Pharmacogenomic characterization of gemcitabine response – a framework for data integration to enable personalized medicine

Michael Harris[a,b], Krithika Bhuvaneshwar[a], Thanemozhi Natarajan[a], Laura Sheahan[a,e], Difei Wang[a], Mahlet G. Tadesse[d], Ira Shoulson[c], Ross Filice[f], Kenneth Steadman[b], Michael J. Pishvaian[b], Subha Madhavan[a,b] and John Deeken[b]

Objectives Response to the oncology drug gemcitabine may be variable in part due to genetic differences in the enzymes and transporters responsible for its metabolism and disposition. The aim of our in-silico study was to identify gene variants significantly associated with gemcitabine response that may help to personalize treatment in the clinic.

Methods We analyzed two independent data sets: (a) genotype data from NCI-60 cell lines using the Affymetrix DMET 1.0 platform combined with gemcitabine cytotoxicity data in those cell lines, and (b) genome-wide association studies (GWAS) data from 351 pancreatic cancer patients treated on an NCI-sponsored phase III clinical trial. We also performed a subset analysis on the GWAS data set for 135 patients who were given gemcitabine + placebo. Statistical and systems biology analyses were performed on each individual data set to identify biomarkers significantly associated with gemcitabine response.

Results Genetic variants in the ABC transporters (*ABCC1, ABCC4*) and the CYP4 family members *CYP4F8* and *CYP4F12*, *CHST3*, and *PPARD* were found to be significant in both the NCI-60 and GWAS data sets. We report significant association between drug response and variants within members of the chondroitin sulfotransferase family (CHST) whose role in gemcitabine response is yet to be delineated.

Conclusion Biomarkers identified in this integrative analysis may contribute insights into gemcitabine response variability. As genotype data become more readily available, similar studies can be conducted to gain insights into drug response mechanisms and to facilitate clinical trial design and regulatory reviews. *Pharmacogenetics and Genomics* 24:81–93 © 2014 Wolters Kluwer Health | Lippincott Williams & Wilkins.

Introduction

Gemcitabine (2′-deoxy-2′,2′-difluorocytidine, dFdC), an analog of deoxycytidine with proven anti-tumorigenic effects, is used in the treatment of solid tumors including pancreatic cancer. Gemcitabine is active toward many solid tumor types but has a narrow therapeutic index and variable responses ranging from lack of efficacy to severe cytotoxicity, which may be attributed to variability in drug

exposure and metabolism [1]. Significant variations in individual response to gemcitabine therapy are common among pancreatic cancer patients. Earlier studies based on cell lines as well as patient–control populations have demonstrated that interindividual variations in germline DNA can impact cellular response to oncology drugs. Identification of such variants with functional/regulatory impact can serve to predict toxicity and efficacy of chemotherapeutic agents [2]. In this study, we have focused on genes encoding drug-metabolizing enzymes and transporters (DMETs), and their association with response to gemcitabine. The products of DMET genes play a substantial role in drug pharmacokinetics, and may have a role in predicting response and clinical outcomes in cancer patients. Some of these variants can not only

---

DOI: 10.1097/FPC.0000000000000015

impact drug metabolism and transport, but also affect the expression of cancer-related signaling proteins in downstream pathways.

Several potentially actionable variants have been reported in DMET genes that affect drug toxicity and efficacy among individuals, yet few have been tested in the clinic. Several variants in genes directly involved in gemcitabine metabolism have been reported to impact gemcitabine response (e.g. deoxycytidine kinase, DCK; DNA polymerase epsilon, POLE; cytidine deaminase, CDA; and transporters: SLC28A1, SLC28A2, SLC28A3, SLC29A1, SLC29A2, ABCB1, ABCC2, and ABCC10). A significant association has been demonstrated between gemcitabine sensitivity and variants in one or more of the above genes [3–5] including associations between single-nucleotide polymorphism (SNP) haplotypes and gemcitabine treatment outcome in pancreatic cancer patients [6–8].

Regulatory viewpoint

Developing innovative clinical evaluation tools and advancing personalized medicine has been identified by the FDA as a core priority area in advancing regulatory science [9]. Understanding the relationship between genetic markers and response to medical products, in terms of both efficacy and toxicity, is a critical component of this priority. First steps toward this understanding include modeling this relationship with a combination of in-vitro and in-vivo genomic and phenotypic markers as described in this paper. Prospective validation of such models along with consideration for regulatory approval and clinical adoption of resultant tools will be required, but work such as this helps lay a strong foundation for more personalized, effective, and safe medical product use.

Importance of data integration to determine clinically actionable variants

Understanding the genetic and molecular mechanisms underlying complex diseases such as cancer is extremely challenging. Genome-wide association studies (GWAS) have been extensively used in the past decade to discover important genetic variants. However, the identified SNPs explain only a small proportion of the phenotypic variation, and the predictive power of these SNPs remains low for many complex diseases [10]. To fully elucidate genetic underpinnings of disease a systems biology approach is necessary to characterize variants, mRNA, copy number, proteins, and metabolites, as well as their cellular interactions [11]. Gene set and pathway association analyses are playing an increasingly important role in explaining disease mechanisms through the identification of functional genetic interactions [12]. Many gene–disease association analyses are based on SNP genotype profiling or gene expression studies. However, SNPs can influence many downstream processes including the expression levels of multiple genes and/or protein

levels, and variations in expression levels can directly or indirectly impact disease progression and even drug response [13]. An integrative approach combining multiple data types can more accurately capture pathway associations [12] for discovery of clinically actionable variants.

Statistical approaches commonly used to associate variants with disease and/or drug response

Fisher's exact test (FET) is commonly used in the association of germline polymorphisms with drug response [14]. The use of probabilistic networks in conjunction with traditional statistical models for mining relationships and associations from genotype–phenotype data is well established [15]. Probabilistic network methods for pharmacogenomics and newer methods such as the Markov Blanket concept may be helpful to better analyze these complex genotype–phenotype associations [16]. Considering the complexity of both cancer prognosis and individual drug response to chemotherapeutics, application of these association methods in conjunction with novel informatics and data integration approaches is necessary to identify clinically relevant variants for validation studies and ultimately testing in the clinic for pharmacogenomics applications.

Methods
Data sets

We analyzed SNPs in DMET genes from gemcitabine-focused studies on cell lines and patients to identify associations with drug response. The analysis workflow is summarized in Fig. 1.

(1) *NCI-60 data sets*:
  (a) *SNP data*: DNA from the NCI-60 cell lines was provided by the Developmental Therapeutics Program (DTP) of the National Cancer Institute (NCI). The NCI-60 are well-characterized tumor cell lines. DNA was analyzed using the Affymetrix Targeted Human Drug Metabolizing Enzymes and Transporters (DMET) 1.0 chip [17], which determines the genotype for 1256 variants in 170 genes involved in drug disposition. An additional 14 variants in DNA repair enzymes of interest outside of the DMET chip were also genotyped due to their potential role in a number of anticancer drug pathways.
  (b) *Gene expression data*: This published data set consists of mRNA expression of the NCI-60 cell lines. Raw data on the Affymetrix U133A gene chip, obtained from the Gene Expression Omnibus (GEO, accession number GSE5720), were used in our analysis [18].
  (c) *Drug sensitivity data*: Drug sensitivity information is denoted by its GI50 value. The GI50 concentration is defined as the concentration required to achieve 50% growth inhibition. A data set containing the

Fig. 1
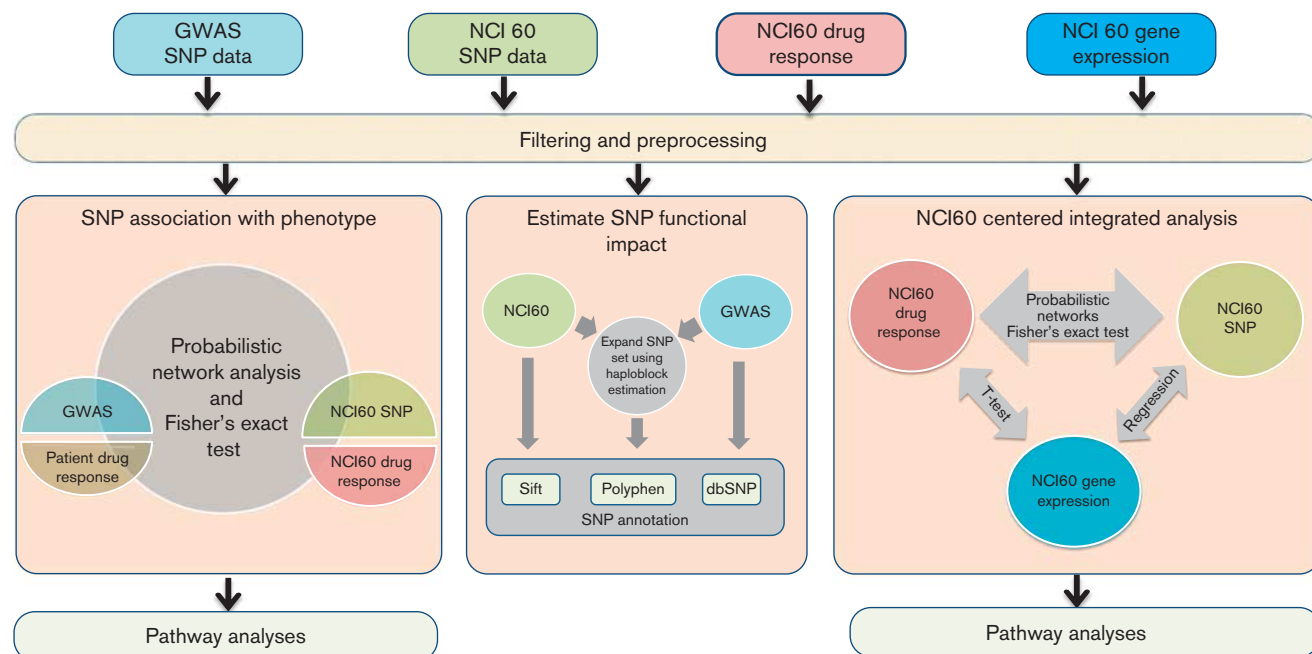
Analysis workflow. GWAS, genome-wide association studies; SNP, single-nucleotide polymorphism.

log 10 of mean GI50 values for the NCI-60 cell lines over seven experiments at a concentration of 10^{-6} mol/l for the drug of interest (gemcitabine, NSC613327) was queried and downloaded from the NCI DTP website *http://dtp.nci.nih.gov*. These data have been provided as Supplementary File S1 (Supplemental digital content 1, *http://links.lww.com/FPC/A660*). These drug sensitivity data for cell lines are analogous to patient outcome data in response to the same drug.

(2) *GWAS: all patients data set*: GWAS data from 351 patients with advanced pancreatic cancer treated on a CALGB phase III clinical trial with gemcitabine with or without bevacizumab (CALGB 80303) were downloaded from dbGaP [dbGaP: phs000250.v1.p1]. Patient outcomes were classified based on RECIST criteria, including CR (complete response), PR (partial response), SD (stable disease), PD (progressive disease), and unevaluated or NA (not applicable). Patients with clinical response reported as 'unevaluated' or NA were excluded from further analysis. Of the 561 466 SNPs in the data, we extracted 2847 SNPs located in genes contained on the Affymetrix DMET platform (Affymetrix Inc., Santa Clara, California, USA) for analysis. Results from the GWAS study showed no significant association between bevacizumab and overall survival [19]. We therefore used the entire cohort for analysis to maximize the input data set.

(3) *GWAS: gemcitabine + placebo data set*: We analyzed the GWAS data set on the subgroup of 135 patients who

were given gemcitabine and placebo to remove the effects of bevacizumab on the results.

Filtering and preprocessing of data sets

The NCI-60 gene expression data were normalized using the Robust Multichip Average (RMA) method in the R Bioconductor package (Affymetrix Inc.) [20]. All SNP data sets were uniformly processed. Samples with more than 60% missing genotype calls were removed. SNPs with the same genotype across all samples were filtered out since they did not contribute new information. Missing calls, denoted as NoCall (NC), PossibleRareAllele (PRA), or NotAvailable (NA) for the NCI-60 SNP data and 0/0 for the GWAS all patients data set were ignored. SNPs with only reference genotype and missing calls across all samples were also filtered out. After filtering, there were 59 cell lines and 432 variants remaining in the NCI-60 SNP genotype data.

For the GWAS: all patients data set, patients with responses CR and PR were aggregated as they were assumed to be more 'sensitive' to drug therapy; similarly, patients with PD and SD were aggregated as they were assumed to be more 'resistant' to drug therapy. The GWAS: all patients data set had 293 patients with 2846 variants following this aggregation and filtering. After preprocessing of the GWAS: gemcitabine + placebo data set, there were 2837 variants and 135 patients for analysis.

We chose two methods to analyze these SNP data sets: FET and Probabilistic Network Analysis (PNA) using the software BayesiaLab 5.0.6 *http://www.bayesia.com/en/ products/bayesialab.php*. FET is a well-known nonparametric test that evaluates the association between categorical variables [21]. FET was implemented in the PLINK Whole Genome Association Analysis Toolset [22], which conducts an allele-based test to identify genotype–phenotype association.

A probabilistic network (PN) is a representation of a joint probability distribution [23]. PNA was conducted using the genotype data and gemcitabine response values for each SNP data set. PNs can learn complex models and therefore offer an attractive option to analyze pharmacogenomic SNP data for discovery and prediction [24,25]. They also allow for the opportunity to build clinical decision support systems in the future. Each SNP was represented as a node in the network, whose states are the genotype values observed for that SNP. Gemcitabine response was represented as a special node called the target node, whose states were the discrete gemcitabine response values (sensitive or resistant).

Probabilistic network analysis requires discrete outcome variables, and hence the GI50 values from the NCI-60 gemcitabine response were dichotomized into two groups that correspond to resistant and sensitive phenotypes. We decided to use two levels to ensure adequate sample size in each group. To understand the distribution of the data, a histogram of the log 10 (GI50) values with the corresponding kernel density estimate were plotted (Fig. 2). The smallest antimode, observed at −6.875, was used as the cutoff value [26] between Resistant and Sensitive groups, as displayed by the vertical red line. After the above filtering steps, we produced a consoli-

dated file containing the variant data and associated drug response values. We then used this as input for the SNP analysis described below.
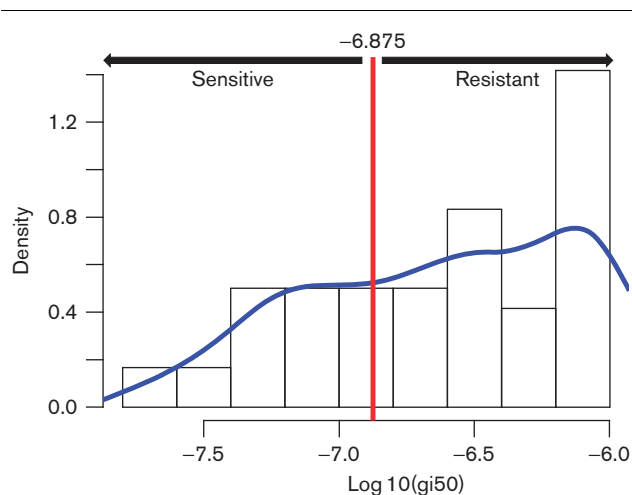
Estimation of haploblocks
We estimated haploblocks using PLINK to estimate the functional impact of the SNPs and to annotate SNPs located in intergenic regions, which can be associated with SNPs in gene or coding regions if they fall in the same haplotype block. SNPs that fall outside of coding regions (e.g. intronic or UTR regions) are more likely to have a significant impact on gene function if they fall in the same haplotype block as other high-impact coding SNPs. Haplotype blocks were not used for the association analysis and were solely used to determine SNP–gene associations and for functional annotation.

SNP comparative analysis of gemcitabine response
We investigated the association between genotype data and the gemcitabine response in each data set. Genes found to be significant were considered for further downstream analysis. We also identified genes found in more than one data set as these are likely to play a critical role in gemcitabine response.

For the FET and haplotype block estimation, a PLINK input file was created for each of the SNP data sets. SNPs that were not biallelic were not considered because PLINK only works on biallelic SNPs. FET results with a *P*-value less than 0.05 were considered significant.

Of the numerous network learning algorithms in PNA, we applied the Augmented Markov Blanket (AMB) algorithm because it has the ability to subset a limited number of SNPs that best predict the outcome [16]. A Markov Blanket corresponds to a set of nodes in the network that make the target independent of all the other nodes conditional on this subset of nodes. The AMB algorithm attempts to find a Markov Blanket for the target node (in this case gemcitabine response) and then uses an unsupervised search to find direct dependencies between each variable belonging to this Markov Blanket. The AMB algorithm implemented in BayesiaLab uses some of the concepts from the Smart-Greedy + algorithm [27]. Specifically, it combines a score-based Augmented Markov Blanket learning algorithm with principles inspired by the constraint-based approaches. The AMB is a nonstochastic learning algorithm that always returns the same result given the same input data and parameters. The AMB algorithm hence created an optimal network with SNPs that were best associated with gemcitabine response [24,28,29].

The genotype data were input 'as is' and no further coding was carried out, nor was any genetic model assumed. The parameter we controlled was the 'structural coefficient,' which allowed for controlling the extent of the blanket. We first found the optimal value of this

Fig. 2

Histogram and density plot for gemcitabine.

coefficient by repeating the AMB algorithm five times with different values of the structural coefficient in a given interval between 0.1 and 1. At each iteration, the network structure was learned on the entire data set with a growing structural coefficient. This analysis, called 'Structural coefficient analysis', plotted the Structure/Data ratio for different values of the structural coefficient (SC). The 'knee area' (the area before the strong increase in the graph) gave the optimal range of the coefficient to be used in the network. From this analysis, we obtained the optimal SC of 0.3 for the NCI-60 and the GWAS: all patients data sets, and 0.35 for the GWAS: gemcitabine + placebo data set. This optimal structural coefficient obtained was used to create the network for each data set. To confirm the stability of the networks, we ran cross-validation with data perturbation to ensure that the frequency measures for arc and node confidence were robust. The structural coefficient analysis, the final network obtained from PNA, and the cross-validation results are shown in Supplementary File S2 (Supplemental digital content 2, *http://links.lww.com/FPC/A661*). Once the network was created a 'target analysis report' was made, which ranked the nodes according to the information they brought to the knowledge of the target variable. This is defined as 'mutual information'. This report also displays *P*-values from an independence test based on the χ^2-distribution computed between each variable in the network and the target variable. The mutual information and the *P*-values were used to further prune the SNPs. Only SNPs with *P*-values less than 0.05 and mutual information of greater than 0 were retained inside the Augmented Markov Blanket.

The resulting SNPs from each test were annotated using the Affymetrix DMET annotation file, and NCBI's dbSNP database. The union of the significant SNPs obtained from FET and PNA was considered for further downstream analysis. By taking the union of the results across data sets used, we generated an accurate and comprehensive set of SNPs that are best associated with gemcitabine response.

NCI-60 integrative analysis of SNP, gene expression, and drug response data in NCI-60 cell lines

We first associated gene expression data with gemcitabine response by applying Student's two sample *t*-tests to the normalized gene expression levels to compare sensitive samples to those that were resistant. The 58 cell lines that had both gene expression and GI50 data were considered.

Next, to understand the association between SNPs and gene expression levels, we applied simple linear regression models relating each gene expression value to each SNP [30]. SNPs were considered as the independent variables and the gene expression data levels were the dependent variables. For each SNP, the most frequent genotype was used as the reference to avoid making genetic model assumptions. If a SNP had k genotypes $(k-1)$, indicator variables were introduced ($k = 2, 3$).

Finally, the above analyses were combined. Using the significant SNPs from PNA and FET, we extracted the significant gene probe sets affected by these SNPs from the regression results. Only those probe sets that were significant with respect to gemcitabine response from the *T*-test results were selected [7], resulting in a consolidated list of SNPs that affect expression of a gene that in turn affects gemcitabine response.

This integrative analysis was performed in R (*http://www.R-project.org*), and can be repeated on any drug of interest for which GI50 data are available and can help identify SNPs and probe sets best associated with drug response. For each analysis we applied multiple testing correction using Benjamini and Hochberg false discovery rates (FDR) [31]. Because of the discovery nature of this analysis, and hence to maximize the inclusion criteria, the adjusted p-values were not used as a filtering threshold for integration.

Systems biology analysis

The significant genes found in each data set were validated against the literature, and pathway analysis was performed to determine their potential role in one or more of the following processes: gemcitabine metabolism, pancreatic cancer, or interaction with other proteins implicated in the response to gemcitabine or other drugs used to treat pancreatic cancer. We also used the results from the NCI-60 integrated analysis to see whether such associations could be explained at the molecular level using pathway analysis. Pathway analysis was performed using IPA (Ingenuity Systems Inc., Redwood City, California, USA; *http://www.ingenuity.com*) and Ariadne's Pathway Studio (Rockville, Maryland, USA).

Results

SNP comparative analysis of gemcitabine response

The PLINK-based FET on the NCI-60 data set resulted in 25 significant SNPs ($P < 0.05$). PNA analysis identified three SNPs and two of these overlapped with the FET results. In total there were 26 unique SNPs corresponding to 14 unique genes. The genes *ABCC1* (rs8187858, FET, $P = 0.001$ [32], cds-synon), *CHST3* (rs4148943, FET, $P = 0.002$, UTR-3), *ABCC4* (rs4148551, FET, $P = 0.01$, UTR-3), *CYP2E1* (rs3813867, FET, $P = 0.01$ [33,34], nearGene-5), and *ALDH3A1* (rs887241, FET, $P = 0.01$, missense [35]) were among the most significant hits. Of the 14 genes identified, three are from the ABCC family (*ABCC1*, *ABCC4*, *ABCC6*), and two are from the CHST family (*CHST3*, *CHST13*). CYP genes identified include: *CYP2D6*, *CYP2E1* and *CYP4F2*, *CYP4F8*, and *CYP4F12*. *ALDH3A1*, *PPARD*, *SLCO1B1*, and *VKORC1* were also identified as significant. Functional analysis of the significant SNPs identified two as damaging: rs2108622,

a missense mutation in the *CYP4F2*; and rs1056522, a missense mutation in the *CHST13* gene.

The FET on the GWAS: gemcitabine + Placebo data set resulted in 123 significant SNPs ($P < 0.05$). PNA analysis identified 19 significant SNPs, of which 12 overlapped with the FET results. In total there were 130 unique SNPs corresponding to 50 unique genes. Top hits include: *CYP39A1* (rs3799884 FET, $P = 2.0E-4$, intron), *ABCC4* (rs7993878, FET, $P = 2.3E-4$, intron), *SLC6A6* (rs2327896, FET, $P = 5.1E-4$, UTR-3), *SLC29A2* (rs2279861, FET, $P = 1.1E-3$, intron), and *ABCB1* (rs17327624, FET, $P = 1.3E-3$, intron). Among the 50 genes identified were three members of the ABCC family (*ABCC1*, *ABCC4*, *ABCC8*) as well as two CHST family members (*CHST8*, *CHST11*). Twelve CYP genes were identified including two members of the CYP2 family (*CYP2B6*, *CYP2J2*) and four members of the CYP4 family (*CYP4B1*, *CYP4F3*, *CYP4F8*, *CYP4F12*). Functional analysis of the significant SNPS identified rs4646487, a missense mutation in the *CYP4B1* gene, as damaging.

The FET on the GWAS: all patients data set resulted in 121 significant SNPs ($P < 0.05$). PNA analysis identified 18 SNPs of which nine overlapped with the FET results. In total there were 130 unique SNPs corresponding to 54 unique genes. Top hits include: *CYP4F10P* pseudogene (rs1543284, FET, $P = 5.0E-5$), *CYP4F10P* pseudogene (rs1543285, FET, $P = 2.2E-4$), rs1063803 (no direct gene association but in the same LD block as SNPs in the *CYP4F3* gene, FET, $P = 3.6E-4$), rs3794989 (no direct gene association but in the same LD block as SNPs in the *CYP4F8* gene, FET, $P = 4.9E-4$), *CHST8* (rs17325358, FET, $P = 1.5E-3$, intron), *SLC6A6* (rs2341985, $P = 1.5E-3$, UTR-3), *CYP39A1* (rs3799884, FET, $P = 1.8E-3$, intron), *CHST11* (rs312172, FET, $P = 2.1E-3$, intron), and *SLC29A2* (rs2279861, FET, $P = 2.1E-3$, intron). The 54 genes identified contained three ABCC family members (*ABCC1*, *ABCC4*, *ABCC8*), five members of the CHST family (*CHST3*, *CHST6*, *CHST8*, *CHST9*, *CHST11*), and 10 CYP genes including three members from the CYP4 family (*CYP4B1*, *CYP4F3*, *CYP4F8*), and 10 SLC genes were identified including two members of the SLC7A subfamily (*SLC7A7*, *SLC7A8*) and two members of the SLC22A subfamily (*SLC22A3*, *SLC22A6*).

Comparison of the significant genes from each data set resulted in the genes *ABCC1*, *ABCC4*, and *CYP4F8* common to all three data sets. The genes *CYP4F12* and *PPARD* were found in both the NCI-60 and the GWAS: gemcitabine + placebo data sets. *CHST3* was found in both the NCI-60 and the GWAS: all patients data set. Thirty-four genes were found in common between the GWAS: all patients and GWAS: gemcitabine + placebo data sets including three members of the UGT2 family (*UGT2A1*, *UGT2A2*, *UGT2B4*).

Table 1 lists the top genes found in each set. Cells in the table are marked 'Yes' if an SNP associated with the gene

Table 1 Genes common to the two study sets after significance analysis

Gene name	NCI-60	GWAS: gemcitabine + placebo	GWAS: all patients
ABCC1	Yes	Yes	Yes
ABCC4	Yes	Yes	Yes
CYP4F8	Yes	Yes	Yes
CYP4F12	Yes	Yes	No
CHST3	Yes	No	Yes
PPARD	Yes	Yes	No
ABCB1	No	Yes	Yes
ABCC8	No	Yes	Yes
ABCG1	No	Yes	Yes
ARNT	No	Yes	Yes
CHST11	No	Yes	Yes
CHST8	No	Yes	Yes
CYP1A1	No	Yes	Yes
CYP4B1	No	Yes	Yes
CYP4F3	No	Yes	Yes
CYP19A1	No	Yes	Yes
CYP20A1	No	Yes	Yes
CYP24A1	No	Yes	Yes
CYP39A1	No	Yes	Yes
DPYD	No	Yes	Yes
EPHX1	No	Yes	Yes
MAT1A	No	Yes	Yes
NR1I2	No	Yes	Yes
PTGIS	No	Yes	Yes
SLC22A3	No	Yes	Yes
SLC29A2	No	Yes	Yes
SLC22A6	No	Yes	Yes
SLC6A6	No	Yes	Yes
SLC7A7	No	Yes	Yes
SLCO3A1	No	Yes	Yes
SLCO4A1	No	Yes	Yes
SULT1E1	No	Yes	Yes
TOP1P1	No	Yes	Yes
UGT2A1	No	Yes	Yes
UGT2A2	No	Yes	Yes
UGT2B4	No	Yes	Yes
XDH	No	Yes	Yes

GWAS, genome-wide association studies.

was found to be significant by either FET or PNA. Detailed results of the SNP comparative analysis using PNA and FET can be found in Supplementary File S3 (Supplemental digital content 3, *http://links.lww.com/FPC/A662*) and Supplementary File S4 (Supplemental digital content 4, *http://links.lww.com/FPC/A663*), respectively.

Haploblock analysis
The haplotype block estimation (Supplementary File S5, Supplemental digital content 5, *http://links.lww.com/FPC/A664*) revealed that SNP rs4148943 in the *CHST3* gene, which was a top hit in the NCI-60 analysis, is found in the same haplotype block as rs4148946, which was previously reported as a top hit in an analysis of long-term survival in stage III myeloma patients (Table 2) [21]. In addition, rs3842 in the *ABCB1* gene has been associated with genetic susceptibility of lung cancer [36] and response to antidepressant treatment [37].

NCI-60 cell line integrative analysis of SNPs, gene expression, and drug response
SNPs in genes that effect their own expression
From the linear regression analysis of NCI-60 SNP and gene expression data, we focused on SNPs that were

Table 2 Estimated haploblocks from NCI-60 and GWAS: all patients data sets for top hits in the association analysis

Block name	Data set	#SNPs	Block size (kb)	Chrom	Start pos	End pos	SNPs	Genes	PMID
B1	NCI-60	10	3.256	10	73769507	73772762	rs4148943[a], rs4148944, rs41448945, rs4148946, rs4148949, rs4148950, rs731027, rs730720	CHST3 (3'-UTR)	18778477
B2	NCI-60	4	0.842	13	95672950	95673791	rs10599751, rs4148553, rs4148551[a], rs3742106	ABCC4 (3'-UTR)	18364470
B3	GWAS	8	55.833	6	46540454	46596286	rs9349368, rs3799884[a], rs3799880, rs12208873, rs3799873, rs3799866	CYP39A1 (spans 4 interior exons)	22562553
B4	GWAS	3	12.539	7	87204279	87216817	rs1202179, rs1202186, rs17327624[a]	ABCB1	15197162, 18504423, 19107781
B5	GWAS	4	3.719	19	15773915	15777633	rs6512050, rs1543284[a], rs1543285[a], rs4808350	CYP4F10P (pseudogene) in region between CYP4F3 and CYP4F12	–
B6	GWAS	2	5.815	19	15766419	15772233	rs718258, rs1063803[a]	CYP4F3 (spans last 4 exons)	–
B7	GWAS	6	2.833	19	15723528	15726360	rs3794989[a], rs3794988, rs3764561, rs1320319, rs16980628, rs4646522	CYP4F8 (spans first exon)	–
B8	GWAS	2	0.229	3	14528028	14528256	rs2341985[a], rs4685165	SLC6A6 (3'-UTR)	–
B9	GWAS	8	55.833	6	46540454	46596286	rs9349368, rs3799884[a], rs3799880, rs12208873, rs3799873, rs3799872, rs7758684, rs3799866	CYP39A1 (spans 4 interior exons)	–

SNP, single-nucleotide polymorphism.
[a]SNP is a top hit in the association analysis.

associated with expression of their own genes. The results, as shown in Table 3, contain 10 unique SNPs in *PPARD*, five in *ABCC4*, two in *CYP2E1*, and one in *SLCO1B1*. We annotated all variants using the SNPnexus annotation tool [38,39].

SNPs that affect gene expression and drug response
The integrative analysis on SNPs, gene expression, and GI50 response revealed significant associations between several variants, and expression of multiple genes that were also associated with GI50 response. Supplementary File S6 (Supplemental digital content 6, *http://links.lww. com/FPC/A665*) shows 25 unique SNPs (14 unique genes) associated with 51 unique probes (39 unique genes), with a total of 129 combinations. Several of the genes include *CHST3* variants affecting *SFPQ*, *SLO4C1*, and *SLC24A3*; *PPARD* variant affecting *SFPQ*; *CYP2D6* variant affecting *FMO5*; and *CYP4F12* variant affecting *FETUB*, *HNRNPU*, and *HNRNPA2B1*.

Pathway analysis
We performed pathway analysis to understand the biological role of genes containing significant variants in a comparative study of the NCI-60 cell lines and GWAS data sets. Our study identified five significant variant genes (*ABCC1*, *ABCC4*, *CYP4F8*, *CYP4F12*, *PPARD*) common to both the NCI-60 and the GWAS: gemcitabine + placebo data sets with respect to gemcitabine response. Pathway analysis identified shared molecular mechanisms for some genes at both the downstream target level and regulatory level (Fig. 3a). Downstream, *ABCC1* and *ABCC4* exert positive regulation on other ABC transporters via shared ligands [40] or inducing agents [41]. At the regulatory level, transcription factors *MYCN* and *NFE2L2* [42–45] and cytokines (*IL1B*) [46,47] play a major role in enhancing the expression of *ABCC1* and *ABCC4*. Pathway analysis of significant genes from NCI-60 (Fig. 3b) and the GWAS: gemcitabine + placebo (Fig. 3c) studies pointed to genes involved in tumor cell proliferation, apoptotic, and inflammatory pathways.

To understand the biological associations between SNPs that affect the expression of target genes as well as drug response, we looked for molecular interactions between genes in which significant SNPs have been identified and target genes that might in-turn affect drug response. The chondroitin sulfotransferase gene, *CHST3*, was strongly associated with multiple differentially expressed genes in the integrated data set (Fig. 4a). Pathway analysis showed that *CHST3* may indirectly influence the RNA-processing gene *SFPQ* (PTB-associated splicing factor) via other matrix proteins and cell adhesion molecules such as versican and vitronectin, whereas *SFPQ* shares common pathways with genes whose expression is associated with *CHST3* variants. In the *PPARD* network (Fig. 4b) *PPARD* may regulate the PTB-associated splicing factor via the

Table 3   SNPs affecting their own gene expression

SNP (predictor variables)	SNP details		Annotation	MAF	Probe details			Linear regression results			ANOVA
	Chr	dbSNP ID			Probe ID (dependent variables)	Gene	Reference genotype	Coefficients	Estimate	SE	P-value
ABCC4_280619>(rs1059751)	13	rs1059751	mRNA-UTR	0.4237	203196_at	ABCC4	C/C	C/T T/T	0.5630 0.3675	0.2133 0.1973	0.0257
ABCC4_280051>(rs4148551)	13	rs4148551	mRNA-UTR	0.3051	203196_at	ABCC4	A/A	A/G G/G	0.5140 0.3614	0.2180 0.2180	0.0419
ABCC4_238593G>A(K1116K)	13	rs1751034	coding-synon	0.1441	203196_at	ABCC4	A/A	A/G G/G	0.0731 −0.7293	0.3048 0.2809	0.0370
ABCC4_279778>(rs3742106)	13	rs3742106	mRNA-UTR	0.2672	203196_at	ABCC4	T/T	G/G G/T	0.4043 0.5318	0.2333 0.2248	0.0354
ABCC4_280434>(rs4148553)	13	rs4148553	mRNA-UTR	0.4211	203196_at	ABCC4	A/A	A/G G/G	0.5243 0.3675	0.2267 0.1992	0.0461
CYP2E1_1295G>C>(rs3813867)	10	rs3813867	Locus-region	0.0339	222100_at	CYP2E1	G/G	C/C	0.3236	0.1467	0.0315
CYP2E1_-1055C>T>(rs2031920)	10	rs2031920	Locus-region	0.0339	222100_at	CYP2E1	C/C	T/T	0.3236	0.1467	0.0315
PPARD_-2578>(rs6911817)	6	rs6911817	Intron	0.0254	208044_s_at	PPARD	C/C	C/T	−0.5434	0.2089	0.0119
PPARD_-66611>(rs2267664)	6	rs2267664	Intron	0.0254	208044_s_at	PPARD	G/G	A/G	0.6551	0.2031	0.0021
PPARD_-66877>(rs6919734)	6	rs6919734	Intron	0.0172	208044_s_at	PPARD	G/G	A/G	−0.8091	0.3461	0.0231
PPARD_-58418>(rs7771474)	6	rs7771474	Intron	0.0170	208044_s_at	PPARD	C/C	C/T	−0.5693	0.2574	0.0311
PPARD_-21172>(rs3798343)	6	rs3798343	Intron	0.0678	210636_at	PPARD	C/C	C/G G/G	0.1324 0.4381	0.0555 0.1298	0.0008
PPARD_12923C>T(N163N)>(rs2076167)	6	rs2076167	Coding-synon	0.2500	210636_at	PPARD	T/T	C/C C/T	0.2193 0.0091	0.0840 0.0397	0.0396
PPARD_-7112>(rs7751481)	6	rs7751481	Intron	0.2672	210636_at	PPARD	G/G	A/A A/G	0.2217 0.0286	0.0840 0.0382	0.0368
PPARD_-16017>(rs6457816)	6	rs6457816	Intron	0.1186	210636_at	PPARD	T/T	C/C C/T	0.4301 0.0284	0.1360 0.0437	0.0093
PPARD_-9059>(rs1883322)	6	rs1883322	Intron	0.2672	210636_at	PPARD	T/T	C/C C/T	0.2217 0.0193	0.0828 0.0377	0.0347
PPARD_-19920>(rs6915115)	6	rs6915115	Intron	0.0172	210636_at	PPARD	C/C	C/T	0.2735	0.0899	0.0036
PPARD_-66611>(rs2267664)	6	rs2267664	Intron	0.0254	37152_at	PPARD	G/G	A/G	0.5711	0.2770	0.0439
SLCO1B1_37041T>C(V174A)	12	rs4149056	Coding-nonsyn	0.1186	210366_at	SLCO1B1	T/T	C/C C/T	0.1357 0.1202	0.0869 0.0420	0.0104

Table includes SNP annotation, probe details, regression coefficient estimates (Estimate), and standard error (SE). It also shows the P-value from the analysis of variance (ANOVA). MAF, minor allele frequency; SNP, single-nucleotide polymorphism.

Fig. 3
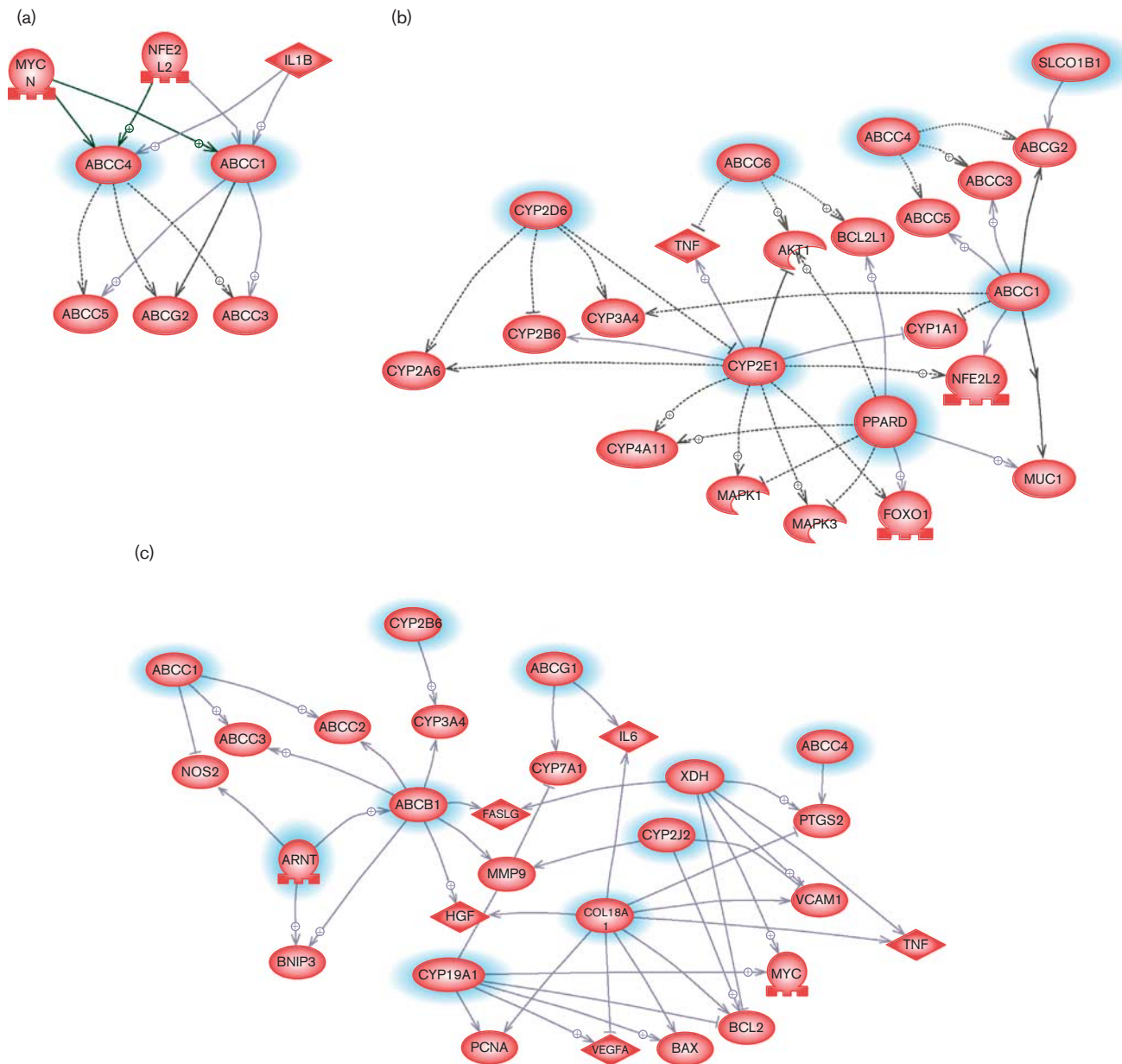
Pathway analysis of top genes from SNP comparative analysis. (a) Network created using genes common to the NCI-60 and GWAS: gemcitabine + placebo data sets; (b) Network using the significant genes in the NCI-60 data set; (c) Network using the significant genes in the GWAS: gemcitabine + placebo data set. The significant genes have been indicated using the blue highlight. GWAS, genome-wide association studies; SNP, single-nucleotide polymorphism.

transcription factor *SP1*. *SP1* and *SFPQ* are known to dynamically regulate growth factor-regulated gene expression response in mammalian cells [48,49]. The *CYP450* network identified potential regulatory mechanisms via the xenosensors (Fig. 4c–e). The *CYP2D6* network showed interaction with other proteins that regulate one of the significantly associated expression genes, Flavin Containing Monooxygenase 5 (*FMO5*), which is known to mediate the oxidative metabolism of several xenobiotics [50]. Cyp enzymes are regulated by

xenosensors such as *NR1I2* (*PXR*) and *NR1H4* (*FXR*) [51], which also induce Fetuin-B, a tumor suppressor, and may contribute to interindividual variability in drug response [52,53] (Fig. 4d). Heterogeneous nuclear ribonucleoproteins (hnRNPU, hnRNPA, hnRNPC) positively regulate the expression of cytokines such as *TNF-α* [54], which in turn can reduce the expression of xenosensors and their CYP enzyme targets [55] and drug transporters such as *SLCO1B1* [56] (Fig. 4e).

Fig. 4
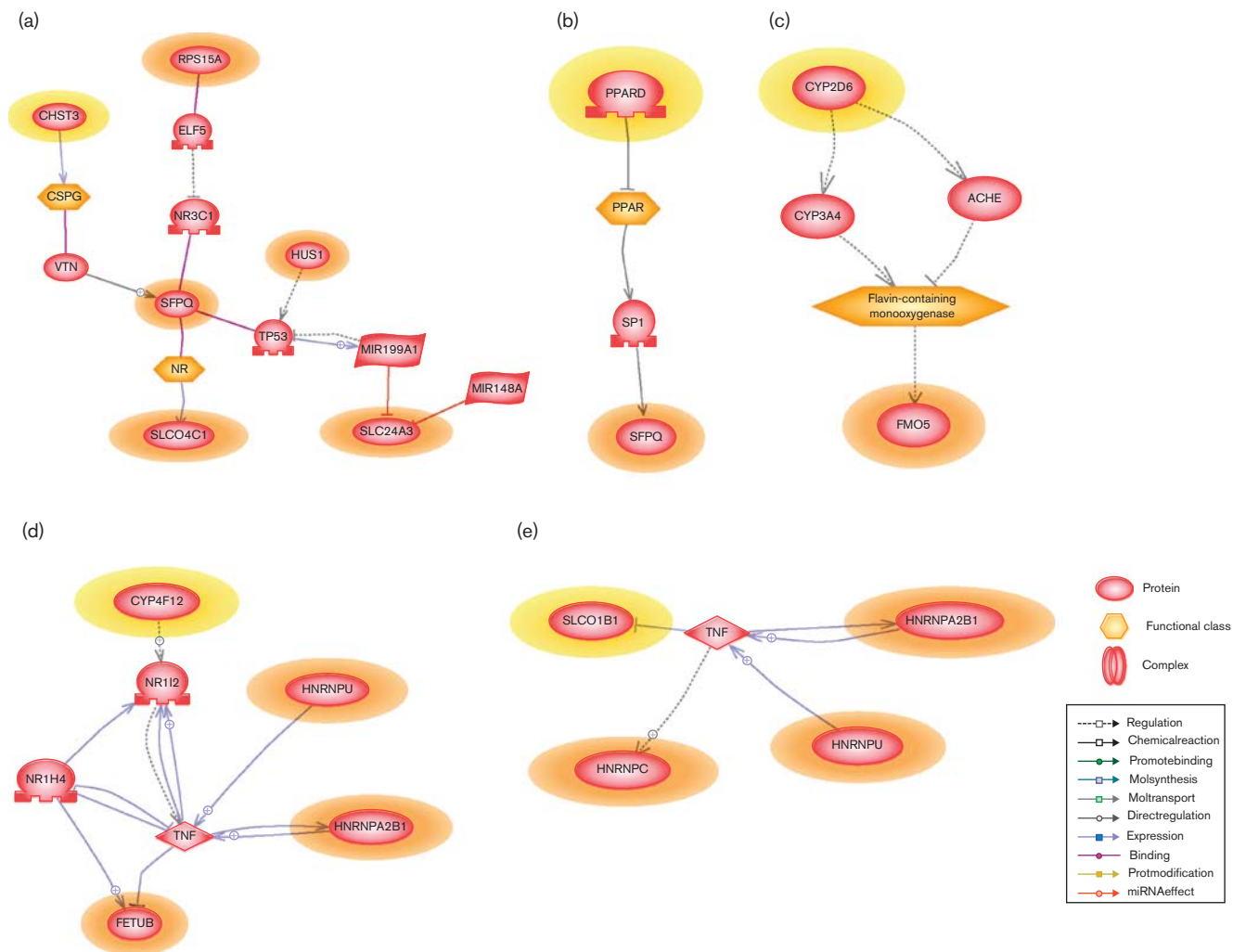
Pathway analysis on SNPs that affect gene expression and drug response. (a) a CHST3 network; (b) a PPARD network; (c−e) a CYP450 network. Entities highlighted in yellow indicate the gene where the SNP is located. Entities highlighted in orange indicate genes from the probes. The legend shows the different types of entities and relationships. GWAS, genome-wide association studies; SNP, single-nucleotide polymorphism.

Discussion

Abnormal expression and activity of drug transporters and metabolizing enzymes may arise from inherent or acquired polymporphisms and lead to undesirable drug response [2]. Identification of such variants can aid in the selection of patients who may benefit from gemcitabine treatment.

For the GWAS data sets we compared patients treated with gemcitabine + bevacizumab and gemcitabine + placebo as the addition of bevacizumab was shown to have no clinical benefit over gemcitabine treatment alone in these patients. Bevacizumab is an anti-angiogenesis drug that inhibits *VEGF-A* and its action is different from gemcitabine's, which inhibits *RRM1*, a gene that encodes the regulatory subunit of ribonucleotide reductase critical to the synthesis of deoxyribonucleotides. The GWAS:

gemcitabine + placebo data set is a smaller, less noisy data set compared with the larger and more diverse GWAS: all patients data set. These differences could be reflected at the molecular level and potentially explain the small difference in results between these two GWAS data sets in the SNP comparative analyses of gemcitabine response and pathway analysis.

ABC transporters have the ability to transport diverse types of anticancer drugs including gemcitabine. Members of ABC transporter family are also involved in diverse processes that impact cell proliferation and apoptotic pathways. Variants in two of these genes, *ABCC1* [57] and *ABCC4* [58], were significant in both studies, the overexpression of which has been observed to contribute to gemcitabine resistance in pancreatic cancer cells. As the functional significance of the SNPs in regulating

these genes is unknown, we looked at molecular interactions that might explain their effect on gemcitabine. We identified shared pathways for these genes, and suspect that one or more of these SNPs could have a combined effect on gemcitabine response. *CYP4F12*, whose role in pancreatic cancer or gemcitabine resistance is yet unknown, is one of the enzymes induced by p53 in response to xenobiotic signals such as chemotherapy [59].

Our analyses also revealed significant variants in CHST genes that have not been reported to be involved in gemcitabine transport or metabolism. These enzymes utilize $3'$-phospho-$5'$-adenylyl sulfate (PAPS) as sulfonate donor and catalyze the transfer of sulfate to selective glycoprotein ligands such as chondroitin, and the resulting chondroitin sulfates play a critical role in oncogenic *HRAS* signaling [60] and diverse biological functions including cell proliferation, adhesion, migration, and differentiation [61,62]. Members of the CHST family also mediate inflammation, immunity, angiogenesis, and extracellular matrix reorganization [63,64]. So far there is one study [65] that reports the involvement of chondroitin sulfates in enhancing antitumor activity of gemcitabine in human bladder cancer cells.

Looking at SNPs within the NCI-60 data that affect their own gene expression we found 10 unique SNPs in *PPARD* and one in *SLCO1B1*. *SLCO1B1* is known to regulate the uptake of drugs and compounds in the liver and SNP rs4149056 is predicted by SIFT and polyphen to be damaging. The C allele of this SNP has been linked to reduced transport activity and hence impaired drug effect [66,67]. Another *SLCO1B1* variant, rs4149015, known to affect its own expression [68], may impact its transporting activity and, consequently, drug response. Variants in *PPARG2* have been known to affect mRNA expression of *PPARG1* and its target genes [69,70], which play a diverse role in cellular energy metabolism and angiogenesis. Activation of PPARβ/δ, in particular, has been reported to favor angiogenesis, and is suggested to play a critical role in inducing angiogenesis in pancreatic cancer [71]. Such alterations in tumor cell characteristics can consequently modify drug response. Variants in *ABCC4* have been observed to affect intracellular concentrations of some drugs even though the SNPs were synonymous or within $3'$UTR [72,73].

SNPs can influence the expression levels of multiple genes either through *cis*-regulatory or *trans*-regulatory effects; and variations in gene expression levels can directly or indirectly impact drug response [74,75]. To understand these effects, we integrated the gene expression data with DMET SNP and drug response data for the NCI-60 cell lines. Associations involving detoxifying proteins such as CYP enzymes and transporters such as *SLCO1B1* revealed pathways closely related to drug response regulation. Associations involving other

variant genes such as *CHST3* and *PPARD* suggest a major role for cell adhesion, cell growth, and tumor-invasive processes in the alteration's drug response. Although *trans*-regulatory effects, if present, will have to be experimentally validated, the pathway analysis provides a potential mechanism to explain how SNPs could have a genome-wide impact on drug response, directly or indirectly, by altering tumor cell characteristics.

Conclusion and future perspective
There is strong agreement among clinicians about the need to implement pharmacogenomic discoveries in clinical practice due to the changing needs in medicine. However, most physicians have limited familiarity with genomics and inadequate knowledge of the pharmacogenomic tests available to help their patients [76,77]. Many associations have been identified between variant genotypes and drug response phenotypes, some of which are now identified in FDA-approved medical product labels [78]. Despite the growing evidence of pharmacogenetic involvement in drug response, physician uptake of pharmacogenetic testing has been poor. We believe this is in part due to lack of strong evidence of clinical utility, evolving clinical guidelines around these markers, and lack of integrated decision support tools. Another major barrier to address is the lack of adequate physician education in the field [76].

To overcome some of these challenges, adequate computational and informatics data analysis and support will be necessary to analyze and interpret results from clinical studies and to help evaluate pharmacogenomics biomarker panels. Multidisciplinary teams of clinicians, geneticists, bioinformaticians, and systems biologists need to partner to apply growing pharmacogenomic knowledge in the clinic. Decision support tools must be tightly coupled to electronic health records and clinical work flows to alert healthcare providers to actionable pharmacogenetic information.

Although we present an integrative systems biology and genetic analysis of 'omics' data for understanding variants that are strongly associated with gemcitabine response (e.g. CHST family of proteins), this report is intended as a framework to systematically integrate varied data types and strengthen the evidence behind pharmacogenomics markers. Such frameworks can ultimately provide motivation to develop clinical decision support tools and interfaces to help clinicians use pharmacogenomics data to improve patient care.

Cancer Institute, or the National Institutes of Health. Support for research computing resources was provided by the Amazon Corporation.

Conflicts of interest

There are no conflicts of interest.

References

1 Baker JA, Wickremsinhe E, Li CH, Oluyedun OA, Dantzig AH, Hall SD, *et al.* Pharmacogenomics of gemcitabine metabolism: functional analysis of genetic variants in cytidine deaminase and deoxycytidine kinase. *Drug Metab Dispos* 2013; **41**:541–545.

2 Deeken JF, Figg WD, Bates SE, Sparreboom A. Toward individualized treatment: prediction of anticancer drug disposition and toxicity with pharmacogenetics. *Anticancer Drugs* 2007; **18**:111–126.

3 Si S, Liao Q, Zhao YP, Hu Y, Zhang Q, You LL. Relationship between single nucleotide polymorphisms in the deoxycytidine kinase gene and chemosensitivity of gemcitabine in six pancreatic cancer cell lines. *Chin Med J (Engl)* 2011; **124**:419–422.

4 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.

5 Fukunaga AK, Marsh S, Murry DJ, Hurley TD, McLeod HL. Identification and analysis of single-nucleotide polymorphisms in the gemcitabine pharmacologic pathway. *Pharmacogenomics J* 2004; **4**:307–314.

6 Okazaki T, Javle M, Tanaka M, Abbruzzese JL, Li D. Single nucleotide polymorphisms of gemcitabine metabolic genes and pancreatic cancer survival and drug toxicity. *Clin Cancer Res* 2010; **16**:320–329.

7 Li L, Fridley BL, Kalari K, Jenkins G, Batzler A, Weinshilboum RM, *et al.* Gemcitabine and arabinosylcytosin pharmacogenomics: genome-wide association and drug response biomarkers. *PLoS One* 2009; **4**:e7765.

8 Kasuya K, Tsuchida A, Nagakawa Y, Suzuki Y, Suzuki M, Aoki T, *et al.* Prediction of a side effect and efficacy of adjuvant chemotherapy with gemcitabine for post operative patient of pancreatic cancer by a genetic polymorphism analysis. *Hepatogastroenterology* 2012; **59**:1609–1613.

9 Advancing Regulatory Science at FDA: A Strategic Plan (August 2011). Available at: http://www.fda.gov/regulatoryscience. [Accessed 14 December 2012].

10 Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, *et al.* Finding the missing heritability of complex diseases. *Nature* 2009; **461**:747–753.

11 Rodin AS, Gogoshin G, Boerwinkle E. Systems biology data analysis methodology in pharmacogenomics. *Pharmacogenomics* 2011; **12**:1349–1360.

12 Xiong Q, Ancona N, Hauser ER, Mukherjee S, Furey TS. Integrating genetic and gene expression evidence into genome-wide association analysis of gene sets. *Genome Res* 2012; **22**:386–397.

13 Turner SD, Bush WS. Multivariate analysis of regulatory snps: empowering personal genomics by considering cis-epistasis and heterogeneity. *Pac Symp Biocomput* 2011; 276–287. Available at: http://www.worldscientific.com/doi/abs/10.1142/9789814335058_0029.

14 Szkandera J, Absenger G, Dandachi N, Regitnig P, Lax S, Stotz M, *et al.* Analysis of functional germline polymorphisms for prediction of response to anthracycline-based neoadjuvant chemotherapy in breast cancer. *Mol Genet Genomics* 2012; **287**:755–764.

15 Rodin A, Mosley TH Jr, Clark AG, Sing CF, Boerwinkle E. Mining genetic epidemiology data with Bayesian networks application to APOE gene variation and plasma lipid levels. *J Comput Biol* 2005; **12**:1–11.

16 Peters BJ, Rodin AS, de Boer A, Maitland-van der Zee AH. Methodological and statistical issues in pharmacogenomics. *J Pharm Pharmacol* 2010; **62**:161–166.

17 Sissung TM, English BC, Venzon D, Figg WD, Deeken JF. Clinical pharmacology and pharmacogenetics in a genomics era: the DMET platform. *Pharmacogenomics* 2010; **11**:89–103.

18 Shankavaram UT, Reinhold WC, Nishizuka S, Major S, Morita D, Chary KK, *et al.* Transcript and protein expression profiles of the NCI-60 cancer cell panel: an integromic microarray study. *Mol Cancer Ther* 2007; **6**:820–832.

19 Van Cutsem E, Vervenne WL, Bennouna J, Humblet Y, Gill S, Van Laethem JL, *et al.* Phase III trial of bevacizumab in combination with gemcitabine and erlotinib in patients with metastatic pancreatic cancer. *J Clin Oncol* 2009; **27**:2231–2237.

20 Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 2003; **31**:e15.

21 Van Ness B, Ramos C, Haznadar M, Hoering A, Haessler J, Crowley J, *et al.* Genomic variation in myeloma: design, content, and initial application of the Bank On A Cure SNP Panel to detect associations with progression-free survival. *BMC Med* 2008; **6**:26.

22 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**:559–575.

23 Pearl J. *Probabilistic reasoning in intelligent systems: networks of plausible inference.* San Francisco, California: Morgan Kaufmann Publishers Inc; 1988.

24 Himes BE, Wu AC, Duan QL, Klanderman B, Litonjua AA, Tantisira K, *et al.* Predicting response to short-acting bronchodilator medication using Bayesian networks. *Pharmacogenomics* 2009; **10**:1393–1412.

25 Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian networks to analyze expression data. *J Comput Biol* 2000; **7**:601–620.

26 Jarjanazi H, Kiefer J, Savas S, Briollais L, Tuzmen S, Pabalan N, *et al.* Discovery of genetic profiles impacting response to chemotherapy: application to gemcitabine. *Hum Mutat* 2008; **29**:461–467.

27 Jouffe L, Munteanu P. New search strategies tailored for learning Bayesian networks. In: Govaert G, Janssen J, Limnios N, editors. *In Proceedings of Tenth International Symposium on Applied Stochastic Models and Data Analysis: 12–15 June 2001; Compiegne.* France: The Societe Francaise de Statistique; 2001. pp. 591–596.

28 Uhl GR, Drgon T, Johnson C, Walther D, David SP, Aveyard P, *et al.* Genome-wide association for smoking cessation success: participants in the Patch in Practice trial of nicotine replacement. *Pharmacogenomics* 2010; **11**:357–367.

29 Han B, Park M, Chen XW. A Markov blanket-based method for detecting causal SNPs in GWAS. *BMC Bioinformatics* 2010; **11 (Suppl 3)**:S5.

30 Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, *et al.* Genetic analysis of genome-wide variation in human gene expression. *Nature* 2004; **430**:743–747.

31 Benjamini Y, Hochberg Y. Controlling the False discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 1995; **57**:289–300.

32 Wang Z, Sew PH, Ambrose H, Ryan S, Chong SS, Lee EJ, *et al.* Nucleotide sequence analyses of the MRP1 gene in four populations suggest negative selection on its coding region. *BMC Genomics* 2006; **7**:111.

33 Anantharaman D, Samant TA, Sen S, Mahimkar MB. Polymorphisms in tobacco metabolism and DNA repair genes modulate oral precancer and cancer risk. *Oral Oncol* 2011; **47**:866–872.

34 Bonaventure A, Goujon-Bellec S, Rudant J, Orsi L, Leverger G, Baruchel A, *et al.* Maternal smoking during pregnancy, genetic polymorphisms of metabolic enzymes, and childhood acute leukemia: the ESCALE study (SFCE). *Cancer Causes Control* 2012; **23**:329–345.

35 Searles Nielsen S, McKean-Cowdin R, Farin FM, Holly EA, Preston-Martin S, Mueller BA. Childhood brain tumors, residential insecticide exposure, and pesticide metabolism genes. *Environ Health Perspect* 2010; **118**:144–149.

36 Wang H, Jin G, Liu G, Qian J, Jin H, Wei Q, *et al.* Genetic susceptibility of lung cancer associated with common variants in the 3′ untranslated regions of the adenosine triphosphate-binding cassette B1 (ABCB1) and ABCC1 candidate transporter genes for carcinogen export. *Cancer* 2009; **115**:595–607.

37 Dong C, Wong ML, Licinio J. Sequence variations of ABCB1, SLC6A2, SLC6A3, SLC6A4, CREB1, CRHR1 and NTRK2: association with major depression and antidepressant response in Mexican-Americans. *Mol Psychiatry* 2009; **14**:1105–1118.

38 Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update). *Nucleic Acids Res* 2012; **40**:W65–W70.

39 Chelala C, Khan A, Lemoine NR. SNPnexus: a web database for functional annotation of newly discovered and public domain single nucleotide polymorphisms. *Bioinformatics* 2009; **25**:655–661.

40 Huang R, Murry DJ, Kolwankar D, Hall SD, Foster DR. Vincristine transcriptional regulation of efflux drug transporters in carcinoma cell lines. *Biochem Pharmacol* 2006; **71**:1695–1704.

41 Jorajuria S, Dereuddre-Bosquet N, Naissant-Storck K, Dormont D, Clayette P. Differential expression levels of MRP1, MRP4, and MRP5 in response to human immunodeficiency virus infection in human macrophages. *Antimicrob Agents Chemother* 2004; **48**:1889–1891.

42 Manohar CF, Bray JA, Salwen HR, Madafiglio J, Cheng A, Flemming C, *et al.* MYCN-mediated regulation of the MRP1 promoter in human neuroblastoma. *Oncogene* 2004; **23**:753–762.

43 Norris MD, Smith J, Tanabe K, Tobin P, Flemming C, Scheffer GL, *et al.* Expression of multidrug transporter MRP4/ABCC4 is a marker of poor prognosis in neuroblastoma and confers resistance to irinotecan in vitro. *Mol Cancer Ther* 2005; **4**:547–553.

44 Loignon M, Miao W, Hu L, Bier A, Bismar TA, Scrivens PJ, *et al.* Cul3 overexpression depletes Nrf2 in breast cancer and is associated with sensitivity to carcinogens, to oxidative stress, and to chemotherapy. *Mol Cancer Ther* 2009; **8**:2432–2440.

45 Xu S, Weerachayaphorn J, Cai SY, Soroka CJ, Boyer JL. Aryl hydrocarbon receptor and NF-E2-related factor 2 are key regulators of human MRP4 expression. *Am J Physiol Gastrointest Liver Physiol* 2010; **299**:G126–G135.

46 Ronaldson PT, Ashraf T, Bendayan R. Regulation of multidrug resistance protein 1 by tumor necrosis factor alpha in cultured glial cells: involvement of nuclear factor-kappaB and c-Jun N-terminal kinase signaling pathways. *Mol Pharmacol* 2010; **77**:644–659.

47 Le Vee M, Gripon P, Stieger B, Fardel O. Down-regulation of organic anion transporter expression in human hepatocytes exposed to the proinflammatory cytokine interleukin 1beta. *Drug Metab Dispos* 2008; **36**:217–222.

48 Okazaki M, Iwasaki Y, Nishiyama M, Taguchi T, Tsugita M, Nakayama S, *et al.* PPARbeta/delta regulates the human SIRT1 gene transcription via Sp1. *Endocr J* 2010; **57**:403–413.

49 Urban RJ, Bodenburg Y. PTB-associated splicing factor regulates growth factor-stimulated gene expression in mammalian cells. *Am J Physiol Endocrinol Metab* 2002; **283**:E794–E798.

50 Yu J, Brown DG, Burdette D. In vitro metabolism studies of nomifensine monooxygenation pathways: metabolite identification, reaction phenotyping, and bioactivation mechanism. *Drug Metab Dispos* 2010; **38**:1767–1778.

51 Waxman DJ. P450 gene induction by structurally diverse xenochemicals: central role of nuclear receptors CAR, PXR, and PPAR. *Arch Biochem Biophys* 1999; **369**:11–23.

52 Habano W, Gamo T, Terashima J, Sugai T, Otsuka K, Wakabayashi G, *et al.* Involvement of promoter methylation in the regulation of Pregnane X receptor in colon cancer cells. *BMC Cancer* 2011; **11**:81.

53 Hariparsad N, Chu X, Yabut J, Labhart P, Hartley DP, Dai X, *et al.* Identification of pregnane-X receptor target genes and coactivator and corepressor binding to promoter elements in human hepatocytes. *Nucleic Acids Res* 2009; **37**:1160–1173.

54 Zhao W, Wang L, Zhang M, Wang P, Qi J, Zhang L, *et al.* Nuclear to cytoplasmic translocation of heterogeneous nuclear ribonucleoprotein U enhances TLR-induced proinflammatory cytokine production by stabilizing mRNAs in macrophages. *J Immunol* 2012; **188**:3179–3187.

55 Moriya N, Kataoka H, Fujino H, Nishikawa J, Kugawa F. Effect of lipopolysaccharide on the xenobiotic-induced expression and activity of hepatic cytochrome P450 in mice. *Biol Pharm Bull* 2012; **35**:473–480.

56 Vee ML, Lecureur V, Stieger B, Fardel O. Regulation of drug transporter expression in human hepatocytes exposed to the proinflammatory cytokines tumor necrosis factor-alpha or interleukin-6. *Drug Metab Dispos* 2009; **37**:685–693.

57 Nath S, Daneshvar K, Roy LD, Grover P, Kidiyoor A, Mosley L, *et al.* MUC1 induces drug resistance in pancreatic cancer cells via upregulation of multidrug resistance genes. *Oncogenesis* 2013; **2**:e51.

58 Zhang Z, Wang J, Shen B, Peng C, Zheng M. The ABCC4 gene is a promising target for pancreatic cancer therapy. *Gene* 2012; **491**:194–199.

59 Goldstein I, Rivlin N, Shoshana OY, Ezra O, Madar S, Goldfinger N, *et al.* Chemotherapeutic agents induce the expression and activity of their clearing enzyme CYP3A4 by activating p53. *Carcinogenesis* 2013; **34**:190–198.

60 Kluppel M, Samavarchi-Tehrani P, Liu K, Wrana JL, Hinek A. C4ST-1/CHST11-controlled chondroitin sulfation interferes with oncogenic HRAS signaling in Costello syndrome. *Eur J Hum Genet* 2012; **20**:870–877.

61 Hollmann J, Thiel J, Schmidt A, Buddecke E. Increased activity of chondroitin sulfate-synthesizing enzymes during proliferation of arterial smooth muscle cells. *Exp Cell Res* 1986; **167**:484–494.

62 Malmstrom A, Bartolini B, Thelin MA, Pacheco B, Maccarana M. Iduronic acid in chondroitin/dermatan sulfate: biosynthesis and biological function. *J Histochem Cytochem* 2012; **60**:916–925.

63 Alberts R, Lu L, Williams RW, Schughart K. Genome-wide analysis of the mouse lung transcriptome reveals novel molecular gene interaction networks and cell-specific expression signatures. *Respir Res* 2011; **12**:61.

64 Mendoza-Londono R, Chitayat D, Kahr WH, Hinek A, Blaser S, Dupuis L, *et al.* Extracellular matrix and platelet function in patients with musculocontractural Ehlers-Danlos syndrome caused by mutations in the CHST14 gene. *Am J Med Genet A* 2012; **158A**:1344–1354.

65 Ferro M, Giuberti G, Zappavigna S, Perdona S, Facchini G, Sperlongano P, *et al.* Chondroitin sulphate enhances the antitumor activity of gemcitabine and mitomycin-C in bladder cancer cells with different mechanisms. *Oncol Rep* 2012; **27**:409–415.

66 Pasanen MK, Backman JT, Neuvonen PJ, Niemi M. Frequencies of single nucleotide polymorphisms and haplotypes of organic anion transporting polypeptide 1B1 SLCO1B1 gene in a Finnish population. *Eur J Clin Pharmacol* 2006; **62**:409–415.

67 Niemi M, Schaeffeler E, Lang T, Fromm MF, Neuvonen M, Kyrklund C, *et al.* High plasma pravastatin concentrations are associated with single nucleotide polymorphisms and haplotypes of organic anion transporting polypeptide-C (OATP-C, SLCO1B1). *Pharmacogenetics* 2004; **14**:429–440.

68 Johnson AD, Wang D, Sadee W. Polymorphisms affecting gene regulation and mRNA processing: broad implications for pharmacogenetics. *Pharmacol Ther* 2005; **106**:19–38.

69 Kolehmainen M, Uusitupa MI, Alhava E, Laakso M, Vidal H. Effect of the Pro12Ala polymorphism in the peroxisome proliferator-activated receptor (PPAR) gamma2 gene on the expression of PPARgamma target genes in adipose tissue of massively obese subjects. *J Clin Endocrinol Metab* 2003; **88**:1717–1722.

70 Costa V, Gallo MA, Letizia F, Aprile M, Casamassimi A, Ciccodicola A. PPARG: gene expression regulation and next-generation sequencing for unsolved issues. *PPAR Res* 2010. Available at: http://www.hindawi.com/journals/ppar/si/987203/.

71 Bishop-Bailey D. PPARs and angiogenesis. *Biochem Soc Trans* 2011; **39**:1601–1605.

72 Anderson PL, Lamba J, Aquilante CL, Schuetz E, Fletcher CV. Pharmacogenetic characteristics of indinavir, zidovudine, and lamivudine therapy in HIV-infected adults: a pilot study. *J Acquir Immune Defic Syndr* 2006; **42**:441–449.

73 Abla N, Chinn LW, Nakamura T, Liu L, Huang CC, Johns SJ, *et al.* The human multidrug resistance protein 4 (MRP4, ABCC4): functional analysis of a highly polymorphic gene. *J Pharmacol Exp Ther* 2008; **325**:859–868.

74 Corse E, Gottschalk RA, Krogsgaard M, Allison JP. Attenuated T cell responses to a high-potency ligand in vivo. *PLoS Biol* 2010; **8**:e1000481.

75 Rotival M, Zeller T, Wild PS, Maouche S, Szymczak S, Schillert A, *et al.* Integrating genome-wide genetic variations and monocyte expression data reveals trans-regulated gene modules in humans. *PLoS Genet* 2011; **7**:e1002367.

76 Stanek EJ, Sanders CL, Taber KA, Khalid M, Patel A, Verbrugge RR, *et al.* Adoption of pharmacogenomic testing by US physicians: results of a nationwide survey. *Clin Pharmacol Ther* 2012; **91**:450–458.

77 Sharp RR, Goldlust ME, Eng C. Addressing gaps in physician education using personal genomic testing. *Genet Med* 2011; **13**:750–751.

78 FDA Table of Pharmacogenomic Biomarkers in Drug Labels. Available at: http://www.fda.gov/drugs/scienceresearch/researchareas/pharmacogenetics/ucm083378.htm (Accessed 4 March 2013).