the disease, with a five-year survival of only 15%, motivates the use of chemotherapy with a risk of life-threatening myelosuppression, but the clinical use of these drugs is limited by their toxicities, and today we lack predictive markers to identify patients at risk for ADRs. Such predictive markers could reduce patient suffering and/or be used to increase dosages for toxicity-tolerant patients.

Several genes influence the pharmacokinetics, disposition, and efficacy of carboplatin and gemcitabine. Gemcitabine is a deoxynucleotide analog that is transported into the cell by SLC28A1 and SLC29A1 and activated intracellularly by deoxycytidine kinase (dCK) to form the monophosphate form of gemcitabine. This monophosphate form is converted to the triphosphate form and it is eventually incorporated into DNA where it inhibits DNA synthesis and subsequent cell division. Gemcitabine is rapidly metabolized by cytidine deaminase (CDA) into inactive metabolites. Carboplatin acts by inducing intra-strand crosslinks and adducts in DNA that cause
conformational changes that inhibit DNA replication. Carboplatin is deactivated by the glutathione-S-transferases, and the DNA damage induced by carboplatin can be repaired by several DNA-repair proteins, e.g. ERCC1 and XPD.

Several candidate gene studies have focused on predicting the myelosuppression after gemcitabine and/or carboplatin chemotherapy based on variants within these genes. However, these studies have been inconclusive and have failed to identify reproducible markers. More recently, Kiyotani et al. published a genome-wide association study to identify single nucleotide variants (SNVs) associated with gemcitabine-induced neutropenia/leukopenia (5). They found an association of myelosuppression with four loci, rs11141915 in DAPK1, rs1901440 at chr2q12, rs12046844 in PDE4B, and rs11719165 at chr3q29. Interestingly, both DAPK1 and PDE4B are expressed in bone marrow and leukocytes and have previously been linked to chemotherapy (6-8).

To make use of a more whole-genome approach, we designed a study to identify genetic variants associated with gemcitabine and carboplatin-induced myelosuppression by exome sequencing of 16 patients with extremely high (CTC grade 3 or 4) and 16 patients with low (CTC grade 0 or 1) neutropenia and thrombocytopenia after a first cycle of chemotherapy. The top candidate variants were validated in an additional 291 patients.
Materials and methods

Patients, treatment, and toxicity evaluation

A total of 323 patients diagnosed with non-small cell lung cancer from the Karolinska University Hospital in Stockholm, Linköping University Hospital in Sweden, Institut de cancérologie Gustave Roussy in Paris, and the France and Christie Hospital in Manchester, UK, were included into the study between 2006 and 2011. The regional ethics committees in Stockholm, Paris, and Manchester approved the study, and all patients provided written and verbal consent as per the Helsinki Declaration before entering the study. All patients were scheduled to be treated with carboplatin and gemcitabine and received at least one cycle of carboplatin AUC = 5 or 6 (depending on the calculation method) and gemcitabine at a dose of 1250–2500 mg/m². After the first chemotherapy cycle, the patients’ blood status was monitored and the blood counts of leucocytes, neutrophils, and platelets were registered at 8, 15, and 21 days after the first chemotherapy treatment. The nadir values after the first chemotherapy cycle were converted to the CTC scale for neutropenia and thrombocytopenia.

From the first 236 patients, we selected 32 patients with extremely high or low myelosuppression for exome sequencing. The 16 patients that suffered extremely high myelosuppression had a CTC grade of 3 to 5 for both neutropenia and thrombocytopenia after the first cycle of chemotherapy (gemcitabine at a dose of 1250–2500 mg/m²). The 16 patients with low toxicity were virtually unaffected by gemcitabine/carboplatin chemotherapy in terms of myelosuppression (gemcitabine dose at 2500 mg/m²). They showed a CTC grade of 0 or 1 for both neutropenia and thrombocytopenia after the first chemotherapy cycle and did not suffer any worse myelosuppression during the following cycles of chemotherapy. The patient characteristics are shown in Table 1.
DNA extraction

DNA was extracted from whole blood using QIAamp® DNA mini-kits (VWR International, Stockholm, Sweden) according to the manufacturer’s protocol. The DNA was quantified using absorbance spectroscopy (260 nm and 280 nm), diluted to 10 ng/µl, and stored at −20°C until analysis.

Exome target enrichment and sequencing

Following extraction, sequence capture was performed and Illumina libraries prepared following the protocol from SeqCap EZ Exome Library (Roche-NimbleGen, Madison, WI, USA) or the protocol from SureSelect Human All Exon V4 (Agilent, Santa Clara, CA, USA). The samples were multiplexed and sequenced using an Illumina HiSeq2000 and paired-end 2 × 100 bp sequencing (Illumina, San Diego, CA, USA).

Alignment, variant calling, and filtering

The overall quality of reads was assessed using FastQC. Before mapping, read quality was improved by 3’ trimming of reads and removing bases with a Phred-score encoded by “B” – which is an Illumina quality-control indicator indicating that the read end should not be used in further analysis. Reads shorter than 40 bases after trimming were discarded as well as reads with four or more uncalled bases. Furthermore, reads with five or more bases with a Phred score of ten or lower, or ten or more bases with a Phred score of 20 or lower, were also discarded. The filtered reads were aligned against the reference genome (hg19) using Mosaik. Variants were called using Samtools mpileup (version 0.1.12-10) and filtered using vcfutils with default parameters apart from the maximum read depth that was set to three times the average exome coverage. Variants
with a total read depth below 10 across all samples were excluded. Variants were annotated using Annovar, and only those that were non-synonymous were selected for further analysis.

After sequencing, comparisons of genetic variants in the high-toxicity group versus the low-toxicity group were made using the following strategies.

**SNV association by Fisher’s exact test**

To identify variants with a difference in allele distribution between the two groups, we used Fisher’s exact test and the statistical software “R” (A Programming Environment for Data Analysis and Graphics, version 2.14.1, [http://www.r-project.org](http://www.r-project.org)).

**Deviation from Hardy–Weinberg equilibrium**

If there is a common genetic variant correlating with toxicity and we have a gene-dose-dependent impact, then we would expect to find mostly homozygous patients (either the reference or the variant genotype) – with heterozygotes mostly absent when we compare groups with extreme phenotype. This is due to the fact that we selected patients with extremely high or extremely low toxicity, i.e. intermediate/heterozygous patients would not be selected. We, therefore identified genetic variants that were out of Hardy–Weinberg equilibrium, using the ‘HWexact’ test and with a significant genetic difference between the groups (Fischer’s exact test as above) using a nominal \( p \)-value < 0.05 as the cutoff.

**Selection of genes by a priori biological knowledge**

We reviewed the literature for genes involved in the elimination, metabolism, target repair, and mechanism of action of gemcitabine and carboplatin. Genes identified by genome-wide
association studies were also included. The difference in allele distribution between patients with high and low toxicity was investigated for all genetic variants within these genes using “R”.

**Analysis of genes selected from gene expression meta-analysis**

We performed a meta-analysis of gene expression data in accordance with our previously described strategy (9). By analyzing all relevant high-quality public-domain gene expression data (from more than 300 microarrays) that concerned platinum analogues and/or gemcitabine, we obtained a list of ranked genes based on the degree of up- or down-regulation induced by the drugs. The genes were also required to be expressed in tissues important for toxicity (i.e. the organ where the toxicity occurs, the bone marrow, and the tissues responsible for the elimination of the drugs such as the kidney). We compared the allele distribution between the two toxicity groups for all SNVs in the top 100 genes generated by this tissue–drug meta-analysis.

**Ingenuity pathway analysis**

All genes with an SNV with a nominal $p$-value (Fischer’s exact test) of less than 0.05 were imported into the Ingenuity database search tool (http://www.ingenuity.com/) to identify enriched pathways. In the analysis, genes from the Ingenuity Knowledge Base were used as the reference set. The analysis included direct and indirect relationships observed in human species only. The enriched pathways were manually curated to include genes with a nominal $p$-value below 0.05 in addition to a selection based on specific relevance for the purpose of this study.

**Network enrichment analysis – FunCoup**

Interactions between genes with individually minor effects could confer non-additive and non-linear effects on the observed toxicity. These effects can be identified by testing for interactions
in the global network of functional coupling using network enrichment analysis (10). We analyzed sets of variants defined at different confidence cutoffs and grouped genes with a higher prevalence of rare variants in the low-toxicity group, the high-toxicity group, and the union of both groups (sign $z = 1.5, 1.7, 1.9, 2.1$; $p$-value $= 0.15–0.04$; number of genes per set $= 29–320$). In parallel, we generated control groups by 1) replacing each gene set member with a random gene with similar topological properties in the network ("null" sets), and 2) taking equally sized groups using the randomly permuted rank of the sign test ("rnd" sets). Each gene set was submitted for network enrichment analysis, and the $z$-scores reported the likelihood of interactions that exist between some (but most likely not all) of the genes in each group.

Validation of candidate genetic markers

From each bioinformatics analysis, a maximum of 20 SNVs with a nominal $p$-value $< 0.05$ (ranked by $p$-value), were selected as candidate markers for prediction of neutropenia and thrombocytopenia and subjected to validation in 291 additional patients treated with gemcitabine and carboplatin (including the remaining 204 patients from the first cohort of patients from which the exome-sequencing patients were selected and an additional 87 patients). In total we ended up with 66 unique genetic variants for the validation. The iPLEX chemistry and the MassArray Analyzer from Sequenom (San Diego, USA) at the Mutation Analysis Facility in Huddinge, Sweden, were used for identification of the SNVs in the validation set. We succeeded in designing assays for 60 genetic variants (using the Sequenom design software) and separated these into three pools. All patients were then sequenced for these 60 SNVs.

The genotypes of the validation patients were compared between patients that suffered grade 0–2 neutropenia and patients that suffered grade 3–5 neutropenia after the first cycle of treatment by
Fischer’s exact test using a dominant model and a recessive model as well as by comparing the allele distribution. The model with the best fit was used based on lowest nominal \(p\)-value, and Hochberg multiple testing was used for correction of the \(p\)-values. The same analysis was performed for patients that suffered grade 0–2 thrombocytopenia versus the group that suffered grade 3–5 thrombocytopenia after the first cycle.

For the top variants based on \(p\)-values in the validation, the association between genotype and neutropenia or thrombocytopenia was also assessed with the Kruskal–Wallis test using the nadir value of neutrophils or platelets during the first cycle.

**Results**

*Patient selection and characteristics*

From the 243 patients treated with gemcitabine and carboplatin, we could identify 32 patients with extremely high or low bone marrow toxicity. For the low-toxicity group, the nadir values for platelets were 2–24 \(\times\) 10\(^9\)/L and neutrophil counts were 0–0.6 \(\times\) 10\(^9\)/L. Patient characteristics, CTC-scale values, and treatments can be found in Table 1. All 32 patients were successfully exome-sequenced. On average, 31 million reads were sequenced per patient, and 99% of the reads mapped to the genome with 78% being on target for an average coverage of 54\(^x\). A summary of the sequencing statistics can be found in Supplemental Table 1. In general, 5000–7000 non-synonymous genetic variants were identified in each patient.

*Exome-wide SNV association testing by Fisher’s test*

When comparing the difference in allele distribution between the high- and low-toxicity groups, a total of 173 non-synonymous variants were found to have a significant difference in allele
distribution between the two groups (nominal p-value < 0.05). The top 20 genetic variants can be found in Supplemental Table 2A. The top three genetic variants were found in PI3 and CSAG1, genes encoding an elastase-specific inhibitor and a member of a family of tumor antigens, respectively.

Deviation from Hardy–Weinberg equilibrium

Of the 173 non-synonymous variants, 38 genetic variants were also found to be out of Hardy–Weinberg equilibrium. The top 20 genetic variants can be found in Supplemental Table 2B. The genetic differences in the CSAG1 gene came up again as one of the top hits. Interestingly, a genetic variant in IL37 was also one of the top hits.

Analysis of genes selected from gene expression meta-analysis

Meta-analysis of gene expression data to identify genes that are up- or down-regulated by carboplatin and/or gemcitabine and are also present in the bone marrow or kidney resulted in a ranked list of 110 genes for gemcitabine and 747 genes for carboplatin. Non-synonymous variants in the top 100 genes from each list were investigated for differences between patients with high or low toxicity. Variants in seven genes had a nominally significant difference between the groups, see Supplemental Table 2C.

Genes previously associated with gemcitabine/carboplatin mechanisms of action

We compiled a list of 60 genes previously known to be involved in the resistance to gemcitabine and carboplatin, their mechanism of action, and the repair of the damage that they cause, see Supplemental Table 3. The only two SNVs in these genes that showed a nominally significant difference were ABCC11 – a gene encoding a drug transporter that might efflux chemotherapeutic
agents out of the cell – and a genetic variant in the DNA-repair gene \textit{ERCC5}, see Supplemental Table 2D.

\textit{Ingenuity pathway analysis}

The highest-ranking pathways in the Ingenuity pathway analysis were associated with general functions such as nucleotide metabolism, cancer, drug metabolism, and hematological system development (Supplemental Table 4). We proceeded with pathways of interest among the top results, and these are highlighted in yellow in Supplemental Table 4. The top 20 genetic variants, with a nominal \textit{p}-value below \textgreater 0.05 from the set of 173 variants, within genes identified in the Ingenuity pathway analysis are presented in Supplemental Table 2E. The top four genetic variants are located in \textit{IL37} (\textit{p} = 0.0015 and \textit{p} = 0.0040), \textit{PLG} (\textit{p} = 0.0054), and \textit{CAST} (\textit{p} = 0.0070).

\textit{Network enrichment analysis – FunCoup}

This analysis clearly demonstrated that genes with a higher degree of rare variants in the low-toxicity group compared to the high-toxicity group ("Real.neg") had enriched network interconnectivity (Supplemental Figure 1). Moreover, the significance of this pattern grew with the confidence of the sign test, i.e. in the series of sign \textit{z}-scores \{1.5, 1.7, 1.9, 2.1\}. A higher degree of rare variants in the high-toxicity group ("Real.pos") did not show any network enrichment. We investigated the known and computationally predicted network relations using the functionality of the FunCoup network (11) and the analytic website HyperSet (http://research.scilifelab.se/andrej_alexeyenko/HyperSet/). This analysis revealed that the top-ranking gene set was in fact a union of 4 or 5 smaller network modules, the members of which could interact with each other.
The network analysis resulted in three different sub-networks that were significantly enriched with genes harboring non-synonymous variants that were different between patients with high and low myelosuppression, see Supplemental Figure 2. The genes that showed the largest difference in variant frequency compared to the human reference genome were OR5M3, CUEDC1, EFCAB7, AKR1E2, GAS2L2, and ITGA8. The genetic variants with a nominal $p$-value below 0.05 in the genes identified by the network analysis are presented in Supplemental Table 2 F.

Validation of top SNVs

We designed assays for 60 genetic variants. Of these, 52 were successfully genotyped in the 291 patients in the validation set (patient characteristics are shown in Supplemental Table 5), while eight variants failed due to only one genotype or to low or unspecific signal (Supplemental Table 6). The SNVs with the strongest association to gemcitabine/carboplatin-induced neutropenia and thrombocytopenia are listed in Table 2 (data shown for all patients). For neutropenia, the marker with the strongest association was rs1453542 located in OR4D6 (Table 2; $p = 0.00087$ nominal and $p = 0.042$ after Hochberg adjustment). At this locus and using a recessive model with the minor allele as the risk allele, 15.8% of patients with grade 3–5 neutropenia were homozygous for the minor allele, whereas only 3.5% of the patients with grade 0-2 neutropenia were homozygous (OR = 5.20, 95% CI = 1.80–18.42). Figure 2A shows the effect of rs1453542 on the neutrophil nadir value after the first chemotherapy cycle, and an association test using nadir values as a quantitative trait also indicated significant association ($p = 0.026$, Kruskal–Wallis test). rs1453542 constitutes a G>T transversion in chr11 position 59224885 (g.11:59224885G>T GRCh37) resulting in an amino acid change from serine to threonine at position 151 (S151T) in
the OR4D6 protein with a scaled CADD score of 22.3, which is considered high and an indication that the shift affects the protein function.

The variant with the strongest association to thrombocytopenia in the validation data set was rs5925720 located in the gene *DDX53* (Table 2; \( p = 0.0015 \) nominal and \( p = 0.07 \) (non-significant) after Hochberg adjustment). Using the allelic counts, the patients with grade 3–5 thrombocytopenia had a minor allele frequency of 5.6% compared to 14.2% in the patients with grade 0–2, indicating that the minor allele is protective (OR = 0.36, 95% CI = 0.17–0.71). Figure 2B shows the effect of the rs5925720 on the platelet nadir value after the first chemotherapy cycle, and this quantitative trait analysis also showed a significant association (\( p = 0.020, \) Kruskal–Wallis test). The genetic variant rs5925720 is a missense variant at position g.X:23019317G>T resulting in a methionine to isoleucine shift at position 381 (M381I) in the DDX53 protein with a scaled CADD score of 14.98.
Discussion

In this study, we performed exome sequencing of 32 patients who experienced either extremely high or low myelosuppression after the first chemotherapy cycle of gemcitabine and carboplatin. Using six different bioinformatics approaches, we identified genetic variants that might be associated with gemcitabine/carboplatin-induced myelosuppression. The top 53 candidates were subjected to validation in an additional 291 patients. The variants rs1453542 in \textit{OR4D6} and rs5925720 in \textit{DDX53} could be validated as predictive markers for gemcitabine/carboplatin-induced neutropenia and thrombocytopenia, respectively.

The validated variants in \textit{OR4D6} and \textit{DDX53} were found in the exome-wide SNV association analysis and the Hardy–Weinberg approach, indicating that these were the most useful bioinformatics analyses. None of the other bioinformatics approaches resulted in a validated biomarker for toxicity, which might indicate that these approaches need to be refined further. Recently, Knights et al. presented work using K-way interaction information (KWII) to identify potential genetic variants by gene-gene and gene-environment interactions associated with gemcitabine-induced neutropenia (12). No variants within the genes found by the KWII approach were among the top variants identified in this study.

The validated genetic variants we present, rs1453542 and rs5925720 for gemcitabine/carboplatin-induced neutropenia and thrombocytopenia, respectively, reside in the genes \textit{OR4D6} and \textit{DDX53}. There is limited information on the function of the olfactory receptor 4D6 (\textit{OR4D6}). Olfactory receptors are G-coupled receptors present in olfactory epithelium where they recognize molecular
motifs and initiate signaling in the nervous system, thus providing the basis for smell. However, recently it has been shown that some of the olfactory receptor genes are expressed in other tissues, such as the liver and kidney, suggesting alternative functions (13). For example, it has been suggested that the expression of olfactory receptors in the testis might be involved in the chemotaxis of sperm (14). Because chemotaxis is an important mechanism in the immune system and especially for neutrophils and their activation, the association of an olfactory gene to drug-induced neutropenia might be biologically reasonable. Furthermore, Zhang et al indicated that \textit{OR4D6} can be expressed in the liver ($p = 0.07$) and kidney ($p = 0.07$). The mechanism by which genetic variants in \textit{OR4D6} might be involved in neutropenia is unclear, but it might be speculated that these variants affect blood flow in the liver or kidney or affect cell death through chemotaxis. The genetic variant rs1453542 was still significantly associated with neutropenia after multiple correction ($p = 0.042$, nominal $p$-value $= 0.00087$) in the validation set. Using a recessive model, the OR for this genetic variant was 5.2. However, we did not correct for multiple testing in the discovery phase, which should be noted. We also investigated the predicted consequence of the genetic variant rs1453542 by calculating its CADD score (15). The score for this variant was 22.3, which suggests that the variant is among the top 1% of genetic variants that are predicted to have a functional effect, and it is especially high for a genetic variant in the olfactory family (15). \textit{DDX53}, DEAD (Asp-Glu-Ala-Asp) box polypeptide 53, also known as \textit{CAGE}, contains several domains found in a protein family of ATP-dependent RNA helicases. The gene is normally mainly expressed in testis but has also been associated with a variety of cancers, especially gastric tumors (16). The function of the gene is not known, but because it is a member of the DEAD gene family it is likely to be involved in the unwinding of RNA prior to protein synthesis and might very well have a regulatory function. Overexpression of DDX53 has been connected to cell proliferation (17). It has also been suggested that DDX53 confers resistance to anti-cancer
drugs through down-regulation of p53 (18), indicating a mechanism in apoptosis. It should be noted that the genetic variant rs5925720 in DDX53 was not significant after correcting for multiple testing (corrected $p$-value of 0.074).

Our approach to identifying variants within genes selected by a priori biological knowledge only resulted in one nominally significant variant in the discovery stage, see Supplemental Table 2D. It should be noted that the validation material used in this study was not completely independent of the screening because the remaining non-extreme patients were used for the validation in addition to an independent patient cohort. Several studies have previously associated toxicity and response with variants of candidate genes involved in metabolism and transport of gemcitabine such as CDA, dCK, SLC28A1, SLC28A3 as well as target molecules RRM1, RRM2 and RRM2b (19-24). In a recent meta-analysis, the genetic variant A79C in the CDA gene was associated with severe anemia, but without being able to show an effect related to neutropenia, thrombocytopenia, or response (25). Furthermore, variants within genes involved in platinum pharmacokinetics and pharmacodynamics have been investigated, such as RRM1, CCNH, XPC, XPD, ERCC1, ERCC5, ERCC6, and XRCC3 (26-32). These variants explain some part of the inter-individual variability seen in gemcitabine/carboplatin-induced toxicity. However the results have been inconclusive, which might explain why we only got two hits in our literature-based approach. DNA-repair genes have previously been shown to play a role in the pharmacogenetics of platinum-based effects (28,29,32). One of our hits in the literature-based approach was a genetic variant in ERCC5, a DNA-repair gene, but this variant the genotyping failed in the validation cohort and it has a low minor allele frequency. A genome-wide study of gemcitabine-induced leukopenia/neutropenia has previously been published (5), and they presented four loci
associated with gemcitabine-induced leukopenia/neutropenia. In our study, we investigated the
coding regions in the suggested genes but could not find a difference in the screening of patients
with high or low toxicity. Innocenti et al. presented a genome-wide association study on overall
survival in pancreatic cancer patients treated with gemcitabine and showed an association
between genetic variants in \textit{IL17F} and the response (33). Interestingly we found several variants
in \textit{IL37} to be associated with toxicity response in the discovery cohort, but none of these were
successfully validated.

One of the limitations of our study is that we only investigated the coding regions of the genome.
It has become more evident that genetic variations in the regulatory regions such as promoters
and enhancers play an important role in regulation of protein function and that as much as 80% of
the genome has a biochemical function as suggested by the ENCODE consortium (34). Most
likely, some of the variability in drug-induced myelosuppression might be explained by variations
in the regulatory regions. In one of our approaches, the analysis of genes selected from gene-
expression meta-analysis, we tried to evaluate some of the regulatory aspects of the genome.
However, because our sequencing data were limited to mainly the coding region we had limited
success with this approach and using whole-genome sequencing would be a way forward.

In conclusion, in this study we exome-sequenced patients with extremely high and low levels of
gemcitabine/carboplatin-induced neutropenia and thrombocytopenia and identified two new
predictive markers, rs1453542 in \textit{OR4D6} and rs5925720 in \textit{DDX53}. In a replication cohort, these
were successfully validated to be associated with gemcitabine/carboplatin-induced neutropenia
and thrombocytopenia, respectively. However, we cannot conclude whether these two genes are
specific for this chemotherapy combination or are generally predictive of toxicity for any chemotherapy agent. Hopefully, with additional work and validation, these variants will prove to be useful as markers for the clinical problem of myelosuppression in lung cancer treatment.
References


Figure legends

Figure 1. The distribution of nadir values for neutrophils and thrombocytes during the first chemotherapy cycle in 234 lung cancer patients who received gemcitabine and carboplatin. The upper right square represents patients unaffected by chemotherapy (note that some will be excluded due to toxicity in later cycles). The lower left square represents patients with grade 4 neutropenia and thrombocytopenia.

Figure 2. The effect of rs1453542 on neutropenia (A) and rs5925720 on thrombocytopenia (B). The numbers of patient for each genotype of rs1453542 were 122 (G/G), 112 (G/C), and 24 (C/C), and for rs5925720 the numbers of patients for each genotype were 245 (G/G), 28 (G/T), and 20 (T/T).
<table>
<thead>
<tr>
<th></th>
<th>Low toxicity group</th>
<th>High toxicity group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Females</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Age, years</td>
<td>Median (range)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>62 (47–74)</td>
<td>66 (51–80)</td>
</tr>
<tr>
<td>Histological type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>Squamous Cell Carcinoma</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Large Cell Carcinoma</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Smoking, n.d. = 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Former</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Never smoked</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Thrombocytopenia</td>
<td>CTCAE grade 0</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>CTCAE grade 1</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>CTCAE grade 3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CTCAE grade 4</td>
<td>16</td>
</tr>
<tr>
<td>Neutropenia</td>
<td>CTCAE grade 0</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>CTCAE grade 1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>CTCAE grade 3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>CTCAE grade 4</td>
<td>12</td>
</tr>
</tbody>
</table>
Table 2 – Statistics for the validated top candidates

<table>
<thead>
<tr>
<th>Neutropenia</th>
<th>Number of patients</th>
<th>Genotype, P-value</th>
<th>MAF</th>
<th>Major allele</th>
<th>Minor allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNV</td>
<td>Gene</td>
<td>CTC grade</td>
<td>Homo. major</td>
<td>Hetero.</td>
<td>Homo. minor</td>
</tr>
<tr>
<td>rs1453542</td>
<td>OR4D6</td>
<td>0-2</td>
<td>76</td>
<td>63</td>
<td>5</td>
</tr>
<tr>
<td>rs6622126</td>
<td>MORC4</td>
<td>0-2</td>
<td>57</td>
<td>44</td>
<td>45</td>
</tr>
<tr>
<td>rs2686817</td>
<td>PKD1L1</td>
<td>0-2</td>
<td>33</td>
<td>68</td>
<td>43</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Thrombocytopenia</th>
<th>Number of patients</th>
<th>Genotype, P-value</th>
<th>MAF</th>
<th>Major allele</th>
<th>Minor allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNV</td>
<td>Gene</td>
<td>CTC grade</td>
<td>Homo. major</td>
<td>Hetero.</td>
<td>Homo. minor</td>
</tr>
<tr>
<td>rs5925720</td>
<td>DDX53</td>
<td>0-2</td>
<td>165</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>rs6896902</td>
<td>MAML1</td>
<td>0-2</td>
<td>94</td>
<td>85</td>
<td>21</td>
</tr>
<tr>
<td>rs13306061</td>
<td>UTS2</td>
<td>0-2</td>
<td>115</td>
<td>78</td>
<td>10</td>
</tr>
</tbody>
</table>

Note: Although not significant, the Hochberg corrected p-value for rs5925720 was 0.074.
Figure 1

Nadir values after first cycle of chemotherapy #/mm^3

Neutropenia

Thrombocytopenia

Nadir values after first cycle of chemotherapy #/mm^3
Figure 2

A

Neutrophil nadir value after first cycle of chemotherapy (K/mm$^3$)

<table>
<thead>
<tr>
<th></th>
<th>G/G</th>
<th>G/C</th>
<th>C/C</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1453542</td>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
<td><img src="image3" alt="Graph" /></td>
</tr>
</tbody>
</table>

$p = 0.026$

B

Thrombocyte nadir value after first cycle of chemotherapy (K/mm$^3$)

<table>
<thead>
<tr>
<th></th>
<th>G/G</th>
<th>G/T</th>
<th>T/T</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs5925720</td>
<td><img src="image4" alt="Graph" /></td>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
</tr>
</tbody>
</table>

$p = 0.020$
Clinical Cancer Research

Using whole exome sequencing to identify genetic markers for carboplatin and gemcitabine-induced toxicities.

Henrik Green, Johanna Hasmats, Ilya Kupershmidt, et al.

Clin Cancer Res  Published OnlineFirst September 16, 2015.

Updated version  Access the most recent version of this article at:

Supplementary Material  Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2015/09/16/1078-0432.CCR-15-0964.DC1

Author Manuscript  Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, use this link http://clincancerres.aacrjournals.org/content/early/2015/09/16/1078-0432.CCR-15-0964. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.