Cancer Chemother Pharmacol (2006) 58: 384–395
DOI 10.1007/s00280-005-0171-8

ORIGINAL ARTICLE

Timothy C. Cheng · Gerald Manorek · Goli Samimi
Xinjian Lin · Charles C. Berry · Stephen B. Howell

Identification of genes whose expression is associated with cisplatin resistance in human ovarian carcinoma cells

Received: 28 September 2005 / Accepted: 8 December 2005 / Published online: 11 January 2006

Abstract The goal of this study was to identify genes consistently differentially expressed in multiple pairs of isogenic cisplatin (DDP)-sensitive and resistant human ovarian carcinoma cell lines using microarray-based expression profiling. Expression profiling was carried out on six pairs of ovarian carcinoma cells lines growing under identical conditions; each cell expression profile was independently replicated six times. No genes were differentially expressed in all six pairs of cells or even in even in any five of the six pairs. Eighteen genes and 1 EST were upregulated, and four genes and 1 EST were downregulated, in at least four cell pairs. Of these, only metallothionein 2A has previously been implicated in DDP resistance. Among the genes identified on the basis of six replicates, an average of 24.8% would have been missed if only five replicates had been performed, and 38.3% would have been missed with only four replicates. The genes did not identify a dominant biochemical pathway or ontology category as being linked to DDP resistance; however, hierarchical clustering provided evidence for two classes DDP-resistant phenotypes within which there are additional cell pair-specific alterations. Many of the genes identified in this study play important roles in cell surface interactions and trafficking pathways not previously linked to DDP resistance. The genes discovered by this extensively replicated analysis are candidates for prediction of DDP responsiveness in ovarian cancer patients.

T. C. Cheng · G. Manorek · G. Samimi · X. Lin
C. C. Berry · S. B. Howell (✉)
Department of Medicine and the Moores USCD Cancer Center,
The University of California Medical Center,
3855 Health Sciences Drive #0819,
La Jolla, San Diego,
CA 92093-0819, USA
E-mail: showell@ucsd.edu
Tel.: +1-858-8221110
Fax: +1-858-8221111

Keywords Cisplatin · Oxaliplatin · Microarray · Drug resistance

Abbreviations SAM: Significance analysis of microarrays · DDP: Cisplatin

Introduction

Cisplatin (DDP) is an effective first-line therapy for many types of cancer but the rapid development of resistance during therapy remains a major clinical challenge. DDP is thought to kill cells predominantly by forming adducts in DNA that block transcription and DNA replication. Mechanisms implicated in cellular resistance include reduced drug uptake, increased drug efflux, increased DNA repair, increased tolerance of DNA damage, and increased levels of intracellular thiols such as glutathione and metallothionein (reviewed in [14]). More recently, the copper transporters have been found to modulate the cellular pharmacology of the platinum drugs [12]. Resistance appears to be multifactorial in origin with no single overarching mechanism predominating even within histologic type of tumor. Identification of patients whose tumors have a high probability of responding to DDP, and avoiding administration of this drug to patients whose tumors have little chance of responding, is an important goal. Application of new genetic and genomic tools offers the promise of a more thorough understanding of the mechanisms that mediate DDP resistance and the possibility of developing strategies for individualizing therapy.

Most attempts to use microarray-based expression profiling to identify genes associated with a drug-resistant phenotype have relied on comparisons between independent cell lines or tumor samples. Although seemingly always successful in identifying differentially expressed genes in these different samples, the confidence that these genes are really markers of resistance has been

compromised by the fact that most studies employed data derived from just one or two independently isolated RNA samples per tumor, or relied on an arbitrarily chosen threshold for identifying genes as being of interest. We report here the identification of genes associated with the DDP-resistant phenotype based on a different approach designed to improve confidence by (1) comparing multiple isogenic pairs of sensitive and stably resistant cells growing under identical conditions, (2) using a large number of independent replicates, and (3) requiring that genes pass a rigorous test of statistical significance in order to be considered associated with the resistant phenotype.

Materials and methods

Cell lines and culture

Six human ovarian carcinoma cell lines (2008, A2780, HEY, IGROV-1, KF28, and UCI 107) were used in this study. Sublines stably resistant to DDP (2008/C13*5.25, A2780/CP70, HEY C2, IGROV-1/CP, KFr13, and UCI CPR) had been prepared from each parental line by repeated in vitro exposure to DDP as previously described [10]. All cell lines were maintained in drug-free RPMI-1640 medium (GIBCO, Inc.) with 10% fetal calf serum (GIBCO, Inc.) at 37°C in a humidified atmosphere containing 5% CO_2. The degree of resistance of each subline was determined at the time RNA was harvested using a clonogenic assay with continuous drug exposure as previously described [10].

Expression profiling

cDNA microarrays were purchased from the Stanford Functional Genomics Facility (http://www.microarray.org) and contained 43,200 elements representing ~29,593 genes as estimated by association with UniGene clusters. When the cell cultures reached ~80% confluence they were lysed with a guanidine isothiocyanate buffer (4 M guanidine isothiocyanate (Gibco, Inc.), 25 mM sodium acetate pH 5.5 (Ambion), 0.5% sarkosyl (Fisher Scientific) and 0.1 M 2-mercaptoethanol (Gibco, Inc.). Total RNA was pelleted through a cesium chloride (Gibco, Inc.) step gradient and reverse transcribed into cDNA with a 2:1 ratio of aminoallyl-dUTP (Sigma) to dTTP (Sigma). cDNA from three of the six pairs of replicates were labeled as Cy3 (sensitive) and Cy5 (resistant) (Amersham); for the remaining three replicates the labeling was reversed. Five hundred nanograms of Cy3-labeled and Cy5-labeled cDNA were hybridized to the cDNA microarrays for 18 h at 42°C. The arrays were washed four times with neat sodium citrate buffer (SSC) containing 0.1% SDS, twice with 1× SSC, once with 0.1× SSC, and spun dry.

Microarray scanning and quality assurance

Features on the microarrays were located and Cy3 and Cy5 fluorescence intensities were analyzed with GenePix Pro 3.0 software (Axon Instruments, Inc.) and a GenePix 4000A scanner (Axon Instruments, Inc.). The data sets were imported into Microsoft Excel spreadsheets for analysis of the quality of each feature. Four parameters were used to assess the quality of each feature, and features were excluded for any of the following conditions: diameter < 50 μm, ≥50% pixels saturated in both channels, $< 54\%$ of the pixels with an intensity greater than the median background intensity plus one standard deviation in either channel; and, if flagged by GenePix as 'not found' or 'absent' or manually flagged as 'bad'. The $\log_2$ (Cy3/Cy5) ratios for replicates 1–3 and $\log_2$ (Cy5/Cy3) for replicates 4–6, hereafter referred to as the $\log_2$ (R/S) values, were calculated for each feature and then normalized in the *R* statistical environment using a within-print tip group normalization method based on locally weighted lowest regression as proposed by Yang et al. [21]. For details of the approach to evaluation of dye bias, see Significance analysis of microarray (SAM) algorithm.

Analysis of reproducibility

Four approaches were taken to assess reproducibility: (1) complete linkage hierarchical clustering of all 36 arrays, (2) correlation plots of $\log_2$ (R/S) for each possible pairwise combination of replicates within each cell line pair, (3) histograms of the difference in $\log_2$ (R/S) between each possible pairwise combination of replicates within each cell line pair, and (4) determination of how many features with a significantly higher or lower expression level would be missed per cell line pair if only four or five replicates had been performed instead of six. Complete linkage hierarchical clustering using Euclidean distance and a heatmap representation for all 36 arrays after quality assurance and normalization was carried out in the *R* statistical environment [6] as were the correlation and histogram plots for the pairwise comparisons of the $\log_2$ (R/S) values. To determine how many differentially expressed features would be missed if less than six replicates had been performed, SAM was applied to all features that had passed quality assurance in all six replicates, followed by features that had met quality assurance criteria in at least four and five replicates. The average number of SAM-identified genes for each possible combination of four or five replicates was then subtracted from the total number of genes found from the six-replicate analysis.

Identification of genes of interest using significance analysis of microarrays

The SAM algorithm works as a Microsoft Excel add-in (http://www-stat.stanford.edu/~tibs/SAM/index.html) and calculates a "*d*" score or modified *t* statistic for the normalized $\log_2 (R/S)$ of each feature. This is the mean $\log_2 (R/S)$ divided by the standard error to which a constant value was added. The addition of a constant value gives the tests more power on average and de-emphasizes large "*d*" score values that arise from genes whose expression level is near zero [16]. The cutoff for significance is determined by the tuning parameter δ, which is chosen by the user based on the estimated false discovery rate (FDR). In these studies the value of δ was always chosen so that the estimated FDR was ~1, which means that on average only a single false positive gene would be identified for each 100 true positives in the analysis. After filtering the data sets to include only those features for which a $\log_2 (R/S)$ value was available in at least four of the six replicates, each pair of isogenic cell line pairs was subjected to SAM [17] with permutations honoring the pairing of resistant/sensitive members of each pair within each replication. SAM analyses were carried out on the data sets from each cell line pair separately and additional SAM analyses were carried out across all six pairs. First, the average $\log_2 (R/S)$ was determined for each feature for the six replicates for a given cell pair. Those features for which an average $\log_2 (R/S)$ was not available for all six pairs were discarded. In addition, the mean of the normalized $\log_2 (R/S)$ ratio for each feature across all six replicates was calculated for each cell pair and SAM was run across all six cell pairs using these six mean values. For evaluation of dye bias between the two groups of differentially labeled samples (replicates 1–3 and 4–6), an unpaired two-class SAM analysis was done using all $\log_2 (R/S)$ data that had passed quality control setting the FDR to 1.

Statistical techniques

The data consist of six replications comparing sensitive and resistant members of six pairs of isogenic cell lines. The SAM analyses generated lists of genes whose modified *t* statistic ("*d*" value) indicated that they were statistically significantly upregulated or downregulated in the resistant member of each pair when the FDR was ~1. These lists were used to tabulate the combinations of cell lines in which each gene was differentially expressed and the number of genes in which exactly two, three, four or five pairs showed differential expression. These tables were then used to calculate the expected number of genes deemed differentially expressed in two, three, four or five cell pairs under the hypothesis of no association of differentially expressed genes between cell lines. The probability that a gene deemed differentially expressed in k cell pairs would be differentially expressed in another pair was calculated.

Ontology and pathways analysis

All SAM-identified genes for features that had passed quality control in four out of six replicates in each cell line were used for the 'in and out' of category calculations wherein a hypergeometric *P*-value was determined given the null hypothesis that the ensemble of SAM-identified genes were not associated with a functional or biochemical relationship as defined by categories in the Kyoto Encyclopedia of Genes and Genomes (http://www.kegg.org) and the Gene Ontology Consortium (http://www.geneontology.org). Categories identified were then further winnowed down using a criterion for low false discovery. All scripts were written in the *R* statistical environment [6]

Quantitative real-time PCR confirmation of differential expression

Forward and reverse primers were designed with the online version of Primer 3 at http://www.frodo.wi.mit.edu/cgi-bin/primer3/primer3.cgi [11]. Candidate sequences were run through a BLAST search http://www.ncbi.nih.gov and discarded if there were non-specific matches: CLDN4 (forward acatcatccaagacttctaca atcc, reverse agcagaatacttggcggagtaag), MT2A (forward caaatgcaaagagtgcaaatg, reverse acggtcagggttgtacata aaaa), ANXA1 (forward gcaggcctggtttattgaaa, reverse gctgtgcattgtttcgctta), F3 (forward ttgaggggctgacttcaatc, reverse gctgcccagaataccaatgt), TIMP1 (forward aattcc gaccztcgtcatcag, reverse tgcagttttccagcaatgag), STARD4 (forward ttacaagttggctgctgtcg, reverse catcagacaggccttcc att), and β-actin (forward ggacttcgagcaagagatgg, reverse agcactgtgttggcgtacag). Triplicate samples of cDNA from DDP-sensitive and resistant cell pairs were prepared as described above but reverse transcribed without amino-allyl-dUTP. Bovine DNase I (2 units, amplification grade, Invitrogen) was added to 50 μg of RNA from a pooled sample of all six replicates and incubated for 15 min at room temperature followed by the addition of 1 μl of 25 mM EDTA and heating at 65°C for 5 min. β-actin was used as an internal no differential expression control and cDNA from HeLa cells was used to construct standard curves from a tenfold dilution series for each primer pair in order to verify that amplification efficiencies were 90–100%. PCR was carried out on a Biorad iCycler (Hercules, CA) with SYBR Green detection (Biorad iQ SYBR Green Supermix). The thermal profile included: 1 cycle at 95°C for 3 min followed by 40 cycles with two steps (95°C for 10 s then 60°C for 60 s). Efficiencies were calculated from three independent experimental sets of standard curves and the mean value used in relative ratio calculations for all subsequent real-time (RT)-PCR runs.

Results

Cisplatin sensitivity

Six pairs of human ovarian carcinoma cell lines were used in this study. Each pair consisted of a DDP-sensitive parental line and a subline that had been selected for stable acquired DDP resistance by repeated in vitro exposure to DDP. Colony formation assays were used to determine the sensitivity of the six pairs of DDP-sensitive and DDP-resistant cell lines to the cytotoxic effect of DDP. The results, presented in Table 1, indicate that the difference in DDP sensitivity, as measured by the ratio of IC_{50} values, ranged from 1.6-fold for the 2008–2008/C13*5.25 pair to 11.9-fold for the A2780-A2780/CP70 pair.

Quality assurance and evaluation of dye effect

RNA was isolated from each member of the six pairs of cell lines, converted to cDNA, labeled with Cy3 or Cy5 and hybridized to cDNA microarrays containing 43,200 elements representing ~29,593 genes as estimated by their association with UniGene clusters. Four parameters were used to assess the quality of each feature: diameter, saturated pixels, intensity of features, and the number of features flagged by the scanning software. The average number of features that passed quality assurance ranged 24.7–38.0%; considering all replicates for all six pairs of cell lines a total of 13,228 features (32.3%) met all quality assurance criteria.

A total of 256 features (0.6%) demonstrated a dye bias as detected by SAM [17]. Among these, an average of only 36.4% were also identified by SAM analysis as exhibiting a significant degree of differential expression. Conversely, among the features that were identified by SAM analysis as exhibiting differential expression, none exhibited dye bias as determined by two-class unpaired SAM occurred in four out of six cell pairs.

Analysis of reproducibility

One of the major challenges with the use of microarrays is to distinguish valid signals from background noise. In order to assess the reproducibility of the results and to generate data with a high signal-to-noise ratio, six replicates were performed using independently isolated RNA samples for each cell line pair such that there were a total of 36 hybridizations when all six pairs were analyzed. The first assessment of reproducibility was based on hierarchical cluster analysis of all 36 samples. As shown in Fig. 1a, the six replicates in each cell line pair merged together in the tree structure before the different cell line pairs merged with the single exception of replicate 6 from the IGROV-1 pair and replicate 4 from the UCI pair. Thus, the expression profile for a replicate of any given cell pair was more closely related to replicates from the same cell pair than to replicates from another cell pair.

The second approach taken to the analysis of reproducibility was to examine histograms of differences in the $\log_2$ (resistant/sensitive) fluorescence ratio [abbreviated here as $\log_2 (R/S)$]. A histogram was produced for each cell line pair using the following technique. For each gene, all possible pairwise differences among the six replicates $\log_2 (R/S)$ were calculated. A histogram was generated from all non-missing $\log_2 (R/S)$ differences for all genes. The frequency in each bin was divided by the total number of $\log_2 (R/S)$ differences to give the relative frequency. The resulting histograms are plotted for each pair of DDP-sensitive and resistant cell lines in Fig. 1b. In these plots, better reproducibility is reflected by tall narrow histograms. Based on inspection of these histograms, the trend in reproducibility is KF > HEY > 2008 > I-GROV-1 > A2780 > UCI.

The third approach used to assess reproducibility was to construct scatterplots of the $\log_2 (R/S)$ values for each feature for each possible pairwise combination of replicates within a cell line pair. If the results of any two replicates were exactly the same, all the data points would lie on a line with a slope of 1.0, while increasing differences between two replicates would be indicated by increasing degrees of scatter around this line. The average of the correlation coefficients for all the cell pairs was 0.71. This analysis yielded that same rank order for reproducibility as the histograms of the $\log_2 (R/S)$ ratio differences.

The question of how many replicates are needed to identify most of the genes that are really differentially expressed between the sensitive and resistant cell lines is important to the design of future studies. Therefore, we asked how many of the genes that were found to be statistically significantly differentially expressed when all six replicates were considered would have been missed if only four or five replicates had been performed on each

Table 1 Cisplatin IC_{50} values as determined by colony formation assay

Parental cells	IC_{50} (μM)[a]	Resistant cells	IC_{50} (μM)[a]	Fold resistant
2008	0.70	2008/C13*5.25	1.10	1.6
A2780	0.21	A2780/CP70	2.50	11.9
HEY	0.76	HEY C2	4.60	6.1
IGROV-1	0.30	IGROV-1/CP	1.08	3.6
KF28	0.18	KFr13	0.40	2.2
UCI 107	1.00	UCI CPR	8.00	8

[a]Continuous drug exposure

cell pair. SAM analysis was carried out for each cell line pair on the features that had passed quality assurance in all six replicates. The numbers of features discovered in the analyses ranged from a low of 488 features for the 2008 and 2008/C13*5.25 pair to a high of 3,797 features for the KF28 and KFr13 pairs. A SAM analysis was then performed on all the features that had passed quality assurance in all four of all possible combinations of four replicates, and on all five of all possible combinations of five replicates. The striking finding was that an average of 24.7% of the genes discovered by SAM when all six replicates were considered would have been missed if only five replicates had been performed, and an average of fully 38.8% would have been missed if only four replicates had been performed. This finding mandates the use of a large number of replicates in future studies aimed at comprehensively identifying truly differentially expressed genes using cDNA arrays.

Identification of genes of interest using SAM

SAM was used to identify genes that were significantly differentially expressed within each cell line pair. In this

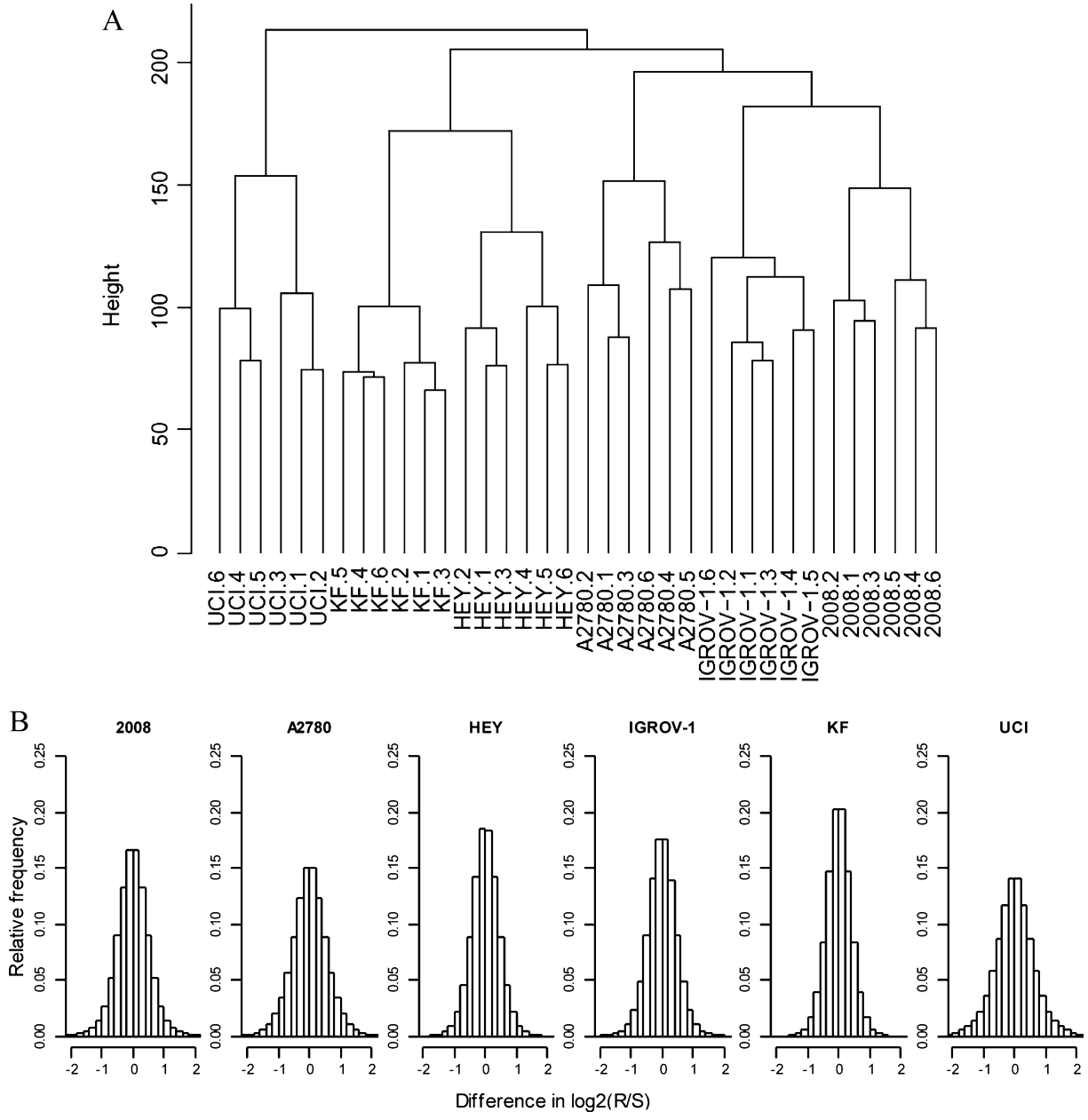

Fig. 1 Analysis of reproducibility. **a** Dendrogram based on complete hierarchical clustering with average linkage of all of the 36 replicates. **b** Histograms of the $\log_2$ (R/S) difference for each possible pairwise combination of replicates per cell line pair

analysis each feature on the array was required to meet the quality assurance criteria in at least four of the six replicates. The average number of features analyzed by SAM across the six cell pairs was 12,962; an average of 959 were found to have higher expression and 909 lower expression levels in the DDP-resistan sublines. Thus, for each cell pair a relatively large number of genes met the statistical criteria of being differentially expressed when the FDR was set at ~1. However, a cross-tabulation analysis indicated that very few of these were consistently significantly differentially expressed in multiple cell pairs. There were no features that were differentially expressed at either a higher or lower level in all six pairs or even in five of the six cell pairs. There were 21 features corresponding to 18 genes and 1 EST whose expression was significantly increased in four of the six pairs. Thioredoxin interacting protein was represented by three features on the microarray and all three features reached statistical significance. There were only five features, corresponding to four genes and 1 EST, whose expression was significantly decreased in four of the six resistant sublines. The gene names and other identifying features of these upregulated and downregulated features are presented in Table. 2 and 3.

An alternative strategy for identifying genes of interest is to average the $\log_2 (R/S)$ values for each feature across the six replicates of each cell pair and then perform a SAM analysis using these mean values across the six cell pairs. Using this approach, metallothionein 2A was the only gene that attained the threshold of statistical significance.

Confirmation of differential expression by quantitative RT-PCR

Quantitative RT-PCR (qRT-PCR) assays were established for 5 of the 26 features differentially expressed in at least four of the six cell pairs (STARD4, F3, ANXA1, CLDN4, and TIMP1) as well as MT2A. These six cell pairs were selected on the basis of their linkage to biochemical mechanisms that could conceivably modulate DDP sensitivity. As shown in Table 4, the qRT-PCR analysis confirmed the microarray analysis for these genes and the two techniques demonstrated a generally similar degree of differential expression. Comparisons were made between the qRT-PCR and microarray data each of the six cell pairs; since metallothionein 2A was not identified as differentially expressed in any one pair but only in the SAM analysis of the mean $\log_2 (R/S)$ values for each pair, it was not included in this analysis. Among 30 possible comparisons, 22 showed complete agreement, 7 were discordant and one (ANXA1 for A2780) could not be evaluated due to lack of signal in several replicates. In only two out of the seven cases of discordance between the two methods were the expression ratios in the opposite direction and in both these

Table 2 Genes whose expression is upregulated in at least four of six DDP-resistant cell pairs

Feature no.	Clone ID	Gene symbol	Gene name	Locus link
12574	IMAGE:554367	TXNIP	Thioredoxin interacting protein	10628
14387	IMAGE:488488	TXNIP	Thioredoxin interacting protein	10628
35178	IMAGE:208718	ANXA1	Annexin A1	301
33376	IMAGE:753610	APOE	Apolipoprotein E	348
5382	IMAGE:346510	CLDN4	Claudin 4	1364
38086	IMAGE:259884	GPR126	G-protein coupled receptor 126	57211
21311	IMAGE:290162	FLJ10726	Hypothetical protein FLJ10726	84250
24341	IMAGE:1928791	F3	Coagulation factor III (thromboplastin, tissue factor)	2152
42503	IMAGE:825461	GADD45B	Growth arrest and DNA-damage-inducible, beta	4616
35087	IMAGE:726035	JUN	v-Jun sarcoma virus 17 oncogene homolog (avian)	3725
34836	IMAGE:609377	KAB	KIAA0470 gene product	9859
40101	IMAGE:280375	ATAD2	ATPase family, AAA domain containing 2	84325
28879	IMAGE:700302	OSTF1	Osteoclast stimulating factor 1	26578
24762	IMAGE:416676	PELI1	Pellino homolog 1 (Drosophila)	57162
2042	IMAGE:235909	STARD4	START domain containing 4, sterol regulated	134429
35082	IMAGE:713696	TIMP1	Tissue inhibitor of metalloproteinase (erythroid potentiating activity, collagenase inhibitor)	7076
41478	IMAGE:840788	TMSB10	Thymosin, beta 10	9168
37772	IMAGE:1473171	TXNIP	Thioredoxin interacting protein	10628
4395	IMAGE:924929	LOC222171	ESTs, weakly similar to putative protein (*Arabidopsis thaliana*) (*A.thaliana*)	222171
12939	IMAGE:79216	AHNAK	AHNAK nucleoprotein (desmoyokin)	79026
30572	IMAGE:1473168	ZC3HDC6	Zinc-finger CCCH-type containing 6	376940

Table 3 Genes whose expression is downregulated in at least four of six DDP-resistant cell pairs

Feature no.	Clone ID	Gene symbol	Gene name	Locus link
30691	IMAGE:878468	DPH2L1	Candidate tumor suppressor in ovarian cancer 2	1801
38803	IMAGE:235882	MRC2	Endocytic receptor (macrophage mannose receptor family), mannose receptor C type 2	9902
10149	IMAGE:509641	IFITM1	Interferon induced transmembrane protein 1 (9–27)	8519
5232	IMAGE:868555	No gene symbol	EST	
22436	IMAGE:1417815	RIMS1	Homo sapiens cDNA FLJ38407 fis, clone FEBRA2008859 regulating synaptic membrane exocytosis 1	22999

cases, for cell lines HEY and UCI, the gene in question was CLDN4.

Ontology/pathway analysis

One reason for identifying genes that are differentially expressed in resistant cells is to use these to point out biochemical mechanisms or cellular functions that might mediate DDP resistance. The 20 genes differentially expressed in at least four of the six cell pairs were examined to determine whether they were associated more frequently than would be expected by chance alone with one of the biochemical pathways defined by the Kyoto Encyclopedia of Genes and Genomes (http://www.KEGG.org) or with one of the ontological categories defined by the Gene Ontology Consortium (http://www.geneontology.org). Genes were identified that had a hypergeometric P-value < 0.05 under the null hypothesis that the genes identified by SAM were not meaningfully associated with known pathways or functional categories. After a further stringent filtering of these candidate categories for low false discovery, none were found to contain a disproportionate number of the genes identified as differentially expressed by SAM analysis.

Classification of cell pairs based on expression of genes identified by SAM analysis

The six pairs of cell lines were clustered on the basis of the Euclidean distance of the average $\log_2 (R/S)$ for the 26 features that were identified by SAM analysis as differentially expressed in four of six cell pairs. As shown in Fig. 2, the six cell pairs resolved into two groups consisting of the KF, HEY, and 2008 pairs and the A2780, IGROV-1, and UCI pairs. The order of the merges in the tree did not correspond to the degree of DDP resistance in each cell pair. Thus, although Fig. 2 suggests the existence of two separate classes of resistant cells, these classes were not distinguished by their degree of DDP resistance.

Co-regulation of genes associated with DDP resistance

Having identified a set of 26 features associated with the DDP-resistant phenotype, it was of interest to determine the extent to which the degree of differential expression of any one gene was correlated with the degree of differential expression of each of the other genes. Figure 3 shows a heat map of the Pearson coefficient for the correlation of each gene with each of the other genes in the set. Within this set there are several genes whose

Table 4 Quantitative real-time PCR of selected genes identified by SAM analysis as being differentially expressed

Gene	Ratio of mRNA abundance in resistant cell line to that in sensitive line[a]						Number of cell lines in which feature altered		Concordance with microarray results
	2008	A2780	HEY	IGROV-1	KF	UCI	Up	Down	
MT2A	2.94	21.25	0.14	987.42	1.37	5.13	5/6	1/6	NE[b]
STARD4	3.12	64.82	2.09	168.51	0.72	14.91	5/6	0/6	5/6
F3	3.88	2.77	3.38	7.41	1.21	1.48	6/6	0/6	4/5
ANXA1	3.19	62.04	0.89	3.15	0.84	1.3	4/6	0/6	4/6
CLDN4	5.34	5.11	0.06	1.37	6.29	0.04	4/6	2/6	4/6
TIMP1	2.86	15.81	2.14	0.83	2.58	34.58	5/6	0/6	5/6

[a]Mean of triplicate measurements
[b]Not evaluable

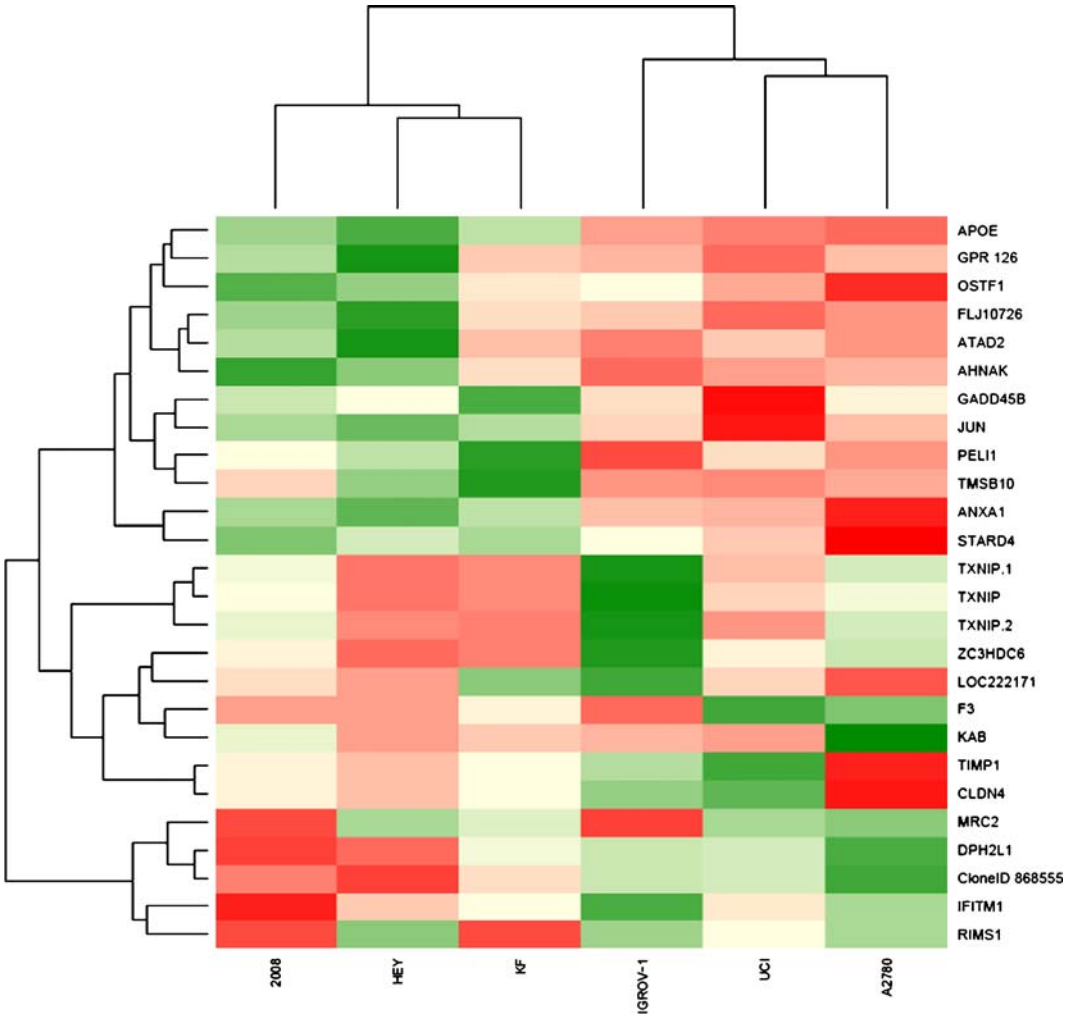

Fig. 2 Average linkage hierarchical clustering by Euclidean distance of the mean $\log_2 (R/S)$ for features that passed quality control in at least four of six replicates for the 26 features that were differentially expressed in four of six cell pairs. Increased expression is depicted as a more intense red and reduced expression as darker green. Missing values and lack of differential expression appear as a mid-tone yellow

degree of differential expression was highly correlated with that of multiple other genes. For example, the degree of differential expression of ApoE was correlated with a large number of other genes; it had the highest mean correlation coefficient ($r = 0.47$). In contrast, the degree of differential expression of CLN4 was highly correlated with that of only one other gene, TIMP1. However, for all the genes in the set there was at least one other gene that exhibited a similar degree of differential expression across the six cell pairs, suggesting that they function in a parallel manner with respect to the resistant phenotype and that the pattern of expression is more closely linked with resistance than that of any one gene.

Discussion

Differential expression of a given gene can arise from genetic or epigenetic changes specific to that gene, secondary effects of mutations in other genes or non-specific effects of random mutations in genes unrelated to the resistant phenotype. To sort these out, one must identify the differentially expressed genes and have a high level of confidence in this identification; this was the primary goal of this study. Regardless of the functional relevance of any particular gene, the validity of the signal is paramount, and it is essential to determine the consistency with which it is differentially expressed both within multiple replicates from a given cell pair and between multiple pairs of sensitive and resistant cell lines. Furthermore, a large number of believably differentially expressed genes is needed to landmark known biochemical pathways and cellular functional categories that can lead to the identification of mechanisms. In this study we approached these challenges by using six cell pairs and a large number of replicates. RNA was independently isolated from separate flasks of cells growing under identical conditions,

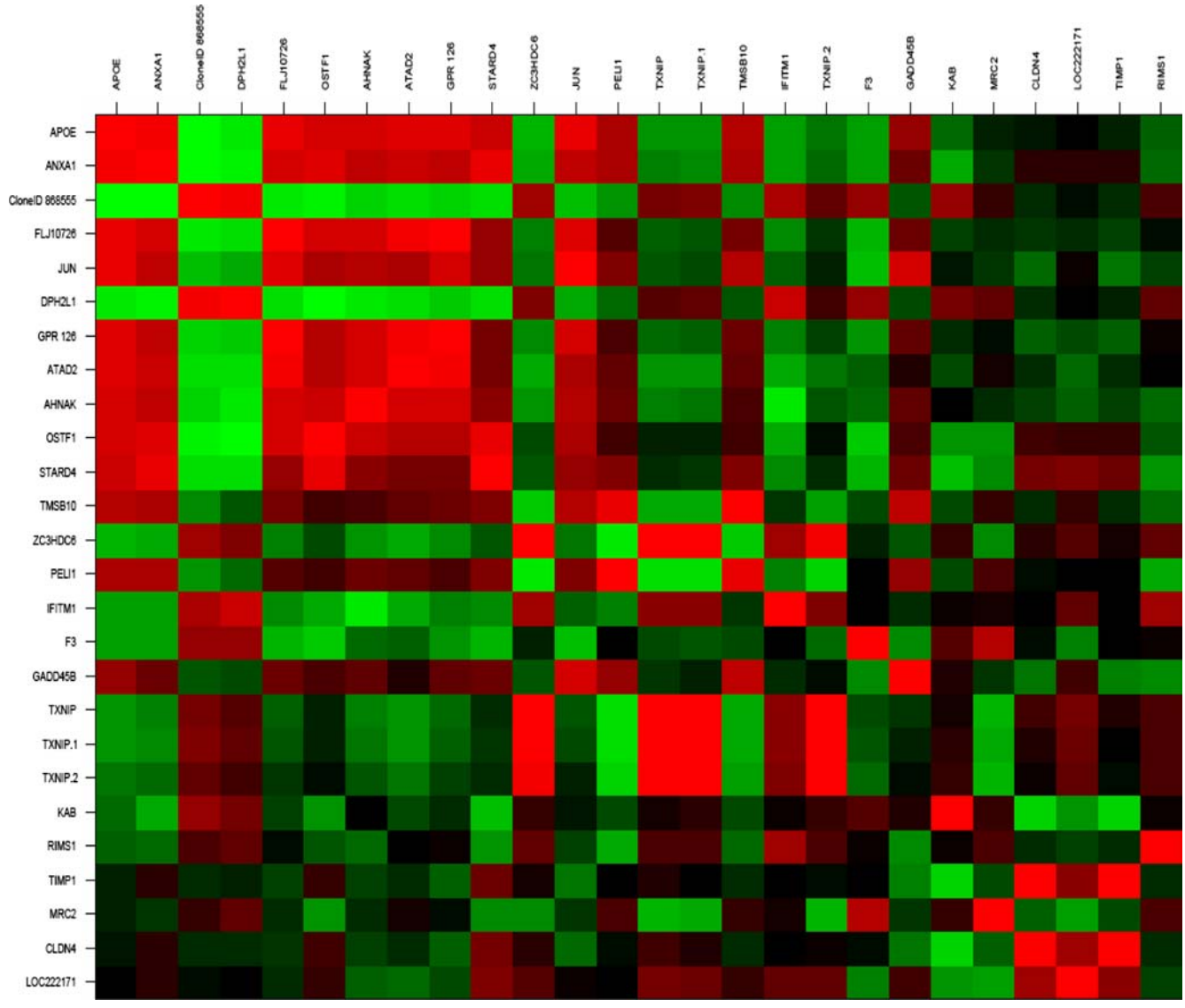

Fig. 3 Image map of Pearson correlation values for the 26 features SAM identified in four of six cell pairs using $\log_2 (R/S)$ for all features passing quality control in four of six replicates. The horizontal axis orders genes from left to right from highest to lowest mean correlation value (mean of correlation coefficients with all other 25 features) and the vertical axis orders genes reading downwards from highest to lowest correlation with ApoE

processed independently, and then analyzed in separate hybridization reactions.

The quality of the data from each array is a crucial factor in searching for genes of interest using cDNA microarrays. The use of multiple replicates permitted a rigorous assessment that disclosed that reproducibility varied significantly among the different cell pairs despite the fact that all were grown under identical conditions and analyzed with identical procedures. Reproducibility across the six replicates was very good for the KF, IGROV-1, and HEY cell line pairs and acceptable for the A2780 and UCI cell line pair. The reproducibility for the 2008 and UCI cell line pairs was poor; however, a careful review of the technical aspects of the experiments performed with the 2008 and UCI cells did not reveal any candidate discrepancies and the source of the additional variance in these pairs remains unidentified. Interestingly, these pairs also demonstrated the smallest number of differentially expressed features (488 for 2008 and 537 for UCI), suggesting that multiple replicates of high reproducibility are necessary for efficient identification of genes for which there is a high level of confidence that they are really differentially expressed. These analyses of reproducibility now provide a benchmark against which future microarray experiments using cDNA microarrays can be compared.

Even when comparisons were made between DDP-sensitive and stably resistant cell lines growing under highly standardized conditions, the noise associated with this cDNA microarray data was high. A large number of

replicates were required in order to identify most of the significantly differentially expressed genes. The fact that fully 38.3% of the genes that were identified by SAM as being significantly differentially expressed would have been missed if only four instead of six replicates had been used has implications for future studies. It is important to find as many of these genes as possible, since false negative results in some cell line pairs will make genes that are found in other pairs seem to be the result of mutations unrelated to resistance. Further, the ability to identify cellular functions or biochemical pathways that are altered in the resistant cells increases when multiple genes are identified that share a pathway or function. Although expensive, there appears to be little alternative to the use of a large number of biologically independent replicates in future studies directed at identifying consistently differentially expressed genes.

One of the major challenges in the analysis of cDNA microarray data is to develop a strategy for judging the statistical significance of a given level of differential expression given the multiple comparisons intrinsic to the microarray approach. In these studies we used the SAM strategy developed by Tusher et al. [17]. The parameters of this analysis were set such that the median number of features identified in error (false positive rate) was ~1. This represents a very stringent threshold, but it is noteworthy that this approach nevertheless identified a large number of differentially expressed genes within each cell pair. Thus, using these cDNA arrays, it was not difficult to identify genes that met the SAM criteria of significant differential expression even in the case where the variance across the six replicates was high as in 2008 and UCI cell pairs.

Overall, there was good agreement between the microarray and qRT-PCR results for the five genes analyzed. The exceptions are 7 out of 29 comparisons in which one method shows differential expression while the other shows none or differential expression in the opposite direction. In the 7 of 29 possible comparisons where the two techniques gave discordant results, one of the techniques simply failed to show differential expression suggesting either inadequate sensitivity or differential splicing of transcripts that affected detection.

It is the genes that are differentially expressed in a large number of cell pairs that are of greatest interest with respect to understanding mechanisms that mediate DDP resistance. However, despite the large number of genes differentially expressed in each cell pair, none of these were consistently differentially expressed in all six, or even in five of the six, cell pairs. A total of 26 features corresponding to 22 known genes and 2 ESTs were differentially expressed in four of six cell pairs; 21 of these features (18 genes and 1 EST) were upregulated and five (four genes and 1 EST) were downregulated. This suggests one of two conclusions: either the mechanisms of resistance in each pair are largely unique to that pair; or, the vast majority of the genes exhibiting differential expression are not primary determinants of the resistant phenotype. Clustering of the six cell pairs based only on the genes identified as significantly differentially expressed by SAM analysis provided some evidence for two classes of DDP-resistant phenotypes within which there are additional cell pair-specific alterations. However, these classes were distinguished by their expression profile rather than by their degree of resistance.

The primary approach taken in this study was to identify the genes that were statistically significantly differentially expressed within each cell pair using SAM analysis, and then ask how many of these genes were consistently differentially expressed in multiple cell pairs. Among the 26 features differentially expressed in four of six isogenic pairs there are a number corresponding to genes whose protein products are known to be involved in functions that might reasonably be expected affect the ability of DDP to kill the cell including DNA damage response signaling (GADD45B, c-JUN), cell–surface interactions and signaling (claudin 4, apolipoprotein E, and annexin A1), lipid transport (STARD4), and cell growth and differentiation (thymosin β10 and tissue factor) as well as protein stability and cell-cycle arrest (thioredoxin interacting protein). An alternative approach to identifying genes differentially expressed in each pair is to average the $\log_2 (R/S)$ for each feature across the six replicates for each cell pair, and then use this mean value to perform a SAM analysis across the six pairs. This alternative and more stringent approach identified only a single gene, metallothionein 2A, as being differentially expressed. The fact that this analysis identified metallothionein 2A if of particular interest since this protein has long been known to modulate sensitivity to DDP. Elevated levels of metallothioneins have been found in some DDP-resistant cell lines and these sulfur-rich proteins are thought to be involved in sequestering DDP by chelation through thiol groups so that interaction of the drug with key cellular targets [5, 20]. The successful identification of this protein as being differentially expressed provides evidence that this type of analysis does in fact identify genes of interest with respect to the DDP-resistant phenotype. However, it is noteworthy that neither approach identified any of the other genes previously implicated in DDP resistance, including those involved in nucleotide excision repair, DNA-damage recognition, DNA mismatch repair, apoptosis, or DDP transport.

Although the resistant phenotype might result from large changes in the expression of a small number of important genes, the results of this study suggest instead that it is the result of subtle changes in an ensemble of genes. The possibility that these function in a common biochemical or signaling pathway was investigated by considering all the genes known to be associated with pathways or cellular functions as defined in the Kyoto Encyclopedia of Genes and Genomes and the Gene Ontology Consortium databases and applying a Fisher exact in/out of category calculation and a Wilcoxon rank sum test using a Komolgorov–Smirnov statistic to SAM-identified genes that passed quality control in

at least four of six replicates. The Komolgorov–Smirnov statistic was considered necessary since the usual reference distributions for these test statistics do not apply as groups of genes tend to be jointly regulated. The failure to identify any category as being statistically significantly associated with the genes identified suggests that DDP resistance may not be the result of co-opting canonical pathways but more a multifactorial phenomenon involving genes and proteins that have as yet uncharacterized behaviors.

Not all differentially expressed genes are expected to be causally linked to the resistant phenotype in ovarian cancer; however, certain new mechanistic insights are suggested by the data from the current study. Among the 26 features of greatest interest, annexin A1, TIMP1, CLDN4, and STARD4 are all implicated in membrane function and intracellular trafficking, functions essential to the uptake of DDP and its access to the nucleus. Drug resistance associated with altered cell–cell contacts and gap junction communication between cells has previously been reported [3, 13]. Loss of an adherens junction protein, β-catenin, presumably through proteolytic degradation, has also recently been associated with DDP resistance [9]. As loss of cell–cell contacts (anoikis) has been noted to induce apoptosis [4], increased contact through tight junctions as a result of upregulation of component proteins such as CLDN4 may favor survival. Recently, sharing of a 'death signal' instigated by the Ku70/Ku80/DNA-dependent protein kinase complex and transmitted from DDP-damaged cells to its neighbors through gap junctions has been demonstrated [7]. The known function of claudin 4 in regulating ion permeabilities [18] may likewise directly modulate either death or survival signals.

Annexin 1 (ANXA1) plays a role in apoptosis, cell adhesion and signaling, and has been linked to resistance to doxorubicin, melphalan, and etoposide in MCF-7 cells via a mechanism that is independent of MRP1 and PgP1 but is linked to enhanced vesicle trafficking [19]. The preponderance of other lipid metabolism and vesicle trafficking genes among the 26 genes identified in four of six cell pairs suggests a role for altered vesicle dynamics in resistant cells. STARD4 is a ubiquitous, cholesterol-regulated member of a family of homologous steroidogenic acute regulatory (STAR)-related lipid transfer proteins that shuttle lipids and sterols intracellularly [15]. Tissue factor has been shown to be regulated by endocytosis through at least two pathways, one of which is dependent on low-density lipoprotein receptor-related proteins such as ApoE. LDL-receptors have cysteine-rich modules and binding of DDP to sulfur moieties is well known. Finally, the data suggest alterations in cytoskeletal genes in DDP-resistant cells. Thymosin beta 10 has been implicated in actin reorganization and is associated with apoptosis [8]. Annexin 1 has been linked to disruption of the actin cytoskeleton through sustained activation of the ERK signaling cascade and inhibition of cyclin D1 expression resulting in reduced cell proliferation [1]. ApoE signaling in neurons is associated with microtubule depolymerization [2]. Furthermore, during the endocytotic recycling of tissue factor–Factor VIIa complexes, Factor VIIa is released from vesicles and is observed to bind to the actin cytoskeleton. At this juncture the association between DDP resistance and any of these potential mechanisms remains highly speculative, and the primary value of the identification of these genes is that they serve as signposts that direct further investigation.

The expression profiling approach used in this study was successful in identifying differentially expressed genes and in providing an unusually high degree of confidence compared to studies based solely on comparison of unrelated cell lines or tumors. To attain this degree of confidence many more technical and biological replicates were required than are conventionally used. The high confidence that the genes identified are really differentially expressed as detected by microarray-based expression profiling suggests these genes as novel markers of DDP resistance in ovarian cancer. Although the differentially expressed genes described in this work do not identify a single biochemical pathway or ontology category, they constitute a set of previously unidentified markers of DDP resistance in ovarian cancer that suggest here-to-fore undescribed mechanisms of resistance that may serve to predict DDP resistance in clinical samples.

Acknowledgement Supported by grants CA78648 and CA95298 from the National Institutes of Health, a post-doctoral fellowship grant from the Marsha Rivkin Foundation, and a grant from the Research Foundation. The authors express their appreciation to Dr. Bernard Palsson for use of the scanner and to Michael Fero and the staff of the Stanford Functional Genomics Facility for supplying the human cDNA microarrays used for this study.

References

1. Alldridge LC, Bryant CE (2003) Annexin 1 regulates cell proliferation by disruption of cell morphology and inhibition of cyclin D1 expression through sustained activation of the ERK1/2 MAPK signal. Exp Cell Res 290:93–107
2. Beffert U, Stolt PC, Herz J (2004) Functions of lipoprotein receptors in neurons. J Lipid Res 45:403–409
3. Croix BS, Rak JW, Kapitain S, Sheehan C, Graham CH, Kerbel RS (1996) Reversal by hyaluronidase of adhesion-dependent multicellular drug resistance in mammary carcinoma cells. J Natl Cancer Inst 88:1285–1296
4. Frisch SM, Francis H (1994) Disruption of epithelial cell–matrix interactions induces apoptosis. J Cell Biol 124:619–626
5. Holford J, Beale PJ, Boxall FE, Sharp SY, Kelland LR (2000) Mechanisms of drug resistance to the platinum complex ZD0473 in ovarian cancer cell lines. Eur J Cancer 36:1984–1990
6. Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. J Comput Graph Stats 5:299–314
7. Jensen R, Glazer PM (2004) Cell-interdependent cisplatin killing by Ku/DNA-dependent protein kinase signaling transduced through gap junctions. Proc Natl Acad Sci USA 101:6134–6139
8. Lee SH, Zhang W, Choi JJ, Cho YS, Oh SH, Kim JW, Hu L, Xu J, Liu J, Lee JH (2001) Overexpression of the thymosin beta-10 gene in human ovarian cancer cells disrupts F-actin stress fiber and leads to apoptosis. Oncogene 20:6700–6706

9. Liang XJ, Shen DW, Gottesman MM (2004) Down-regulation and altered localization of gamma-catenin in cisplatin-resistant adenocarcinoma cells. Mol Pharmacol 65:1217–1224
10. Mishima M, Samimi G, Kondo A, Lin X, Howell SB (2002) The cellular pharmacology of oxaliplatin resistance. Eur J Cancer 38:1405–1412
11. Rozen S, Skaletsky HJ (2000) Bioinformatics methods and protocols: methods in molecular biology. In: Krawetz S (ed) Primer3 on the WWW for general users and for biologist programmers. Humana, NJ
12. Safaei R, Howell SB (2005) Copper transporters regulate the cellular pharmacology and sensitivity to Pt drugs. Crit Rev Oncol Hematol 53:13–23
13. Shain KH, Dalton WS (2001) Cell adhesion is a key determinant in de novo multidrug resistance (MDR): new targets for the prevention of acquired MDR. Mol Cancer Ther 1:69–78
14. Siddik ZH (2003) Cisplatin: mode of cytotoxic action and molecular basis of resistance. Oncogene 22:7265–7279
15. Soccio RE, Adams RM, Romanowski MJ, Sehayek E, Burley SK, Breslow JL (2002) The cholesterol-regulated StarD4 gene encodes a Star-related lipid transfer protein with two closely related homologues, StarD5 and StarD6. Proc Natl Acad Sci USA 99:6943–6948
16. Storey JD, Tibshirani R (2003) SAM thresholding and false discovery rates for detecting differential gene expression in DNA microarrays. In: Parmigiani G, Garrett ES, Irizarry RA, Zeber SL (eds) The analysis of gene expression data: methods and software. Springer, Berlin Heidelberg New York
17. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA 98:5116–5121
18. Van Itallie CM, Fanning AS, Anderson JM (2003) Reversal of charge selectivity in cation or anion-selective epithelial lines by expression of different claudins. Am J Physiol Renal Physiol 285:F1078–1084
19. Wang Y, Serfass L, Roy MO, Wong J, Bonneau AM, Georges E (2004) Annexin-I expression modulates drug resistance in tumor cells. Biochem Biophys Res Commun 314:565–570
20. Yang CP, Galbiati F, Volonte D, Horwitz SB, Lisanti MP (1998) Upregulation of caveolin-1 and caveolae organelles in Taxol-resistant A549 cells. FEBS Lett 439:368–372
21. Yang Y, Dudoit S, Luu P, Lin D, Peng V, Ngai J, Speed T (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 30:e15–200