Intercohort Gene Expression Co-Analysis Reveals Chemokine Receptors as Prognostic Indicators in Ewing's Sarcoma

Idriss M. Bennani-Baiti, Aaron Cooper, Elizabeth R. Lawlor, Maximilian Kauer, Jozef Ban, Dave N.T. Aryee, and Heinrich Kovar

Abstract

Purpose: We report a novel analytic method, named intercohort co-analysis or Ican, which aids in the discovery of genes with predictive value for the progression or outcome of diseases from small-size cohorts. We tested this premise in Ewing's sarcoma (ES), a highly metastatic cancer of bone and soft tissues that lacks validated molecular metastasis and prognostic indicators.

Experimental Design: To uncover genes significantly expressed in ES patient subsets, we first determined a nonarbitrary gene expression significance cutoff based on expression levels in validated expressing and nonexpressing tissues. We next searched for genes that were consistently significantly expressed in several ES cohort and cell line datasets. Significantly expressed genes were independently validated by quantitative reverse transcription-PCR in an additional ES cohort.

Results: Analysis of ES cohorts revealed marked intercohort gene expression variability. After filtering out the intercohort variability, CXCR4 and CXCR7 were found to be consistently associated with specific ES subsets. Pairwise analyses showed CXCR4 to correlate with ES metastases, and CXCR4 and CXCR7 to patient survival, but not with several other clinicopathological variables.

Conclusion: Ican is a powerful novel method to identifying genes consistently associated with particular disease states in cancers for which large cohorts are not available, currently the case of most cancers. We report for the first time that high CXCR4 expression preferentially associates with metastatic ES, and that of CXCR7 with poor patient survival.

Cancer genomics is a powerful approach to uncovering genes that may be involved in the pathogenesis of cancer as well as those that may serve to predict cancer spread or disease outcome. This has, for instance, been the case of several adult tumors for which large cohorts were available (1–3). Despite low incidence and major advances in diagnosis and treatment, cancer remains the leading cause of death by disease in children in developed countries (4). This is in part due to the fact that these cancers consist of a large and heterogeneous group of low-incidence neoplasms, thus hampering large cohort studies aimed at uncovering markers of disease occurrence, progression, recurrence, spread, or outcome (5–7).

In recent years, several small-size cohort childhood cancer genomic studies have been reported. We hypothesized that a co-analysis of unrelated small-size cohorts that filters out the high intercohort variability inherent to small-size cohorts may help uncover genes associated with subsets of tumors with common clinical features. We named this approach intercohort co-analysis, henceforth dubbed Ican, and tested this premise in Ewing tumor and related sarcomas of the bone (7, 8), a family of tumors that lack a reliable molecular prognostic or metastasis indicator, and which we refer to herein as Ewing's sarcoma (ES). As an example of analysis, we used Ican to screen for chemokines and chemokine receptors significantly associated with ES subsets (flowchart in Supplementary Fig. S1). Chemokines are 8- to 10-kDa secreted chemoattractant cytokines, and chemokine receptors are seven-transmembrane G-protein-coupled cell surface proteins. These receptors are defined by their ability to induce the directional migration of cells toward a cognate chemotactic cytokine gradient in a process known as chemotaxis. Chemokine receptors, therefore, support several physiological and pathological processes underlying cell motility, egress, and homing to distant sites. For the purpose of this study, we chose to survey chemokines and their receptors because they have been associated with tumor growth, metastasis, or neoangiogenesis in several adult cancers (9, 10). They can also serve as therapeutic targets to curtail tumor growth or metastasis (9, 11, 12) and may thus offer novel treatment options for ES.

We report here that Ican successfully identified CXCR4 (also known as CD184, LESTR, WHIM, or SDF-1 receptor;
**Translational Relevance**

Childhood cancers consist of a large number of low-incidence histogenetically and molecularly distinct neoplasms, allowing only for small-size cohort trials. These, however, lack the necessary statistical power for population-based genomic studies, thus hampering the discovery of markers. Using Ican, we uncovered two chemokine receptors that can serve as metastasis or prognosis indicators in Ewing's sarcoma (ES), a highly metastatic childhood cancer. These markers may help in ES metastasis diagnosis and patient treatment stratification. Additionally, both chemokine receptors are drug targets, suggesting possible novel treatment options for poor-prognosis ES. Our findings in ES serve as a proof of principle for Ican translational applications in childhood and other low-incidence diseases.

Entrez gene ID no. 7852) and CXCR7 (also known as CMKOR1, GPR139, RDC1, or chemokine orphan receptor 1; Entrez gene ID no. 57007), two chemokine receptors known to associate with metastasis and/or poor prognosis in other cancers (10, 13), as significantly associated with ES subsets. We show for the first time that high CXCR4 expression marks ES metastases and that high expression of both CXCR4 and CXCR7 serves as a predictor of adverse patient survival. Expression of these genes in the analyzed tumors is unlikely to stem from nontumoral sources, as we found them to be also expressed in ES-derived cell lines (EDCL). In EDCL, high CXCR4 expression was restricted to those derived from metastases, and both CXCR4 and CXCR7 expression were dependent on EWS-FLI1, an ES-specific transcription factor. We present these findings as a proof of principle that Ican of publicly available gene expression profiles can yield genes with predictive value for the progression or outcome of low-incidence diseases.

**Materials and Methods**

**ES-derived cell lines**

EDCL culture and transfection conditions have been described (14). Cell lines derived from ES metastases (M-EDCL) consisted of SK-N-MC, SMB, STA-ET-2.2, STA-ET-3, STA-ET-7.3, STA-ET-8.1, TC252, VH64, WE-68-M1, and WE-68-M2, and cell lines derived from localized ES (L-EDCL) consisted of A673, RD-ES, RM82, STA-ET-2.1, STA-ET-4, STA-ET-6, STA-ET-7.1, STA-ET-7.2, STA-ET-11, and WE-68. Additional details on these EDCL were previously reported (15).

**EWS-FLI1–dependent gene regulation assessed by RNA interference**

Knockdown studies were carried out using EWS-FLI1 type I– or type II–specific small-hairpin RNA (shRNA) in EDCL and compared with nontargeting shRNA–transfected cells as previously reported (14).

**Gene expression profiling of EDCL**

Gene expression profiles of cell lines derived from localized ES or from ES metastases were surveyed by Affymetrix HG-U133A arrays (Affymetrix, Inc.) as previously described (14), and analyses done in R statistical environment using Bioconductor packages (see below).

**Bioinformatic analyses and procedures**

**Ican of gene expression datasets.** Cancer genomic datasets may be retrieved from public gene expression profile repositories such as GEO (http://www.ncbi.nlm.nih.gov/geo; ref. 16) or ArrayExpress (http://www.ebi.ac.uk/arrayexpress; ref. 17). These may contain both normalized and raw gene expression values, but usually are accompanied by little, if any, biographical or clinical data. For the purpose of this study, we considered those obtained on Affymetrix GeneChip human U133A arrays, which consist of more than 22,000 probe sets, covering about 14,500 human genes. As outlined below, we analyzed only probe sets that passed strict selection criteria and run several quality control checks before the analyses. Finally, to further increase dataset good quality odds, we limited our analyses to peer-reviewed, published datasets (Supplementary Fig. S1).

ES cohort, EDCL, and normal tissue (NT) CEL file data (accession nos. E-MEXP-353, E-MEXP-1142, GSE7007, GSE12102, GSE14543, and GSE1133) were extracted from GEO or ArrayExpress. Similarly to the vast majority of gene expression datasets currently deposited in these databases, all four ES cohorts tested lacked biographical or clinical data. ES cohorts were normalized using the *germa* algorithm (18). Intensity distributions from all samples were normalized using quantile normalization and the Bioconductor package *limma* (19). Using the Bioconductor package *pwrnpl*, probe sets expressed at levels comparable to the probability density distribution of probes with no EST hits were deemed negative. Affymetrix probe sets are known to have widely varying hybridization characteristics resulting in nonuniform thresholds of expression across probe sets that can be labeled expressed versus nonexpressed (20). Thus, after the initial normalization steps, we empirically searched for a threshold of gene expression for each probe set that differentiated known expressing from nonexpressing tissues. We found that, in our datasets, the median of the probe density distribution across NTs separated known expressing from nonexpressing tissues (9, 21). This was determined following systematic testing of cutoffs in *in silico* analyses of gene expression in tissues shown either to express or not biologically relevant levels of genes (52 genes in 75 different tissues and 2 biological replicates; ref. 22). To reduce the probability of spurious data points, only expression values exceeding median expression in NT in at least two biological replicates were deemed significant. Whenever multiple probe sets probing the same gene satisfied the aforementioned criteria, we chose those whose expression was most variable across the spectrum of ES and NT. This was carried out to maximize the likelihood to pick probe sets that best...
differentiated between expressing and nonexpressing ES (23). Based on these criteria, we selected 34 of 40 chemokines and 18 of 22 known chemokine receptors for analysis.

**Differential gene expression analysis of M-EDCL and L-EDCL.** Differentially expressed genes between cell lines established from localized ES or from ES metastases were determined using a moderated two-sample t test in the R package limma (17). To this end, three L-EDCL and three M-EDCL were treated as biological replicates for either condition (i.e., localized ES or ES metastases). Thus, for each gene, average expression values of replicates were determined and the significance of the difference between the means of the two groups was determined using the moderated t test in the limma package. P values were corrected for multiple testing using the Benjamini-Hochberg correction method.

**Gene set enrichment analysis.** Gene set enrichment analysis (GSEA) is a bioinformatic method to assess whether genes with known biological/molecular functions are concomitantly upregulated or downregulated in a certain gene expression dataset (24). The Molecular Signatures Database (MSigDB) hosted by the Broad Institute currently provides the most comprehensive collection of gene sets with annotated/known functions. The collection of Gene Ontology (http://www.geneontology.org/) terms available from the Broad Institute offers the advantage of manual curation thereby enhancing the information content of these gene sets (a detailed description of the curation process may be found at http://www.broad.mit.edu/gsea/msigdb/). The Gene Ontology (GO) gene set collection C5 of the MSigDB was downloaded from the Broad Institute and loaded into R statistical environment. GSEA was used to calculate whether certain GO gene sets, and herefore functional gene categories, were differentially regulated between cell lines established from localized ES or from ES metastases. To this end, LogFD = Log2 (M–EDCL)–Log2 (L–EDCL) was calculated and used as input for the PGSEA Bioconductor-package (25). GO gene sets larger than 200 genes were omitted from the analysis because they contained largely nonspecific terms. All C5 MSigDB gene sets, gene IDs, and P values from the tests are listed in Supplementary Tables S1 and S2.

**Quantitative real-time reverse transcription-PCR**

Analysis of EDCL by quantitative reverse transcription-PCR (RT-PCR) was as described earlier (14). Archived fresh-frozen ES samples were obtained from the tumor bank at the Children’s Hospital Los Angeles or from St. Anna Children’s Hospital in Vienna. All specimens were obtained in compliance with Health Insurance Portability and Accountability Act regulations and with Institutional Review Board ethics committee approval. Total RNA was extracted from fresh-frozen tumor sections using Qiagen columns, and first-strand cDNA was synthesized from 250 ng of RNA using the iScript Reverse Transcriptase kit (Bio-Rad). Quantitative RT-PCR was done using validated proprietary TaqMan Gene Expression Assays (Applied Biosystems) for CXCR4 (Hs00607978_s1), CXCR7 (Hs00664172_s1), GAPDH (Hs99999905_m1), and ACTB (Hs99999903_m1) on an Applied Biosystems 7900HT system. Expression of CXCR4 or CXCR7 was normalized relative to the average expression of ACTB and GAPDH in the same sample using the following formula: % expression = 2^[-ΔCt × 100].

**Statistical analyses of clinical data**

Correlation of gene expression to clinical variables in ES was tested using nonparametric two-tailed unpaired Mann-Whitney U tests (significant if P < 0.05) or one-way ANOVA and Bonferroni’s correction for multiple comparison testing (significant if P < 0.01), as indicated. Patient survival significance was tested using Mantel-Cox log-rank (significant if P < 0.05).

**Results**

**High intercohort variability hampers gene-disease association studies of low-incidence cancers**

Discovery of gene-disease associations that can serve as markers of certain disease states, say of good or poor disease outcome, is greatly facilitated by the availability of large cohorts and, conversely, severely compromised by the small size of cohorts in low-incidence diseases (5, 26). To illustrate this problem, there currently are two reported small-size cohort studies of gene expression signatures associated with ES metastases (27, 28). We surveyed the two signatures by calculating LogFD, the Log2 fold difference between the mean expression of genes in metastatic and localized ES, and compared the extent of convergence between the two signatures. As shown in Fig. 1, there was little overlap of differentially expressed genes between the two studies (10,047 genes analyzed; Pearson correlation coefficient r = 0.111; 95% confidence interval, 0.092-0.130). To avoid the possibility that a correlation between the two signatures might have been underestimated due to the large size of the two datasets considered, we also restricted the analysis to the expression profiles of 426 genes corresponding to the gene population subsets preferentially expressed in metastatic or localized ES in one cohort (LogFD ≥ 1 or LogFD ≤ −1, respectively; ref. 28) and compared them to those in the second cohort. Again, little correlation was observed between the two cohorts (426 genes analyzed; r = 0.117; 95% confidence interval, 0.023-0.209). These data exemplify the high intercohort variability manifest in a genome-wide analysis of two small-size cohorts.

**Chemokine and chemokine receptor gene expression profiles in ES**

To solve the problem posed by intercohort variability, we tested Ican on several publicly available gene expression profiles. In Ican, gene expression levels in biologically relevant NTIs are used as a benchmark of gene expression significance in test samples in each independent cohort, and a subtractive overlay of cohort-restricted associations is carried out to uncover genes consistently associated with sample subsets in all cohorts. To test feasibility, we retrieved gene expression profiles derived from four independent ES cohorts gathered at ES reference centers at the
United Kingdom, France, Germany, and Italy (27–30). We also included gene expression profiles of 6 EDCL and those of 75 different NTs (14, 22) and researched the expression of chemokine ligands and receptors in these samples.

Gene expression significance is usually inferred from gene expression profiles by comparing the ranking of gene probeset to an arbitrarily chosen gene expression cutoff. The subjectiveness of this approach has resulted in cutoff values varying widely from one study to another, which, in addition, do not define the biological significance of gene expression. To rationally assess the biological significance as determined by gene expression profiling, we first analyzed gene expression in NT. Tissues known to express biologically relevant levels of chemokines or chemokine receptors (9, 21) displayed expression values above the median expression in NT (Supplementary Tables S3 and S4). We, therefore, selected the median expression in NT as a cutoff for biological gene expression significance. We next reasoned that the expression of genes that mark a specific fraction of a cohort must be restricted to that fraction. To uncover chemokines and chemokine receptors that may serve as markers of ES metastases, there-fore, we searched for those that are significantly expressed in ES/EDCL (Fig. 1, right). Strikingly, these genes encode receptors that mediate the effects of CXCL11 and CXCL12 uncovered in the chemokine screen. Thus, two independent Ican tests uncovered three receptor/ligand pairs (CXCR4/CXCL12, CXCR7/CXCL11, and CXCR7/CXCL12) highlighting Ican analytical power. Significant expression of CXCL11 and CXCL12 in both ESFT and EDCL denotes possible autocrine and/or paracrine mechanisms, although in tumors, expression of these chemokines may also stem from tumor stroma leading to strictly paracrine activation of CXCR4 or CXCR7. A fourth receptor/ligand pair consisting of CCR1/CCL3 was also found significant in ESFT, but not in EDCL, and was not further pursued in this study.

Next, to uncover chemokine ligands and receptors that may correlate to prognosis, we considered the fact that about half of all ES patients succumb to disease within 5 years of diagnosis (31). We thus used Ican to search for genes that associate with half of ES and which may serve as indicators of patient survival. As in the top-quartile Ican tests, this analysis revealed considerable intercohort expression variability for many genes (e.g., Fig. 3). Following Ican, only CXCR7 expression distinctly associated with half of samples in ES cohorts and EDCL. CXCR4 and CXCR7, therefore, show an association with distinct ES/EDCL subsets and were selected for further investigation.

**ES-specific CXCR4 and CXCR7 expression in EDCL is EWS-FLI1 dependent**

To test for ES relevance, we assayed CXCR4 and CXCR7 dependence on EWS-FLI1, an ETS (E26 transformation-specific) domain–containing fusion oncogene. EWS-FLI1 results from t(11;22)(q24;q12) chromosomal translocations harbored by the vast majority of ES and is a transcription fac-tor that imparts ES-characteristic gene expression programs causative of ES pathogenesis. We carried out shRNA-mediated EWS-FLI1 knockdown experiments in six EDCL and found expression of both CXCR4 and CXCR7 to be highly dependent on EWS-FLI1, showing that CXCR4 and CXCR7 expression is integral to EWS-FLI1–imposed, ES-specific transcriptome (Fig. 4A). No other chemokine receptor showed EWS-FLI1–dependent expression (Supplementary Fig. S2).

**CXCR4 expression levels are associated with metastatic phenotype in ES**

As indicated in Fig. 4A, CXCR4 expression was higher in cell lines established from metastases (M-EDCL) than in those derived from localized tumors (L-EDCL), suggesting a possible association with ES metastases. We extended the analysis to 20 EDCL and confirmed the higher expression of CXCR4 in M-EDCL (Fig. 4B). Because direct comparison of cell lines can be hampered by genetic background differences inherent to cells derived from different patients, we also assessed CXCR4 expression in three L-EDCL/M-EDCL.
Fig. 2. *Ican* uncovers chemokines and chemokine receptors significantly expressed in ES subsets. Expression levels of chemokines (left) and chemokine receptors (right) in ES cohorts (A) are shown. Genes whose expression in the top quartile of a given cohort surpassed the median expression in NTs (C) were deemed significant and depicted in pale blue. Genes consistently expressed in all cohorts are shown in dark blue, and those that are also significantly expressed in ECDL (B) are labeled in red. Expression of all genes tested could readily be detected in cognate NT (C; see also Supplementary Tables S3 and S4).
pairs, each derived from a single patient (Fig. 4C). In these isogenic EDCL, CXCR4 was again expressed at higher levels in M-EDCL. For one L-EDCL/M-EDCL pair for which corresponding ES samples were available, analysis of CXCR4 expression in the original noncultured tumors confirmed these findings, indicating that CXCR4 correlation to cells derived from ES metastases was neither due to an interpatient genetic bias nor to a cell culture artifact (Fig. 4C). Finally, a comparison of M-EDCL and L-EDCL transcriptomes showed that CXCR4 ranked as one of the most preferentially expressed genes in M-EDCL, further validating the association of CXCR4 expression with ESFT metastases (Fig. 4D, left). A comparison of gene expression profiles in M-EDCL versus L-EDCL showed that genes (Fig. 4D, left) and gene groups (Fig. 4D, right) that can support metastasis were over-represented in M-EDCL, suggesting a functional relevance to ES metastasis (see also Supplementary Data and Supplementary Tables S1 and S2).

To further test the significance of CXCR4 and CXCR7 association to ES, we researched the correlation of CXCR4 or CXCR7 to several clinicopathological variables traditionally used in ES prognosis and management (32–35). As shown in Fig. 5, neither gene correlated to any of the parameters tested (See Supplementary Table S5 for additional details). We next asked whether CXCR4 associated with ES metastases in an independent U.S. cohort of 49 patients. Again, high CXCR4 (but not CXCR7) expression correlated to ES metastases (Mann-Whitney two-tailed test, \( P = 0.0209 \)), validating the \( Ican \) ES/EDCL findings (Fig. 6A).

**High CXCR4 and CXCR7 expression demarcates poor prognosis in ES patients**

Finally, we researched the correlation of CXCR4 and CXCR7 expression to patient survival. ES patients whose tumors expressed low levels of CXCR4 and CXCR7 had the highest 5-year overall survival (OS) rates (90%; Fig. 6B). These rates were diminished for patients with high CXCR4 (OS, 54.5%) or CXCR7 (OS, 45.4%) tumor content and were drastically reduced for patients with high-expressing CXCR4 and CXCR7 tumors (OS, 28.6%; log-rank test, \( P = 0.0161 \); hazard ratio 8.0). Analysis of event-free survival rates suggested a trend wherein high CXCR4 and CXCR7 expression might have negatively affected event-free survival (median event-free survival in high CXCR4/CXCR7: 1.5 versus 5.4 years for patients with low CXCR4/CXCR7; hazard ratio, 0.34; 95% confidence interval, 0.11–1.0), but this correlation did not reach statistical significance (\( P = 0.0586 \)), presumably owing to the small sample size analyzed. Except for metastasis addressed above (Fig. 6A), neither CXCR4 nor CXCR7 expression levels correlated with other variables that may influence disease outcome (pairwise univariate analyses; Fig. 5A–D). CXCR4 and CXCR7 thus seem to serve as prognostic factors of metastasis and OS, respectively.

**Discussion**

Cancer is the leading cause of death by disease of children aged 1 to 14 years in the developed world (4). This fact underscores the urgency of uncovering markers that can help in the prognosis of tumor progression and outcome. These markers would aid in patient treatment stratification and help curtail the high mortality that afflicts subsets of cancer patients. Childhood cancers regroup a very large number of different low-incidence cancers, thus preventing large cohort studies and compromising marker discovery (5–7). Although several multicenter cancer consortia have formed to consolidate cohorts (5, 6), these cannot make their resources available to most nonconsortium proposals without compromising critical deliverables and deadlines. These and other considerations render investigations aimed at uncovering molecular prognostic factors in childhood cancers impractical for most laboratories or even single institutions (6). We present here an analytic method, \( Ican \), which enables to uncover genes that
are consistently associated with patient subsets with common features from small-size cohorts.

Using Ican, we found the CCL3/CCR1, CXCL12/CCR4, CXCL11/CXCR7, and CXCL12/CXCR7 chemokine/chemokine receptor pairs to be significantly associated with ES subsets. The relevance of these chemokine receptors to cancer is well established: CCR1 and CXCR7 contribute to growth or metastasis of breast, liver, lung, oral, and prostate cancers (9, 36, 37), whereas CXCR4 supports tumor mitogenesis, chemoresistance, or metastasis of more than 20 neoplasms of different etiologies and tissues of origin (9, 10). These considerations suggest that these chemokines and chemokine receptors may also be involved in ES. Activation of the CXCL12/CXCR4 axis, a prototypical metastasis-mediating chemokine/chemokine receptor pair, can lead to metastatic processes involving at least three steps. First, cells within the tumor (tumor cells and also stroma cells such as tumor-associated fibroblasts and endothelial cells of the tumor vasculature) can secrete CXCL12. This leads to increased CXCR4-dependent motility and, for some CXCR4-expressing tumor cells, egress into tumor vasculature. This first step is of paracrine and/or autocrine nature. Next, CXCR4-expressing tumor cells migrate toward a CXCL12 chemotactic gradient in the bloodstream to target organs producing high levels of CXCL12 (e.g., bone marrow and lungs). This second step uses endocrine mechanisms. Finally, CXCR4-expressing tumor cells invade CXCL12-expressing organs. The very high CXCL12 levels in these

**Fig. 4.** CXCR4 expression correlates to EDCL metastatic origin. A, EWS-FLI1 knockdown decreased CXCR4 and CXCR7 expression in six EDCL (three L-EDCL and three M-EDCL). CXCR4 was highly expressed only in M-EDCL (M). B, quantitative RT-PCR analysis of 20 EDCL validated the preferential high-level CXCR4 expression in M-EDCL. C, quantitative RT-PCR analysis of CXCR4 in cell lines and tumors derived from single patients further confirmed the CXCR4-metastasis association. Blue, localized; red, metastases. D, gene expression enrichment analysis yielded genes (top and bottom 50 are shown; left) and gene sets (top GO-groups are shown; right enriched in either M-EDCL or L-EDCL. P values and number of genes/gene sets are shown on split x-axis. LogFD = \( \frac{1}{2} \log_2 (M \text{–EDCL}) - \log_2 (L \text{–EDCL}) \). Co, control; Kd, knockdown.
organs lead to CXCR4 internalization, decreased cell motility, and attachment of tumor cells to the site of metastasis. Paracrine mechanisms prevail at this step. These three steps are detailed along with some of the supporting published evidence in the excellent review by Burger and Kipps (10).

In the subsequent analyses, we chose not to further investigate the CCL3/CCR1 pair because it was not significantly expressed in EDCL. While this may be due to a lack of expression of both genes following cell culture, this may also signify that CCL3/CCR1 expression in the tumor samples stemmed from nontumoral sources such as tumor stroma, tumor-infiltrating lymphoid cells, or tumor-associated vasculature endothelial cells, as previously reported (38).

Expression of both CXCR4 and CXCR7 was highly dependent on EWS-FLI1 in EDCL. CXCR4 contains several ETS binding sites within close proximity to its transcription start site that may allow ETS-specific transcriptional regulation in NTs (39). CXCR4 promoter region also contains hypoxia-inducible factor 1 (HIF1α) binding sites (39). These mediate CXCR4 transcriptional activation in

Fig. 5. CXCR4 and CXCR7 expression does not correlate to several molecular and clinical variables. Several clinicopathologic parameters were surveyed in connection with CXCR4 or CXCR7 expression. CXCR4 and CXCR7 did not correlate to (A) tumor location (osseous (OS) versus extra-osseous (EOS) sites, axial (AX) versus extremity (EXT) locations, and pelvis (PL) versus nonpelvic (NPL) locations); (B) EWS-ETS fusions: EWS-FLI1 type I (EF I), type II (EF II), or other EWS-ETS (OEE) fusions; (C) treatment status and outcome: relapsed (RL) versus nonrelapsed (NRL) tumors; pre-chemotherapy (PRE) versus post-chemotherapy (POST); or (D) biographical data such as age (children <13 y old versus patients ≥13 y old) or gender (M, males; F, females). Pairwise univariate analyses (nonparametric two-tailed unpaired Mann-Whitney U tests) were carried out for all analyses to assess significance (P < 0.05), except for data in B, wherein Bonferroni's multiple comparison test (significant if P < 0.01) and one-way ANOVA (P values shown) were used.
response to hypoxia, one of the earliest initiating events of metastasis (10). Similarly to CXCR4, we find that CXCR7 promoter proximal region contains several ETS binding sites (not shown). It is thus plausible that EWS-FLI1 activates CXCR4 and CXCR7 transcription by binding to these ETS binding sites, leading to ES-specific expression. Furthermore, we recently uncovered a two-way crosstalk between EWS-FLI1 and HIF1α that modulates the ES gene signature during hypoxia (40). Here, we find CXCR4 to be preferentially expressed in ES metastases as well as in cell lines derived from ES metastases. The association of CXCR4 with ES metastases, the presence of multiple ETS and HIF1α transcription factor binding sites at CXCR4 promoter, and the EWS-FLI1-HIF1α crosstalk in ES suggest that CXCR4 preferential expression in ES metastases might be the result of transcriptional activation by EWS-FLI1 and possibly HIF1α. Further research into these and other possible molecular mechanisms of CXCR4 activation in ES are warranted because they may lead to means of targeting and disrupting the CXCR4/CXCL12 axis and possibly ES metastasis.

The CXCR4 association to ES metastases is also consistent with the previous finding that CXCR4 supports EDC1 cellular migration in response to CXCL12 (41), an in vitro characteristic indicative of in vivo metastatic potential (42). Moreover, the bone/bone marrow and lungs are the richest sources of CXCL12 (Supplementary Table S3) and are also organs of predilection for ES metastasis (43). The strong correlation between organs that produce the most CXCL12 and those to which ES metastasize further supports the CXCR4/ES metastasis association. CXCR4 correlation to ES metastases may be of clinical relevance as metastasis represents the single most significant adverse clinicopathological variable in ES patients (31). The dismal survival rates of ES patients are, in part, due to a lack of metastasis and prognostic indicators that would help in risk assessment and patient stratification. Our findings suggest that CXCR4 and CXCR7 may serve that purpose. Furthermore, CXCR4 and CXCR7 antagonists that can effectively reduce the growth or spread of CXCR4- and CXCR7-expressing tumors in experimental models and patients have been reported (9, 12). These drugs may also be effective in treating CXCR4/CXCR7-expressing ES and merit investigation.

In the present study, we used Ican of public expression datasets to uncover gene-ES associations. Ican may be applied to any disease for which only small-size cohorts are available, currently the case of most diseases. In the case of childhood cancers, several are already represented by more than 100 arrays each in GEO database alone, making Ican feasible. This approach is poised to gain further momentum as many journals now request the mandatory submission of gene expression profiles to public repositories. These policies, along with the growing number of gene expression profiling investigations, we predict, will swiftly drive upward the number of publicly available low-incidence disease expression datasets and, consequently, of Ican applications.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Acknowledgments**

We thank CHLA, COG, and Peter Ambros for tumor samples; Ynnez Goye and Betty Schaub for technical assistance; Ulrike Pötschger for help with statistics; and Barbara Bennani-Baiti for suggestions throughout this work and for critical reading of the manuscript.

**Grant Support**

NIH SPECS grant 1R01CA114757-04 (E.R. Lawlor), European Community FP6-project EET-pipeline grant 037260 (H. Kovar), and National Austrian Bank Medical Sciences Fund grant OeNB-12765 (I.M. Bennani-Baiti). 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.

Received 03/03/2010; revised 05/12/2010; accepted 05/12/2010; published OnlineFirst 06/04/2010.
References
