Lung Cancer 147 (2020) 106–114

Contents lists available at ScienceDirect

Lung Cancer

journal homepage: www.elsevier.com/locate/lungcan

Genetic association of gemcitabine/carboplatin-induced leukopenia and neutropenia in non-small cell lung cancer patients using whole-exome sequencing

Anna Svedberg[a,1], Niclas Björn[a,1], Benjamín Sigurgeirsson[b,c], Sailendra Pradhananga[b], Eva Brandén[d,e], Hirsh Koyi[d,e], Rolf Lewensohn[f], Luigi De Petris[f], María Apellániz-Ruiz[g], Cristina Rodríguez-Antona[g], Joakim Lundeberg[b,2], Henrik Gréen[a,b,h,2,*]

[a] Clinical Pharmacology, Division of Drug Research, Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden
[b] Science for Life Laboratory, School of Engineering Sciences in Chemistry, Biotechnology and Health, Department of Gene Technology, KTH Royal Institute of Technology, Solna, Sweden
[c] School of Engineering and Natural Sciences, University of Iceland, Reykjavík, Iceland
[d] Department of Respiratory Medicine, Gävle Hospital, Gävle, Sweden
[e] Centre for Research and Development, Uppsala University/Region Gävleborg, Gävle, Sweden
[f] Thoracic Oncology Unit, Tema Cancer, Karolinska University Hospital, and Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden
[g] Hereditary Endocrine Cancer Group, Spanish National Cancer Research Centre (CNIO), Madrid, Spain
[h] Department of Forensic Genetics and Forensic Toxicology, National Board of Forensic Medicine, Linköping, Sweden

ARTICLE INFO

Keywords:
Non-small cell lung cancer
Gemcitabine
Carboplatin
Whole-exome sequencing
Adverse drug reactions
Hematological toxicity
Toxicity
Chemotherapy

ABSTRACT

Objectives: Gemcitabine/carboplatin treatment is known to cause severe adverse drug reactions which can lead to the need for reduction or cessation of chemotherapy. It would be beneficial to identify patients at risk of severe hematological toxicity in advance before treatment start. This study aims to identify genetic markers for gemcitabine/carboplatin-induced leukopenia and neutropenia in non-small cell lung cancer patients.

Material and methods: Whole-exome sequencing was performed on 215 patients. Association analysis was performed on single-nucleotide variants (SNVs) and genes, and the validation was based on an independent genome-wide association study (GWAS). Based on the association and validation analyses the genetic variants were then selected for and used in weighted genetic risk score (wGRS) prediction models for leukopenia and neutropenia.

Results: Association analysis identified 50 and 111 SNVs, and 12 and 20 genes, for leukopenia and neutropenia, respectively. Of these SNVS 20 and 19 were partially validated for leukopenia and neutropenia, respectively. The genes *SVIL* (p = 2.48E-06) and *EFCAB2* (p = 4.63E-06) were significantly associated with leukopenia contain the partially validated SNVs rs3740003, rs10160013, rs1547169, rs10927386 and rs10927387. The wGRS prediction models showed significantly different risk scores for high and low toxicity patients.

Conclusion: We have identified and partially validated genetic biomarkers in SNVs and genes correlated to gemcitabine/carboplatin-induced leukopenia and neutropenia and created wGRS models for predicting the risk of chemotherapy-induced hematological toxicity. These results provide a strong foundation for further studies of chemotherapy-induced toxicity.

1. Introduction

Lung cancer is the second most common cancer in the United States with a 5-year relative survival rate of only 18 % [1]. A widely used non-small cell lung cancer (NSCLC) chemotherapy treatment is gemcitabine in combination with carboplatin which is also used in the treatment of other solid tumors such as bladder, ovarian and breast cancer. Gemcitabine/carboplatin treatment is known to cause severe toxicity that can

* Corresponding author at: Clinical Pharmacology, Division of Drug Research, Department of Biomedical and Clinical Sciences, Linköping University, SE-58185 Linköping, Sweden.
E-mail address: henrik.green@liu.se (H. Gréen).
[1] AS and NB contributed equally to this work.
[2] JL and HG contributed equally to this work.

https://doi.org/10.1016/j.lungcan.2020.07.005
Received 25 February 2020; Received in revised form 26 June 2020; Accepted 3 July 2020
Available online 12 July 2020

lead to the need for postponed treatment, reduced doses and in some cases even discontinuation of treatment [2].

Hematological toxicities, such as leukopenia and neutropenia, are common adverse drug reactions (ADRs) induced by gemcitabine/carboplatin [3,4]. The large variation in toxicity can cause some patients to display no or moderate symptoms while others experience severe ADRs or even death. Severe ADRs of grade 3–4 according to the common terminology criteria for adverse events (CTCAE) have been reported in 20–70 % of patients in clinical studies [2,5–8].

The underlying mechanisms for induced ADRs in patients treated with chemotherapy are to date not fully understood. Efforts have been made to find ways to predict hematological toxicity in chemotherapy treatment using candidate gene approaches and genome-wide association studies (GWAS) [9–14]. Present GWAS have either focused on one therapy towards mixed tumor types or mixed therapies against one tumor type and are mainly of Asian origin.

The aim of this study is to gain a better understanding of the genetic variability affecting chemotherapy-induced leukopenia and neutropenia in patients diagnosed with NSCLC undergoing gemcitabine/carboplatin treatment. Furthermore, the goal is to identify toxicity associated SNVs that can be used for prediction of ADRs using weighted genetic risk score (wGRS) models. Therefore, we performed whole-exome sequencing of gemcitabine/carboplatin treated NSCLC patients using state of the art high-throughput next-generation sequencing and association analyses of single-nucleotide variants (SNVs) and genes with leukopenia and neutropenia. The findings were partially validated using external GWAS data sets and lastly used for constructing wGRS prediction models.

2. Materials and methods

An overview of the presented study is illustrated in Fig. 1.

2.1. Study population

Starting in 2006, 215 patients were included, during two consecutive years, to the study which was approved by the regional ethics committee in Stockholm (DNR-0 3-4 13 and DNR-2016/2585-32/1). Written informed consent was obtained from each patient prior to enrollment in accordance with the Helsinki Declaration. The patients were diagnosed with NSCLC and scheduled for four treatment cycles of gemcitabine/carboplatin. The included patients received at least one cycle of gemcitabine (1250 mg/m^2) on days 1 and 8, and carboplatin (AUC = 5) on day 1. Parts of this study population has previously been described in an extreme-phenotype study [15] and the entire study population has been used to address genetic variation and thrombocytopenia [16].

2.2. Hematological toxicity

Leukocyte and neutrophil counts were monitored during the first treatment cycle and recorded at baseline and days 8, 15 and 21. The severity of leukopenia and neutropenia was measured using the nadir parameter, defined as the lowest measured blood count on day 8, 14 and 21, and the decrease parameter, defined as baseline adjusted nadir values (nadir/baseline).

2.3. DNA extraction, exome enrichment and sequencing

QIAamp® DNA mini-kits (VWR International, Stockholm, Sweden) were used to extract DNA from whole blood according to the manufacturer's protocol. Nextera Rapid Capture Exome kit (Illumina FC-140-1003, San Diego, USA) was used for target enrichment and library preparation according to the manufacturer's instructions. The DNA samples, 215 in total, were sequenced on Illumina HiSeq 2500 v4 at Science for Life Laboratory, Stockholm, Sweden.

2.4. Quality control, alignment and variant calling

The sequenced reads were processed using TrimGalore! (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and cutadapt [17] for quality and adapter trimming, removing reads with Phred quality score < 25 and pairs with one read length < 25. Further data processing involved Bowtie2 [18] for alignment to reference genome GRCh37.72, SAMtools [19] for quality filtering, Picard Tools [20] for duplicate removal and GATK for variant calling applying their best

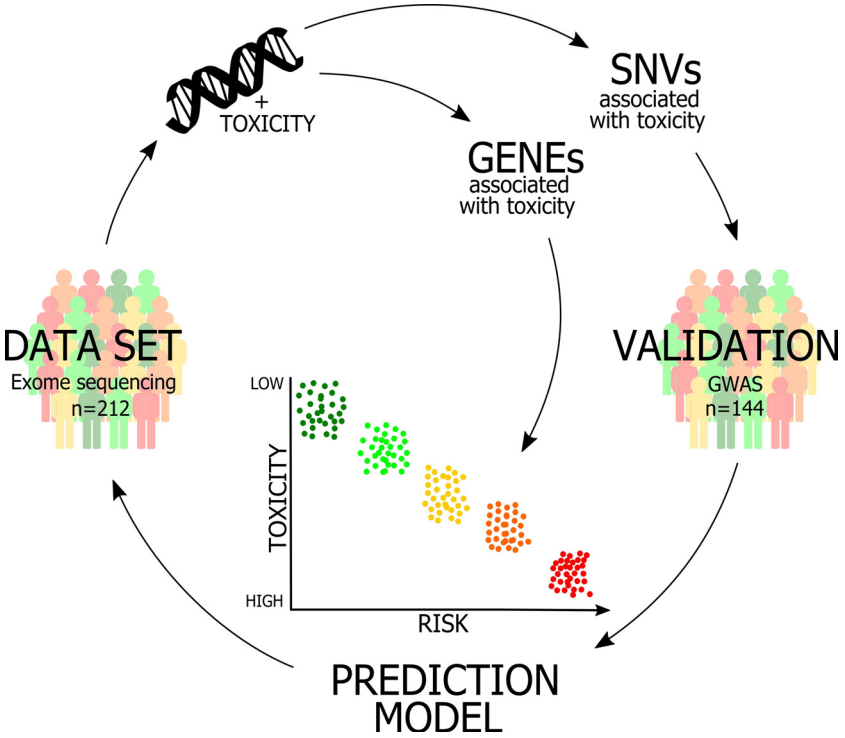

Fig. 1. Overview over the different steps involved in this project. First, the included NSCLC patients were whole-exome sequenced. Genetic variation was analyzed to find leukopenia and neutropenia associated SNVs and genes. The toxicity associated SNVs were partially validated with an external GWAS data set. A prediction model was thereafter set up using validated SNVs and SNVs located in toxicity associated genes and applied on the existing data set.

practices [21].

2.5. Post variant calling, quality control and outliers

Variants not labeled as PASS, with a genotyping rate < 0.95, with a mean coverage < 10 in all samples and out of Hardy Weinberg ($p < 0.0001$) were discarded using VCFtools [22].

PLINK was used for detection of outliers [23]. Identity by descent identified two samples (S0580 and S0664) that contained a large number of shared variants with the rest of the study population, indicating contamination. Identity by missingness identified two samples (S0328 and S0664) containing a large number of missing genotypes probably due to inadequate sequencing. These three samples were removed from the analysis leaving 212 of the 215 samples to remain for downstream analysis. The three outlier samples also diverged from the rest of the study population in the number of variants and usable reads, Supplementary Figure S1.

2.6. Phenotype considerations

The phenotype parameters, nadir and decrease, for leukopenia and neutropenia were normalized using van der Waerden rank scores implemented in the R package multic [24]. Transformation using natural logarithm was initially tested but rejected due to the high influence of one patient with extreme phenotype values.

Initial association analyses with permutations revealed that the nadir phenotype parameters generated a negligible signal compared to the background and were therefore not studied further, Supplementary Figure S2. The association analysis was thereafter only carried out for for the leukopenia and neutropenia decrease phenotype parameters. Leukopenia included 212 patients while neutropenia only included 129 patients due to missing baseline values.

2.7. Statistical analysis

Association analysis, adjusted for age and gender, was performed on SNVs and genes for leukopenia and neutropenia. No influence from other parameters such as stage or histology were identified. The SNV association analysis was performed on common SNVs (minor allele frequency (MAF) ≥ 0.01) in PLINK [23] using an additive genetic model. A p-value cut-off for indicating suggestive association was set to ≤ 2.2E-04, for leukopenia, and ≤ 1.35E-03, for neutropenia, after evaluation in PLINK using 1000 permutations to find the most advantageous false discover rate (FDR) in the p-value interval 0.00001-0.01, Supplementary Figure S2.

The gene association analysis was performed on all SNVs using SKATO in the R package SKAT [25,26]. The SNVs were assigned equal weight otherwise default settings were applied. RefSeq GRCh37/hg19 gene information was obtained from the UCSC table browser [27] and SNVs were mapped to the corresponding gene (exon region +/- 6 bases) using PLINK [23]. Genes that only contained one rare SNV (MAF < 0.01) were removed from the analysis.

2.8. Validation

2.8.1. Validation data set

GWAS data from Leandro-García et al. (2013) [28] was used as a validation data set. The study was performed on 144 patients with European origin homogenously treated with paclitaxel/carboplatin. The majority of the patients were diagnosed with ovarian (70 %) and lung cancer (19 %). Leukocyte and neutrophil counts were monitored throughout the treatment and leukopenia and neutropenia were defined as time to first toxic event (CTCAE ≥ 1). SNV association analysis, adjusted for age, was performed with Cox regression using an additive genetic model in PLINK (version 1.07). The validation data set ($p < 0.005$, MAF ≥ 0.01) consisted of 2965 SNVs associated with leukopenia and 3437 SNVs associated with neutropenia.

2.8.2. Validation method

Validation of SNVs was performed by examining pairwise linkage disequilibrium (LD) between SNVs of suggestive association in the presented study population with the SNVs in the validation GWAS data set using Ensembl REST API (version 6.3). Genotype data originated from the 1000 Genomes Project Phase 3 and pair-wise LD within a distance of 500 kb were studied. A European population panel was selected including CEU (Utah residents with Northern and Western European ancestry), FIN (Finnish in Finland), GBR (British in England and Scotland), IBS (Iberian populations in Spain) and TSI (Toscani in Italy). SNV pairs with a D' above 0.33 were considered to be in LD and thereby also support the validity of the SNV pair [29].

2.9. Assembly of prediction models

2.9.1. SNV selection

For SNVs to be included in the wGRS prediction models, the SNVs were required to be a partially validated SNVs or represented in both the SNV and gene association analyses. For leukopenia, eight other top SNVs were additionally included due to that to few eligible SNVs were otherwise available. In case the SNVs were in LD, only the SNV with the highest beta value was selected [30].

2.9.2. Weighted genetic risk scores (wGRS)

The toxicity risk prediction model was based on wGRSs created using the R package PredictABEL [31]. Beta values acquired in the SNV association analysis were used as weights for the selected SNVs. The wGRSs were determined by multiplying the beta value with the number of risk alleles (0, 1 or 2) and then summarized across all included SNVs. The wGRS were divided into five risk groups based on the distribution of the wGRS. Risk group 5 was defined as wGRS ≤ (mean wGRS - 1.5 SD), risk group 4 as (mean wGRS – 1.5 SD) < wGRS ≤ (mean wGRS – 0.5 SD), risk group 3 as (mean wGRS – 0.5 SD) < wGRS ≤ (mean wGRS + 0.5 SD), risk group 2 as (mean wGRS + 0.5 SD) < wGRS ≤ (mean wGRS + 1.5 SD) and risk group 1 as wGRS > (mean wGRS + 1.5 SD).

Differences in toxicity between the risk groups were determined by one-way ANOVA followed by Sidak's multiple comparison test using GraphPad Prism version 6.0 (GraphPad Software, La Jolla California USA).

3. Results

3.1. Patient characteristics

Patient and toxicity characteristics are outlined in detail in Supplementary Table S1. The median age of the study population was 64 (45–82) years, with a gender distribution of 53 % females and 47 % males (confirmed with sex check in PLINK). Tumor stage was distributed over stages I (18.9 %), II (13.2 %), III (29.7 %), IV (37.3 %) and not specified (0.9 %). The majority of the patients were diagnosed with adenocarcinoma (62.3 %) followed by squamous cell carcinoma (18.9 %), unspecified NSCLC (13.7 %), large cell carcinoma (4.7 %) and other (0.9 %). The highest proportion of the patients were former smokers (46.7 %) followed by current smokers (43.4 %) and never smokers (9.9 %). Leukopenia and neutropenia were assessed after the first treatment cycle according to CTCAE v4.03. Leukopenia of CTCAE 0, CTCAE 1–2 and CTCAE 3–4 were identified in 29.7 %, 47.2 % and 23.1 % of the patients, respectively. Neutropenia of CTCAE 0, CTCAE 1–2 and CTCAE 3–4 were identified in 36.7 %, 13.8 % and 49.5 % of the patients, respectively.

3.2. Sequencing characteristics

For the 215 exome sequenced samples, the average number of

Table 1
Top associated SNVs for leukopenia and neutropenia (p ≤ 1E-04).

	Chr	rsID	Alleles (ref/alt)	Gene symbol	Annotation	MAF[a]	CADD	P-value
Leukopenia	1	rs10927386	T/C	EFCAB2	intron	0.3632	7.609	2.52E-06
	1	rs10927387	C/T	EFCAB2	synonymous	0.3632	15.44	2.52E-06
	3	rs17854381	G/A	NEK10	synonymous	0.2807	8.163	3.12E-05
	3	rs17680166	G/C	NEK10	synonymous	0.2759	10.93	3.61E-05
	14	rs17128572	C/G	GOLGA5	missense	0.1132	24.5	5.17E-05
	14	rs17128593	C/G	GOLGA5	intron	0.1132	9.083	5.17E-05
	18	rs4065379	CTCTG/C	DLGAP1-AS1	intron	0.1108	6.561	6.09E-05
	14	rs17128583	A/G	GOLGA5	synonymous	0.1156	7.938	6.21E-05
	22	rs192201349	G/A	DMC1	intron	0.01415	3.217	6.25E-05
	17	rs79350244	A/C	DNAH2	missense	0.03774	24.2	6.28E-05
	17	rs117465420	A/T	DNAH2	missense	0.03774	19.64	6.28E-05
	17	rs117985215	G/A	KDM6B	intron	0.03774	6.304	6.28E-05
	3	rs1550769	C/G	NEK10	intron	0.2736	1.164	7.42E-05
	3	rs1550768	A/T	NEK10	intron	0.2736	1.964	7.42E-05
	3	rs11129280	G/T	NEK10	synonymous	0.2736	11.98	7.42E-05
	3	rs10510592	A/G	NEK10	missense	0.2736	9.834	7.42E-05
	8	rs7826836	T/G	KIAA1456	missense	0.2119	1.595	8.70E-05
Neutropenia	1	rs17131429	G/C	HFM1	intron	0.2618	7.811	1.53E-05
	2	rs2540923	A/G	STRN	intron	0.02358	9.572	2.50E-05
	1	rs11165778	A/G	HFM1	missense	0.2995	0.256	3.49E-05
	1	rs281992	T/G	HFM1	synonymous	0.3679	11.2	4.39E-05
	1	rs1800822	C/T	FMO3	synonymous	0.04717	10.34	4.84E-05
	19	rs429358	T/C	APOE	missense	0.1353	0.007	5.11E-05
	1	rs10493845	G/A	HFM1	synonymous	0.3656	7.491	6.62E-05
	19	rs3764605	A/G	ATP8B3	synonymous	0.4505	0.005	8.97E-05

Chr, Chromosome; Ref Reference allele; Alt, Alternative Allele; CADD Combined Annotation Dependent Depletion; MAF, Minor allel frequency.
[a] MAF was calculated within the cohort.

paired reads was 38.7 million. On average, the mapping yield was 99.2 % and the coverage was 74 X over the target region. The number of variants saturated around 45 million reads, Supplementary Figure S1. In the 212 samples, after removal of three outliers (see methods), 149713 filtered variants were identified and of which 71374 variants were common (MAF ≥ 0.01).

3.3. Genetic association to hematologic toxicity

The SNV association analysis identified 50 and 111 SNVs to be suggestively associated with leukopenia (p ≤ 2.2E-04, Supplementary Tables S2) and neutropenia (p ≤ 1.35E-03, Supplementary Tables S3), respectively. Of these, the top associated SNVs (p ≤ 1E-04) for leukopenia and neutropenia are listed in Table 1. The two SNVs, rs10927386 (p = 2.52E-06) and rs10927387 (p = 2.52E-06), with the lowest p-value associated with leukopenia reside in *EFCAB2*. Three (p ≤ 4.39E-05) out of the top four SNVs for neutropenia are in *HFM1*.

The gene association analysis identified 12 and 20 genes to be suggestively associated with leukopenia (p ≤ 2.2E-04, Supplementary Tables S4) and neutropenia (p ≤ 1.35E-03, Supplementary Tables S5), respectively. Of these the top associated genes (p ≤ 1E-04) for leukopenia and neutropenia are listed in Table 2. For leukopenia two genes were statistically significant after FDR correction for multiple testing, *EFCAB2* (FDR adjusted p = 0.047) and *SVIL* (FDR adjusted p = 0.047). The gene *EFCAB2* also harbors the two top SNVs, rs10927386 and rs10927387, identified in the SNV association analysis. The gene *SVIL* contains 36 SNVs in the gene association analysis of which four were also identified in the SNV association analysis with p ≤ 2.2E-04.

3.4. Validation

When comparing the validation GWAS data set with our data set, 112 and 155 SNV pairs within 500 kb were identified for leukopenia and neutropenia, respectively. LD was confirmed in 38 SNV pairs for leukopenia and 33 SNV pairs for neutropenia that resulted in partial validation of 20 unique SNVs for leukopenia and 19 unique SNVs for neutropenia. All the validated SNVs from our study for leukopenia and neutropenia and their corresponding GWAS validation SNVs are listed in Tables 3 and 4, respectively.

The majority of the validated SNVs are located within the genes, *EFCAB2*, *SVIL*, *NEK10*, *SERAC1* and *GTPBP1*, which were identified in the gene association analysis for leukopenia. The SNVs rs10927386 and rs10927387 in *EFCAB2* were validated with rs4658733 located in the neighboring gene *KIF26B*. In *NEK10*, rs17854381, rs17680166, rs1550769, rs1550768, rs11129280 and rs10510592 were validated with rs11129290 in *UBA52P4*, rs6768214 in *SLC4A7* and rs6774965 in *LOC105377005*. The SNV rs6929274 in *SERAC1* was validated with rs1744178, rs1750040 and rs17489570 located in the neighboring gene *SYNJ2*. For *SVIL*, the SNVs rs3740003, rs10160013 and rs1547169 were validated with the GWAS SNVs, rs1624281 and rs1148226, both also in SVIL.

3.5. Toxicity risk prediction model

Two wGRS toxicity risk prediction models were created, one for leukopenia including 20 SNVs and one for neutropenia including 28 SNVs, Supplementary Table S6. The wGRSs categorize patients into five risk groups based on their genetic variation in the model SNVs and their corresponding beta values as weights. The performance of the models is illustrated in, Fig. 2A and B, for leukopenia and neutropenia, respectively. Statistically significant differences in toxicity were observed between all risk groups except between risk group 4 and 5 for neutropenia.

4. Discussion

In the current study, we have shown how genetic variation correlates to leukopenia and neutropenia in NSCLC patients treated with gemcitabine/carboplatin, Fig. 1. Whole-exome sequencing was performed to identify SNVs, within the coding region of the genome, associated with chemotherapy-induced hematological toxicity. Genetic variation associated with leukopenia and neutropenia were studied on both SNV and gene level. Suggestively associated SNVs were partly validated with SNVs identified in an independent GWAS data sets by

Table 2
Top associated genes for leukopenia and neutropenia (p ≤ 1E-04).

	Chr	Gene symbol	Gene name	P-value	Number of variants	Variants in gene
Leukopenia	10	SVIL	Supervillin	**2.48E-06***	35	rs56022643, rs7921306, rs11007607, 10:29762907, rs61737920, rs10763720, rs11007612, rs1056782, rs56817459, rs146267453, rs7070135, rs145392867, rs7070678, 10:29813439, 10:29820187, rs17756919, rs41284748, rs1328323, rs147010426, rs150826046, rs1247696, rs7076239, rs142262993, rs138539716, rs143011277, rs141506698, 10:29839785, rs10160013, rs17834991, rs1270874, 10:29839886, rs3740003, rs3740002, rs1547169, rs375845375
	1	EFCAB2	EF-hand calcium binding domain 2	**4.63E-06***	6	rs111647414, rs56285780, rs10927386, rs10927387, rs142888002, rs115556370
	11	PNPLA2	Patatin like phospholipase domain containing 2	1.34E-05	9	rs145999340, rs10902224, rs140201358, rs56152088, rs1135628, rs137866968, rs150770244, rs11554663, 11:824625
	1	PIAS3	Protein inhibitor of activated STAT 3	2.70E-05	5	rs143409313, rs201406617, rs114124194, rs17354559, 1:145585452
	14	GOLGA5	Golgin A5	4.63E-05	9	rs17128572, rs76861952, rs34515753, rs17128583, rs1040835, rs17128593, rs117821736, rs79947836, rs374813643
	18	DLGAP1-AS1	DLGAP1 antisense RNA 1	6.85E-05	2	rs11661062, rs4065379
Neutropenia	1	HFM1	HFM1, ATP dependent DNA helicase homolog	1.55E-05	16	rs201592712, rs281992, rs74843031, rs11584478, rs10493845, rs143399622, 1:91816318, rs117177978, rs113908392, 1:91850794, rs12144808, rs11165778, rs199964473, rs376533937, 1:91870382, rs17131429
	22	MCM5	Minichromosome maintenance complex component 5	3.53E-05	10	rs34391116, 22:35802602, rs2307340, rs2230932, rs2230933, rs133417, 22:35809897, rs200017918, rs200883424, rs133427

Chr, Chromosome.
* Surpassed correction for multiple testing, false discovery rate (FDR) adjusted p < 0.05.

LD. The obtained result did not indicate that variation in toxicity is due to one SNV but rather suggested that several SNVs contribute with small effects. Therefore, the cumulative effect of several genetic variants was merged in wGRSs to create toxicity risk prediction models.

This study included 212 NSCLC patients homogeneously treated with gemcitabine/carboplatin, which to the best of our knowledge, is the largest study performed using whole-exome sequencing with drug-induced leukopenia and neutropenia as primary outcomes.

Leukopenia and neutropenia were monitored at several time points during the first cycle. The phenotype information for neutropenia was unfortunately not as comprehensive as for leukopenia. For neutropenia, the decrease phenotype parameter was only available for 129 patients since the neutrophil blood counts weren't routinely analyzed at the time of the study and the neutrophil count at baseline was therefore easily missed.

An optimal threshold for indicating suggestive association was determined using permutations for leukopenia and neutropenia to enable further studies of SNVs since no SNVs reached statistical significance after correction for multiple testing, Supplementary Figure S2. These permutation results indicated that the signal from the nadir phenotype parameters, for both leukopenia and neutropenia, were weak or absent compared to background noise and therefore the nadir parameters were excluded from the study. A prominent signal was on the other hand identified for the leukopenia decrease parameter at the cut-off at p ≤ 2.2E-04 (FDR = 28.5 %), which suggest that around 30 of the 50 identified SNVs should be true positives and of relevance for the phenotype. An FDR minimum was also observed for the neutropenia decrease parameter at p ≤ 1.35E-03 (FDR = 79.5 %) even though not as pronounced as for leukopenia. Hence, more focus has been directed towards the leukopenia phenotype as it should contain a larger proportion of true positive SNVs compared to neutropenia.

Despite that the nadir phenotype parameter is frequently used to assess toxicity during treatment, it might be of advantage to also add additional information about the initial blood count with the possibility to enable a more accurate and precise prediction when trying to individualize treatment before treatment start.

Due to the lack of additional patients in this study, and no previous whole-exome sequencing studies with a similar research question, we turned to a currently partly unpublished GWAS that has studied leukopenia and neutropenia in paclitaxel/carboplatin treated NSCLC patients. The validation data set is not ideal, as the treatment regimens and technique of retrieving genetic information differs between the studies. Despite that, our validation approach was to compare LD between tag-SNPs from the GWAS with the SNVs obtained in our whole-exome sequencing study. The approach has limitations, for instance, the SNVs causality cannot be determined, however, the regions identified in both studies and validated with LD indicate the SNVs importance and their corresponding genetic regions should be of broad interest to the research community.

It is worth mentioning that 70 % (14 out of 20 SNVs) of the partially validated leukopenia associated SNVs reside in genes that were also found in the gene association analysis. Two of the recurrently identified genes were *SVIL* and *EFCAB2*. They harbor SNVs identified among the top SNVs in the SNV association analysis, their gene association test were significant after correction for multiple testing and two SNVs, rs10927386 and rs10927387, in *EFCAB2* and three SNVs, rs3740003, rs10160013 and rs1547169, in *SVIL* were also partially validated with the GWAS data set.

SVIL encodes the peripheral membrane protein supervillin, expressed for instance in the bone marrow and in neutrophilic lipid rafts [32,33]. SVIL is known to bind both myosin II and filamentous actin, and interact with several cytoskeletal proteins [34,35]. The protein regulates all stages of cell motility and is involved in early cytokinesis [35]. Knockdown experiments of *SVIL* have shown to reduce cell division and increase cell death in HeLa and U2OS cell lines [36]. This indicates that genetic variation in *SVIL* might be important for and

Table 3
Validation results for leukopenia.

	Original Data					Validation Data					CEU		FIN		IBS		TSI		GBR	
CHR	rsID	Location	Gene symbol	P-value	MAF	rsID	Location	Nearest Gene	P-value	MAF	r^2	D'	r^2	D'	r^2	D'	r^2	D'	r^2	D'
1	rs10927386	245245378	EFCAB2	2.52E-06	0.36	rs4658733	245420757	KIF26B	9.24E-04	0.15	–	–	0.06	0.64	–	–	–	–	–	–
	rs10927387	245245402	EFCAB2	2.52E-06	0.36						–	–	0.06	0.64	–	–	–	–	–	–
3	rs17854381	27204052	NEK10	3.12E-05	0.28	rs11129290	27531796	UBA52P4	8.03E-04	0.16	0.23	0.84	0.36	0.79	0.27	0.87	0.37	0.91	0.38	0.88
						rs6768214	27456543	SLC4A7	1.79E-03	0.16	0.23	0.84	0.39	0.84	0.27	0.87	0.37	0.91	0.38	0.88
						rs6774965	27548813	LOC105377005	2.03E-03	0.14	0.23	0.84	0.32	0.77	0.25	0.82	0.38	0.95	0.34	0.87
	rs17680166	27243045	NEK10	3.61E-05	0.28	rs11129290	27531796	UBA52P4	8.03E-04	0.16	0.28	0.92	0.40	0.94	0.28	0.87	0.41	0.95	0.46	0.94
						rs6768214	27456543	SLC4A7	1.79E-03	0.16	0.28	0.92	0.43	1.00	0.28	0.87	0.41	0.95	0.46	0.94
						rs6774965	27548813	LOC105377005	2.03E-03	0.14	0.28	0.92	0.36	0.94	0.25	0.82	0.38	0.95	0.41	0.94
	rs1550769	27296796	NEK10	7.42E-05	0.27	rs11129290	27531796	UBA52P4	8.03E-04	0.16	0.29	0.92	0.46	0.95	0.33	0.94	0.45	1.00	0.50	0.95
						rs6768214	27456543	SLC4A7	1.79E-03	0.16	0.29	0.92	0.49	1.00	0.33	0.94	0.45	1.00	0.50	0.95
						rs6774965	27548813	LOC105377005	2.03E-03	0.14	0.25	0.85	0.42	0.94	0.30	0.88	0.42	1.00	0.45	0.94
	rs1550768	27296822	NEK10	7.42E-05	0.27	rs11129290	27531796	UBA52P4	8.03E-04	0.16	0.29	0.92	0.46	0.95	0.33	0.94	0.45	1.00	0.50	0.95
						rs6768214	27456543	SLC4A7	1.79E-03	0.16	0.29	0.92	0.49	1.00	0.33	0.94	0.45	1.00	0.50	0.95
						rs6774965	27548813	LOC105377005	2.03E-03	0.14	0.25	0.85	0.42	0.94	0.30	0.88	0.42	1.00	0.45	0.94
	rs11129280	27326097	NEK10	7.42E-05	0.27	rs11129290	27531796	UBA52P4	8.03E-04	0.16	0.29	0.92	0.46	0.95	0.33	0.94	0.45	1.00	0.50	0.95
						rs6768214	27456543	SLC4A7	1.79E-03	0.16	0.29	0.92	0.49	1.00	0.33	0.94	0.45	1.00	0.50	0.95
						rs6774965	27548813	LOC105377005	2.03E-03	0.14	0.25	0.85	0.42	0.94	0.30	0.88	0.42	1.00	0.45	0.94
	rs10510592	27332820	NEK10	7.42E-05	0.27	rs11129290	27531796	UBA52P4	8.03E-04	0.16	0.29	0.92	0.46	0.95	0.33	0.94	0.44	1.00	0.50	0.95
						rs6768214	27456543	SLC4A7	1.79E-03	0.16	0.29	0.92	0.49	1.00	0.33	0.94	0.44	1.00	0.50	0.95
						rs6774965	27548813	LOC105377005	2.03E-03	0.14	0.25	0.85	0.42	0.94	0.30	0.88	0.41	1.00	0.45	0.94
6	rs6929274	158571501	SERAC1	1.52E-04	0.32	rs1744178	158496856	SYNJ2	4.26E-04	0.29	0.13	0.88	–	–	0.13	0.83	0.12	0.65	0.16	0.90
						rs1750040	158498555	SYNJ2	2.88E-03	0.10	0.06	1.00	–	–	–	–	–	–	0.05	1.00
						rs17489570	158487953	SYNJ2	2.94E-03	0.09	0.06	1.00	–	–	–	–	–	–	–	–
10	rs3740003	29840038	SVIL	1.12E-04	0.25	rs1624281	29989182	SVIL	1.27E-03	0.48	0.11	0.58	0.11	0.41	–	–	–	–	–	–
						rs1148226	29970252	SVIL	2.28E-03	0.44	0.07	0.40	0.13	0.38	–	–	–	–	–	–
	rs10160013	29839787	SVIL	1.16E-04	0.25	rs1624281	29989182	SVIL	1.27E-03	0.48	0.11	0.58	0.11	0.43	–	–	–	–	–	–
						rs1148226	29970252	SVIL	2.28E-03	0.44	0.07	0.40	0.13	0.40	–	–	–	–	–	–
	rs1547169	29843833	SVIL	1.63E-04	0.25	rs1624281	29989182	SVIL	1.27E-03	0.48	0.11	0.58	0.11	0.43	–	–	–	–	–	–
						rs1148226	29970252	SVIL	2.28E-03	0.44	0.07	0.40	0.13	0.40	–	–	–	–	–	–
17	rs75664430	8064779	PER1	2.11E-04	0.25	rs4468690	7940754	ALOX15B	1.52E-04	0.17	0.07	0.34	–	–	–	–	–	–	–	–
						rs4287627	7932788	GUCY2D	4.06E-03	0.13	0.06	0.34	0.08	0.34	–	–	–	–	–	–
22	rs4987164	38934515	DMC1	2.16E-04	0.02	rs16998847	38646900	TMEM184B	4.72E-03	0.01	–	–	–	–	0.33	1.00	–	–	–	–
	rs11570392	38958260	DMC1	2.16E-04	0.02						–	–	–	–	0.33	1.00	–	–	–	–
	rs192201349	38964167	DMC1	6.25E-05	0.02						–	–	–	–	0.50	1.00	–	–	–	–
	rs16999173	39034786	FAM227A	2.16E-04	0.02						–	–	–	–	0.33	1.00	–	–	–	–
	rs16999297	39122038	GTPBP1	2.16E-04	0.02						–	–	–	–	0.25	1.00	–	–	–	–
	rs16999301	39125529	GTPBP1	2.16E-04	0.02						–	–	–	–	0.25	1.00	–	–	–	–
	rs138708	39138332	SUN2	1.43E-04	0.02						–	–	–	–	0.50	1.00	–	–	–	–

CHR, Chromosome; MAF, Minor allele frequency; OR, Odds ratio; CEU, Utah residents (CEPH) with northern and western european ancestry; FIN, Finnish in Finland; IBS, Iberian population in Spain; TSI, Toscani in Italy; GBR, British in England and Scotland.

Table 4
Validation results for neutropenia.

	Original Data					Validation Data					CEU		FIN		IBS		TSI		GBR	
CHR	rsID	Location	Gene symbol	P-value	MAF	rsID	Location	Nearest Gene	P-value	MAF	r^2	D'	r^2	D'	r^2	D'	r^2	D'	r^2	D'
2	rs497692	169789016	ABCB11	9.67E-04	0.46	rs831017	170145426	LRP2	4.06E-04	0.43	0.08	0.40	–	–	–	–	–	–	–	–
						rs831022	170149613	LRP2	7.52E-04	0.38	0.05	0.36	–	–	–	–	–	–	–	–
3	rs7432838	88190809	CGGBP1	7.29E-04	0.12	rs12106679	88003281	HTR1F	1.29E-03	0.17	0.15	0.47	0.11	0.36	0.21	0.56	0.15	0.42	0.29	0.53
4	rs34840208	79387586	FRAS1	1.05E-03	0.19	rs6847138	79319844	FRAS1	2.87E-03	0.36	0.18	1.00	0.17	1.00	0.18	1.00	0.20	0.90	0.25	1.00
						rs12505864	79384865	FRAS1	3.66E-03	0.35	0.18	1.00	0.17	1.00	0.19	1.00	0.23	1.00	0.26	1.00
						rs931606	79443850	FRAS1	4.11E-03	0.42	0.18	0.76	0.20	0.86	0.19	0.84	0.22	0.83	0.24	0.88
						rs1496598	79444029	FRAS1	4.11E-03	0.42	0.18	0.76	0.20	0.86	0.19	0.84	0.22	0.83	0.24	0.88
6	rs282117	89981413	GABRR2	1.14E-03	0.35	rs3777530	89909557	GABRR1	2.61E-04	0.17			–	–			0.11	0.82	–	–
						rs4519988	90087650	RRAGD	1.41E-03	0.26	0.08	0.58	–	–	0.05	0.47	–	–	–	–
						rs2038392	89711694	CYCSP16	2.51E-03	0.47	–	–	–	–	–	–	–	–	0.07	0.35
						rs205223	89743488	LOC100131124	3.38E-03	0.28	–	–	–	–	–	–	–	–	0.08	0.60
						rs12198870	89910715	GABRR1	4.13E-03	0.11	–	–	–	–	–	–	0.07	0.77	–	–
7	rs12536928	21765452	DNAH11	3.19E-04	0.48	rs2893061	21993767	CDCA7L	4.59E-05	0.02	–	–	–	–	–	–	0.07	1.00	–	–
	rs2288551	128049573	IMPDH1	7.97E-04	0.02	rs6467202	128308386	FAM71F2	4.84E-03	0.45	–	–	–	–	–	–	0.06	1.00	0.05	1.00
	rs6967301	141420768	WEE2	8.36E-04	0.47	rs217020	140991143	TMEM178B	4.39E-03	0.07	–	–	–	–	0.05	1.00	–	–	–	–
10	rs11817589	134219036	PWWP2B	9.46E-04	0.07	rs954820	133957761	JAKMIP3	2.44E-03	0.43	–	–	0.10	0.66	–	–	–	–	0.08	1.00
12	rs1048371	55248357	MUCL1	8.87E-04	0.46	rs17113238	54926329	NCKAP1L	2.74E-03	0.23	0.06	0.52			–	–	–	–	–	–
						rs1874309	54950227	PDE1B	4.51E-03	0.18	–	–	0.06	0.51	–	–	–	–	–	–
	rs34849596	109217007	SSH1	1.26E-04	0.31	rs803576	108828237	LINC01498	9.10E-04	0.19	–	–	–	–	0.06	0.35	–	–	–	–
16	rs191207351	56543827	BBS2	1.20E-03	0.01	rs4784683	56594499	MT4	4.72E-03	0.11	–	–	–	–	–	–	–	–	0.07	0.62
18	rs138484926	74728800	MBP	1.99E-04	0.01	rs13381277	74318610	LINC00683	1.21E-03	0.06	0.07	1.00	–	–	0.06	0.29	–	–	–	–
						rs17651075	74471817	CCND3P2	1.72E-03	0.22	–	–	–	–	–	–	–	–	0.10	1.00
						rs2036663	74363062	LOC107985151	2.68E-03	0.34	–	–	–	–	–	–	–	–	0.08	1.00
19	rs11085822	12985576	MAST1	1.18E-03	0.26	rs8107173	13035638	FARSA	7.17E-05	0.02	0.12	1.00	–	–	0.17	1.00	0.10	0.73	0.10	1.00
20	rs11780	43566787	PABPC1L	1.12E-03	0.42	rs6094035	43400262	RIMS4	2.06E-03	0.28	0.05	0.26	–	–	0.06	0.30	0.08	0.30	0.19	0.55
						rs4812861	43402541	RIMS4	2.06E-03	0.28	0.05	0.26	–	–	0.06	0.30	0.08	0.30	0.19	0.55
						rs1555300	43430490	RIMS4	4.78E-03	0.31	–	–	–	–	–	–	–	–	0.11	0.35
	rs495337	48522330	SPATA2	9.09E-04	0.49	rs6020166	48594678	SNAI1	3.63E-04	0.10	0.10	1.00	0.21	1.00	0.09	1.00	0.05	1.00	0.08	1.00
22	rs133427	35815880	MCM5	1.10E-03	0.08	rs9619494	35356763	ISX-AS1	1.62E-03	0.05	–	–	–	–	0.07	0.39	–	–	–	–
	rs133885	26159289	MYO18B	3.30E-04	0.41	rs574946	25843884	CRYBB2P1	9.33E-04	0.36	0.06	0.35	–	–	0.05	0.23	–	–	–	–
						rs4822669	26280942	MYO18B	2.02E-03	0.27	–	–	–	–	0.06	0.34	–	–	–	–
	rs4820044	31687313	PIK3IP1	3.64E-04	0.07	rs2413045	31665862	LIMK2	2.39E-03	0.15	0.36	1.00	0.22	1.00	0.35	1.00	0.17	1.00	0.31	1.00
	rs9609261	31743032	PATZ1	3.64E-04	0.07						0.36	1.00	0.22	1.00	0.30	0.89	0.13	0.79	0.31	1.00

CHR, Chromosome; MAF, Minor allele frequency; OR, Odds ratio; CEU, Utah residents (CEPH) with northern and western european ancestry; FIN, Finnish in Finland; IBS, Iberian population in Spain; TSI, Toscani in Italy; GBR, British in England and Scotland.

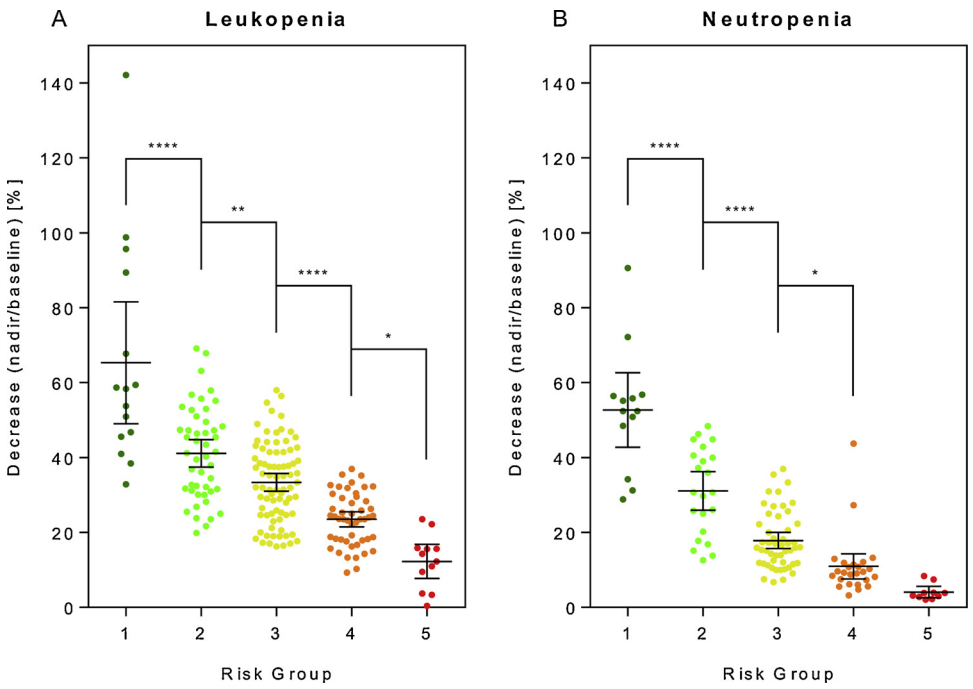

Fig. 2. Weighted genetic risk score (wGRS) models for leukopenia (A) and neutropenia (B). The patients are categorized in risk groups based on their wGRS that is determined based on the patients genetic variation in 20 and 28 SNVs associated with leukopenia and neutropenia, respectively. Risk group 5 represents the highest risk group that primarily contain patients that experienced a large decrease, having a low percentage of leukocytes and neutrophils remaining compared to their baseline levels. The error bars illustrate the average decrease with 95 % confidence interval (CI) for each risk group. Differences between the risk groups were analyzed using ANOVA with Sidak's multiple comparisons test. Notes: *p < 0.05; **p < 0.01; ****p < 0.0001.

could affect the recovery of blood cells after chemotherapy and thereby also the risk of receiving chemotherapy-induced leukopenia.

In the previous extreme-phenotype study [15], where 32 patients in this study with extreme phenotypes were participating, rs10160013 in *SVIL* was identified to be associated with thrombocytopenia. The SNV did, however, not pass the validation step. Otherwise, no clear overlap between the extreme phenotype study and this study was found. Reasons for this could be that different phenotypes and phenotype parameters were studied and the previous study only used parts of the patient material.

The second recurrently identified gene, *EFCAB2,* has been identified to be upregulated in zebrafish, *Danio rerio*, after addition of the transcription factor Ets1-related protein (etsrp) which is required for the formation of myeloid cells in zebrafish [37,38]. To date, not much is known about EFCAB2 but the upregulation after etsrp stimulation indicates that it can possess an important role in myelopoiesis.

To utilize the cumulative effect of the several associated SNVs, wGRS prediction models were created with the intention to, in advance, be able to predict which patients in risk of receiving severe gemcitabine/carboplatin-induced leukopenia or neutropenia, Fig. 2. The models rely on the ability to categorize the patient into a risk group based on a set of SNVs in combination with baseline leukocyte and neutrophil particle count. When the models were applied to our data the models were able to distinguish between different degrees of leukopenia and neutropenia decrease between the risk groups. Further validation of these models is, however, desirable to evaluate and validate their performance.

In conclusion, we have in this study identified and partly validated genetic biomarkers, SNVs and genes, that are associated with gemcitabine/carboplatin-induced leukopenia and neutropenia and created wGRS models for toxicity risk prediction. Of particular interest are the genes, *EFCAB2* and *SVIL*, since they were identified as significant in the gene association test and contain several top hits in the SNV association analysis that were partially validated against the independent GWAS data set. These results generate solid support for further investigation into chemotherapy-induced leukopenia and neutropenia. The results can also be important for other drugs were hematological toxicities are prominent adverse drug reactions or in other cancers where gemcitabine or carboplatin are also used.

Author contributions

JL and HG conceived and designed the study. AS, NB, BS, JL and HG developed the methodology used in the study and interpreted the data. EB, HK, RL, LDP, CR-A, MA-R, JL and HG acquired and managed patients, data and/or provided facilities. AS, NB, BS and SP conducted the bioinformatics and statistical analyses in the study. JL and HG supervised the study. All authors held administrative roles for reporting and organizing data as well as read and approved the final manuscript.

Funding

This work was supported by grants from the Swedish Cancer Society [Grant CAN 2016/602], the Swedish Research Council, Linköping University, ALF grants Region Östergötland, the Funds of Radiumhemmet, Marcus Borgströms stiftelse and the Spanish Ministry of Economy and Competitiveness (SAF2015-64850-R).

Ethics approval and consent to participate

The study was approved by the regional ethics committee in Stockholm (DNR-0 3-4 13 and DNR-2016/2585-32/1). Written informed consent was obtained from each patient prior to enrollment in accordance with the Helsinki Declaration.

Availability of data and material

The raw datasets generated and/or analyzed during the current study are not publicly available due to that this is not permitted by the ethical approval of the study but the data is available from the corresponding author upon reasonable request with the appropriate ethical approval.

Declaration of Competing Interest

The authors declare no potential conflicts of interest.

Acknowledgement

We gratefully 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 NBIS (National Bioinformatics Infrastructure Sweden) is gratefully acknowledged.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.lungcan.2020.07.005.

References

[1] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, CA Cancer J. Clin. 65 (1) (2015) 5–29 2015.
[2] C. Sederholm, et al., Phase III trial of gemcitabine plus carboplatin versus single-agent gemcitabine in the treatment of locally advanced or metastatic non-small-cell lung cancer: the Swedish Lung Cancer study Group, J. Clin. Oncol. 23 (33) (2005) 8380–8388.
[3] M. Barton-Burke, Gemcitabine: a pharmacologic and clinical overview, Cancer Nurs. 22 (2) (1999) 176–183.
[4] A.H. Calvert, et al., Early clinical studies with cis-diammine-1,1-cyclobutane dicarboxylate platinum II, Cancer Chemother. Pharmacol. 9 (3) (1982) 140–147.
[5] B.H. Gronberg, et al., Phase III study by the Norwegian lung cancer study group: pemetrexed plus carboplatin compared with gemcitabine plus carboplatin as first-line chemotherapy in advanced non-small-cell lung cancer, J. Clin. Oncol. 27 (19) (2009) 3217–3224.
[6] F. Imamura, et al., Randomized phase II study of two schedules of carboplatin and gemcitabine for stage IIIB and IV advanced non-small cell lung cancer (JACCRO LC-01 study), Chemotherapy 57 (4) (2011) 357–362.
[7] R.M. Rudd, et al., Gemcitabine plus carboplatin versus mitomycin, ifosfamide, and cisplatin in patients with stage IIIB or IV non-small-cell lung cancer: a phase III randomized study of the London Lung Cancer group, J. Clin. Oncol. 23 (1) (2005) 142–153.
[8] P. Zatloukal, et al., Gemcitabine plus cisplatin vs. Gemcitabine plus carboplatin in stage IIIb and IV non-small cell lung cancer: a phase III randomized trial, Lung Cancer 41 (3) (2003) 321–331.
[9] S. Cao, et al., Genome-wide association study of myelosuppression in non-small-cell lung cancer patients with platinum-based chemotherapy, Pharmacogenomics J. (2015).
[10] B. Han, et al., Association of ABCC2 polymorphisms with platinum-based chemotherapy response and severe toxicity in non-small cell lung cancer patients, Lung Cancer 72 (2) (2011) 238–243.
[11] K. Kiyotani, et al., A genome-wide association study identifies four genetic markers for hematological toxicities in cancer patients receiving gemcitabine therapy, Pharmacogenet. Genomics 22 (4) (2012) 229–235.
[12] J.K. Lamba, et al., Genetic variation in platinating agent and taxane pathway genes as predictors of outcome and toxicity in advanced non-small-cell lung cancer, Pharmacogenomics 15 (12) (2014) 1565–1574.
[13] S.K. Low, et al., Genome-wide association study of chemotherapeutic agent-induced severe neutropenia/leucopenia for patients in Biobank Japan, Cancer Sci. 104 (8) (2013) 1074–1082.
[14] J. Qian, et al., Association between CASP8 and CASP10 polymorphisms and toxicity outcomes with platinum-based chemotherapy in Chinese patients with non-small cell lung cancer, Oncologist 17 (12) (2012) 1551–1561.
[15] H. Green, et al., Using whole-exome sequencing to identify genetic markers for carboplatin and gemcitabine-induced toxicities, Clin. Cancer Res. (2015).
[16] N. Björn, et al., Genes and variants in hematopoiesis-related pathways are associated with gemcitabine/carboplatin-induced thrombocytopenia, Pharmacogenomics J. (2019).
[17] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, EMB.net J. 17 (1) (2011) 2011.
[18] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (4) (2012) 357–359.
[19] H. Li, et al., The sequence Alignment/Map format and SAMtools, Bioinformatics 25 (16) (2009) 2078–2079.
[20] P. Harter, et al., Is there a role for HIPEC in ovarian cancer? Arch. Gynecol. Obstet. 298 (5) (2018) 859–860.
[21] M.A. Depristo, et al., A framework for variation discovery and genotyping using next-generation DNA sequencing data, Nat. Genet. 43 (5) (2011) 491–501.
[22] P. Danecek, et al., The variant call format and VCFtools, Bioinformatics 27 (15) (2011) 2156–2158.
[23] S. Purcell, et al., PLINK: a tool set for whole-genome association and population-based linkage analyses, Am. J. Hum. Genet. 81 (3) (2007) 559–575.
[24] E. Lunde, et al., Multic: Quantitative Linkage Analysis Tools Using the Variance Components Approach, (2014).
[25] I. Ionita-Laza, et al., Sequence kernel association tests for the combined effect of rare and common variants, Am. J. Hum. Genet. 92 (6) (2013) 841–853.
[26] S. Lee, M.C. Wu, X.H. Lin, Optimal tests for rare variant effects in sequencing association studies, Biostatistics 13 (4) (2012) 762–775.
[27] D. Karolchik, et al., The UCSC table browser data retrieval tool, Nucleic Acids Res. 32 (DATABASE ISS) (2004) D493–D496.
[28] L.J. Leandro-García, et al., Genome-wide association study identifies ephrin type a receptors implicated in paclitaxel induced peripheral sensory neuropathy, J. Med. Genet. 50 (9) (2013) 599–605.
[29] T. Strachan, A.P. Read, Human Molecular Genetics, 4 ed, (2011) p. 482.
[30] M. Kircher, et al., A general framework for estimating the relative pathogenicity of human genetic variants, Nat. Genet. 46 (3) (2014) 310–315.
[31] S. Kundu, et al., PredictABEL: an R package for the assessment of risk prediction models, Eur. J. Epidemiol. 26 (4) (2011) 261–264.
[32] T. Nebl, et al., Proteomic analysis of a detergent-resistant membrane skeleton from neutrophil plasma membranes, J. Biol. Chem. 277 (45) (2002) 43399–43409.
[33] R.K. Pope, et al., Cloning, characterization, and chromosomal localization of human superillin (SVIL), Genomics 52 (3) (1998) 342–351.
[34] Y. Chen, et al., F-actin and myosin II binding domains in supervillin, J. Biol. Chem. 278 (46) (2003) 46094–46106.
[35] T.C. Smith, Z. Fang, E.J. Luna, Novel interactors and a role for supervillin in early cytokinesis, Cytoskeleton (Hoboken) 67 (6) (2010) 346–364.
[36] Z. Fang, E.J. Luna, Supervillin-mediated suppression of p53 protein enhances cell survival, J. Biol. Chem. 288 (11) (2013) 7918–7929.
[37] G. Gomez, et al., Identification of vascular and hematopoietic genes downstream of etsrp by deep sequencing in zebrafish, PLoS One 7 (3) (2012).
[38] S. Sumanas, et al., Interplay among Etsrp ER71, Scl, and Alk8 signaling controls endothelial and myeloid cell formation, Blood 111 (9) (2008) 4500–4510.