

## Using Whole-Exome Sequencing to Identify Genetic Markers for Carboplatin and Gemcitabine-Induced Toxicities

Henrik Gréen<sup>1,2,3</sup>, Johanna Hasmats<sup>1</sup>, Ilya Kupersmidt<sup>1,4</sup>, Daniel Edsgård<sup>1</sup>, Luigi de Petris<sup>5</sup>, Rolf Lewensohn<sup>5</sup>, Fiona Blackhall<sup>6,7</sup>, Svante Vikingsson<sup>3</sup>, Benjamin Besse<sup>8</sup>, Andrea Lindgren<sup>9,10,11</sup>, Eva Brandén<sup>5</sup>, Hirsh Koyi<sup>5</sup>, Curt Peterson<sup>3</sup>, and Joakim Lundeberg<sup>1</sup>

### 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 is 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: (i) whole-exome nonsynonymous single-nucleotide variants association analysis, (ii) deviation from Hardy-Weinberg equilibrium, (iii) analysis of genes selected by *a priori* biologic knowledge, (iv) analysis of genes

selected from gene expression meta-analysis of toxicity datasets, (v) Ingenuity Pathway Analysis, and (vi) 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. *Clin Cancer Res*; 22(2); 366–73. ©2015 AACR.

<sup>1</sup>Science for Life Laboratory, School of Biotechnology, Division of Gene Technology, Royal Institute of Technology, Solna, Sweden. <sup>2</sup>Department of Forensic Genetics and Forensic Toxicology, National Board of Forensic Medicine, Linköping, Sweden. <sup>3</sup>Division of Drug Research, Clinical Pharmacology, Department of Medical and Health Sciences, Linköping University, Linköping, Sweden. <sup>4</sup>NextBio, Cupertino, California. <sup>5</sup>Department of Oncology and Pathology, Karolinska Institute and Oncology Clinic, Karolinska University Hospital, Stockholm, Sweden. <sup>6</sup>Department of Medical Oncology, Christie Hospital, Manchester, United Kingdom. <sup>7</sup>Institute of Cancer Sciences, University of Manchester, Manchester, United Kingdom. <sup>8</sup>Unité INSERM U 981, Université Paris Sud, Département de Médecine, Institut Gustave Roussy, Villejuif, France. <sup>9</sup>Department of Clinical Physiology, University Hospital, Linköping, Sweden. <sup>10</sup>Department of Medical and Health Sciences, Linköping University, Linköping, Sweden. <sup>11</sup>Department of Pulmonary Medicine, University Hospital, Linköping, Sweden.

**Note:** Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

H. Green and J. Hasmats contributed equally to this article.

Current address for Eva Brandén and Hirsh Koyi: Department of Respiratory Medicine, Gävle Hospital, Gävle, Sweden.

**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. Phone: 46-13-252157; Fax: 46-13-136005; E-mail: henrik.green@liu.se

doi: 10.1158/1078-0432.CCR-15-0964

©2015 American Association for Cancer Research.

### 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 (ADR) are mainly hematologic toxicities, including neutropenia, leukopenia, and thrombocytopenia (4). There is also large interindividual variability when it comes to the severity and frequency of these ADRs. Our data show that 15% to 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% to 37% are basically unaffected (grade 0) when treated with the same drug regimen. The severity of the disease, with a 5-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.

### 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.

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 intrastrand 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 and colleagues published a genome-wide association study to identify single-nucleotide variants (SNV) 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, United Kingdom, 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 1,250–2,500 mg/m<sup>2</sup>. 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 1,250–2,500 mg/m<sup>2</sup>). The 16 patients with low toxicity were virtually unaffected by gemcitabine/carboplatin chemotherapy in terms of myelosuppression (gemcitabine dose at 2500 mg/m<sup>2</sup>). 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) according to the manufacturer's protocol. The DNA was quantified using absorbance spectroscopy

**Table 1.** Patient characteristics in the discovery data set

	Low-toxicity group	High-toxicity group
Gender		
Males	8	7
Females	8	9
Age, y, median (range)	62 (47–74)	66 (51–80)
Histologic type		
Adenocarcinoma	11	13
Squamous cell carcinoma	1	2
Large cell carcinoma	1	1
Other	3	0
Smoking, n.d. = 2		
Current	10	8
Former	4	6
Never smoked	0	2
Thrombocytopenia		
CTCAE grade 0	9	
CTCAE grade 1	6	
CTCAE grade 3		0
CTCAE grade 4		16
Neutropenia		
CTCAE grade 0	13	
CTCAE grade 1	2	
CTCAE grade 3		4
CTCAE grade 4		12

(260 nm and 280 nm), diluted to 10 ng/ $\mu$ L, and stored at  $-20^{\circ}\text{C}$  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) or the protocol from SureSelect Human All Exon V4 (Agilent Technologies). The samples were multiplexed and sequenced using an Illumina HiSeq2000 and paired-end  $2 \times 100$  bp sequencing (Illumina).

#### 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 10 or lower, or 10 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 exact test

To identify variants with a difference in allele distribution between the two groups, we used Fisher exact test and the statistical software "R" (A Programming Environment for Data Analysis and Graphics, version 2.14.1, <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 exact test as above) using a nominal  $P$  value  $< 0.05$  as the cutoff.

#### Selection of genes by a priori biologic 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 analogs and/or gemcitabine, we obtained a list of ranked genes based on the degree of up- or downregulation 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 (Fisher 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 nonadditive and nonlinear 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  $z = 1.5, 1.7, 1.9, 2.1$ ;  $P = 0.15-0.04$ ; number of genes per set = 29–320). In parallel, we generated control groups by (i) replacing each gene set member with a random gene with similar topological properties in the network ("null" sets), and (ii) 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 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 Fisher 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.

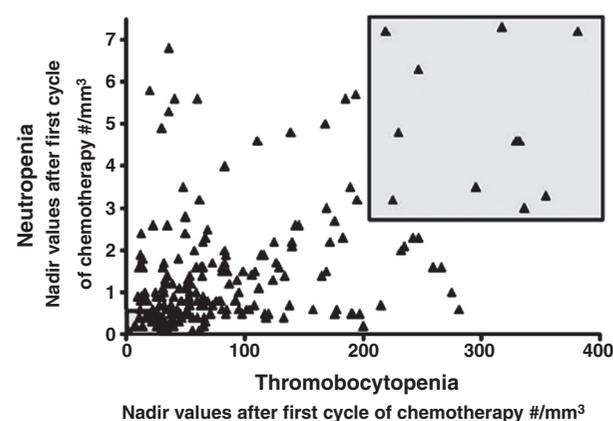
## Results

### 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, see Figure 1. For the low-toxicity group, the nadir values for platelets were  $2\text{--}24 \times 10^9/\text{L}$  and neutrophil counts were  $0\text{--}0.6 \times 10^9/\text{L}$ . 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  $\times 54$ . A summary of the sequencing statistics can be found in Supplementary Table S1. In general, 5,000–7,000 nonsynonymous genetic variants were identified in each patient.

### Exome-wide SNV association testing by Fisher test

When comparing the difference in allele distribution between the high- and low-toxicity groups, a total of 173 nonsynonymous variants were found to have a significant difference in allele distribution between the two groups (nominal *P* value  $< 0.05$ ).



**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. Top right square, patients unaffected by chemotherapy (note that some will be excluded due to toxicity in later cycles). Bottom left square, patients with grade 4 neutropenia and thrombocytopenia.

The top 20 genetic variants can be found in Supplementary Table S2A. 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 nonsynonymous variants, 38 genetic variants were also found to be out of Hardy–Weinberg equilibrium. The top 20 genetic variants can be found in Supplementary Table S2B. 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 downregulated 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. Nonsynonymous 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 Supplementary Table S2C.

### 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 Supplementary Table S3. 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 *ERCC5*, see Supplementary Table S2D.

### 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 hematologic system development (Supplementary Table S4). We proceeded with pathways of interest among the top results, and these are highlighted in yellow in Supplementary Table S4. The top 20 genetic variants, with a nominal *P* value below  $>0.05$  from the set of 173 variants, within genes identified in the Ingenuity pathway analysis are presented in Supplementary Table S2E. The top four genetic variants are located in *IL37* ( $P = 0.0015$  and  $P = 0.0040$ ), *PLG* ( $P = 0.0054$ ), and *CAST* ( $P = 0.0070$ ).

### Network enrichment analysis–FunCoup

This analysis clearly demonstrated that genes with a higher degree of rare variants in the low-toxicity group compared with the high-toxicity group ("Real.neg") had enriched network interconnectivity (Supplementary Fig. S1). Moreover, the significance of this pattern grew with the confidence of the sign test, i.e., in the series of sign *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 ([https://research.scilifelab.se/andrej\\_alexeyenko/HyperSet/Ashwini/DE\\_conditions.html](https://research.scilifelab.se/andrej_alexeyenko/HyperSet/Ashwini/DE_conditions.html)). This analysis revealed that the top-ranking gene set was in fact a union of 4 or

Gréen et al.

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 nonsynonymous variants that were different between patients with high and low myelosuppression, see Supplementary Fig. S2. The genes that showed the largest difference in variant frequency compared with 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 Supplementary Table S2F.

### 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 Supplementary Table S5), whereas eight variants failed due to only one genotype or to low or unspecific signal (Supplementary Table S6). 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 (nonsignificant) after Hochberg adjustment). Using the allelic counts, the patients with grade 3–5 thrombocytopenia had a minor allele frequency of 5.6% compared with 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

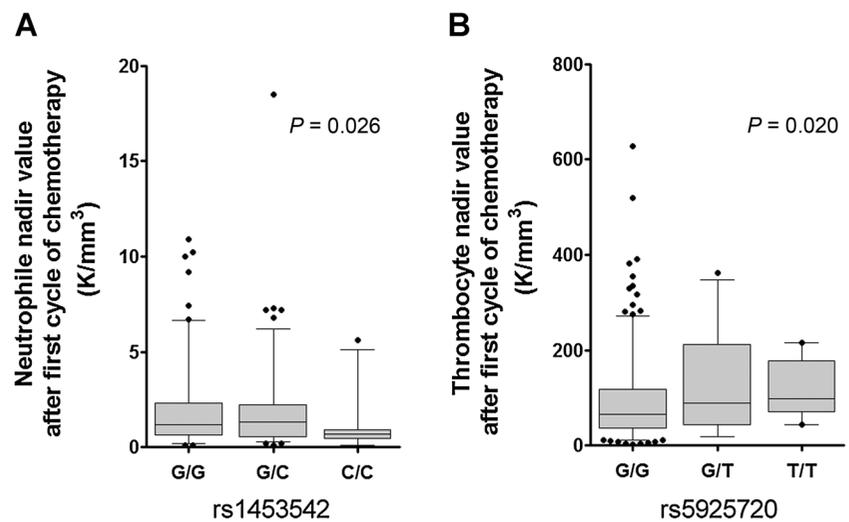
**Table 2.** Statistics for the validated top candidates

SNV	Gene	CTC grade	Number of patients			Model	OR (95% CI)	Genotype, <i>P</i>		MAF	Major allele	Minor allele
			Homo. major	Hetero.	Homo. minor			Nominal	Hochberg			
rs1453542	OR4D6	0–2	76	63	5	Recessive	5.2 (1.8–18)	0.00087	0.042	31%	G	C
		3–5	50	51	19							
			57	44	45	Dominant	0.54 (0.32–0.91)	0.014	n.s.	42%	G	A
rs6622126	MORC4	0–2	66	22	34							
		3–5	33	68	43	Allele	0.65 (0.46–0.94)	0.018	n.s.	49%	A	C
			40	57	23							
rs2686817	PKDIL1	0–2	115	78	10							
		3–5	69	25	4							
			165	20	19	Allele	0.36 (0.17–0.71)	0.0015	n.s.	11%	G	T
rs5925720	DDX53	0–2	88	9	1							
		3–5	94	85	21	Allele	1.5 (1.1–2.2)	0.021	n.s.	35%	G	A
			33	46	17							
rs6895902	MAML1	0–2	115	78	10							
		3–5	69	25	4	Dominant	0.55 (0.32–0.94)	0.023	n.s.	22%	C	T
			165	20	19							
rs13306061	UTS2	0–2	115	78	10							
		3–5	69	25	4							
			165	20	19							

NOTE: Although not significant, the Hochberg corrected *P* value for rs5925720 was 0.074.

**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). (K/mm<sup>3</sup> denotes cell count per mm<sup>3</sup>.)



candidates were subjected to validation in an additional 291 patients. The variants rs1453542 in *OR4D6* and rs5925720 in *DDX53* could be validated as predictive markers for gemcitabine/carboplatin-induced neutropenia and thrombocytopenia, respectively.

The validated variants in *OR4D6* and *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 and colleagues 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 *OR4D6* and *DDX53*. There is limited information on the function of the olfactory receptor 4D6 (*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 and colleagues (13) indicated that *OR4D6* can be expressed in the liver ( $P = 0.07$ ) and kidney ( $P = 0.07$ ). The mechanism by which genetic variants in *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).

*DDX53*, DEAD (Asp-Glu-Ala-Asp) box polypeptide 53, also known as *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 anticancer 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* biologic knowledge only resulted in one nominally significant variant in the discovery stage, see Supplementary Table S2D. 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 and colleagues presented a genome-wide association study on overall survival in pancreatic cancer patients treated with gemcitabine and showed an association between genetic variants in *IL17F* and the response (33). Interestingly we found several variants in *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 *OR4D6* and rs5925720 in *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 chemo-

therapy 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.

### Disclosure of Potential Conflicts of Interest

J. Hasmats is an employee of Agilent Technologies. No potential conflicts of interest were disclosed by the other authors.

### Authors' Contributions

**Conception and design:** H. Gréen, I. Kupersmidt, R. Lewensohn, H. Koyi, C. Peterson, J. Lundeberg

**Development of methodology:** H. Gréen, J. Hasmats, I. Kupersmidt, J. Lundeberg

**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** H. Gréen, L. de Petris, R. Lewensohn, F. Blackhall, B. Besse, A. Lindgren, E. Brandén, H. Koyi, C. Peterson, J. Lundeberg

**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** H. Gréen, J. Hasmats, I. Kupersmidt, D. Edsgård, F. Blackhall, S. Vikingsson, C. Peterson, J. Lundeberg

**Writing, review, and/or revision of the manuscript:** H. Gréen, J. Hasmats, I. Kupersmidt, D. Edsgård, L. de Petris, R. Lewensohn, F. Blackhall, S. Vikingsson, A. Lindgren, E. Brandén, H. Koyi, J. Lundeberg

**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** H. Gréen, L. de Petris, S. Vikingsson, B. Besse, E. Brandén, H. Koyi, J. Lundeberg

**Study supervision:** H. Gréen, H. Koyi, C. Peterson, J. Lundeberg

### Acknowledgments

The authors thank Ellen Sherwood, Magnus Bjursell, Henrik Stranneheim, Nathaniel Street, and Mikael Huss for technical assistance during the genotyping and bioinformatics analysis. The authors 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.

### Grant Support

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.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 20, 2015; revised July 24, 2015; accepted August 11, 2015; published OnlineFirst September 16, 2015.

### References

- Eichelbaum M, Ingelman-Sundberg M, Evans WE. Pharmacogenomics and individualized drug therapy. *Annu Rev Med* 2006;57:119–37.
- Evans WE, McLeod HL. Pharmacogenomics—drug disposition, drug targets, and side effects. *N Engl J Med* 2003;348:538–49.
- Robert J, Morvan VL, Smith D, Pourquier P, Bonnet J. Predicting drug response and toxicity based on gene polymorphisms. *Crit Rev Oncol Hematol* 2005;54:171–96.
- Cortes-Funes H, Martin C, Abratt R, Lund B. Safety profile of gemcitabine, a novel anticancer agent, in non-small cell lung cancer. *Anticancer Drugs* 1997;8:582–7.
- Kiyotani K, Uno S, Mushiroda T, Takahashi A, Kubo M, Mitsuhata N, et al. A genome-wide association study identifies four genetic markers for hematological toxicities in cancer patients receiving gemcitabine therapy. *Pharmacogenet Genomics* 2012;22:229–35.
- Holleman A, den Boer ML, de Menezes RX, Cheok MH, Cheng C, Kazemier KM, et al. The expression of 70 apoptosis genes in relation to lineage, genetic subtype, cellular drug resistance, and outcome in childhood acute lymphoblastic leukemia. *Blood* 2006;107:769–76.
- Zhang X, Yashiro M, Qiu H, Nishii T, Matsuzaki T, Hirakawa K. Establishment and characterization of multidrug-resistant gastric cancer cell lines. *Anticancer Res* 2010;30:915–21.
- Tooker P, Yen WC, Ng SC, Negro-Vilar A, Hermann TW. Bexarotene (LGD1069, Targretin), a selective retinoid X receptor agonist, prevents and reverses gemcitabine resistance in NSCLC cells by modulating gene amplification. *Cancer Res* 2007;67:4425–33.
- Hasmats J, Kupersmidt I, Rodriguez-Antona C, Su QJ, Khan MS, Jara C, et al. Identification of candidate SNPs for drug induced toxicity from differentially expressed genes in associated tissues. *Gene* 2012;506:62–8.
- Alexeyenko A, Lee W, Pernemalm M, Guegan J, Dessen P, Lazar V, et al. Network enrichment analysis: extension of gene-set enrichment analysis to gene networks. *BMC Bioinformatics* 2012;13:226.

11. Alexeyenko A, Sonnhammer EL. Global networks of functional coupling in eukaryotes from comprehensive data integration. *Genome Res* 2009;19:1107–16.
12. Knights J, Sato Y, Kaniwa N, Saito Y, Ueno H, Ramanathan M. Genetic factors associated with gemcitabine pharmacokinetics, disposition, and toxicity. *Pharmacogenet Genomics* 2014;24:15–25.
13. Zhang X, De la Cruz O, Pinto JM, Nicolae D, Firestein S, Gilad Y. Characterizing the expression of the human olfactory receptor gene family using a novel DNA microarray. *Genome Biol* 2007;8:R86.
14. Spehr M, Gisselmann G, Poplawski A, Riffell JA, Wetzell CH, Zimmer RK, et al. Identification of a testicular odorant receptor mediating human sperm chemotaxis. *Science* 2003;299:2054–8.
15. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46:310–5.
16. Cho B, Lim Y, Lee DY, Park SY, Lee H, Kim WH, et al. Identification and characterization of a novel cancer/testis antigen gene CAGE. *Biochem Biophys Res Commun* 2002;292:715–26.
17. Por E, Byun HJ, Lee EJ, Lim JH, Jung SY, Park I, et al. The cancer/testis antigen CAGE with oncogenic potential stimulates cell proliferation by up-regulating cyclins D1 and E in an AP-1- and E2F-dependent manner. *J Biol Chem* 2010;285:14475–85.
18. Kim Y, Park H, Park D, Lee YS, Choe J, Hahn JH, et al. Cancer/testis antigen CAGE exerts negative regulation on p53 expression through HDAC2 and confers resistance to anti-cancer drugs. *J Biol Chem* 2010;285:25957–68.
19. Tanaka M, Javle M, Dong X, Eng C, Abbruzzese JL, Li D. Gemcitabine metabolic and transporter gene polymorphisms are associated with drug toxicity and efficacy in patients with locally advanced pancreatic cancer. *Cancer* 2010;116:5325–35.
20. Chew HK, Doroshow JH, Frankel P, Margolin KA, Somlo G, Lenz HJ, et al. Phase II studies of gemcitabine and cisplatin in heavily and minimally pretreated metastatic breast cancer. *J Clin Oncol* 2009;27:2163–9.
21. Sugiyama E, Kaniwa N, Kim SR, Kikura-Hanajiri R, Hasegawa R, Maekawa K, et al. Pharmacokinetics of gemcitabine in Japanese cancer patients: the impact of a cytidine deaminase polymorphism. *J Clin Oncol* 2007;25:32–42.
22. Rha SY, Jeung HC, Choi YH, Yang WI, Yoo JH, Kim BS, et al. An association between RRM1 haplotype and gemcitabine-induced neutropenia in breast cancer patients. *Oncologist* 2007;12:622–30.
23. Wong A, Soo RA, Yong WP, Innocenti F. Clinical pharmacology and pharmacogenetics of gemcitabine. *Drug Metab Rev* 2009;41:77–88.
24. Khatri A, Williams BW, Fisher J, Brundage RC, Gurvich VJ, Lis LG, et al. SLC28A3 genotype and gemcitabine rate of infusion affect dFdCTP metabolite disposition in patients with solid tumours. *Br J Cancer* 2014;110:304–12.
25. Li H, Wang X, Wang X. The impact of CDA A79C gene polymorphisms on the response and hematologic toxicity in gemcitabine-treated patients: A meta-analysis. *Int J Biol Markers* 2014;29:e224–32.
26. Zhu XL, Sun XC, Chen BA, Sun N, Cheng HY, Li F, et al. XPC Lys939Gln polymorphism is associated with the decreased response to platinum based chemotherapy in advanced non-small-cell lung cancer. *Chin Med J (Engl)* 2010;123:3427–32.
27. Wei SZ, Zhan P, Shi MQ, Shi Y, Qian Q, Yu LK, et al. Predictive value of ERCC1 and XPD polymorphism in patients with advanced non-small cell lung cancer receiving platinum-based chemotherapy: a systematic review and meta-analysis. *Med Oncol* 2011;28:315–21.
28. Ren S, Zhou S, Wu F, Zhang L, Li X, Zhang J, et al. Association between polymorphisms of DNA repair genes and survival of advanced NSCLC patients treated with platinum-based chemotherapy. *Lung Cancer* 2012;75:102–9.
29. Tzvetkov MV, Behrens G, O'Brien VP, Hohloch K, Brockmoller J, Benohr P. Pharmacogenetic analyses of cisplatin-induced nephrotoxicity indicate a renoprotective effect of ERCC1 polymorphisms. *Pharmacogenomics* 2011;12:1417–27.
30. Khrunin AV, Moisseev A, Gorbunova V, Limborska S. Genetic polymorphisms and the efficacy and toxicity of cisplatin-based chemotherapy in ovarian cancer patients. *Pharmacogenomics J* 2010;10:54–61.
31. Iranzo V, Sirera R, Bremnes RM, Blasco A, Jantus-Lewintre E, Taron M, et al. Chemotherapy-induced neutropenia does not correlate with DNA repair gene polymorphisms and treatment efficacy in advanced non-small-cell lung cancer patients. *Clin Lung Cancer* 2011;12:224–30.
32. Sullivan I, Salazar J, Majem M, Pallares C, Del Rio E, Paez D, et al. Pharmacogenetics of the DNA repair pathways in advanced non-small cell lung cancer patients treated with platinum-based chemotherapy. *Cancer Lett* 2014;353:160–6.
33. Innocenti F, Owzar K, Cox NL, Evans P, Kubo M, Zembutsu H, et al. A genome-wide association study of overall survival in pancreatic cancer patients treated with gemcitabine in CALGB 80303. *Clin Cancer Res* 2012;18:577–84.
34. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74.

# Clinical Cancer Research

## Using Whole-Exome Sequencing to Identify Genetic Markers for Carboplatin and Gemcitabine-Induced Toxicities

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

*Clin Cancer Res* 2016;22:366-373. Published OnlineFirst September 16, 2015.

**Updated version** Access the most recent version of this article at:  
[doi:10.1158/1078-0432.CCR-15-0964](https://doi.org/10.1158/1078-0432.CCR-15-0964)

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

**Cited articles** This article cites 34 articles, 11 of which you can access for free at:  
<http://clincancerres.aacrjournals.org/content/22/2/366.full#ref-list-1>

**Citing articles** This article has been cited by 1 HighWire-hosted articles. Access the articles at:  
<http://clincancerres.aacrjournals.org/content/22/2/366.full#related-urls>

**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](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, use this link  
<http://clincancerres.aacrjournals.org/content/22/2/366>.  
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.