A framework to select clinically relevant cancer cell lines for investigation by establishing their molecular similarity with primary human cancers

Garrett M. Dancik 1,§, Yuanbin Ru 1,§, Charles Owens 1, and Dan Theodorescu 1,2,3,*

1 Department of Surgery, University of Colorado, Aurora, Colorado, USA
2 Department of Pharmacology, University of Colorado, Aurora, Colorado, USA
3 University of Colorado Comprehensive Cancer Center, Aurora, Colorado, USA
§ Both authors contributed equally to this work
* Corresponding author

Type: Regular Article

Running Title: Selection of clinically relevant cancer cell lines

Abbreviations: adeno: adenocarcinoma, scc: squamous cell carcinoma, NMI: non-muscle invasive, MI: muscle invasive, SRCCM, Spearman’s rank correlation classification method, KM: Kaplan-Meier, HR: hazard ratio, DSS: disease specific survival

Key Words: cell lines, tissue of origin, tumor grade, survival analysis, DNA microarray

Acknowledgements: This work was supported in part by National Institutes of Health grants CA075115 and CA143791 to DT.

Corresponding author: Dan Theodorescu, Departments of Surgery and Pharmacology and University of Colorado Comprehensive Cancer Center, Aurora, CO 80045, Tel: 303-724-7135, Fax: 303-724-3162, E-Mail: dan.theodorescu@ucdenver.edu.

Conflicts of Interest: none

Manuscript info:
Word Count (excluding references): 5009
Number of Figures: 6
Number of Supplementary Figures: 1
Total Number of Tables: 1
Number of Supplementary Tables: 7
ABSTRACT

Experimental work using human cancer cell lines often does not translate to the clinic. We posit 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 SRCCM (Spearman’s Rank Correlation Classification Method) 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 demonstrate utility, we used SRCCM to assess similarity of 36 bladder cancer lines with 10 epithelial human tumor types (N=1630 samples) and with bladder tumors of different stages and grades (N=144 samples). While 34/36 lines aligned to bladder tumors rather than other histologies, only 16/28 had SRCCM assigned grades identical to that of their original source tumors. To evaluate the clinical relevance of this approach, we demonstrate that gene expression profiles of aligned cell lines stratify survival in an independent cohort of 87 bladder patients (HR = 3.41, logrank p = 0.0077), while 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, logrank p = 0.0019). By selecting cell lines that reflect human tumors, our technique promises to improve the clinical translation of laboratory investigations in cancer.
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 anti-cancer agents (7) and to identify novel multi-agent 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 non-muscle invasive tumors progress to muscle invasion (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 anti-cancer agents depends on stage and grade, with resistance more common in poorly differentiated tumors (16, 17). Lastly, 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 demonstrate that gene expression signatures of cancer cell lines selected on the basis of their alignments to human cancer phenotypes predict patient survival outcomes while 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. Since 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) (21), accession #GSE5845, Ref #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 (SRCCM)

The 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 (Figure 1A). The gene expression profile is based on a gene signature that is unique to each phenotype (Figure 1B-C). The tissue of origin gene signature consists of 12,402 “high fidelity” probes whose expression values are consistent across formalin fixed, paraffin embedded and fresh frozen 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 logrank p-values < 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 \( P = (P_1, ..., P_k) \) be the set of training phenotypes of interest with known gene expression profiles. In Figure 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 \( p, i = 1, ..., n_p \). 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 \( \bar{r}^{(p)} = \frac{1}{n_p} \sum_{i=1}^{n_p} r_i^{(p)} \) 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 \( \bar{r}^{(p)} \). For purposes of ranking the test samples when \( k = 2 \) phenotypes, we use a correlation score \( = r^{(P_2)} - r^{(P_1)} \), 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. Software for the SRCCM is available to the scientific community through the Comprehensive R Archive Network (28) as part of the correlation classification method (CCM) contributed package.

**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 non-muscle 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) (25), validation (MSKCC), and BLA-36 datasets, with 46 genes common to both
signatures. The two signatures were further validated using the Chungbuk National University Hospital (CNUH) cohort (29) and the Stransky cohort of patients from the Henri Mondor Hospital in France (30) using the genes common to all datasets (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, respectfully.

**Bladder and colorectal disease specific survival 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) (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 using the Kaplan-Meier 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 Spearman's Rank Correlation Classification Method (SRCCM) and first applied it to 36 bladder cancer cell lines (BLA-36) (Supplementary Table S1). These lines were assigned to the following tumor phenotypes: tissue of origin (from 10 epithelial cancers), bladder cancer stage (invasive vs. non-muscle invasive), grade
(low vs. high), and disease specific survival (DSS, long-term vs. short-term survivors). SRCCM (Figure 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) (Figure 1B-C). The SRCCM consists of three steps: 1) 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), 2) assign the phenotype with the highest mean correlation to the test sample, producing the alignment, 3) repeat the procedure for all test samples (Figure 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 one independent test dataset 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) while the secondary output is a quantification of the relative strength of one 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 disease specific survival 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 et al. (33) but is modified in three important ways. The first modification is that we use Spearman's 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 probes or samples on the array, whereas the expression value is highly dependent on the pre-processing and normalization methods used and can be affected by additional probes and samples (e.g., when robust multichip average (RMA) 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 (Figure 1C). For tissue alignment, because bladder samples (N=350) were both formalin fixed, paraffin embedded (FFPE) and fresh frozen (FF) (Supplementary Table S2), the gene signature used consists of 12,402 probes whose expression values were preserved across FF and FFPE (24). For stage and grade alignment, two previously published and independently validated gene signatures are used (25) (Supplementary Tables S3-S4). For disease specific survival alignment, the gene signature consists of the 181 probes that significantly correlate with survival in the training dataset (MSKCC, Supplementary Table S5) with univariate Cox proportional hazards model logrank p-values < 0.01. The final modification is that for each phenotype, we ensure that the SRCCM molecular alignment is accurate through validation on at least one independent dataset. References for all datasets are provided in Supplementary Tables S1-S5 and datasets 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 performs slightly better in terms of overall accuracy and consistency across multiple datasets (described in Supplemental Information and Supplementary Tables S6-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 1) epithelial histology, 2) availability of at least one public expression microarray dataset containing ≥ 30 tumor samples, and 3) 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 3320 tumor samples were used in this analysis with 1630 used in training and 1690 in validation of the SRCCM algorithm (Supplementary Table S2). When applied on the 1690 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 (Figure 2A). The overall accuracy for bladder cancer was 91% (Figure 2A). Interestingly, the majority of misaligned samples are aligned with lung (adenocarcinoma), including 12% of all ovary samples and 11% of all lung (squamous cell carcinoma) samples (Figure 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%) cells align with tumors of bladder origin. Of the two not predicted to be of bladder tumor origin, one (CubIII) is predicted to be colorectal and the other (SW1710) ovarian (Figure 2C).

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 (25) (Figure 1C). Stage alignment is to non-muscle 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 three additional datasets for independent validation: patient profiles from the Memorial Sloan-Kettering Cancer Center (MSKCC) (26), the Chungbuk National University Hospital (CNUH) (29), and the Stransky cohort of patients from the Henri Mondor Hospital in France (30) (Tables S3-S4). The SRCCM performs 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 and 16 with non-muscle-invasive tumors (Figure 3A). For grade, 22 and 14 aligned with high-grade and low-grade tumors,
respectively (Figure 3B). The rankings of the cell lines in stage and grade alignment are highly correlated (Spearman $r = 0.89$, $p < 0.0001$), with 32/36 cell lines resembling the muscle-invasive high-grade and non-muscle-invasive low-grade tumor types (Figure 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 (Figure 3A,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 three in aggressive disease and the association of the latter two with a non-muscle invasive phenotype (37-39). Notably, the correlation with $TP53$ reaches a marginal significance (Fisher’s 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’s exact $p = 0.024$), $RB1$ $(r = 0.667$, Fisher’s 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, while 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, Table S4) and 12 (63%) of the 19 cell lines from high grade tumors are aligned with
high grade lesions. Strikingly, 43% (12/28) of the cell lines don’t 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 (Figure 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 dataset (Lindgren, Supplementary Table S4) or the BLA-36 cell lines, using their original or reassigned grades, and then generated Kaplan-Meier (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 a hazard ratio (HR) of 11.29 and a logrank p-value of 0.0026 (Figure 4B). Use of the Lindgren cohort establishes the efficacy of the SRCCM algorithm to predict MSKCC grade and then generate KM survival curves since this resulted in excellent stratification (HR = 5.62, logrank p = 0.0013, Figure 4C). Using the grades of the original tumors from where BLA-36 were derived, the relationship between grade and survival is lost (HR = 0.91, logrank p = 0.79, Figure 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, logrank p = 0.0077, Figure 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 dataset. 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, logrank \( p \)-value < 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 a LOOCV on MSKCC, the long- and short-term survivors produce survival curves with a hazard ratio of 3.75 and a logrank \( p \)-value = 0.00011 (Figure 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 (Figure 5B).

Next we compared this ranking to the tumor of origin alignment (Figure 2C) and to the grade ranking (Figure 3B) that was found clinically relevant (Figure 4E) and selected only cells aligning with bladder tumors that were either in high-grade/short-term survivors or low-grade/long-term survivor groups. This provides us with 15 and 4 cell lines in each group respectively (Figure 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 hazard ratio of 4.16 and a logrank \( p \)-value = 0.0040 (Figure 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, where the SRCCM alignment accuracy for colorectal tumors in the independent test datasets was 99% (Figure 2A). All CO-22 cell line samples aligned with colorectum with the exception of COLO-320HSR, whose samples all aligned with uterus, and RKO-E6, where one sample aligned with bladder (Supplementary Figure 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 dataset (VMC cohort, Ref #27). The VMC cohort has a median survival time of 38.0 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, the long- and short-term survivors produce survival curves with a hazard ratio of 8.00 and a logrank p-value = 0.0225 (Figure 5C). We then used the SRCCM to align the CO-22 cell lines with long- and short-term survivors in VMC, ranked the cell lines according to their correlation score, and used the median correlation score to separate long- and short-term survivors (Figure 5D).

For cell line selection, a cell line was deemed clinically relevant if all of its replicates aligned with colorectum and all replicates had the same DSS alignment (Figure 6C). Cell lines were removed if they had ambiguous DSS alignments defined as alignments with one or more replicates aligned with short-term survivors and at least one aligned with long-term survivors. Our selection resulted in the removal of the following lines: COLO-320HSR, which aligned with uterus; LS1034, SW1463, and SW948, which had ambiguous DSS alignments, and RKO-E6, which had an ambiguous DSS alignment and also had one replicate aligning with bladder.
The clinical relevance of the remaining 17 cell lines was then evaluated in an independent cohort of patients from the Moffitt Cancer Center (MCC, N = 177, 5 year DSS rate = 67.0%, Ref #27, Supplementary Table S5). Alignment of the MCC patients to the selected colorectal cell lines produces survival curves with a hazard ratio of 2.35 and a logrank p-value of 0.0019, establishing their clinical relevance (Figure 6D). Importantly, when all 22 colorectal cell lines were used in this analysis (i.e., without the selection step), the hazard ratio falls to 1.69 and the separation of survival curves is no longer statistically significant (logrank p-value = 0.054, Supplementary Figure S1B).
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 paper 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 (42, 43) and with cell passage number, the cell lines selected here as most clinically relevant may not be so in another lab with the same lines. Instead, we recommend SRCCM alignment and cell line selection using gene expression profiles obtained in the specific lab. 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 (44, 45). 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 datasets 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 datasets are prone to over-fitting (46), and technical batch effects can correlate with biological outcomes of interest (47), independent validation is essential if reported classification accuracies are expected to be representative of performance in additional datasets. We
validated the tissue of origin, stage, and grade SRCCM alignments on multiple independent datasets, and achieved excellent overall accuracies. We validated the DSS SRCCM alignments on an independent dataset 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 34/36 (94%) of the bladder cell lines and 20/22 (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 where squamous cell histologies are common (e.g., cervix and ovary) than with tissues where squamous cell histologies are rare (e.g., breast, thyroid, prostate, and kidney) (Figure 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 2/36 (6%) of the cell lines are misaligned in the BLA-36 panel is within the expected margin of error. Further analysis of these two 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 while FGFR3 mutations are associated with aligned low grade, non-muscle invasive tumors, which is consistent with the prevalence of TP53, RB1, and FGFR3 mutations in patients (37-39). Since tumor grade is an important clinical parameter that correlates with both stage and survival in patients (40, 48), 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 (Figure 4D,E).

Cell line models have also been selected based on their xenograft behavior. However, animal models do not always recapitulate human physiology and behavior (49, 50). In a meta-analysis looking at 39 pharmacological agents with both xenograft data and results from Phase II clinical trials, Johnson et al. found that out of six 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 (51). In contrast, xenografts derived from aligned cell lines validated for their ability to reflect human tumor biology, using alignment and selection process shown here (Figure 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, 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 demonstrated 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.
### Table 1. Performance of stage and grade classification using the SRCCM alignment algorithm

<table>
<thead>
<tr>
<th></th>
<th>Stage</th>
<th></th>
<th></th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Accuracy</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ta-T1</td>
<td>T2-T4</td>
<td>Overall</td>
</tr>
<tr>
<td>LOOCV in Lindgren*</td>
<td>0.722</td>
<td>0.933</td>
<td>0.789</td>
<td>0.930</td>
</tr>
<tr>
<td>Independent Validation in MSKCC†</td>
<td>0.880</td>
<td>0.818</td>
<td>0.835</td>
<td>0.895</td>
</tr>
<tr>
<td>Independent Validation in CNUH‡</td>
<td>0.702</td>
<td>0.721</td>
<td>0.709</td>
<td>0.813</td>
</tr>
<tr>
<td>Independent Validation in Stransky§</td>
<td>0.919</td>
<td>0.750</td>
<td>0.817</td>
<td>0.920</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Grade</th>
<th>Accuracy</th>
<th></th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Low grade</td>
<td>High grade</td>
<td>Overall</td>
</tr>
<tr>
<td>LOOCV in Lindgren*</td>
<td>0.861</td>
<td>0.875</td>
<td>0.868</td>
<td>0.942</td>
</tr>
<tr>
<td>Independent Validation in MSKCC†</td>
<td>0.722</td>
<td>0.836</td>
<td>0.813</td>
<td>0.814</td>
</tr>
<tr>
<td>Independent Validation in CNUH‡</td>
<td>0.724</td>
<td>0.867</td>
<td>0.776</td>
<td>0.871</td>
</tr>
<tr>
<td>Independent Validation in Stransky§</td>
<td>0.790</td>
<td>0.849</td>
<td>0.824</td>
<td>0.877</td>
</tr>
</tbody>
</table>

Abbreviation: LOOCV, leave-one-out cross validation; AUC, area under curve.

*Profiled on two specialized platforms, Swegene Human 27K RAP UniGene 188 and SWEGENE H_v3.0.1. 35K arrays; available from the Gene Expression Omnibus (21)
†Profiled on the HG-U133A microarray platform, available as supplementary material to publication (24).
‡Profiled on the Illumina human-6 v2.0 microarray platform, available from Gene Expression Omnibus at Accession #GSE13507
§Profiled on the HG-U195 and HG-U195Av2 microarray platforms, available from Array Express (52) at Accession #E-TABM-147
FIGURE LEGENDS

Figure 1. Overview and use of the Spearman’s Rank Correlation Classification Method (SRCCM). A, overview of the SRCCM algorithm for cell line / test sample alignment. A test sample is aligned to the phenotype with the highest mean correlation between test and training gene expression profiles using a relevant gene signature. B, alignment of a test sample with a clinical phenotype using SRCCM is based on the gene expression profiles of the training and test samples and a relevant gene signature for the desired phenotype. C, published and derived gene signatures and phenotypes used by SRCCM for tissue of origin, and bladder cancer stage, grade, and disease specific survival (DSS) alignment. ^ signatures from Ref #24; * stage and grade signatures from Ref #25 are different, despite the same number of genes; ** univariate Cox proportional hazards model, logrank p-value < 0.01 in training cohort from Ref #26, see Materials and Methods for details. Abbreviations: FF, fresh frozen; FFPE, formalin fixed and paraffin embedded.

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 dataset including bladder, breast, cervix, colorectum, kidney, lung adenocarcinoma (adenocarcinoma), lung squamous cell carcinoma (squamous cell carcinoma), ovary, prostate, and thyroid samples (N = 1630). A, tissue specific accuracy of SRCCM alignment algorithm on independent datasets (N= 1690). 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 CubIII and SW1710 (arrows) which are aligned with colorectum and ovary, respectively.

Figure 3. A, Stage and B, grade alignment heatmaps. Cell lines are ranked by their correlation scores (see Materials and Methods) from high (red) to low (green) and classified as either muscle invasive (MI, T2-T4) or non-muscle invasive (NMI, Ta-T1) for stage and high (G3-G4) or low (G1-G2) grade, separated by a blue horizontal line. The heatmap also contains the documented gene mutations according to the COSMIC database (53), with purple, yellow, and black indicating known mutation, wild-type, and unknown mutation status, respectively. The histograms below the heatmaps show correlations of mutations with each alignment. P-values of Fisher exact tests: *, 0.05≤p<0.1; **, p<0.05. C, Plot of correlation scores for SRCCM assigned stage and grade. Red points denote cell lines aligned with high grade MI tumors, green points denote cell lines aligned with low grade NMI tumors, and all other cell lines are plotted in black.

Figure 4. Original cell line grades no longer correlate with disease specific survival but correlation is restored via SRCCM alignment to patient tumors. A, Overview of the methodology used. MSKCC tumors are aligned to the grades of Lindgren tumors or BLA-36 cell lines using the grades of the tumors they were derived from (original grades) or the assigned grades from alignment with the Lindgren patients (aligned grades). We report the accuracy of MSKCC grade assignment. We then generate KM survival curves based on B, the original patient grades in MSKCC (positive control), C, MSKCC grades based on alignment to Lindgren patient tumors (second positive control), D, MSKCC grades based on alignment to BLA-36 (original grades), and E, MSKCC grades based on alignment to BLA-36 (aligned grades).

Figure 5. Disease specific survival (DSS) in BLA-36 and CO-22 as a function of long- and short-term survivors determined by “median cut” labeling (see Materials and Methods). A, KM survival curves for bladder cancer patients in the training MSKCC dataset obtained through SRCCM alignment and LOOCV. B, DSS alignment heatmap for BLA-36 cell lines. Cell lines are ranked from high (red) to low (green) by their correlation scores (see Materials and Methods) and classified as either long-term or
short-term survivors, separated by a horizontal blue line, based on their alignment to long- and short-term survivors in MSKCC. C, KM survival curves for colorectal cancer patients in the training VMC dataset obtained through SRCCM alignment and LOOCV. D, DSS alignment heatmap for CO-22 cell lines. Cell lines are ranked from high (red) to low (green) by their correlation scores (see Materials and Methods) and classified as either long-term or short-term survivors, separated by a horizontal blue line, based on their alignment to long- and short-term survivors in VMC.

**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).
REFERENCES


12. MacLeod RAF, Dirks WG, Matsuo Y, Kaufmann M, Milch H, Drexler HG. Widespread intraspecies cross-contamination of human tumor cell lines arising at source. International Journal of Cancer 1999;83:555-63.


34. Garber ME, Troyanskaya OG, Schluens K, et al. Diversity of gene expression in


Step 1: Calculate Spearman correlation between the gene expression profile of test sample (X) and gene expression profiles of known classes (classes A and B here) using a relevant gene signature.

Step 2: The test sample is aligned to phenotype with the highest mean correlation (class B here).

Step 3: Repeat steps 1-2 to classify all test samples.

Training, Class A (e.g., low grade)
Training, Class B (e.g., high grade)

A

Alignment

B

Training samples (expression profiles, phenotypes of human tumors)
Gene signature
Cell line / test samples (expression profiles)

C

Tissue of origin
Stage
Grade
DSS

Gene signature

12,402 FF/FFPE high fidelity probes^*  
84 genes*

84 genes*

181 probes**

SRCCM

Bladder, breast, cervix, colorectal, kidney, lung (adeno), lung (squamous), ovary, prostate, thyroid
Non-muscle invasive (Ta-T1) vs. muscle invasive (T2-T4)
Low grade vs. high grade
Long- vs short-term survivors

Downloaded from cancerres.aacrjournals.org on April 21, 2017. © 2011 American Association for Cancer Research.
Figure 4

A

Training dataset

- Lindgren
- BLA-36 (original grades)
- BLA-36 (aligned grades)

SRCCM

MSKCC original grades

MSKCC SRCCM aligned grades

Grade accuracy

- 1
- 0.813
- 0.582
- 0.802

Kaplan-Meier curves stratified by grade

B

DSS in MSKCC (MSKCC patient grades)

- Low grade: HR = 11.29, P = 0.0026

C

DSS in MSKCC (Lindgren alignment)

- Low grade: HR = 5.62, P = 0.0013

D

DSS in MSKCC (BLA-36 alignment, original grades)

- Low grade: HR = 0.91, P = 0.79

E

DSS in MSKCC (BLA-36 alignment, aligned grades)

- Low grade: HR = 3.41, P = 0.0077
Figure 5

A

Alignment of BLA-36 to MSKCC

B

Alignment of CO-22 to VMC

C

D
Figure 6

A) BLA-36 cell lines
1) SRCCM Alignment (Figures 2C, 3B, 5A,B)
2) Cell Line Selection (see text)

Clinical relevance of alignment and selection in independent dataset

B) Bladder, high grade, short-term survivors
DSS(%) vs Months
HR = 4.16
P = 0.0040

C) CO-22 cell lines
1) SRCCM Alignment (Figures S1A, 5C,D)
2) Cell Line Selection (see text)

Clinical relevance of alignment and selection in independent dataset

D) Colorectal, short-term survivor
dS(%) vs Months
HR = 2.35
P = 0.0019
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.

Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-11-2427

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2011/10/19/0008-5472.CAN-11-2427.DC1

Author Manuscript
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

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.