Cross-cancer Prediction: A Novel Machine Learning Approach to Discover Molecular Targets for Development of Treatments for Multiple Cancers

Katie Gao[1], Dayong Wang[2] and Yi Huang[3]

[1]Harvard University, Boston, MA, USA. [2]PathAI Inc., Boston, MA, USA. [3]Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, MD, USA.

ABSTRACT: Conventional cancer drug development has long been limited to organ- or tissue-specific cancer types. However, it has become increasingly known that specific genetic abnormalities are responsible for the carcinogenesis of multiple cancers. The recent US Food and Drug Administration (FDA) approval of the first multi-cancer drug, Keytruda, has demonstrated the feasibility of developing new drugs that target multiple cancers. Despite a promising future, methodological development for identifying multi-cancer molecular targets remains encumbered. This study developed a novel machine learning approach to identify such genes responsible for multiple cancers by synthesizing salient genomic information from cancer-specific classification models. This approach centered on the cross-cancer prediction method for identifying groups of cancers with high cross-cancer predictability. Furthermore, a robust hybrid classifier, comprising Prediction Analysis for Microarrays and Random Forest, was developed to integrate predictive models for gene inference. This approach has successfully identified key genes shared by endometrial cancer, mammary gland ductal carcinoma, and small cell lung cancer. The results are supported by published experimental evidence. This framework holds potential to transform the current methods of discovering multi-cancer molecular targets for clinical oncology.

KEYWORDS: cross-cancer prediction, machine learning, molecular target, multi-cancer

Introduction

Cancer, one of the leading causes of death worldwide, is conventionally categorized and treated by organ- or tissue-specific disease types.[1] Accordingly, current cancer drug development is often limited to individual cancer types. However, several genetic factors have been observed to be shared by multiple cancers. For instance, germline *BRCA1* and *BRCA2* mutations are known to be associated with breast cancer and epithelial ovarian cancer.[2,3] Recent studies have also suggested that specific genomic loci are associated with breast, ovarian, and prostate cancer.[4] These findings support the emerging concept that specific genetic abnormalities are critical to the pathogenesis, progression, and metastasis of multiple cancers. The discovery of multi-cancer genes will facilitate promising translational applications, including the development of novel pharmaceutical treatments that target multiple cancers. Recently, the US Food and Drug Administration (FDA) approved the first multi-cancer treatment, Keytruda, which targets a specific genetic feature, the microsatellite instability high or mismatch repair deficient, to treat several cancers.[5] Such breakthroughs will provide new opportunities to treat patients with different cancers who share genetic disposition but respond poorly to organ-specific treatment.

High-throughput genomics technology, capable of analyzing tens of thousands of genes simultaneously, has presented a unique means to search for multi-cancer molecular targets on the genome scale. The vast majority of oncological genomics research has aimed to identify genes associated with specific cancer types.[6,7] Despite the rapidly accumulating data from such studies, the lack of a robust analytical methodology has become a bottleneck for data integration from individual cancers to identify multi-cancer molecular targets. Methods such as meta-analysis and gene network analysis have been proposed as solutions.[8,9] Although these methods are useful tools to compare and summarize lists of differentially expressed genes derived from individual cancer datasets, they are not designed to computationally validate candidate genes. Moreover, they often rely solely on expensive and labor-intensive experimental tests for large-scale screening or validation.

We hypothesized that cross-cancer predictability indicates the potential presence of common molecular mechanisms and, more importantly, predictive genes with therapeutic significance among sets of cancers. This study presents a novel machine learning approach to identify multi-cancer molecular targets based on cross-cancer predictability, with built-in independent computational validation that reduces false findings by ensuring the high relevancy of candidate targets. This approach consists of 3 key elements: (1) cross-cancer prediction to identify groups of cancers with high cross-cancer predictability and to serve as independent computational validation; (2) Prediction Analysis for Microarrays and Random Forest (PAM-RF) hybrid classification to synthesize salient genomic information from predictive models for gene inference; and (3) gene pathway enrichment analysis to

Table 1. Description of datasets.

DATASET	DESCRIPTION	TOTAL NUMBER OF SAMPLES
GDS4824	Malignant and benign prostate tissues	21
GDS4794	Small cell lung cancer (SCLC)	65
GDS4589	Stage I endometrial cancer	103
GDS4382	Colorectal cancer	34
GDS4103	Pancreatic ductal adenocarcinoma	78
GDS4102	Pancreatic tumor	52
GDS3837	Non-small cell lung carcinoma in female non-smokers	120
GDS3341	Nasopharyngeal carcinoma	41
GDS2635	Invasive ductal and lobular breast carcinomas	30
GDS2250	Basal-like and non-basal-like breast cancer tumors	47
GDS1732	Papillary thyroid cancer	14
GDS3853	Mammary gland ductal carcinoma in situ	19
GDS2609	Early-onset colorectal cancer	22
GDS1439	Prostate cancer progression	19

facilitate a biological understanding of the potential molecular mechanisms underlying cross-cancer predictability. This framework can be generalized to all types of cancers and different experimental genomics platforms. As an application, this framework was used to identify common molecular targets for stage I endometrial cancer, mammary gland ductal carcinoma, and small cell lung cancer.

Methods

Dataset description

This study used 14 publicly available cancer microarray gene expression datasets downloaded, in the format of SOFT file, from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus database (Table 1). The datasets originated from the Affymetrix GeneChip Human Genome U133 Plus 2.0 Array platform (GPL570). The samples were classified by sample status (normal vs cancerous). The data in the SOFT file was preprocessed and normalized by the submitting laboratory.

Cross-cancer prediction: overall framework

The cross-cancer prediction approach was developed to discover key molecular targets that are significant for multiple cancers based on cross-cancer predictability. Cross-cancer predictability is defined as the ability of a genomics classifier, derived from cancer A, to predict cancer B. We hypothesized that cross-cancer predictability is based on the expression of common genes shared by multiple cancers. The overall framework is illustrated in Figure 1. Cancer types were paired based on genomics-driven cross-cancer predictability. Predictive inference of salient molecular targets was conducted using the PAM-RF hybrid classification method. The individual components of this framework are elaborated on in the following sections.

Grouping cancer types based on cross-cancer predictability

Based on the hypothesis that cross-cancer predictability indicates the potential existence of common predictive genes among cancers, the goal of this step was to group individual cancers based on cross-predictability. This was executed by applying the cross-cancer prediction method to evaluate the predictability of a cancer-specific classifier, built from the genomic dataset of 1 cancer, on different cancer datasets, in other words applying the cancer A–specific classifier to the datasets of cancers B, C, D, etc. The method was twofold. First, cancer-specific classifiers were constructed from single-cancer gene expression training sets via the Prediction Analysis for Microarrays (PAM) method, which employs the nearest shrunken centroid method.[10] Performance of the cancer-specific classifier was evaluated with fivefold cross-validation. Subsequently, cross-prediction was conducted through the application of each cancer-specific classifier to independent test datasets of other cancers. The cancer-specific classifiers were directed to predict the clinical status of the tumor samples (cancerous vs normal). Cross-cancer predictability, measured by sensitivity, specificity, and receiver operating characteristic (ROC) curves, was used to determine cancer subset grouping for further analysis.

Key gene inference: PAM-RF hybrid classifier

Gene inference, conducted on the cancer subsets identified via cross-cancer predictability from the previous step, aimed to discover genes common across multiple cancers. PAM tends to generate large quantities of genes, which may cause further experimental validation to be costly and time-consuming. To improve the gene inference method, we developed a 2-layer cascading hybrid classifier. In our implementation, PAM was deployed on the lower layer, whereas Random Forest (RF)[11] was deployed on the upper layer and concatenated to PAM. This hybrid classifier provides 2 functions (Figure 2):
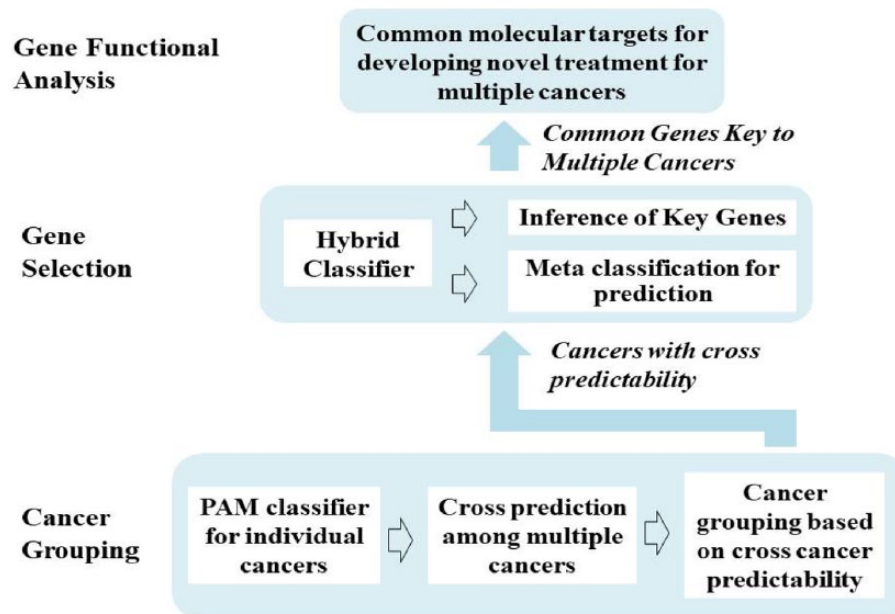
Figure 1. Schema of the cross-cancer prediction approach.
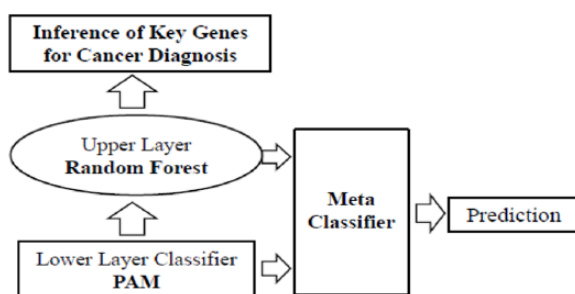PAM: Prediction Analysis for Microarrays.

Figure 2. Schema of the 2-layer hybrid classifier with PAM as the lower layer and RF as the upper layer.
PAM: Prediction Analysis for Microarrays; RF: Random Forest.

1. *Gene inference.* PAM performs the first round of feature selection on the entire set of 54 000+ probesets at the first stage. The genes selected from PAM were fed to RF to further infer key genes.
2. *Cancer prediction.* Both PAM and RF serve as base classifiers. Meta-classification is conducted by integrating the predictive probabilities from the base classifiers. The median score from PAM and RF is calculated for each sample; performance of meta-classification is evaluated with the ROC curve.

Gene pathway enrichment analysis

Identification of the common biological pathways can facilitate an understanding of the genomic basis of cross-cancer predictability. Through PAM's automatic gene selection feature, a pool of genes was generated from cancer-specific PAM classifiers. Gene pathway enrichment analysis was then performed on the selected genes to identify pathways that were expressed more frequently than expected by chance, with the cutoff

P-value < .05.[12] The Kyoto Encyclopedia of Genes and Genomes (KEGG)[13] was used as the molecular pathway database for enrichment analysis. Subsequently, we created a matrix representing the absence/presence (coded as 0/1) of the identified gene pathways across all related cancers (rows represent gene pathways and columns represent cancer types). Two-dimensional hierarchal clustering was conducted to identify patterns among clustered pathways and cancer subsets in the pathway matrix.

Results

Cross-cancer prediction: connecting cancers based on cross-predictability

We first explored whether multiclass PAM, an established machine learning technique, was a viable option for cross-cancer prediction. Multiclass PAM was applied to the 14 pooled cancer datasets to evaluate classification performance and gene selection. Cross-validation results showed a high classification error rate (>50%). In addition, the multiclass classifier was unable to effectively select signature genes from the input set of 4220 probesets (Figure 3). These results suggest that the direct use of multiclass PAM may not be suitable for cross-prediction on a large number of classes, partly because different cancer datasets may express varying levels of cross-cancer predictability. To overcome this challenge, we developed a pairwise cross-cancer prediction approach to identify groups of cancers with high cross-predictability.

We used the binary-class PAM algorithm to build cancer-specific classifiers based on the 14 individual datasets, with fivefold cross-validation for performance assessment. The cancer-specific classifiers were then applied to the test datasets of other cancers for cross-cancer prediction. The prediction
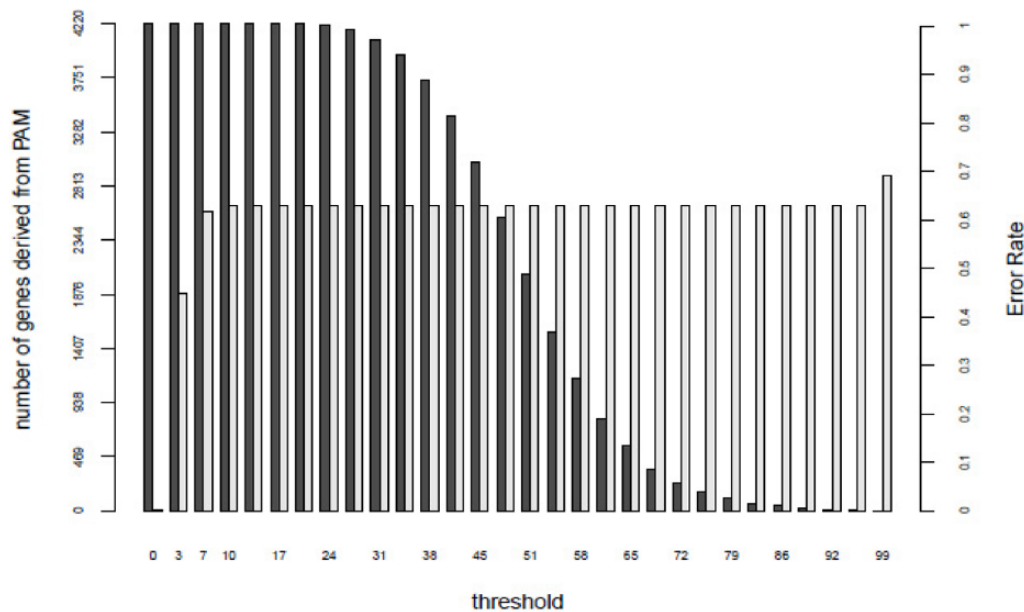
Figure 3. Classification and gene selection performance of multiclass PAM on the 14 cancer datasets. The direct application of multiclass PAM was unable to select signature genes at an acceptable classification error rate across all thresholds. Note that the input combined dataset for PAM was composed of 4220 probesets that were filtered from approximately 54 000 probesets through non-specific gene filtering.
PAM: Prediction Analysis for Microarrays.

performance, measured by area under curve (AUC) of ROC, was summarized in the heat map for all cancer pairs; clusters of cancers with high cross-predictability were highlighted (Figure 4). It was noted that the pattern of high cross-predictability does not appear to be symmetric across the heat map. This is likely caused by the imbalanced distribution of salient predictive genes between cancers. Nevertheless, because the datasets were mutually independent, cross-cancer prediction was robust and unaffected by issues such as model overfitting.

The cross-prediction heat map was used to inform the grouping of individual cancers with high cross-prediction performance. For example, the classifier C3837, built from the training set of non-small cell lung carcinoma in female non-smokers, demonstrated high cross-cancer predictability on test sets P2635 (invasive ductal and lobular breast carcinomas), P3853 (mammary gland ductal carcinoma in situ), P4589 (stage I endometrial cancer), and P4794 (small cell lung cancer) (Figure 4). ROC analysis indicated that the classifier built from the dataset of non-small cell lung carcinoma was capable of predicting other types of cancer with high sensitivity and specificity (Figure 5).

In another cluster, the classifier C4589, built from the training set of stage I endometrial cancer, showed high cross-cancer predictability on the test sets P4794 (small cell lung cancer) and P3853 (mammary gland ductal carcinoma in situ). Further analysis of this tri-cancer cluster was conducted in section "PAM-RF hybrid classifier" to infer shared key genes. The results demonstrate that cross-cancer predictability can be used to effectively identify multiple subsets of cancers with the underlying genomic relationships, as indicated by cross-predictability.

Gene pathways shared by multiple cancers

To understand the potential biological mechanisms of cross-cancer predictability, we conducted pathway enrichment analysis on the genes identified from the cancer-specific PAM classifier gene selection. The resulting heat map revealed several clusters of cancers that shared gene pathways (Figure 6). Among them, a cluster of 3 cancer datasets with high cross-predictability (including mammary gland ductal carcinoma in situ, non-small cell lung carcinoma in female non-smokers, and invasive ductal and lobular breast carcinomas) appears to share 2 common pathways. One is the focal adhesion pathway, which is a key determinant for the regulation of cancer cell migration. Previous research suggests that this pathway is related to breast cancer,[14] lung cancer,[15] and pancreatic cancer.[16] The second pathway is based on extracellular matrix (ECM) receptor interaction, which is known to push the progression of cancer cells along the metastatic cascade.[17] Overall, these results suggest that common biological pathways exist in cancer datasets with high cross-predictability.

PAM-RF hybrid classifier: key gene inference for developing treatments targeting multiple cancers

Referencing the cancer clusters generated from cross-cancer prediction, we identified shared gene expression signatures that hold high predictive power for multiple cancers. Because PAM often selects large numbers of genes, we developed a 2-layer PAM-RF hybrid classifier focused on gene selection and inference. PAM-selected genes were inputted into the RF layer for further selection. The final list of genes was inferred from RF, which was then used to reconstruct the RF classifier. The
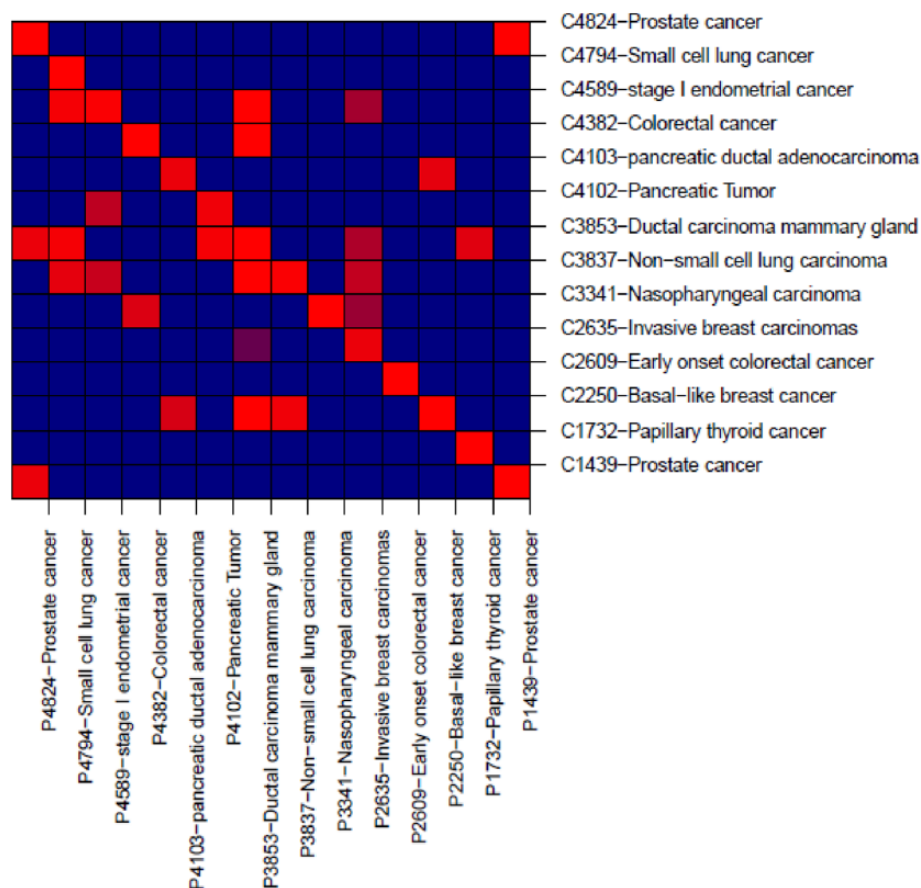
Figure 4. Cross-cancer prediction performance measured by AUC ROC, revealing cross-predictability among different cancers. Rows represent cancer-specific classifiers built from individual training datasets; columns represent test datasets from different types of cancers. The color scale indicates AUC ROC, a measure of prediction performance. Areas in red represent high-accuracy prediction among a dataset pair; areas in blue represent low-accuracy cross-prediction.
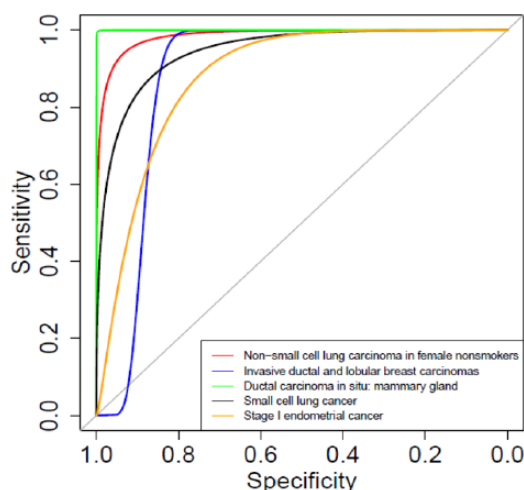

Figure 5. Example of cross-cancer prediction with performance assessed by ROC curves. The prediction classifier was developed from the training dataset C3837 (non-small cell lung carcinoma in female non-smokers). The cancer-specific classifier was then applied to the following independent test datasets: P2635 (invasive ductal and lobular breast carcinomas), P3853 (mammary gland ductal carcinoma in situ), P4589 (stage I endometrial cancer), and P4794 (small cell lung cancer). The training set P3837 (non-small cell lung carcinoma in female non-smokers) was included as a cross-application reference set. ROC: receiver operating characteristic.

hybrid classifier prediction method was based on meta-classification, which combined the predictive information of PAM and RF to improve cross-cancer predictive performance.

As depicted in the pairwise prediction heat map (Figure 4), the classifier C4589, built from the training set of stage I endometrial cancer, demonstrated high cross-cancer predictability on the test sets P4794 (small cell lung cancer) and P3853 (mammary gland ductal carcinoma in situ). Furthermore, to test the hybrid classifier's gene inference capabilities, we trained the PAM-RF hybrid classifier with the C4589 (stage I endometrial cancer) training dataset. The lower layer PAM classifier selected 66 genes, which were subsequently used as the input data for the upper layer RF classifier. The RF classifier further selected 6 candidate genes, which included dual specificity phosphatase 1 (DUSP1), transient receptor potential cation channel subfamily C member 1 (TRPC1), isocitrate dehydrogenase 2, mitochondrial (IDH2), alcohol dehydrogenase iron containing 1 (ADHFE1), histamine N-methyltransferase (HNMT), and mitochondrial calcium uptake family member 3 (MICU3). The RF classifier was then reconstructed using the 6 genes. The 6-gene PAM-RF hybrid classifier outperformed the PAM classifier informed by the same 6 genes (Figure 7).
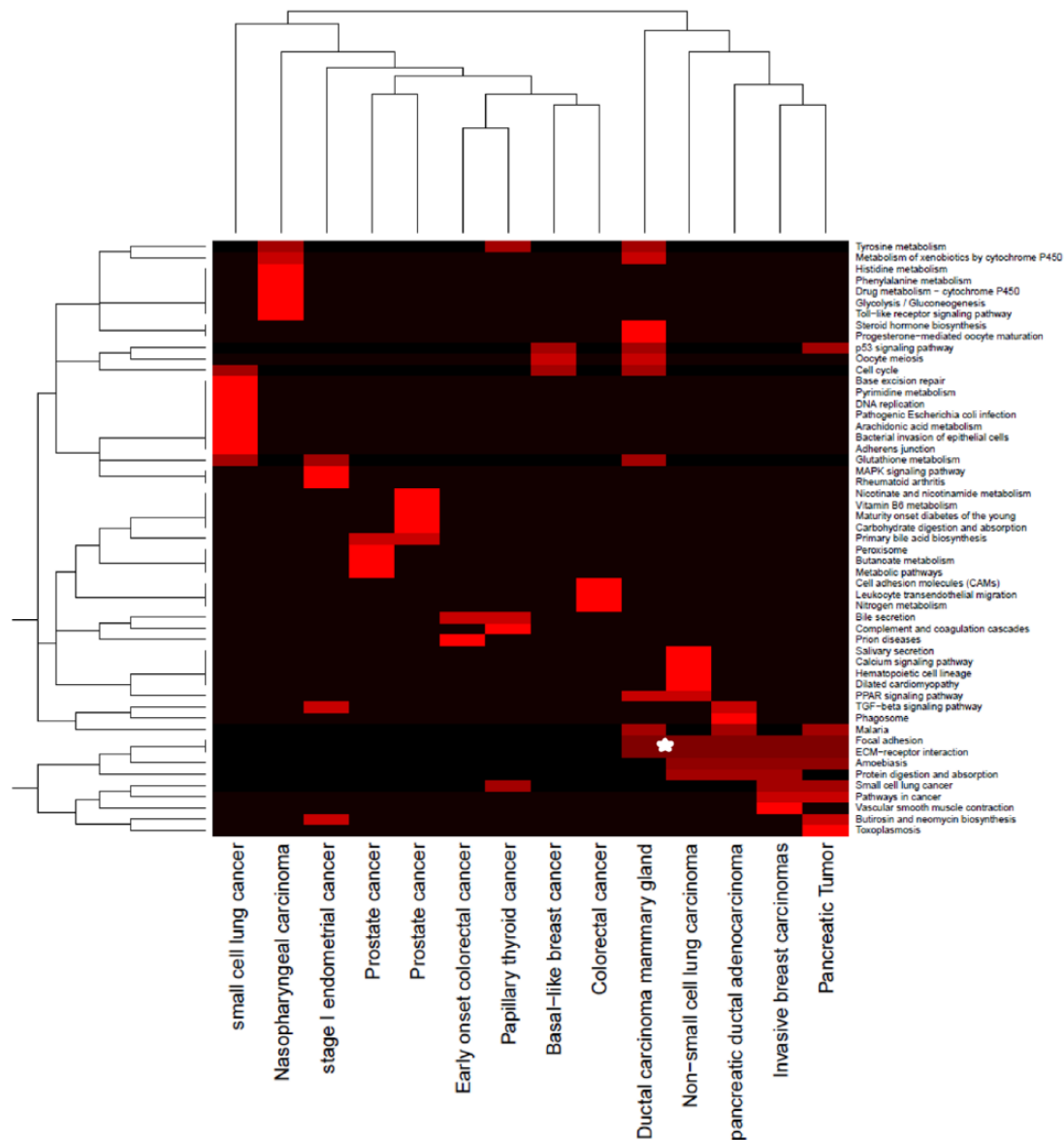
Figure 6. Patterns of KEGG pathways among different cancers identified using 2-dimensional hierarchical clustering. The horizontal dimension represents the cancer types and the vertical dimension represents the KEGG pathways identified from the cancer-specific signature genes. The clustering was based on the absence or presence of KEGG pathways in cancer types. The purpose of the clustering was to explore potential commonalities in KEGG pathways among different cancer types, which we believe is relevant to cross-cancer predictability. The figure reveals that a cluster of 3 cancer datasets with high cross-predictability (composed of mammary gland ductal carcinoma in situ, non-small cell lung carcinoma in female non-smokers, and invasive ductal and lobular breast carcinomas) shares 2 pathways: focal adhesion pathway and extracellular matrix (ECM) receptor interaction. In addition, the datasets from pancreatic cancer and pancreatic ductal adenocarcinoma share these 2 pathways. The cluster is indicated by the white star. KEGG: Kyoto Encyclopedia of Genes and Genomes; PPAR: peroxisome proliferator-activated receptor; TGF: transforming growth factor.

We searched the biomedical literature to better understand the biological functions of the 6 inferred genes across the 3 cancers. Based on existing experimental evidence, the genes DUSP1 and TRPC1 are linked to all 3 cancers; IDH2 is associated with mammary gland ductal carcinoma in situ and small cell lung cancer; ADHFE1, HNMT, and MICU3 are linked to mammary gland ductal carcinoma in situ (Table 2). These findings support the biological carcinogenic relevance of the identified genes.

Discussion and Conclusions

This study presented a novel machine learning approach—cross-cancer prediction—to discover shared multi-cancer

molecular targets that are crucial to the carcinogenesis and progression of several cancers. We have demonstrated the capabilities of cross-cancer prediction in identifying groups of cancers with the underlying biological connections that inform cross-cancer predictability. Furthermore, a novel hybrid classifier, integrating PAM and RF classification models, has been developed to refine gene selection and improve prediction performance. As an application, this approach has successfully identified key genes shared by endometrial cancer, mammary gland ductal carcinoma, and small cell lung cancer. The resulting molecular target candidates were supported by high cross-cancer predictability analysis results, as
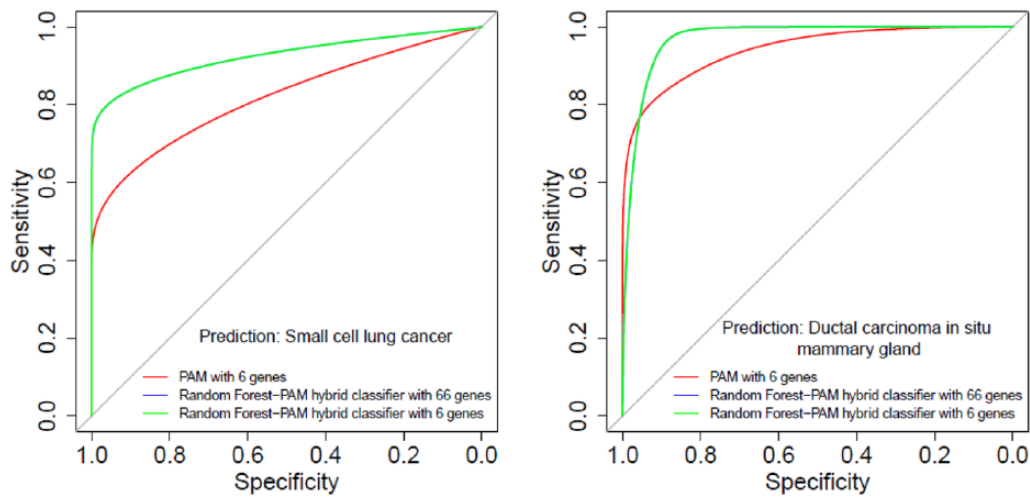
Figure 7. The PAM-RF classifier informed with 6 candidate genes, built from the C4589 training set (stage I endometrial cancer), showed higher prediction performance on P4794 (small cell lung cancer, left panel) and P3853 (mammary gland ductal carcinoma in situ, right panel) than the PAM classifier informed with the same 6 genes. Performance was measured by ROC.
PAM-RF: Prediction Analysis for Microarrays and Random Forest; PAM: Prediction Analysis for Microarrays; ROC: receiver operating characteristic.

Table 2. Biological function of the 3 key genes shared by stage I endometrial cancer, ductal carcinoma in situ, and small cell lung cancer.

GENE	STAGE I ENDOMETRIAL CANCER	DUCTAL CARCINOMA IN SITU: MAMMARY GLAND	SMALL CELL LUNG CANCER
Dual specificity phosphatase 1 (DUSP1)	DUSP1 deficiency promotes endometrial cancer progression via the MAPK/ERK pathway, suggesting that DUSP1 may serve as a potential therapeutic target for the treatment of endometrial cancer.[18]	DUSP1 is a key downstream target of HER2 in breast cancer cells and can prevent apoptotic induction by limiting the accumulation of phosphorylated active forms of the stress kinase JNK.[19]	DUSP1 in lung cancer cells contributes to tumor growth, tumor invasion, and angiogenesis.[20]
Transient receptor potential cation channel, subfamily C, member 1 (TRPC1)	Upregulation of TRPC1 and consequent enhancement of SOC-mediated Ca^{2+} influx play a crucial role via p-CREB-mediated transcription, to which the activation of FOXO1 might contribute. This finding may provide a new therapeutic strategy for endometrial cancers.[21]	TRPC1 is a differential regulator of hypoxia-mediated events and Akt signaling in PTEN-deficient breast cancer cells.[22]	siRNA-mediated TRPC1 depletion in non-small cell lung carcinoma cell lines induced G_0/G_1 cell cycle arrest, resulting in a dramatic decrease in cell growth.[23]
Isocitrate dehydrogenase 2 (NADP+), mitochondrial (IDH2)	To be determined.	Knockdown of IDH2 markedly decreased intracellular onco-metabolite 2-hydroxyglutarate in breast cancer cells with aberrant 2HG accumulation.[24]	Functional IDH2 genetic variant is associated with the risk of lung cancer.[25]

well as existing experimental evidence published by other researchers.

The cross-cancer prediction solution addresses the shortcomings of existing methods for analyzing common genes among cancers. First, the direct use of multiclass PAM classification performed poorly in cancer prediction when a large number of cancer types/classes were included. This is likely because cross-cancer predictability may only be present in specific subsets, not consistently across all cancer types. To solve this problem, we developed the cross-cancer pairwise prediction method, which effectively identified cancer subsets with high cross-predictability. Second, meta-analysis and gene network–based approaches have been proposed to identify common biomarkers from genomics datasets of multiple cancers.[8,9] These methods focus on the comparative summarization of differentially expressed gene lists from individual cancer types, but are unable to computationally validate these candidate genes. In addition, these methods often generate a large number of genes for expensive and labor-intensive experimental validation. Our approach resolves the aforementioned problems using built-in computational validation—cross-cancer prediction based on independent datasets—to ensure that classifiers with shared discriminative genes hold high predictive power for multiple cancer types. Therefore, the candidate molecular targets identified from this approach are chosen based on biological commonalities as well as novel data-driven cross-cancer predictability.

Although this study demonstrated promising results, there are a few external limitations. Tumor pathogenesis is complex and multifaceted, and may involve factors such as gene–environment interaction and stochastic processes. Hence, this approach needs to be further developed to adapt to these complex scenarios. In addition, more validation efforts are needed to evaluate the potential impact of organ/tissue type variability and heterogeneous sample processing methods. Nevertheless, this approach is generalizable to all cancer types and different genomic platforms. In the future, we will expand the approach to other genomic platforms, like RNA-Seq and Proteomics data.

This research holds significant translational potential in clinical oncology and pharmaceutical development for new cancer drugs. The proposed approach will open up new opportunities to discover molecular targets that are salient to the pathogenesis, progression, and metastasis of multiple cancers. Medication targeted towards multiple cancers is an important scientific and medical breakthrough that will reshape the landscape of clinical oncology. It provides new opportunities to treat patients of different cancers who share genetic predisposition but respond poorly to organ-specific treatment. In the world of pharmaceutical development, it will also potentially mitigate costs and streamline the cancer drug development cycle.

Author Contributions

KG, DW, and YH conceived and designed the experiments. KG analyzed the data. KG wrote the first draft of the manuscript. KG, DW, and YH contributed to the writing of the manuscript, agreed with manuscript results and conclusions, jointly developed the structure and arguments for the paper, made critical revisions, and approved the final version. All authors reviewed and approved the final manuscript.

Disclosures and Ethics

As a requirement of publication, the author(s) have provided to the publisher signed confirmation of compliance with legal and ethical obligations including but not limited to the following: authorship and contributorship, conflicts of interest, privacy and confidentiality, and (where applicable) protection of human and animal research subjects. The authors have read and confirmed their agreement with the ICMJE authorship and conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material. Any disclosures are made in this section. The external blind peer reviewers report no conflicts of interest.

REFERENCES

1. World Health Organization. Cancer: Fact Sheet. 2017. http://www.who.int/news-room/fact-sheets/detail/cancer.
2. Bolton KL, Chenevix-Trench G, Goh C, et al. Association between BRCA1 and BRCA2 mutations and survival in women with invasive epithelial ovarian cancer. *JAMA*. 2012;307:382–390.
3. Bordeleau L, Panchal S, Goodwin PJ. Prognosis of BRCA related breast cancer: a summary of evidence. *Breast Cancer Res Treat*. 2010;119:13–24.
4. Kar SP, Beesley J, Olama AAA, Michailidou K. Genome-wide meta-analyses of breast, ovarian and prostate cancer association studies identify multiple new susceptibility loci shared by at least two cancer types. *Cancer Discover*. 2016;6:1052–1067.
5. Food and Drug Administration. FDA approves first cancer treatment for any solid tumor with a specific genetic feature. *FDA Press Release*. May 23, 2017. https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm560167.htm.
6. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J*. 2015;13:8–17.
7. Ow GS, Kuznetsov VA. Big genomics and clinical data analytics strategies for precision cancer prognosis. *Sci Rep*. 2016;6:36493.
8. Pihur V, Datta S, Datta S. Finding common genes in multiple cancer types through meta-analysis of microarray experiments: a rank aggregation approach. *Genomics*. 2008;92:400–403.
9. Martinez-Ledesma E, Verhaak RG, Treviño V. Identification of a multi-cancer gene expression biomarker for cancer clinical outcomes using a network-based algorithm. *Sci Rep*. 2015;5:11966.
10. Tibshirani R, Hastie T, Narasimhan B, et al. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A*. 2002;99:6567–6572.
11. Ho TK. The random subspace method for constructing decision forests. *IEEE T Pattern Anal*. 1998;20:832–844.
12. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102:15545–15550.
13. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 1999;27:29–34.
14. Bijian K, Lougheed C, Su J, et al. Targeting focal adhesion turnover in invasive breast cancer cells by the purine derivative reversine. *Br J Cancer*. 2013;109:2810–2818.
15. Mukhopadhyay NK, Gordon GJ, Chen CJ, et al. Activation of focal adhesion kinase in human lung cancer cells involves multiple and potentially parallel signaling events. *J Cell Mol Med*. 2005;9:387–397.
16. Jiang H, Hegde S, Knolhoff BL. Targeting focal adhesion kinase renders pancreatic cancers responsive to checkpoint immunotherapy. *Nat Med*. 2016;22:851–860.
17. Venning FA, Wullkopf L, Erler JT. Targeting ECM disrupts cancer progression. *Front Oncol*. 2015;5:224.
18. Yang Y, Zhou JY, Zhao LJ, et al. Dual-specificity phosphatase 1 deficiency induces endometrioid adenocarcinoma progression via activation of mitogen-activated protein kinase/extracellular signal-regulated kinase pathway. *Chin Med J (Engl)*. 2016;129:1154–1160.
19. Candas D, Lu CL, Fan M, et al. Mitochondrial MKP1 is a target for therapy resistant HER2 positive breast cancer cells. *Cancer Res*. 2014;74:7498–7509.
20. Moncho-Amor V, Ibañez de Cáceres I, Bandres E, et al. DUSP1/MKP1 promotes angiogenesis, invasion and metastasis in non-small-cell lung cancer. *Oncogene*. 2011;30:668–678.
21. Kawarabayashi Y, Hai L, Honda A, et al. Critical role of TRPC1-mediated Ca^{2+} entry in decidualization of human endometrial stromal cells. *Mol Endocrinol*. 2012;26:846–858.
22. Azimi I, Milevskiy MJG, Kaemmerer E, et al. TRPC1 is a differential regulator of hypoxia-mediated events and Akt signaling in PTEN-deficient breast cancer cells. *J Cell Sci*. 2017;130:2292–2305.
23. Tajeddine N, Gailly P. TRPC1 protein channel is major regulator of epidermal growth factor receptor signaling. *J Biol Chem*. 2012;287:16146–16157.
24. Terunuma A, Putluri N, Mishra P, et al. MYC-driven accumulation of 2-hydroxyglutarate is associated with breast cancer prognosis. *J Clin Invest*. 2014;124:398–412.
25. Li J, Lu J, He Y, et al. A new functional IDH2 genetic variant is associated with the risk of lung cancer. *Mol Carcinog*. 2017;56:1082–1087.