New Potential Ligand-Receptor Signaling Loops in Ovarian Cancer Identified in Multiple Gene Expression Studies

Giancarlo Castellano,1 James F. Reid,2,4 Paola Alberti,1 Maria Luisa Carcangiu,3 Antonella Tomassetti,1 and Silvana Canevari1

1Unit of Molecular Therapies, Department of Experimental Oncology, 2Department of Experimental Oncology, and 3Unit of Pathology C, Department of Pathology, Istituto Nazionale Tumori; and 4Molecular Genetics of Cancer Group, Fondazione Istituto FIRC di Oncologia Molecolare, Milan, Italy

Abstract

Based on the hypothesis that gene products involved in the same biological process would be coupled at transcriptional level, a previous study analyzed the correlation of the gene expression patterns of ligand-receptor (L-R) pairs to discover potential autocrine/paracrine signaling loops in different cancers (Graeber and Eisenberg. Nat Genet 2001; 29:295). By refining the starting database, a list of 511 L-R pairs was compiled, combined to eight data sets from a single pathology, epithelial ovarian cancer, and examined as a proof-of-principle of the statistical and biological validity of the correlation of the L-R gene expression patterns in cancer. Analysis revealed a Bonferroni-corrected significant correlation of 105 L-R pairs in at least one data set and, by systematic analysis, identified 39 more frequently correlated L-R pairs, 7 of which were already biologically confirmed. In four data sets examined for an L-R correlation associated with patient survival time, 15 L-R pairs were significantly correlated in short surviving patients in two of the data sets. EFNB3-EPHB4 pair was one of the newly identified L-R pairs (i.e., EFNB3-EPHB4) revealed the correlated expression of ephrin-B3 and EphB4 proteins in 45 of 55 epithelial ovarian tumor samples (P < 0.0001). Together, these data not only support the validity of cross-comparison analysis of gene expression data because known and expected correlations were confirmed but also point to the promise of such analysis in identifying new L-R signaling loops in cancer. (Cancer Res 2006; 66(22): 10709-19)

Introduction

Traditional hypothesis-driven strategies for identifying molecular markers of a disease state were based on individual gene analysis. Although useful, these approaches could fail to identify biological relevant differences that are based on subtle but multiple and coordinated gene alterations more than on quantitative expression differences of a single gene. Several recent critical advances, such as sequencing of the human genome and the development of high-throughput techniques for identifying global gene expression, have dramatically accelerated the speed of research. Although extrapolation of biological mechanisms from the generated gene lists remains a major challenge, the availability of numerous published microarray analyses, rich in the amount of high-quality data (1), and the public access to the original data sets have accelerated the developments of new types of analysis. In fact, the combination of hypothesis- and discovery-based research resulted in the development of new techniques, based on aggregated gene sets (reviewed in ref. 2), to extract useful information from microarray gene expression data sets (3, 4) and to interpret genome-wide expression profiles (5).

The use of pathway-oriented approaches has enabled the interrogation and dissection of multiple disrupted signaling pathways during oncogenesis. Accordingly, an algorithm was designed (6) that is suitable for detecting dysregulation of autocrine/paracrine ligand-receptor (L-R) signaling loops. This approach was based on the hypothesis that two gene products participating in a common function show correlated expression as reflected in their correlated transcription levels. However, to date, this algorithm has only been applied in a single study, in which five cancer-based gene expression data sets originated from different cancers were analyzed separately (6). In principle, this type of analysis could provide a tool to compare independently derived gene expression data sets, even those obtained from different platforms, and to obtain more consistent results than those from single gene analysis.

Here, we examined patterns of correlated gene expression of ligands and receptors with respect to their role as possible activated signaling pathways involved in epithelial ovarian cancer (EOC). The unfavorable statistics in EOC patients reflects, in part, the poor understanding of the molecular pathogenesis and progression of the disease. As a step toward gaining insight into the mechanisms underlying this pathology and toward identifying potentially meaningful activated signaling pathways, we exploited a previously described L-R database (6) to select frequently correlated L-R pairs by a “systematic” analysis of EOC publicly available data sets of gene expression. Analysis across eight selected EOC microarray data sets gave 39 L-R pairs with significant and consistent correlation in at least three data sets. In four data sets, analysis of samples from EOC patients with short-term versus long-term survival showed that 15 L-R pairs were associated to short-term survival in two of the data sets. EFNB3-EPHB4 pair was one of the newly identified L-R pairs and the coexpression was confirmed at the protein level by immunohistochemistry on epithelial ovarian tumors.

Materials and Methods

Public gene expression data. Twenty-five publications on microarray analysis of gene expression profiling of EOC samples were recorded from PubMed5 (from January 2000 to May 2005). Gene expression data were

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Requests for reprints: Antonella Tomassetti, Unit of Molecular Therapies, Department of Experimental Oncology, Istituto Nazionale Tumori, 20133 Milan, Italy. Phone: 39-02-23902568; Fax: 39-02-23903073; E-mail: antonella.tomassetti@istitutotumori.mi.it.

©2006 American Association for Cancer Research. doi:10.1158/0008-5472.CAN-06-1327

www.aacrjournals.org
Table 1. Characteristics of the explored EOC data sets

<table>
<thead>
<tr>
<th>Code</th>
<th>Author (ref.)</th>
<th>Platform Name</th>
<th>No. cDNA or probesets</th>
<th>No. samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Burchuck et al. (11)</td>
<td>Oligo (U133A)</td>
<td>22,283</td>
<td>68</td>
</tr>
<tr>
<td>II</td>
<td>De Cecco et al. (12)</td>
<td>cDNA</td>
<td>4,451</td>
<td>81</td>
</tr>
<tr>
<td>III</td>
<td>Lancaster et al. (13)</td>
<td>Oligo (Hu GeneFL)</td>
<td>7,070</td>
<td>34</td>
</tr>
<tr>
<td>IV</td>
<td>Spentzos et al. (14)</td>
<td>Oligo (U95Av2)</td>
<td>12,625</td>
<td>68</td>
</tr>
<tr>
<td>V</td>
<td>Schaner et al. (15)</td>
<td>cDNA</td>
<td>42,000</td>
<td>59</td>
</tr>
<tr>
<td>VI</td>
<td>Jazaeri et al. (16)</td>
<td>cDNA</td>
<td>7,651</td>
<td>61</td>
</tr>
<tr>
<td>VII</td>
<td>Schwartz et al. (17)</td>
<td>Oligo (Hu GeneFL)</td>
<td>7,070</td>
<td>113</td>
</tr>
<tr>
<td>VIII</td>
<td>Welsh et al. (18)</td>
<td>Oligo (Hu GeneFL)</td>
<td>7,070</td>
<td>38</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intersections of L-R pairs in data sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>147</td>
</tr>
<tr>
<td>68</td>
</tr>
<tr>
<td>387</td>
</tr>
<tr>
<td>89</td>
</tr>
</tbody>
</table>

B.

<table>
<thead>
<tr>
<th>Clinical characteristics*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data set</td>
</tr>
<tr>
<td>I</td>
</tr>
<tr>
<td>II</td>
</tr>
<tr>
<td>III</td>
</tr>
<tr>
<td>IV</td>
</tr>
<tr>
<td>V</td>
</tr>
<tr>
<td>VI</td>
</tr>
<tr>
<td>VII</td>
</tr>
<tr>
<td>VIII</td>
</tr>
</tbody>
</table>

Abbreviations: Undif, undifferentiated; oligo, oligonucleotide; n.a., not available.

*Due to missing values in a few cases, numbers do not add up to the total number of patients in every category.

†The respective authors defined the response to treatment as complete versus partial or no response in data set II and complete and partial versus no response in data set IV.

Available in only eight of these publications and used for our analysis (Table 1A). These data sets were generated by hybridization on cDNA and oligonucleotide DNA chips in three and five cases, respectively. No additional data manipulation was done to the downloaded processed gene expression matrices, except for the thresholding of negative values to 0 for MAS4-processed data (data sets III and VII). All probe sets and cDNA clones from each platform were assigned National Center for Biotechnology Information (NCBI) gene identifications (7), which were used to match common genes across data sets based on the most recent platform annotation from either NetAffx® (release 2005-12-20; ref. 8) or Stanford SOURCE® (release 2005-10-26; ref. 9). All data (clinical, platform annotation and expression) were stored in BioConductor® (10) expression sets.

L-R pair compilation. The list of L-R pairs was manually curated by a PubMed search for proteins interacting on the cell surface. The list was integrated to database of L-R partners (DLRP) contained in the database of interacting proteins® (6), yielding 199 ligands and 157 receptors for a total of 511 L-R pairs.

Correlation analysis of L-R pairs. Expression measurements of each L-R pair were extracted from each data set through their respective NCBI gene identification. When more than one cDNA clone or probe set matched a given gene, all possible pairs where considered. Pearson and Spearman correlation coefficients were computed for each L-R pair across each data set; Ps for each correlation were computed using the function cor.test of the software package R® and adjusted for multiple testing using the Bonferroni method. The complete list of the extracted correlations for each data set is available.

Tissue samples and study subjects. The pathologic and clinical characteristics of EOC patients (Table 1B) were derived from published data (11–18) or, for our previous study (data set II), by up-to-date clinical information. Data sets I to IV reported clinical information (Table 1B) that allowed subgrouping of the samples according to overall survival. Criteria for subcategorization in short-term and long-term survivors have been reported for data sets I, III, and IV (11, 13, 14); for data set II (12), the 36 samples from patients with available follow-up data were split into two groups to obtain median overall survival times similar to those reported in ref. 11.

Immunohistochemistry. All clinical specimens used in this study were obtained with Institutional Review Board approval and informed consent to use excess biological material for investigative purposes from all participating patients. Immunohistochemistry was done using nine routine

7 http://source.stanford.edu/.
9 http://dip.doe-mbi.ucla.edu/dip/DLRP.cgi.
tissue blocks and two commercially available tissue arrays (Ovary cancer, AccuMax Array, Petagen and C1 Human, Ovary cancer, and Super Bio Chips) essentially as described. Fifty-five samples were selected by the pathologist (M.L.C.) and classified in tumors of low malignant potential (LMP) and carcinomas of either grade II and III. Conditions for antigen retrieval and dilution for each antibody were optimized as suggested by the manufacturer. For staining with mouse monoclonal anti-EphB4 antibody (Zymed, San Francisco, CA), tissue sections were heated for 20 minutes in citrate buffer (pH 6) in a pressure cooker; for staining with the rabbit polyclonal anti-ephrin-B3 antisemum (Zymed), antigen retrieval was by pepsin digestion (0.4% in 0.01 N HCl for 30 minutes at 37°C). Negative control slides omitting primary antibody were included in all assays. Two observers (M.L.C. and A.T.) classified independently as slides as negative, weak/moderate, or strong staining based on intensity and percentage of positive cells. Statistical analysis of L-R correlated expression levels in the same sample was done using $\chi^2$ test. $P < 0.05$ (two sided) were considered significant.

Results

Generation of DLRP-rev1. The original list of L-R pairs present in the DLRP database was refined based on a literature search for proteins interacting on the cell surface and whose interactions were experimentally proved, which yielded 25 ligands and 33 receptors for a total of 42 added L-R pairs. The revised list (DLRP-rev1) used in this study consisted of 199 ligands, 157 receptors, and 511 L-R pairs involved in autocrine/paracrine signaling in eukaryotic cells (Supplementary Table S1). This final set of L-R pairs was subdivided according to functional consequence of the L-R interactions [i.e., angiogenesis factor (4%), chemokine (15%), cytokine (23%), growth factor (42%), motility/adhesion factor (14%), and others (2%)].

Identification of correlated L-R pairs in EOC. Analysis of the eight data sets revealed extensive variability in the L-R pairs coverage according to platform type and time of manufacture (Table 1A). The L-R pairs from DLRP-rev1 extractable from the eight selected EOC microarray data sets (shaded values) ranged from 68 to 417 (excluding redundant pairs) and represented a maximum of 82.5% DLRP-rev1 pairs in data set I to <13.5% DLRP-rev1 pairs in data set II. The matrix measuring the overlap provides the possible cross-comparisons between the data sets (Table 1A).

When multiple probe sets or clones matching a single gene (ligand or receptor) were present, each one was considered separately. Both Spearman and Pearson correlation coefficients were computed for each L-R pair in each data set. Because correlation coefficients ($\rho$ and $r$) and the associated $P$s assigned for each L-R pair were very similar (data not shown), only the Pearson correlation values were considered. A total of 105 L-R pairs had significant correlation after Bonferroni correction in at least one data set (Supplementary Table S2). These L-R pairs were subsequently analyzed across the other data sets for concordant correlations. After several arbitrary selections (e.g., Bonferroni-corrected significant correlations in two or three data sets and/or significant correlations in at least half of the informative data sets), the L-R pairs that showed a significant correlation after Bonferroni correction in at least one data set and have significant correlations, without Bonferroni correction, in at least two other data sets were considered potentially meaningful. Forty-one L-R pairs showed a significant correlation in at least three data sets (Table 2). This arbitrary cutoff produced a substantial number of L-R pairs shown to be implicated in EOC molecular signaling [PDGFA-PDGFR (19), CXCL12-CXCR4 (20), CSF1-CSF1R (21, 22), FGF2-FGFR4 (12), IGF2-IGF2R (23), HGF-MET (24), and PLAU-PLAUR (25)]. Correlations were consistently positive or negative for 36 and 3 L-R pairs, respectively, whereas 2 L-R pairs, CCL4-CXCR5 and TGFB1-TGFBR2, showed a contradictory trend of correlation in different data sets and were excluded from further analysis. Among the 39 consistently correlated L-R pairs, 3 pairs are involved in angiogenesis, 13 are chemokine family members, 9 are cytokine family members, 11 mediate growth, and 5 regulate cell migration/adhesion.

Identification of L-R pairs correlated in EOC subgroups according to survival. Not all tested samples were from surgical specimens (the number ranged from 22 in data set V11 to 113 in data set VII) and the type of clinical information was variable in the different data sets. Based on the available clinical information, EOC samples could be subdivided according to histotype, grading, staging, response to treatment, and outcome (Table 1B). Because the identification of predictive markers of poor outcome is still a major challenge in ovarian cancer, we focused our further analysis on data sets I to IV where outcome data were available. The samples used for this analysis were all, but three in data set IV, at late stage of the disease and several other clinical variables are known and recorded (see respective references for the criteria for subcategorization in short-term and long-term survivors and further clinical information). Samples with complete clinical history were classified according to length of survival (Table 3A), and the correlation analysis was focused on L-R correlation in short-term and long-term survivors. Initially, we selected L-R pairs that were significantly correlated in short-term survivors and not or inversely correlated in long-term survivors from the same data set. When no correction for multiple test was done, 166 L-R pairs were found significantly correlated (Supplementary Table S3). No L-R pairs showed a significant correlation after Bonferroni correction, probably due to the limited number of samples for each subgroup (range, 16-37). When we arbitrarily based our external statistical validation on L-R pairs showing a concordant correlation in at least two data sets in short-term survivors, the number of correlated L-R pairs dropped to 15 pairs (Table 3B). A support to our selection criteria comes from the observation that in about a third of the pairs significantly correlated in short-term survivors (LIF-IL6ST, FGFL1-FGFR1, FGF4-FGFR3, and FGF7-FGFR2) we observed in the long-term survivors from the same data set an opposite correlation. The ability to discriminate between patient subsets, as well as the inclusion of the IGF2-IGF2R pair already associated with a worse prognosis (23), further supported the validity of our selection. Correlations in short-term survivors were consistently positive or negative for 10 and 5 L-R pairs, respectively. Of the selected L-R pairs, only IGF2-IGF2R and EFN33-EphB4 were identified as correlated also in the entire case material (see Table 2), suggesting that the weight of correlation in short-term survivors only was very strong.

Among the 15 correlated L-R pairs, 1 is involved in angiogenesis, 1 and 2 belong to the chemokine and cytokine families, respectively, 6 mediate growth, and 5 regulate cell migration/adhesion. Furthermore, we found that four of the six growth-associated L-R pairs exhibited a negative correlation and four of the five motility/adhesion factors belong to the ephrin-Eph receptor family.

EFNB3-EphB4 correlated protein expression in epithelial ovarian tumor specimens. Emerging information on the roles of ephrin proteins and their receptors links them to tumorigenesis and invasion (26). Although the EFN3-EphB4 L-R pair is tagged in the original DLRP as experimentally determined, it is not published...
described as a canonical pair in normal or cancer signaling (27).
In our analysis, the EFNB3-EPHB4 pair was significantly correlated in four of seven EOC data sets and, in two of them, the correlation was also significant only in the short-term survivor subgroup. Thus, immunohistochemical analysis was carried out to determine whether these proteins are coexpressed in epithelial ovarian tumors. Consistent with previous reports (28, 29), anti-ephrin-B3 antibodies selectively reacted with arterioles, whereas they did not stain venous vessels and the surrounding stromal cells (Fig. 1A, 1B). Anti-EphB4 was strongly

<table>
<thead>
<tr>
<th>Table 2. L-R pairs with statistically significant correlation coefficient (Pearson r) after Bonferroni correction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data set</strong></td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>Angiogenesis factors</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Chemokines</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Cytokines</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Growth factors</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Motility/adhesion factors</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

NOTE: Shaded areas identify significant correlation (P < 0.05); bold characters identify a significant correlation after Bonferroni correction. *Discordant correlations observed for the indicated L-R pairs in different data sets.
reactive with the cell membrane of a colonic adenocarcinoma but was negative on stromal cells surrounding the tumor cells (Fig. 1A, 2). Representative staining with anti-ephrin-B3 and anti-EphB4 antibodies on different histotypes of EOCs is shown (Fig. 1B, 1-8). Anti-ephrin-B3 staining was strong and well defined on the tumor cell membrane of all the different histotypes with some staining in the cytoplasm of tumor and stromal cells; some tumor nuclei also showed weak/moderate reactivity (Fig. 1B, 1, 3, and 7). Anti-EphB4 antibody strongly stained the membrane of tumor cells of all the different histologic types (Fig. 1, 5-7), with the exception of clear cell carcinoma where was observed only a weak/moderate cytoplasmic staining of the tumor cells. Basically, no reactivity was observed with anti-EphB4 antibody on stromal cells surrounding EOC cells. On the mucinous-type tumor, the staining with both anti-ephrin-B3 and anti-EphB4 appeared reduced in some cells because of the cytoplasmic filling by mucous (Fig. 1B, 3 and 4). The anti-ephrin-B3 and anti-EphB4 staining on 8 LMP ovarian tumors and 47 EOC is summarized in Table 2.

<table>
<thead>
<tr>
<th>Data set</th>
<th>L-R pairs with statistically significant correlation coefficient (Pearson r) after Bonferroni correction (Cont’d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angiogenesis factors</td>
<td>ANGPT1 TEK 0.12 0.46809 0.11 0.39983 0.15 0.10732 0.00069 0.40 0.00682</td>
</tr>
<tr>
<td>PDGFA* PDGFRA 0.34 0.33128 0.22 0.08806 0.31 0.00069 0.40 0.00682</td>
<td></td>
</tr>
<tr>
<td>PFG FLT1 0.22 0.16112 0.15 0.10732 0.22 0.08806 0.31 0.00069 0.40 0.00682</td>
<td></td>
</tr>
<tr>
<td>Chemokines</td>
<td>CCL19 CCR7 0.24 0.01008 0.08 0.54674 0.32 0.03172 0.32 0.03172</td>
</tr>
<tr>
<td>CCL2 CCR1 0.23 0.13944 0.15 0.11226 0.22 0.15490 0.12 0.39983 0.31 0.00069 0.40 0.00682</td>
<td></td>
</tr>
<tr>
<td>CCL21 CCR7 0.24 0.01008 0.08 0.54674 0.32 0.03172 0.32 0.03172</td>
<td></td>
</tr>
<tr>
<td>CCL25 CCR9 0.41 0.00676 0.11 0.42422 0.47 0.00000 0.40 0.00682</td>
<td></td>
</tr>
<tr>
<td>CCL3 CCR1 0.41 0.00676 0.11 0.42422 0.47 0.00000 0.40 0.00682</td>
<td></td>
</tr>
</tbody>
</table>
| Anti-EphB4 antibody strongly stained the membrane of tumor cells of all the different histologic types (Fig. 1, 5-7), with the exception of clear cell carcinoma where was observed only a weak/moderate cytoplasmic staining of the tumor cells. Basically, no reactivity was observed with anti-EphB4 antibody on stromal cells surrounding EOC cells. On the mucinous-type tumor, the staining with both anti-ephrin-B3 and anti-EphB4 appeared reduced in some cells because of the cytoplasmic filling by mucous (Fig. 1B, 3 and 4). The anti-ephrin-B3 and anti-EphB4 staining on 8 LMP ovarian tumors and 47 EOC is summarized in Table 2.

L-R Signaling Loops in Ovarian Cancer

www.aacrjournals.org 10713 Cancer Res 2006; 66: (22). November 15, 2006

Published OnlineFirst November 6, 2006; DOI: 10.1158/0008-5472.CAN-06-1327
Table 3. L-R pairs differentially correlated between short-term and long-term survivors in at least two EOC data sets

<table>
<thead>
<tr>
<th></th>
<th>Short-term survivors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
</tr>
<tr>
<td><strong>A. Clinical information</strong></td>
<td></td>
</tr>
<tr>
<td>No. patients</td>
<td>30</td>
</tr>
<tr>
<td>Median survival</td>
<td>17.5</td>
</tr>
<tr>
<td><strong>B. Ligand receptor</strong></td>
<td>r</td>
</tr>
<tr>
<td>Angiogenesis</td>
<td></td>
</tr>
<tr>
<td>VEGF*</td>
<td></td>
</tr>
<tr>
<td>NRP2</td>
<td>-0.47</td>
</tr>
<tr>
<td>Chemokines</td>
<td></td>
</tr>
<tr>
<td>CCL11 *</td>
<td></td>
</tr>
<tr>
<td>CCR5</td>
<td>0.37</td>
</tr>
<tr>
<td>Cytokines</td>
<td></td>
</tr>
<tr>
<td>IL15</td>
<td>0.53</td>
</tr>
<tr>
<td>IL2RG</td>
<td>0.41</td>
</tr>
<tr>
<td>LIF</td>
<td>0.41</td>
</tr>
<tr>
<td>Growth factors</td>
<td></td>
</tr>
<tr>
<td>FGF1</td>
<td>0.39</td>
</tr>
<tr>
<td>FGF13</td>
<td>0.55</td>
</tr>
<tr>
<td>FGF4</td>
<td>-0.38</td>
</tr>
<tr>
<td>FGF7 *</td>
<td>-0.44</td>
</tr>
<tr>
<td>FGF8</td>
<td>-0.50</td>
</tr>
<tr>
<td>IGF2 *</td>
<td>0.37</td>
</tr>
<tr>
<td>EFNA5 *</td>
<td>0.39</td>
</tr>
<tr>
<td>EFNB2</td>
<td>0.41</td>
</tr>
<tr>
<td>Motility adhesion factors</td>
<td></td>
</tr>
<tr>
<td>EFNB3</td>
<td>0.38</td>
</tr>
<tr>
<td>TNC *</td>
<td>0.45</td>
</tr>
</tbody>
</table>

**NOTE:** Shaded areas identify significant correlations ($P < 0.05$) that differentiate short-term from long-term survivors.

*Discordant correlations observed for the indicated L-R pairs in different data sets.

*Discordant correlations observed for the indicated L-R pairs in the same data set.

Table 4. Twenty-seven of the 30 serous tumors showed correlated expression of both ephrin-B3 and EphB4 proteins. Within LMP serous tumors, one showed weak staining with anti-ephrin-B3 antibody and one was not reactive with anti-EphB4 antibody. Within serous EOCs, only one showed no reactivity with either antibody, and two samples with weak reactivity with anti-EphB4 antibody showed strong and negative staining with anti-ephrin-B3 antibody, respectively. In all mucinous (6) and endometroid (9) tumors, the staining intensities of both antibodies were correlated. By contrast, 4 of 10 clear cell EOCs were not reactive with anti-EphB4 antibody and 4 samples showed strong reactivity; only 1 sample showed no reactivity with anti-ephrin-B3 antibody. Contingency analysis by $\chi^2$ test indicated a significant ($P < 0.0001$) correlation between ephrin-B3 and EphB4 protein expression levels.

**Discussion**

Initial evidence in *Saccharomyces cerevisiae* (30) indicated that genes with similar expression profiles were more likely to encode interacting proteins and, very recently, similar results were obtained for the human genome (31). A pioneering study analyzed the correlation of the gene expression patterns of L-R pairs with the aim of discovering potential autocrine/paracrine signaling loops in different cancers (6). Although limited by the data sets available at the time (one for each pathology) and by the relative paucity of L-R pairs extractable from them (range, 33-68), that study identified a large number (>30) of known and new signaling loops as potentially active in diffuse large B-cell lymphoma, leukemia, and colon and breast cancer. The systematic evaluation of multiple data sets promises to yield more reliable and more valid results because it is based on a larger number of samples and the effects of individual study-specific biases are weakened (32).

In the present study, we refined the starting list of L-R pairs and applied it to numerous data sets from a single pathology, EOC, as proof-of-principle that the correlation of gene expression patterns of L-R pairs is statistically and biologically valid. Integration of the refined DLRP-rev1 database with eight EOC data sets of gene expression matched a larger number of L-R pairs (range, 68-417) compared with the initial study and enabled the identification of 105 L-R pairs showing significant correlation after Bonferroni correction. To select potential candidates, the 105 L-R pairs identified as significantly correlated in a single data set were analyzed across the other data sets to obtain independent statistical validation of their correlation. Several arbitrary selection
Only advanced stage EOC patients were selected for analysis of L-R pairs. The limited knowledge of EOC progression, we should rely only on correlations, possibly including numerous false-positives, due to the smaller number of patients included in each data set. When a less stringent level of significance ($P < 0.05$) was not informative, statistical cutoffs were applied and some were also tested experimentally. In our analysis, we gaged statistical validation as well as the existing knowledge about EOC biology. For the entire case material, we focused on EOC biology, selecting a cutoff based on the correct identification of all seven L-R pairs already implicated in EOC (12, 19–25) and on means of down-weighting data sets representing fewer initial L-R pairs and/or samples. Our selection criteria reduced the potential L-R candidates to 39 pairs, 7 of which already biologically confirmed.

Because the case materials analyzed were biased toward advanced-stage disease (see Table 1B), the identified L-R pairs are likely associated to EOC biology as well as to EOC progression. The identification of pathways involved in epithelial ovarian oncogenesis awaits gene expression analysis of a larger series of early-stage case material.

The availability of clinical information about overall survival in four data sets provided an opportunity to evaluate correlations potentially associated with late-stage tumor progression. The statistical cutoff adopted for the entire case material (correlation significance after Bonferroni correction) was not informative, probably due to the smaller number of patients included in each data set. When a less stringent level of significance ($P < 0.05$) was considered, 166 L-R pairs were identified. To validate these correlations, possibly including numerous false-positives, due to the limited knowledge of EOC progression, we should rely only on validation using statistical independent data sets. Our selected cutoff reduced the list of potential candidates to 15 L-R pairs. Only advanced stage EOC patients were selected for analysis of L-R correlation with survival, but the limited number of available cases precluded a further subcategorization. Thus, several clinical variables, such as difference in age, histology, residual disease after debulking, and type of treatment, could confound the results. Some of these factors have been already taken into account in the original articles describing the subcategorization in short-term and long-term survivors (11, 13, 14) and seemed not to play an important role in determining patient’s outcome. However, the relevance of candidate L-R pairs in disease progression should be interpreted with caution and await validation in larger data sets and confirmation in biological/functional assays.

To retrieve further information about the biological significance of our data, we evaluated the distribution of the identified L-R pairs in functional classes (Fig. 2). The distribution of the L-R pairs potentially correlated to EOC biology after cross-analysis (Fig. 2C), compared with their relative presence in the DLRP-rev1 database and in correlation analysis of single data sets (Fig. 2D), clearly indicated a significant increase in L-R pairs involved in chemokine signaling (33% versus 15% and 23%, respectively) accompanied by a decrease in growth factors and cognate relevant receptor L-R pairs (28% versus 42% and 37%, respectively). By contrast, analysis of L-R pairs potentially correlated to EOC progression in late stage implicated mainly motility/adhesion signaling molecules [33% after cross-analysis (Fig. 2D) versus 14% in DLRP-rev1 database and in correlation analysis (Fig. 2B)] and suggested a switch toward a negative correlation in the class of growth factors (see Table 3). These observations are consistent with current hypotheses linking

### Table 3. L-R pairs differentially correlated between short-term and long-term survivors in at least two EOC data sets (Cont’d)

<table>
<thead>
<tr>
<th>No. patients</th>
<th>Median survival</th>
<th>Long-term survivors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>I</td>
</tr>
<tr>
<td>24</td>
<td>107.5</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>&gt;84</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>&gt;47</td>
<td></td>
</tr>
</tbody>
</table>

| Angiogenesis | VEGF* | NRP2 | −0.31 | 0.13698 | −0.16 | 0.62375 | −0.48 | 0.00467 |
| Chemokines   | CCL11 | CCR5 | 0.19  | 0.38031 | 0.57  | 0.02524 | 0.25  | 0.15438 |
| Cytokines    | IL15  | IL2RG| 0.35  | 0.09470 | 0.31  | 0.61209 | 0.15  | 0.58851 |
|             | LIF   | IL6ST| 0.29  | 0.16797 |       |         |       |         |
| Growth factors | FGF1  | FGRFR1| 0.41  | 0.04807 | 0.12  | 0.66879 | 0.23  | 0.20495 |
|              | FGF13 | FGRFR1| 0.27  | 0.20500 | 0.04  | 0.87376 | 0.17  | 0.33862 |
|              | FGF4  | FGRFR3| 0.33  | 0.11983 | 0.65  | 0.00929 | 0.37  | 0.03483 |
|              | FGF7  | FGRFR2| 0.60  | 0.00179 | 0.31  | 0.25912 | 0.53  | 0.00142 |
|              | FGF8  | FGRFR3| 0.26  | 0.21225 | 0.39  | 0.15425 | 0.38  | 0.02975 |
|              | IGF2  | IGF2R| 0.22  | 0.30655 | 0.80  | 0.01670 | −0.17 | 0.54510 |
| Motility adhesion factors | EFNA5 | EPHA5 | 0.32  | 0.12698 | 0.17  | 0.53370 | 0.37  | 0.03245 |
|              | EFNB2 | EPHA4 | 0.07  | 0.74726 | 0.46  | 0.08787 | 0.15  | 0.39864 |
|              | EFNB2 | EPHB2 | 0.39  | 0.06175 | 0.21  | 0.45219 | −0.31 | 0.08244 |
|              | EFNB3 | EPHB4 | −0.09 | 0.69181 | 0.28  | 0.38473 | 0.12  | 0.50100 |
|              | TNC   | ANXA2 | 0.10  | 0.63268 | −0.46 | 0.08542 | 0.39  | 0.02466 |

Published OnlineFirst November 6, 2006; DOI: 10.1158/0008-5472.CAN-06-1327
epithelial ovarian oncogenesis and progression to inflammation (33) and EOC progression to a dysregulation of cell-cell and cell-stroma interactions (34).

The relevance of DLRP database revision based on new biological knowledge is supported by the observation that at least two of the significantly correlated L-R pairs, PLAU-PLAUR (identified in the EOC biology analysis) and TNC-ANXA2 (identified in the EOC progression analysis), could be retrieved by implementation of the originally published database. That database was focused mainly on autocrine signaling, although it contained pairs able to signal also or only through paracrine interactions, and our implementation was partially dedicated to increasing identification of motility/adhesion-involved molecules. Our L-R pair selection, together with the observation that the cancer specimens in all data sets contained 70% to 80% tumor cells but were not micro-dissected, enabled retrieval of most of the potential tumor-tumor and tumor-stroma autocrine/paracrine interactions.

Despite the improvements relative to the initial study, several potential limitations and bias, as outlined in the original article (6), could also affect our analyses. In fact, the intra-data set evaluation was strongly limited by the number of samples considered (<40 in data sets III and VIII), and the inter-data set comparison might be biased by the type and size of the platform, by the selected genes present on each array, and by the type of samples included in each data set.

Both previous (6) and present studies identified some negatively correlated L-R pairs. By cross-comparison analysis of L-R pairs correlated with EOC biology, only 3 pairs with consistently negative correlation versus 36 with consistently positive correlation were observed. When only the L-R pairs correlated with short-term survivors were considered, the percentage of negatively correlated pairs strongly increased (5 of 15). This may reflect either a lack of autocrine/paracrine signaling due to decreased levels of a ligand/receptor whenever its cognate receptor/ligand is produced or the transcriptional activation of an alternative ligand/receptor to compensate for the absence of the physiologic signaling. Biological/functional validation is necessary to identify the underlying mechanism.

Figure 1. Expression of ephrin-B3 and EphB4 proteins on normal tissues and ovarian tumor specimens. Immunostaining with anti-ephrin-B3 and EphB4 antibodies was done on paraffin-embedded tissues. Positive controls were as follows: A, 1, normal ovarian arterioles (anti-ephrin-B3); 2, colonic adenocarcinoma (anti-EphB4). B, 1 and 5, serous ovarian carcinoma; 2 and 6, mucinous ovarian carcinoma; 3 and 7, endometroid ovarian carcinoma; 4 and 8, clear cell ovarian carcinoma. Original magnification, ×200.

Table 4. Immunostaining results in 55 paraffin-embedded samples from epithelial ovarian tumors

<table>
<thead>
<tr>
<th>Tumor histotype*</th>
<th>Total cases (%) correlated expression</th>
<th>Ephrin-B3</th>
<th>EphB4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Negative</td>
<td>Weak/moderate</td>
</tr>
<tr>
<td>Serous</td>
<td>30 (90)</td>
<td>—</td>
<td>1</td>
</tr>
<tr>
<td>LMP</td>
<td>4</td>
<td>—</td>
<td>2</td>
</tr>
<tr>
<td>EOC</td>
<td>26</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Mucinous</td>
<td>6 (100)</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>LMP</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>EOC</td>
<td>3</td>
<td>—</td>
<td>1</td>
</tr>
<tr>
<td>Endometrioid</td>
<td>9 (100)</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>LMP</td>
<td>1</td>
<td>—</td>
<td>2</td>
</tr>
<tr>
<td>EOC</td>
<td>8</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Clear cell</td>
<td>10 (30)</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

NOTE: Ephrin-B3 and EphB4 protein expression levels were significantly correlated ($P < 0.0001$) by $\chi^2$ test.

* Routine tissue blocks and commercially available tissue arrays.

† Weak/moderate or strong staining of 80% of tumor cells in the tissue section.

‡ EOCs comprise grade II and III tumors.
The identification of known and expected correlations in our analysis strongly supports the validity of our cross-comparison analysis and potentially implicated the newly identified correlated L-R pairs in EOC biology and progression. Among the newly identified correlated L-R pairs, we focused on EFNB3-EPHB4. Eph receptors, divided into A and B type based on interaction with their ligands, comprise the largest group of membrane tyrosine kinase receptors, and their ligands, ephrins, are also membrane bound (for review, see ref. 26). The paracrine/juxtacrine signaling is cell contact dependent and can potentially trigger a bidirectional response leading to either cell repulsion or invasion. At present, only the EphA2 protein has been reported to be associated with aggressive ovarian carcinomas (35, 36). Although very little is known about how Eph receptors contribute to the oncogenic process, EphB4 overexpression has been reported in colon (37), breast (38), and prostate carcinomas (39) and positively associated with malignant potential, clinical grade and stage in endometrial carcinoma (40, 41). Furthermore, EphB4 de novo expression in a breast carcinoma cell line contributes to tumor progression by attracting endothelial cells and inducing neovascularization, thus promoting tumor cell proliferation and survival (42). Signaling through a paracrine loop between EphB4 and any member of B-class ephrins is required for directional growth of developing vasculature, confirming that the EphB4 receptor can interact with and signal through an overexpressed ephrin-B3 ligand (28). Our immunohistochemical analysis showed

Figure 2. Flow chart diagram of the experimental design showing the steps of analysis and the percentage distribution of L-R pairs according to functional consequence of the L-R interaction. After extraction of L-R pairs from EOC data sets, all tumor samples were examined with respect to EOC biology (A and C) and only the samples from patients with short-term survival were used for EOC progression analysis (B and D). A and B, the class distribution of L-R pairs from correlation analysis of eight and four data sets, respectively. C and D, the class distribution of selected L-R pairs after cross-analysis. See results for cutoff criteria.
the correlated expression of ephrin-B3 and EphB4 in 45 of 55 ovarian tumor specimens. Clear cell was the only EOC histotype, in which the correlation was absent. A potential explanation rests in recent evidence, by gene expression profiling, that EOC clear cell histotype can be reconstituted to normal uterine endometrium (43), instead of ovarian surface epithelium from which the vast majority of EOC originate (44). The correlated expression in all the other histotypes was similarly irrespective of grading and malignant potential. Indeed, seven of eight LMP tumors resulted not to express EphB4. LMP ovarian tumors represent a subset of EOC with a very good prognosis, and most of them show molecular characteristics distinct from carcinomas (45, 46). Hence, analysis of a larger number of samples would allow evaluating the significance of ephrin-B3/EphB4 correlated expression in this subset of tumors. The Eph-ephrin signaling occurs at the membrane level, but it was shown recently that the L-R pair ephrin-B2-EphB4, once activated by cell-cell contact, is endocytosed as a consequence of a cytoskeletal rearrangement that requires RAC function (47). Accordingly, the cytoplasmic localization of both ephrin-B3 and EphB4 in our analysis might reflect internalization after signaling activation. Overall, our observations suggest the involvement of paracrine/juxtacrine signaling through ephrin-B3-EphB4 in EOC progression. Further studies are needed to address this possibility and their potential usefulness as phenotypic and/or prognostic markers.

Together, our data point to the feasibility of a cross-platform analysis of gene expression data to identify L-R signaling loops in other oncoproteins, provided that a sufficient number of data sets are available and that the principle of external validation in independent data sets is upheld. The proposed bioinformatic systematic search of L-R coexpression might also prove useful in conjunction with conventional protein-protein interaction methods of prediction. Once supported and validated by biological/functional assays, the L-R coexpression search might provide insight into the biology of a specific oncotype and could open avenue to the design of specifically targeted new diagnostic and therapeutic tools.

Acknowledgments

Received 4/17/2006; revised 7/14/2006; accepted 8/30/2006.

Grant support: Associazione Italiana Ricerca Cancro and CARIPLO Foundation (S. Canevari).

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

We thank M.A. Pierotti for helpful and stimulating discussion, Y. Yarden for critical reading of the article, and Gloria Bosco for secretarial assistance.

References


New Potential Ligand-Receptor Signaling Loops in Ovarian Cancer Identified in Multiple Gene Expression Studies

Giancarlo Castellano, James F. Reid, Paola Alberti, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-06-1327

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2006/11/09/66.22.10709.DC1

Cited articles
This article cites 46 articles, 22 of which you can access for free at:
http://cancerres.aacrjournals.org/content/66/22/10709.full.html#ref-list-1

Citing articles
This article has been cited by 4 HighWire-hosted articles. Access the articles at:
/content/66/22/10709.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.