Application Note

Identification of Gene Expression Biomarkers Associated with Clinical Response in Ovarian Cancer

Finding Gene Sets that Discriminate Between Responding and Non-responding Patients Treated with Combination Platinum/Paclitaxel Chemotherapy

Eric Kaldjian, M.D.
Senior Scientific Director, Pharmacogenomics and Oncology

GENE LOGIC

Identification of Gene Expression Biomarkers Associated with Clinical Response in Ovarian Cancer

Finding gene sets that discriminate between responding and non-responding patients treated with combination platinum/paclitaxel chemotherapy

Background

The World Health Organization estimates that 190,000 new cases of ovarian cancer – and more than 110,000 deaths resulting from it – occur each year. Most patients with epithelial ovarian cancer, the most common form, are diagnosed after their cancer has already reached an advanced stage, having spread beyond the ovary and often the pelvis. A standard first line therapy for ovarian cancer is surgery followed by combination chemotherapy with a platinum-based compound and a taxane, such as Taxol (paclitaxel). Although the initial response rate to treatment is high, the long-term survival of patients with stage III and IV is poor – between 10 and 30% at five years. It would therefore be of considerable value to determine a molecular biomarker profile that can discriminate between tumors that will or will not yield lasting responses to this combination therapy so that alternative therapy could be pursued.

In this Application Note, data from Gene Logic's BioExpress® System are used to identify potential gene expression biomarkers associated with clinical response in ovarian cancer. The BioExpress® System contains the gene expression profiles of over 18,000 clinical and research samples processed using Affymetrix GeneChip® arrays. It includes more than 2,400 human malignancies, with extensive clinical data annotation. Using the Genesis Enterprise System® Software to mine the database, patients with ovarian cancer were identified that had associated clinical follow-up data. The cancer samples from these patients were grouped into responder and non-responder sets based on clinical outcome, and a simple but powerful analysis was performed to identify a set of genes that discriminate the two patient groups. The results demonstrate that a gene set may be rapidly identified in a small cohort of patients that can predict therapeutic outcomes. A similar approach to biomarker gene set identification may be used for discrimination between other clinically definable cancer sub-populations.

Materials and Methods

Sample sets were curated from Gene Logic's BioExpress® System by selecting samples of epithelial ovarian carcinoma, both primary and metastatic, mainly of serous cystadenoacarcinoma type. Clinical data fields in BioExpress® System demonstrated those samples that came from patients who had been treated with platinum-based (cis-platin or carboplatin) and paclitaxel combination chemotherapy (Figure 1).

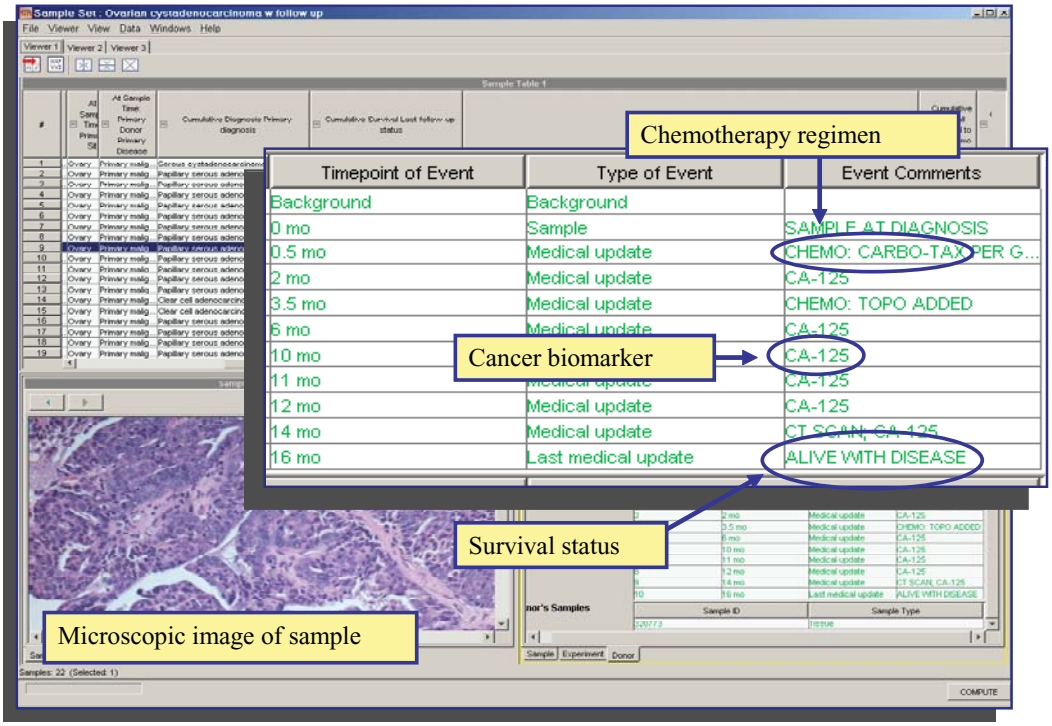
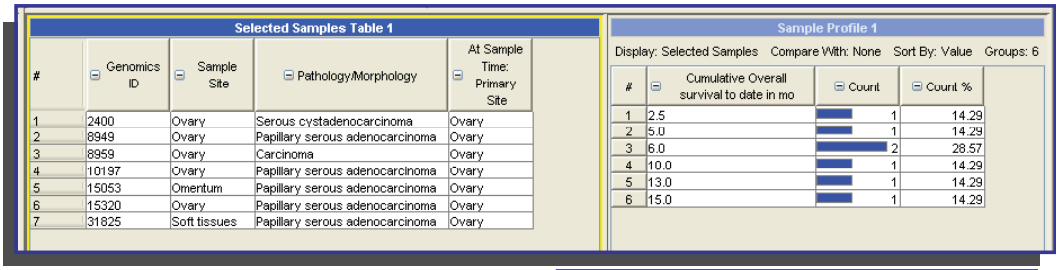


Figure 1: Curation of ovarian cancer sample sets in the BioExpress® System using the Genesis Enterprise System® Software. Each sample has links to extensive laboratory and clinical data. Shown here are histology and clinical course data, including treatment type, serologic biomarker evaluations and survival status. All green text is linked to further information within the BioExpress® System for that data field.

Clinical outcome annotations were used to determine response categories for the samples. Patient samples with clinical descriptions such as “no evidence of disease” following chemotherapy were placed in the complete response (CR) category. Patients whose tumors showed no clinical evidence of reduction in mass in the treatment course were placed in the non-response (NR) category. Samples with inconclusive response data were omitted from further consideration. The CR sample set totaled 9 samples and the NR totaled 7 samples. The median cumulative overall survival to date was 18 months in the CR group and 6 months in the NR group, consistent with a better clinical outcome (Figure 2).

NR set



Sample profile window demonstrates cumulative overall survival in each patient group.

CR set

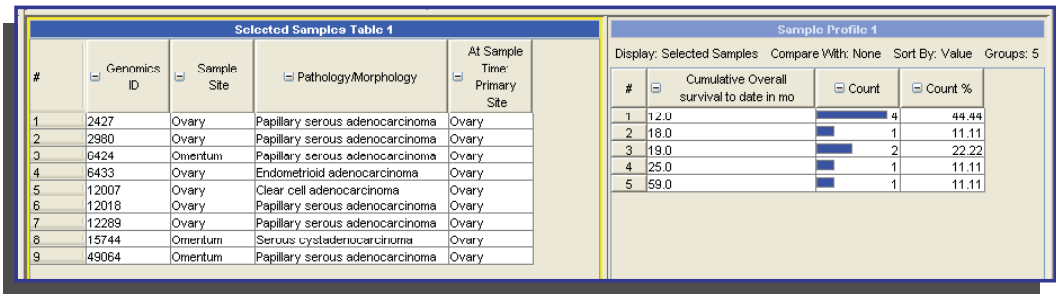


Figure 2: Summary data for samples sets. Displays with the Genesis Enterprise System® Software can be tailored for the presentation of data according to many different parameters. Shown are sample site and pathology/morphology. The majority of samples are of papillary serous adenocarcinoma, the most common primary ovarian epithelial malignancy. The linked sample profile window creates a histogram of the selected samples, in this case displaying cumulative overall survival.

The Genesis Enterprise System® Software incorporates a Contrast Analysis tool that allows rapid identification of genes which are differentially expressed between sample sets. This is performed by calculating an F-score, which is a statistical measure of the ability of each individual gene to discriminate between sample groups; higher F-scores result when sample values are more tightly grouped within each set as determined by variance. Using the Contrast Analysis tool, genes were identified by creating an analysis filter to select those genes showing differential expression between the two Sample Sets that were present in at least 20% of samples in one of the two sample sets (% present filter).

Results

When the F-score filter was set at 12.0 or greater, 54 genes were identified that discriminated between the NR and CR sample groups. A Principal Components Analysis (PCA) was performed in the Genesis Enterprise System® Software to visualize the separation between the clustered sample groups. To investigate whether fewer genes could discriminate between the sample sets, the F-score filter was increased to 14.0, which yielded a set of 23 genes. The

PCA plot of this set indicated that it could also separate NR and CR sample sets. The first principle component incorporated ~60% of the discriminatory power of the analysis, distinctly separating the sample sets into two groups (Figure 3). The overall contribution of each gene to the PCA analysis is listed in the "component loading" column. Sorting the display on this column allows genes to be rapidly ranked according to the ability to discriminate between the two sample sets. Note that the component loading ranking is not identical to the F-score ranking; this is because the PCA analysis utilizes information from the entire 53 gene set, whereas the F-score is an individual gene-based assessment of sample set discrimination.

Gene Table 1

#	FragmentID (ChipID)	Fragment Name	Sequence Clusters: Cluster Title	Known Genes: Gene Symbol	F-Score	Component Loading (1)
1	234212(28)	203550_s_at	chromosome 1 open reading frame 2	C1orf2	14.07	0.23
2	249501(28)	219002_at	hypothetical protein FLJ21901	FLJ21901	14.58	0.23
3	243061(28)	212541_at	FAD-synthetase	PP591	14.93	0.22
4	255188(29)	224714_at	MKI67 (FHA domain) interacting nucleolar phosphoprotein	MKI67IP	17.25	0.22
5	259368(29)	228899_at	cullin 1	CUL1	15.19	0.22
6	265451(29)	234987_at	Similar to dJ132F21.2 (Contains a novel protein similar to the L82E ...	na	20.75	0.21
7	258330(29)	227861_at	hypothetical protein MGC33214	MGC33214	15.08	0.21
8	239972(28)	209344_at	(tropomyosin 3, tropomyosin 4)	(TPM3, TPM4)	16.89	0.21
9	235767(28)	205105_at	mannosidase, alpha, class 2A, member 1	MAN2A1	16.53	0.21
10	266733(29)	236269_at	hypothetical protein LOC89887	LOC89887	14.66	0.19
11	268726(29)	238262_at	speedy	SPDY1	16.14	0.19
12	274338(29)	243874_at	MRNA; cDNA DKFZp686L23125 (from clone DKFZp686L23125)		16.46	-0.18
13	261801(29)	231332_at	Transcribed sequence with strong similarity to protein sp:P00722 ...		14.45	-0.19
14	247632(28)	217130_at	chromosome 9 open reading frame 33	C9orf33	14.59	-0.20
15	249472(28)	218973_at	elongation factor Tu GTP binding domain containing 1	EFTUD1	14.42	-0.20
16	243264(28)	212744_at	Bardet-Biedl syndrome 4	BBS4	17.05	-0.20
17	262518(29)	232049_at	CDNA: FLJ23065 fis, clone LNG04894		16.84	-0.20
18	245293(28)	214782_at	Transcribed sequence with moderate similarity to protein sp:P391 ...		14.17	-0.20
19	268908(29)	238444_at	zinc finger protein 618	ZNF618	20.32	-0.21
20	264186(29)	233720_at	Arg/Abl-interacting protein ArgBP2	ARGBP2	16.55	-0.21
21	236263(28)	205601_s_at	homeo box B5	HOXB5	14.30	-0.21
22	236262(28)	205600_x_at	homeo box B5	HOXB5	17.69	-0.22
23	239358(28)	208728_s_at	cell division cycle 42 (GTP binding protein, 25kDa)	CDC42	15.92	-0.23

Figure 3: Table of 23 genes identified in the biomarker set, with F-score and component loading fields. Genes with negative component loading values have the reverse expression pattern in the sample groups as genes with positive values. At the bottom is cdc42, the expression of which is shown in Figure 5.

The sample correlation Heat Map linked to the PCA plot allows simple visualization of the relatedness of each sample within the analysis to every other sample (Figure 4). In general the NR and CR samples show greater similarity to other samples within the same category than to those in the other category. Individual samples may be selected in the PCA plot to immediately view their relatedness to other samples in the correlation analysis.

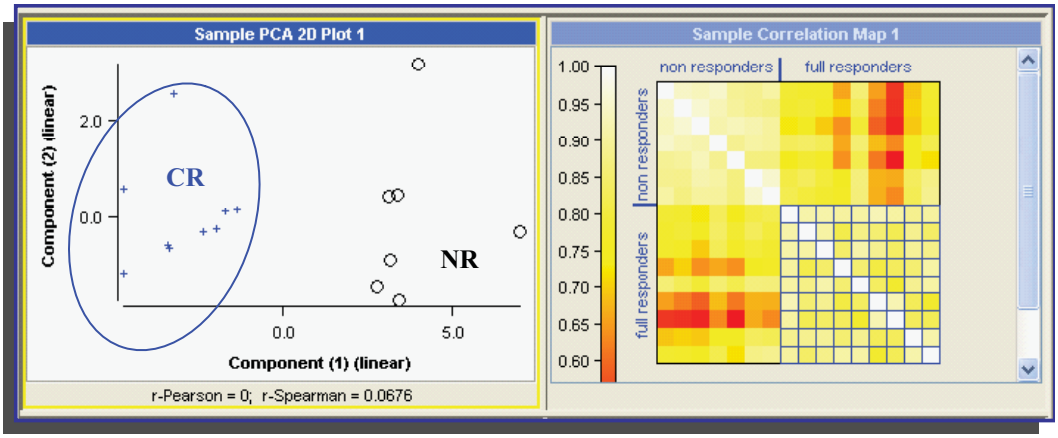


Figure 4: Principal Components Analysis and correlation mapping of responder sample groups using the 23 gene set. The NR and CR amples groups are separated distinctly by the first prinicpal component.

The e-Northern® tool provides Genesis Enterprise System® Software users with a powerful way to visualize the expression profile of a gene across an entire sample set. The range of expression of the gene in each sample set can be visualized to provide an assessment of ability to separate groups according to expression intensity. Figure 5 demonstrates the e-Northern® analysis for cdc42, one of the genes in the 23 gene set, cdc42 is a plasma membrane-associated small GTPase which in its active state binds to a variety of effector proteins, promoting cell cycle progression from G1 to S and regulating pathways considered essential to tumor growth. Tumors that responded to platinum/paclitaxel treatment had higher cdc42 expression than did the non-responding tumors. This supports the hypothesis that more actively cycling tumors are more susceptible to this therapy than those that are not proliferating as rapidly. Placement of a separating threshold (dashed bar in Figure 5) correctly categorizes 6 of 7 (86%) NR and all 9 (100%) CR samples for cdc42.

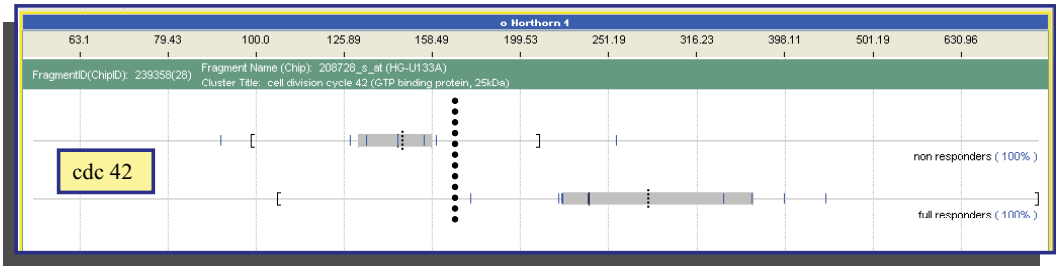


Figure 5: e-Northern® analysis of cdc42. The mean expression value for cdc42 is higher in the CR group than in NR group.

Identified genes of interest are also linked by the Pathway Viewer tool to a library of pathways in which the genes participate. Figure 6 shows a flag box around cdc42 indicating its involvement downstream in the ras signaling pathway. Use of these tools can guide the further selection, prioritization and development of biomarkers and assay platforms for application in the clinical context.

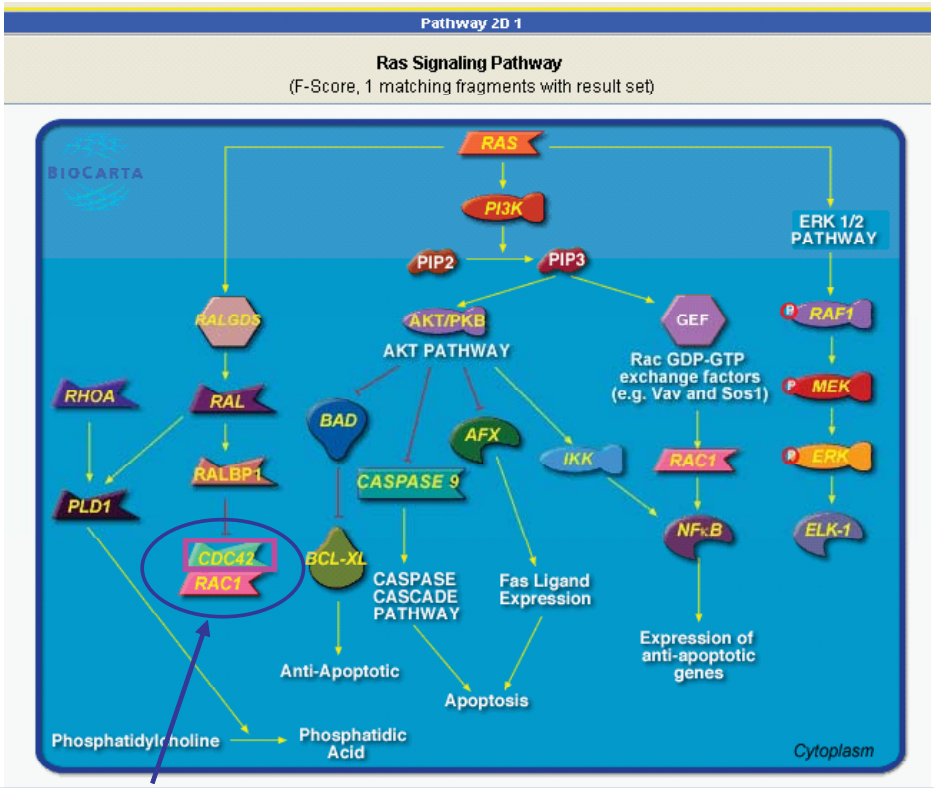


Figure 6: Linked Biocarta™ pathway map of ras signalling. Selected genes from analyses are flagged in pathway maps. In this example cdc42 is highlighted.

Conclusions

The extensive clinical annotation of samples within the BioExpress® System allows the curation of sample groups for biomarker investigation according to a wide variety of parameters. In this Application Note, clinical treatment regimen and outcome response were used to define groups of ovarian cancer samples from patients treated with a standard chemotherapy. Contrast Analysis within the Genesis Enterprise System® Software rapidly produced a reasonably sized candidate biomarker gene set that discriminated between responding and non-responding tumors. One of the genes was further investigated using associated tools, including the e-Northern® and Pathway Viewer, allowing rapid generation of a mechanistic hypothesis for how the genes may be reflecting response to treatment. A similar approach to biomarker gene set discovery may be applied to any sub-populations definable within the BioExpress® System, in cancer and in other disease areas.

References

Olson MF, et al. An essential role for Rho, Rac, and Cdc42 GTPases in cell cycle progression through G1. Science. 1995 Sep 1;269(5228):1270-2.

Spentzos D, et al. Gene expression signature with independent prognostic significance in epithelial ovarian cancer. J Clin Oncol. 2004 Dec 1;22(23):4648-58.

Stewart BW and Kleihues P (Eds): World Cancer Report. IARC Press. Lyon, 2003

610 Professional Drive • Gaithersburg, Maryland 20879 • Phone: 1.800.GENELOGIC • Fax: 301.987.1701
Email: info@genelogic.com • Web: www.genelogic.com