A Framework to Select Clinically Relevant Cancer Cell Lines for Investigation by Establishing Their Molecular Similarity with Primary Human Cancers

Garrett M. Dancik¹, Yuanbin Ru¹, Charles R. Owens¹, and Dan Theodorescu¹,²,³

Abstract

Experimental work on human cancer cell lines often does not translate to the clinic. We posit that this is because some cells undergo changes in vitro that no longer make them representative of human tumors. Here, we describe a novel alignment method named Spearman's rank correlation classification method (SRCCM) that measures similarity between cancer cell lines and human tumors via gene expression profiles, for the purpose of selecting lines that are biologically relevant. To show utility, we used SRCCM to assess similarity of 36 bladder cancer lines with 10 epithelial human tumor types (N = 1,630 samples) and with bladder tumors of different stages and grades (N = 144 samples). Although 34 of 36 lines aligned to bladder tumors rather than other histologies, only 16 of 28 had SRCCM assigned grades identical to that of their original source tumors. To evaluate the clinical relevance of this approach, we show that gene expression profiles of aligned cell lines stratify survival in an independent cohort of 87 bladder patients (HR = 3.41, log-rank P = 0.0077) whereas unaligned cell lines using original tumor grades did not. We repeated this process on 22 colorectal cell lines and found that gene expression profiles of 17 lines aligning to colorectal tumors and selected based on their similarity with 55 human tumors stratified survival in an independent cohort of 177 colorectal cancer patients (HR = 2.35, log-rank P = 0.0019). By selecting cell lines that reflect human tumors, our technique promises to improve the clinical translation of laboratory investigations in cancer. Cancer Res; 71(24): 1–12. ©2011 AACR.

Introduction

Cell lines derived from human tumors serve as model systems that have greatly increased our understanding of cancer biology. These are routinely used to characterize molecular mechanisms of disease (1–3) and therapeutic agents (4–6). High-throughput cell line screening programs have been used to characterize the efficacy of anticancer agents (7) and to identify novel multiagent therapies (8). Nevertheless, the validity of a cell line model depends on how representative it is of the tumor type under investigation. Notably, cell lines can lose molecular features that drive clinically relevant characteristics as they adapt to culture conditions (9). An analysis of NCI-60 cell lines found that the proportion of cell lines most similar to their presumed tissue of origin could be as low as 57% and that the molecular profiles of patient tumors are more similar to normal tissues of the same type than their derived cell lines (10). Cross-contamination of cell lines is well documented and also contributes to reduce the number of relevant cancer cell line models for research (11–13).

Stage and grade are cornerstone prognostic and predictive factors of tumor aggressiveness and disease outcome. In the case of bladder cancer, high-grade cancers are more likely to metastasize than low grade ones and up to 30% of nonmuscle-invasive (NMI) tumors progress to muscle invasion (MI; ref. 14). Metastatic risk correlates with stage, with up to 50% of patients harboring muscle-invasive cancers developing metastatic disease during follow-up (15). Furthermore, tumor sensitivity to anticancer agents depends on stage and grade, with resistance more common in poorly differentiated tumors (16, 17). Last, the evaluation of chemotherapy and radiation treatments often occurs in subsets of cancer patients having tumors of a specific stage or grade (18–20). For all of these reasons, the choice of an appropriate cell line model that aligns with the specific human tumor characteristics being investigated is critical if results from cell line experiments are to have clinical relevance.

Here, we develop a novel framework for classifying or aligning human cancer cell lines as a function of their similarity to several important phenotypes of human tumors via gene expression profiles. We then use the gene expression profiles of these aligned cancer cells as biomarkers to determine which
cells predict clinical outcomes in human cancer patients (i.e., are clinically relevant). As a proof of principle, we used bladder and colorectal cancer cell line panels and patient tumor cohorts to show that gene expression signatures of cancer cell lines selected on the basis of their alignments to human cancer phenotypes predict patient survival outcomes whereas

the signatures of unaligned cells do not. Our practical approach can guide the selection of “clinically relevant” cell lines for use in experimental studies which may enhance the likelihood of clinical translation of laboratory investigations. Because this methodology is applicable to any cancer cell line and measurable phenotype (e.g., grade, invasiveness, and metastasis), it has broad applicability in cancer research.

Materials and Methods

The bladder and colorectal cell line panels

Gene expression profiles (CEL files) for 36 bladder cancer cell lines (BLA-36) are available from the Gene Expression Omnibus (GEO; ref. 21), accession #GSE5845 (22). The database contains expression profiles for 40 cell lines, including several cell lines of the same lineage. In our analysis, we chose to analyze unique cell lines or related cell lines that were derived in separate laboratories. As a result, we removed the cell lines FL3, SLT4, T24T (all derived from T24), and 253J-BV (derived from 253J) prior to our analysis.

Information about the BLA-36 cell lines along with our alignments are provided in Supplementary Table S1. Gene expression profiles (CEL files) for 22 colorectal cell lines (CO-22) profiled in triplicate from GlaxoSmithKline are available through the National Cancer Institute’s cancer Bioinformatics Grid (caBIG; ref. 23).

The Spearman’s rank correlation classification method

The Spearman’s rank correlation classification method (SRCCM) algorithm classifies a test sample based on the Spearman’s rank correlation between its gene expression profile and the gene expression profile of a set of training samples with known phenotypes (Fig. 1A). The gene expression profile is based on a gene signature that is unique to each phenotype (Fig. 1B and C). The tissue of origin gene signature consists of 12,402 “high fidelity” probes in which expression values are consistent across formalin-fixed, paraffin-embedded (FFPE), and fresh-frozen (FF) preservation methods (24).

The bladder stage and grade gene signatures have been previously published and independently validated (25). The bladder disease-specific survival (DSS) gene signature consists of the 181 probes with univariate Cox proportional hazard model log-rank \( P < 0.01 \) in the training cohort from Memorial Sloan-Kettering Cancer Center (MSKCC; ref. 26). The colorectal DSS gene signature is a previously published and validated signature consisting of 34 genes that correlate with recurrence and survival (27).

The test sample is aligned to (i.e., classified as or assigned) the phenotype with the highest average correlation. Formally, let \( P = (P_1, \ldots, P_k) \) be the set of training phenotypes of interest with known gene expression profiles. In Fig. 1A, \( k = 2 \) and \( P = \) (low grade, high grade). Let \( t \) be the gene expression profile of the test sample and \( x_i^{(p)} \) be the gene expression profile of the \( i \)th training sample of phenotype \( p \), \( i = 1, \ldots, n_p \). Let \( r_{i}^{(p)} = \text{cor} [t, x_i^{(p)}] \) be the Spearman’s rank correlation between gene expression profiles \( t \) and \( x_i^{(p)} \), which is the standard Pearson product moment correlation calculated using the ranks of the data. Finally, let \( \tilde{r}_p = \frac{1}{n_p} \sum_{i=1}^{n_p} r_{i}^{(p)} \) be the average Spearman’s rank correlation between the test sample’s gene expression profile and the gene expression profiles of phenotype \( p \). Then the classification of the test sample is the phenotype \( p \in P \) that maximizes \( \tilde{r}_p \).

For purposes of ranking the test samples when \( k = 2 \) phenotypes, we use a correlation score \( \rho = \tilde{r}_p^{(S)} - \tilde{r}_p^{(N)} \), where \( p = \) (non-muscle invasive, muscle invasive), (low grade, high grade), and (long-term survivor, short-term survivor) for stage, grade, and DSS alignment, respectively.

Bladder stage and grade alignments

The SRCCM algorithm was used along with separate published and validated stage and grade gene signatures (25) to classify tumors as muscle invasive (stages T2–T4) or nonmuscle invasive (stages Ta–T1), and as high grade (G3–G4) or low grade (G1–G2). Each gene signature consists of 100 genes, with 54 genes common to both signatures. For each signature, we used the 84 genes common to the training (Lindgren; ref. 25), validation (MSKCC), and BLA-36 data sets, with 46 genes common to both signatures. The 2 signatures were further validated using the Chungbuk National University Hospital (CNUH) cohort (29) and the Stansky cohort of patients from the Henri Mondor Hospital in France (30) using the genes common to all data sets (72 for stage and 70 for grade). For visualization purposes, the correlation score (described above) was used to rank the cell lines, with higher scores corresponding to muscle-invasive and high-grade tumors, respectively.

Bladder and colorectal DSS alignments

We use DSS in all of our analyses with survival time equal to the time from diagnoses to time of death or last follow-up. All individuals not dead from disease are censored at the time of last follow-up or time of death from another cause. For DSS alignment, the “median cut” labeling method was used to identify long-term survivors (i.e., low-risk individuals) and short-term survivors (i.e., high-risk individuals; ref. 31).

Patients without censored survival times are assigned to the long-term survivor group if the survival time exceeds the median survival time of all patients in the training cohort, and assigned to the short-term survivor group otherwise. For a patient with an unobserved survival time of \( T \) that is censored at \( C \), we calculate

\[
P(\text{long-term survivor}) = P(T > \text{median survival time} | T > C),
\]

where \( P(T > \text{median survival time}) \) and \( P(T > C) \) are estimated by the Kaplan–Meier (KM) survival curve for all patients (32).
The patient is then assigned to the long-term survivor group if \( P(\text{long-term survivor}) > 0.5 \), and assigned to the short-term survivor group otherwise.

### Results

**A new method that aligns cell lines to human tumors based on gene expression similarity**

We developed a new classification method, herein referred to as the SRCCM and first applied it to 36 bladder cancer cell lines (BLA-36; Supplementary Table S1). These lines were assigned the following tumor phenotypes: tissue of origin (from 10 epithelial cancers), bladder cancer stage (invasive vs. nonmuscle invasive), grade (low vs. high), and DSS (long-term vs. short-term survivors). SRCCM (Fig. 1A) aligns cell lines to human tumor phenotypes based on the gene expression similarity between cell lines and human tumors calculated using gene signatures, described below and in Materials and Methods, that consists of a subset of all genes profiled in both groups (cell lines and human tumors) (Fig. 1B and C). The SRCCM consists of 3 steps: (i) for each test sample, calculate the rank correlation between its gene expression profile and the gene expression profiles of each training sample with known phenotype (e.g., grade), (ii) assign the phenotype with the highest mean correlation to the test sample producing the alignment, (iii) repeat the procedure for all test samples (Fig. 1A). Therefore, each test sample (e.g., cell line in a set) is aligned to the human tumor phenotype it is most similar to, based on the similarity between gene expression profiles.

For each tumor phenotype, the SRCCM algorithm is validated on at least 1 independent test data set and then applied to the BLA-36 cell lines. For each test sample, the primary output of the SRCCM is its molecular alignment (i.e., a qualitative phenotype), whereas the secondary output is a quantification of the relative strength of 1 alignment (e.g., to high grade) to another (e.g., to low grade) by subtracting the corresponding mean correlations, producing a correlation score. For stage, grade, and DSS alignments, a higher correlation score indicates greater similarity with muscle invasive, high grade, and short-term survivors, respectively (see Materials and Methods).
Our SRCCM algorithm is based on the classification method of Wang and colleagues (33) but is modified in 3 important ways. The first modification is that we use the Spearman rank correlation rather than Pearson product moment correlation to measure gene expression similarity. Our choice of a rank-based classifier is motivated by the fact that the rank value of a probe’s expression level within a sample is insensitive to standard within-sample preprocessing methods and is not directly affected by other samples on the array, whereas the expression value is highly dependent on the preprocessing and normalization methods used and can be affected by additional probes and samples (e.g., when robust multichip average normalization is used). The second modification is that instead of measuring similarity based on a sample’s global gene expression profile (i.e., the expression levels of all common microarray probes) as used previously (33), we measure similarity based on gene signatures appropriate for the microarray data sets under study and the phenotype of interest (Fig. 1C). For tissue alignment, because bladder samples \((N = 350)\) were both FFPE and FF (Supplementary Table S2), the gene signature used consists of 12,402 probes in which expression values were preserved across FF and FFPE (24). For stage and grade alignment, 2 previously published and independently validated gene signatures are used (ref. 25; Supplementary Tables S3 and S4). For DSS alignment, the gene signature consists of the 181 probes that significantly correlate with survival in the training data set (MSKCC; Supplementary Table S5) with univariate Cox proportional hazards model log-rank \(P < 0.01\). The final modification is that for each phenotype, we ensure that the SRCCM molecular alignment is accurate through validation on at least 1 independent data set. References for all data sets are provided in Supplementary Tables S1 to S5, and data sets are processed as described in Supplementary Materials. The SRCCM algorithm is comparable with common classification methods, including support vector machines, nearest centroid classification, and \(k\)-nearest neighbor, but carries out slightly better in terms of overall accuracy and consistency.

![Figure 2](image-url)

**Figure 2.** Independent validation and BLA-36 cell line alignment to tissue of origin using the SRCCM algorithm. Test samples and BLA-36 cell lines are aligned to 10 epithelial cancers from a training data set including bladder, breast, cervix, colorectum, kidney, lung adeno, lung scc, ovary, prostate, and thyroid samples \((N = 1,690)\). A, tissue-specific accuracy of SRCCM alignment algorithm on independent data sets \((N = 1,690)\). B, confusion matrix for independent validation presented as a heatmap, with green indicating correct alignment and red indicating incorrect alignment. For each tissue type (i.e., row of the matrix), the proportion of samples aligned with each tissue type is reported. C, BLA-36 tissue of origin alignment heatmap. For each cell line, the color represents the average correlation with each tissue type, with red indicating strong positive correlation and green indicating weak positive correlation. All cell lines are most strongly correlated (i.e., aligned) with bladder, with the exception of Cubill and SW1710 (arrows) which are aligned with colorectum and ovary, respectively.
across multiple data sets (described in Supplemental Information and Supplementary Tables S6 and S7).

**Molecular alignment of human bladder cell lines as a function of tumor tissue of origin**

We first determined whether the BLA-36 cell lines had gene expression profiles that were more similar to bladder cancer or other common carcinomas of cervical, prostate, ovarian, breast, thyroid, kidney, colorectal, and lung origin. These tumor types were selected based on the following criteria (i) epithelial histology, (ii) availability of at least 1 public expression microarray data set containing 30 or more tumor samples, and (iii) profiled on an Affymetrix platform, the platform used for BLA-36 profiling. Because distinct molecular profiles for lung cancer subtypes have been well characterized (34), we considered lung adenocarcinoma (adeno) and lung squamous cell carcinoma (scc) separately. A total of 3,320 tumor samples were used in this analysis with 1,630 used in training and 1,690 in validation of the SRCCM algorithm (Supplementary Table S2). When applied on the 1,690 validation samples from 10 tumor types, the SRCCM algorithm had a mean accuracy of 89%, with tumor-specific accuracy ranging from 77% for cervical tumors to 99% for both colorectal and prostate tumors (Fig. 2A). The overall accuracy for bladder cancer was 91% (Fig. 2A). Interestingly, the majority of misaligned samples are aligned with lung (adeno), including 12% of all ovary samples and 11% of all lung (scc) samples (Fig. 2B). The majority of misaligned cervical samples are aligned with bladder, which comprise 15% of all cervical samples. Importantly, when we applied the SRCCM to the BLA-36 panel, 34 of 36 (94%) cell lines align with tumors of bladder origin. Of the 2 not predicted to be of bladder tumor origin, one (CuhIII) is predicted to be colorectal and the other (SW1710) ovarian (Fig. 2C).

**Table 1. Performance of stage and grade classification using the SRCCM alignment algorithm**

<table>
<thead>
<tr>
<th>Stage</th>
<th>Accuracy</th>
<th>Ta–T1</th>
<th>T2–T4</th>
<th>Overall</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOOCV in Lindgren(^a)</td>
<td>0.722</td>
<td>0.933</td>
<td>0.789</td>
<td>0.930</td>
<td></td>
</tr>
<tr>
<td>Independent validation in MSKCC(^b)</td>
<td>0.880</td>
<td>0.818</td>
<td>0.835</td>
<td>0.895</td>
<td></td>
</tr>
<tr>
<td>Independent validation in CNUH(^c)</td>
<td>0.702</td>
<td>0.721</td>
<td>0.709</td>
<td>0.813</td>
<td></td>
</tr>
<tr>
<td>Independent validation in Stransky(^d)</td>
<td>0.919</td>
<td>0.750</td>
<td>0.817</td>
<td>0.920</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade</th>
<th>Accuracy</th>
<th>Low grade</th>
<th>High grade</th>
<th>Overall</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOOCV in Lindgren(^a)</td>
<td>0.861</td>
<td>0.875</td>
<td>0.868</td>
<td>0.942</td>
<td></td>
</tr>
<tr>
<td>Independent validation in MSKCC(^b)</td>
<td>0.722</td>
<td>0.836</td>
<td>0.813</td>
<td>0.814</td>
<td></td>
</tr>
<tr>
<td>Independent validation in CNUH(^c)</td>
<td>0.724</td>
<td>0.867</td>
<td>0.776</td>
<td>0.871</td>
<td></td>
</tr>
<tr>
<td>Independent validation in Stransky(^d)</td>
<td>0.790</td>
<td>0.849</td>
<td>0.824</td>
<td>0.877</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Profiled on 2 specialized platforms, Swegene Human 27K RAP UniGene 188 and SWEGENE H_v3.0.1. 35K arrays; available from the GEO (21).

\(^b\)Profiled on the HG-U133A microarray platform, available as supplementary material to publication (24).

\(^c\)Profiled on the Illumina human-6 v2.0 microarray platform, available from GEO at Accession #GSE13507.

\(^d\)Profiled on the HG-U195 and HG-U195Av2 microarray platforms, available from Array Express (51) at Accession #E-TABM-147.

**Molecular alignment of human bladder cell lines as a function of tumor stage and grade**

We next used the SRCCM to align the 36 bladder cancer cell lines as a function of human bladder tumor stage and grade, using previously published and validated stage and grade gene signatures (ref. 25; Fig. 1C). Stage alignment is to nonmuscle- or muscle-invasive (Ta–T1 or T2–T4) tumors and grade alignment is to high- or low-grade (G3–G4 or G1–G2) tumors. For both stage and grade alignments, we use the Lindgren cohort (25) for training and leave-one-out cross-validation (LOOCV) and use 3 additional data sets for independent validation: patient profiles from the MSKCC (26), the CNUH (29), and the Stransky cohort of patients from the Henri Mondor Hospital in France (ref. 30; Supplementary Tables S3 and S4). The SRCCM carries out well in LOOCV and independent validation for stage and grade (Table 1). The high area under the curve (AUC) in LOOCV (0.93 ∼ 0.94) and independent validation (0.81 ∼ 0.92) indicates the SRCCM algorithm accurately separates stage and grade subgroups in tumor samples. When the SRCCM was applied to the 36 cell lines, 20 aligned with muscle-invasive tumors and 16 with nonmuscle-invasive tumors (Fig. 3A). For grade, 22 and 14 aligned with high-grade and low-grade tumors, respectively (Fig. 3B). The rankings of the cell lines in stage and grade alignment are highly correlated (Spearman $r = 0.89$, $P < 0.0001$), with 32 of 36 cell lines resembling the muscle-invasive high-grade and nonmuscle-invasive low-grade tumor types (Fig. 3C) commonly seen in patients (25, 35, 36).

To determine the value of the stage and grade realignment, we evaluated the correlation between SRCCM assigned stages and grades and the status of several common gene mutations (see Supplementary Materials) known to be associated with biological phenotypes in human bladder cancer (Fig. 3A and B).
For assigned cell line stages, we observed positive correlations with TP53 ($r = 0.671$), RB1 ($r = 0.408$), PTEN ($r = 0.354$), and KRAS ($r = 0.277$) mutants, and negative correlations with CDKN2A ($r = -0.408$) and FGFR3 ($r = -0.181$) mutants, consistent with the role of the first 3 in aggressive disease and the association of the latter 2 with a nonmuscle-invasive phenotype ($37–39$). Notably, the correlation with TP53 reaches a marginal significance (Fisher exact $P = 0.06$), despite the fact that TP53 mutation status is known for just 11 cell lines. For assigned cell line grades, we observe positive correlations with TP53 ($r = 0.810$, Fisher exact $P = 0.024$), RB1 ($r = 0.667$, Fisher exact $P = 0.076$), PTEN ($r = 0.304$), and KRAS ($r = 0.207$) mutants, and negative correlations with CDKN2A ($r = -0.310$) and FGFR3 ($r = -0.305$) mutants.

**Cell lines assigned their original tumor grades do not correlate with survival but correlation is restored using molecularly aligned cell line grades**

We collected stage and grade information for the tumor samples used to derive the BLA-36 lines (Supplementary Table S1) and compared the original tumor information with the SRCCM assigned grades described above. Only 8 cell lines had stage information from their original tumors, whereas 28 cell lines had grade information so only the latter was used in the analysis below. Of the 9 cell lines originating from low-grade tumors, 4 (44%) aligned with low-grade tumors in the Lindgren bladder tumor set ($N = 144$; Supplementary Table S4) and 12 (63%) of the 19 cell lines from high-grade tumors are aligned with high-grade lesions. Strikingly, 43% (12 of 28) of the cell lines do not resemble their tumor origins with respect to grade gene signatures.

Because of the disagreement between BLA-36 original tumor grade and SRCCM assigned grade and because grade is a strong indicator of patient survival (40), we next investigated the clinical relevance of this alignment (Fig. 4A). Specifically, we asked whether the grades of the original tumors from which the cell lines were derived or the grades assigned via SRCCM were more predictive of DSS. To evaluate this, we aligned 87 tumor samples from an independent patient cohort (MSKCC; Supplementary Table S4) to either a patient data set (Lindgren; Supplementary Table S4) or the BLA-36 cell lines, using their original or reassigned grades, and then generated KM DSS curves for the assigned low- and high-grade samples. In the MSKCC cohort, patient grade is a very strong predictor of survival, producing KM survival curves with an HR of 11.29 and a log-rank $P = 0.0026$ (Fig. 4B). Use of the Lindgren cohort establishes the efficacy of the SRCCM algorithm to predict MSKCC grade and then generate
KM survival curves because this resulted in excellent stratification (HR = 5.62, log-rank P = 0.0013, Fig. 4C). Using the grades of the original tumors from where BLA-36 was derived, the relationship between grade and survival is lost (HR = 0.91, log-rank P = 0.79; Fig. 4D). In contrast, when the MSKCC samples are aligned with SRCCM assigned cell line grades, the relationship between grade and survival is restored (HR = 3.41, log-rank P = 0.0077; Fig. 4E). Together, these results indicate that cell lines no longer resemble the tumors they were derived from and cell lines selected by their original grades no longer correlate with survival. In contrast, the SRCCM is effective in identifying lines that correlate with clinical outcome.

The clinical impact of selecting human bladder cancer cell lines based on their similarity with human tumors

The overarching objective of this work is to identify the most clinically relevant cancer cell lines, those most strongly correlated with disease outcome. Above, we noted that the aligned grades were more clinically significant than the actual grades of the original tumors from which the cell lines were derived. To define the most clinically relevant cell lines, here, we first align the BLA-36 cell lines directly with DSS, using the MSKCC cohort as the training data set. The MSKCC cohort contains 87 tumor samples and has a median survival time of 26.0 months and a 5-year DSS rate of 56.3% (Supplementary Table S5). Patients were assigned to risk groups (i.e., long- and short-term survivors) using the “median cut” labeling method (31) as described in Materials and Methods. For SRCCM analysis, we use a gene signature consisting of all probes correlated with DSS in MSKCC (univariate Cox proportional hazards model, log-rank P < 0.01), and applied the SRCCM algorithm to align the BLA-36 cell lines to long- and short-term survivors in the MSKCC cohort. In an LOOCV on MSKCC, the long- and short-term survivors produce survival curves with a HR of 3.75 and a log-rank P = 0.00011 (Fig. 5A). We then used the SRCCM to align the BLA-36 cell lines with long- and short-term survivors in MSKCC and ranked the cell lines according to their correlation score (Fig. 5B).

Next, we compared this ranking to the tumor of origin alignment (Fig. 2C) and to the grade ranking (Fig. 3B) that was found clinically relevant (Fig. 4E) and selected only cells aligning with bladder tumors that were either in high-grade/short-term survivor groups or low-grade/long-term survivor groups. This provides us with 15 and 4 cell lines in each group, respectively (Fig. 6A).
We then evaluated the clinical relevance of this selection by using it to align an independent cohort of patients from CNUH (N=129, 5-year DSS rate = 80.5%; ref. 29; Supplementary Table S5). Patients aligned to these high-grade/short-term and low-grade/long-term survivors produced survival curves with a HR of 4.16 and a log-rank \(P=0.0040\) (Fig. 6B) indicating the selection of cell lines based on their tissue of origin, grade, and DSS alignments has biological significance.

**Selecting human colorectal cancer cell lines based on their similarity to human tumors leads to the identification of clinically relevant lines**

To evaluate the generalizability of our SRCCM algorithm and cell line selection concept, we analyzed a panel of 22 colorectal cell lines (CO-22) profiled by GlaxoSmithKline. The CO-22 cell lines were first aligned to tissue of origin, in which the SRCCM alignment accuracy for colorectal tumors in the independent test data sets was 99% (Fig. 2A). All CO-22 cell line samples aligned with colorectum with the exception of COLO-320HSR, in which all three replicates aligned with uterus, and RKO-E6, in which 1 replicate aligned with bladder (Supplementary Fig. S1A).

We next aligned the CO-22 cell lines directly to DSS, using a set of 55 patients with colorectal cancer from the Vanderbilt Medical Center as the training data set (VMC cohort; ref. 27). The VMC cohort has a median survival time of 38.0 months and a 5-year DSS rate of 73.4% (Supplementary Table S5). Patients were assigned to risk groups using the "median cut" labeling method as was done for the BLA-36 cell lines (see Materials and Methods). For SRCCM analysis, we used a previously published and validated gene signature consisting of 34 genes that correlates with recurrence and survival (27). In a LOOCV on VMC,
Selection of Clinically Relevant Cancer Cell Lines

Figure 6. Selection and validation of clinically relevant BLA-36 and CO-22 cell lines. A, SRCCM alignment of BLA-36 cell lines to tissue of origin, grade, and DSS. Heatmap of the clinically relevant BLA-36 cell lines aligning with bladder, high grade, and short-term survivors or bladder, low grade, and long-term survivors. The cell lines are ranked by their DSS alignment score. B, validation of the clinical relevance of the selected cell lines in an independent dataset. KM survival curves are generated for patients in CNUH following alignment to the clinically relevant cell lines in A. C, SRCCM alignment of CO-22 cell lines to tissue of origin and DSS. Heatmap of the clinically relevant CO-22 cell lines having all replicates aligning with colorectum and no ambiguous DSS alignments. The cell lines are ranked by their average DSS alignment score. D, validation of the clinical relevance of the selected cell lines in an independent dataset. KM survival curves are generated for patients in MCC following alignment to the clinically relevant cell lines in C.

Discussion

For a cell line model to have clinical utility it must be representative of the human tumor of interest. Importantly, cell line models selected based on the characteristics of the tumors from which they are derived, their xenograft behavior, or their technical convenience may not be clinically relevant and therefore lack clinical translatability. In this article, we describe a novel framework for the selection of the most clinically relevant cell lines. At the heart of this approach is the SRCCM alignment algorithm, with software we have made available to the scientific community, and which is applicable to any cancer cell line. Although we have selected clinically relevant cell lines for bladder and colorectal cancers, we caution that because gene expression profiles vary across culture conditions (41, 42) and with cell passage number, the cell lines selected here as most clinically relevant may not be so in another laboratory.
with the same lines. Instead, we recommend SRCCM alignment and cell line selection using gene expression profiles obtained in the specific laboratory. In time, we will determine whether this possible variation in the repertoire of the "most clinically relevant" lines does indeed occur.

In developing SRCCM, a variety of classification methods for aligning cell lines with clinical phenotypes could have been implemented, including standard machine learning algorithms such as nearest neighbor and support vector machines (43, 44). However, the SRCCM algorithm was developed after carefully considering its ease of interpretation and its applicability. Correlation is an intuitive measure of similarity, while the specific use of Spearman’s correlation ensures broad applicability to additional data sets and cell lines, as the choice of data normalization and processing methods will have relatively little impact on the output of the algorithm.

Independent validation of SRCCM was a critical feature of our work. Because classifiers developed from high dimensional microarray data sets are prone to overfitting (45), and technical batch effects can correlate with biological outcomes of interest (46), independent validation is essential if reported classification accuracies are expected to be representative of performance in additional data sets. We validated the tissue of origin, stage, and grade SRCCM alignments on multiple independent data sets and achieved excellent overall accuracies. We validated the DSS SRCCM alignments on an independent data set by first aligning the BLA-36 cell lines to long- and short-term survivors in the MSKCC training cohort and then aligning 129 patients in an independent cohort (CNUH) to the 19 clinically relevant cell lines to produce KM survival curves with clear stratification.

The SRCCM tissue alignment finds that 34 of 36 (94%) of the bladder cell lines and 20 of 22 (91%) of the colorectal cell lines resemble their tissue of origin more closely than the 9 additional carcinomas we considered. This is reassuring, as a previous tissue classification of the NCI-60 cell line panel suggested that as few as 57% of cell lines most closely resembled their presumed tissue of origin (10). In general, the BLA-36 cell lines correlate more strongly with tissue types in which squamous cell histologies are common (e.g., cervix and ovary) than with tissues in which squamous cell histologies are rare (e.g., breast, thyroid, prostate, and kidney; Fig. 2C), suggesting that it may be relatively difficult to distinguish between tissue types if their histologies are similar. In the independent validation on over 300 bladder samples, 9% were misaligned, and the fact that 2 of 36 (6%) of the cell lines are misaligned in the BLA-36 panel is within the expected margin of error. Further analysis of these 2 lines, CubIII and SW1710, is required to determine whether these are of bladder origin.

Most striking is our finding that the grade alignment of the BLA-36 cell lines is consistent with only 44% of the low-grade tumors and 63% of the high-grade tumors that the cell lines were derived from. One possible explanation for these results is that the alignment is inaccurate. However, our findings support the validity of the alignment and loss of relevance of the original tumor grade. For example, we found that TP53 and RB1 mutations are associated with aligned high-grade muscle-invasive tumors whereas FGFR3 mutations are associated with aligned low-grade, nonmuscle-invasive tumors, which is consistent with the prevalence of TP53, RB1, and FGFR3 mutations in patients (37–39). Because tumor grade is an important clinical parameter that correlates with both stage and survival in patients (40, 47), for a cell line model to have clinical relevance, it should properly represent these clinical phenotypes. Importantly, we have found that selection of cell line models based on clinical information from their original tumors will fail to properly represent these phenotypes for the cell lines we analyzed. In contrast, once aligned using SRCCM, the new cell line grades relate to clinical outcomes (Fig. 4D and E).

Cell line models have also been selected based on their xenograft behavior. However, animal models do not always recapitulate human physiology and behavior (48, 49). In a meta-analysis looking at 39 pharmacologic agents with both xenograft data and results from phase II clinical trials, Johnson and colleagues found that out of 6 xenograft histologies, only lung xenografts had the ability to predict clinical activity in the same human cancer histology, though interestingly several were predictive of clinical activity in other histologies (50). In contrast, xenografts derived from aligned cell lines validated for their ability to reflect human tumor biology, using alignment and selection process shown here (Fig. 6), may provide a readout of drug activity in xenografts that parallel that seen in human tumors.

In conclusion, we describe a framework for aligning cell lines with the clinicopathologic phenotypes of human tumors such as tissue of origin, stage, and grade, and DSS that can be used to select the most clinically relevant cell line models. Our technique has the potential to reduce the discrepancy often encountered between in vitro and in vivo laboratory data and patient interventions. Although the approach is shown in bladder and colorectal cancer cells, SRCCM provides a general framework that is applicable to all cancer types and is freely available to the scientific community.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Grant Support

This work was supported in part by NIH grants CA075115 and CA143791 to D. Theodorescu.

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 July 22, 2011; revised October 5, 2011; accepted October 14, 2011; published OnlineFirst October 19, 2011.

References


A Framework to Select Clinically Relevant Cancer Cell Lines for Investigation by Establishing Their Molecular Similarity with Primary Human Cancers


Cancer Res Published OnlineFirst October 19, 2011.