Clinical Cancer Research

Deep Learning–Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer

Kumardeep Chaudhary[1], Olivier B. Poirion[1], Liangqun Lu[1,2], and Lana X. Garmire[1,2]

Abstract

Identifying robust survival subgroups of hepatocellular carcinoma (HCC) will significantly improve patient care. Currently, endeavor of integrating multi-omics data to explicitly predict HCC survival from multiple patient cohorts is lacking. To fill this gap, we present a deep learning (DL)–based model on HCC that robustly differentiates survival subpopulations of patients in six cohorts. We built the DL-based, survival-sensitive model on 360 HCC patients' data using RNA sequencing (RNA-Seq), miRNA sequencing (miRNA-Seq), and methylation data from The Cancer Genome Atlas (TCGA), which predicts prognosis as good as an alternative model where genomics and clinical data are both considered. This DL-based model provides two optimal subgroups of patients with significant survival differences ($P = 7.13e-6$) and good model fitness [concordance index (C-index) = 0.68]. More aggressive subtype is associated with frequent *TP53* inactivation mutations, higher expression of stemness markers (*KRT19* and *EPCAM*) and tumor marker *BIRC5*, and activated Wnt and Akt signaling pathways. We validated this multi-omics model on five external datasets of various omics types: LIRI-JP cohort ($n = 230$, C-index = 0.75), NCI cohort ($n = 221$, C-index = 0.67), Chinese cohort ($n = 166$, C-index = 0.69), E-TABM-36 cohort ($n = 40$, C-index = 0.77), and Hawaiian cohort ($n = 27$, C-index = 0.82). This is the first study to employ DL to identify multi-omics features linked to the differential survival of patients with HCC. Given its robustness over multiple cohorts, we expect this workflow to be useful at predicting HCC prognosis prediction. *Clin Cancer Res; 24(6); 1248–59. ©2017 AACR.*

Introduction

Liver cancer is the second leading cancer responsible for the mortality in men worldwide (1). In the United States, more than 40,000 people are estimated to be diagnosed with liver cancer in 2017, according to the American Cancer Society (2). It is one of the few cancer types with increase in both incidence and mortality rates, by approximately 3% per year in the United States (3). Hepatocellular carcinoma (HCC) is the most prevalent type (70%–90%) of liver cancer. It is aggravated by various risk factors, including hepatitis B virus/hepatitis C virus (HBV/HCV) infection, nonalcoholic steatohepatitis (NASH), alcoholism, and smoking. The 5-year survival rate of HCC varies greatly across different populations, with an average rate of less than 32% (4–9). The high level of heterogeneity in HCC, along with the complex etiologic factors, makes the prognosis prediction very challenging (10, 11). Moreover, treatment strategies in HCC are very limited, imposing an additional urgent need for developing tools to predict patient survival (12).

To understand the HCC heterogeneity among patients, a considerable amount of work has been done to identify the HCC molecular subtypes (13–19). A variety of numbers of subtypes were identified, ranging from two to six, based on various omics data types, driving hypotheses and computational methods. Besides most commonly used mRNA gene expression data, a recent study integrated copy number variation (CNV), DNA methylation, mRNA, and miRNA expression to identify the five HCC molecular subtypes from 256 samples from The Cancer Genome Atlas (TCGA; ref. 20). However, most of the studies explored the molecular subtypes without relying on survival during the process of defining subtypes (21). Rather, survival information was used *post hoc* to evaluate the clinical significance of these subtypes (20). As a result, some molecular subtypes showed converging and similar survival profiles, making them redundant subtypes in terms of survival differences (16). New approaches to discover survival-sensitive and multi-omics data-based molecular subtypes are much needed in HCC research.

To address these issues, for the first time, we utilized deep learning (DL) computational framework on multi-omics HCC datasets. We chose the autoencoder framework as the implementation of DL for multi-omics integration. Autoencoders aim to reconstruct the original input using combinations of nonlinear functions that can then be used as new features to represent the dataset. These algorithms have already been proven to be efficient approaches to produce features linked to clinical outcomes (22). Autoencoders were successfully applied to analyze high-dimensional gene expression data (23, 24) and to integrate heterogeneous data (25, 26). Notably, autoencoder transformation tends to aggregate genes sharing similar pathways (27), therefore making it appealing to interpret the biological functions. The contributions of this study to the HCC field is manifested not only in its thorough and integrative computational rigor but also in its unification of the discordant molecular subtypes into robust subtypes that withstand the testing of various cohorts, even when they are in different omics forms.

[1]Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii. [2]Molecular Biosciences and Bioengineering Graduate Program, University of Hawaii at Manoa, Honolulu, Hawaii.

K. Chaudhary and O.B. Poirion contributed equally to this article.

Corresponding Author: Lana X. Garmire, University of Hawaii Cancer Center, 701 Ilalo Street, Honolulu, HI 96813. Phone: 808-441-8193; Fax: 808-586-2982; E-mail: lgarmire@cc.hawaii.edu

AACR

We derived the model from 360 HCC samples in the TCGA multi-omics cohort, which have mRNA expression, miRNA expression, CpG methylation, and clinical information. We discovered two subtypes with significant differences in survival. These subtypes hold independent predictive values on patient survival, apart from clinical characteristics. Most importantly, the two subtypes obtained from our DL framework are successfully validated in five independent cohorts, which have an miRNA or mRNA or DNA methylation dataset. Functional analysis of these two subtypes identified that gene expression signatures (*KIRT19*, *EPCAM*, and *BIRC5*) and Wnt signaling pathways are highly associated with poor survival. In summary, the survival-sensitive subtype model reported here is significant for both HCC prognosis prediction and therapeutic intervention.

Materials and Methods

Datasets and study design

In this study, we used a total of six cohorts, and the descriptions of these are detailed below. We used the TCGA data in two steps: The first step is to obtain the labels of survival-risk classes, using the whole TCGA dataset; the second is to train a Support Vector Machine (SVM) model by splitting the samples 60%/40% to training and held-out testing data (detailed in "Data partitioning and robustness assessment" subsection). We used five additional confirmation datasets to evaluate the prediction accuracy of the DL-based prognosis model.

TCGA set. We obtained multi-omics HCC data from the TCGA portal (https://tcga-data.nci.nih.gov/tcga/). We used the R package TCGA-Assembler (v1.0.3; ref. 28) and obtained 360 samples with RNA sequencing (RNA-Seq) data (UNC IlluminaHiSeq_RNASeqV2; Level 3), miRNA sequencing (miRNA-Seq) data (BCGSC IlluminaHiSeq_miRNASeq; Level 3), DNA methylation data (JHU-USC HumanMethylation450; Level 3), and the clinical information. For the DNA methylation, we mapped CpG islands within 1,500 base pairs (bp) ahead of transcription start sites (TSS) of genes and averaged their methylation values. In dealing with the missing values (preprocessing of data), three steps were performed as elsewhere (29). First, the biological features (e.g., genes/miRNAs) were removed if having zero value in more than 20% of patients. The samples were removed if missing across more than 20% features. Second, we used the *impute* function from the R impute package (30) to fill out the missing values. Third, we removed input features with zero values across all samples.

Confirmation cohort 1 (LIRI-JP cohort, RNA-Seq). A total of 230 samples with RNA-Seq data were obtained from the ICGC portal (https://dcc.icgc.org/projects/LIRI-JP). These samples belong to a Japanese population primarily infected with HBV/HCV (31). We used the normalized read count values given in the gene expression file.

Confirmation cohort 2 (NCI cohort, microarray gene expression). A total of 221 samples with survival information were chosen from GSE14520 Affymetrix high-throughput GeneChip HG-U133A microarray dataset from an earlier study of patients with HCC (32). This is a Chinese population primarily associated with HBV infection. Log₂ Robust Multi-array Average (RMA)–calculated signal intensity values provided by the authors were used for analysis.

Confirmation cohort 3 (Chinese cohort, miRNA expression array). A total of 166 pairs of HCC/matched noncancerous normal tissue samples were downloaded, with CapitalBio custom Human miRNA array data (GSE31384; ref. 33). As the data were already log₂ transformed, we used unit-scale normalization.

Confirmation cohort 4 (E-TABM-36, gene expression microarray) Forty HCC samples were used, with survival information and transcriptional profiling from Affymetrix HG-U133A GeneChip arrays platform (16). We used the CHPSignal values for the further processing as a measure of gene expression.

Confirmation cohort 5 (Hawaiian cohort, DNA methylation array). Twenty-seven samples were used, with genome-wide methylation profiling from Illumina HumanMethylation450 BeadChip platform (34). Probe-to-gene conversion was done the same way as for the TCGA HCC methylation data.

All the available clinical information for the confirmation cohorts is listed in Supplementary Table S1. These cohorts were used to test the SVM-based machine-learning models.

Transformed features using a DL framework

We used the three preprocessed TCGA HCC omics datasets of 360 samples as the input for the autoencoders framework. We stacked the three matrices that are unit-norm scaled by sample to form a unique matrix as reported before (35). An autoencoder is an unsupervised feed-forward, nonrecurrent neural network (36). Given an input layer taking the input $x = (x_1, \ldots, x_n)$ of dimension n, the objective of an autoencoder is to reconstruct x by the output x' (x and x' have the same dimension) via transforming x through successive hidden layers. For a given layer i, we used *tanh* as activation function between input layer x and output layer y. That is:

$$y = f_i(x) = \tanh(W_i.x + b_i)$$

Where x and y are two vectors of size d and p, respectively, and Wi is the weight matrix of size $p \times d$, b_i an intercept vector of size p and $W_i.x$ gives a vector of size p. For an autoencoder with k layers, x' is then given by:

$$x' = F_{1 \rightarrow k}(x) = f_1 {}^{\circ} \ldots {}^{\circ} f_{k-1} {}^{\circ} f_k(x)$$

Where $f_{k-1}{}^{\circ}f_k(x) = f_{k-1}(f_k(x))$ is the composed function of f_{k-1} with f_k. To train an autoencoder, the objective is to find the different weight vectors W_i minimizing a specific objective function. We chose *logloss* as the objective function, which measures the error between the input x and the output x':

$$logloss(x, x') = \sum_{k=1}^{d} (x_k \log(x'_k) + (1 - x_k) \log(1 - x'_k))$$

To control overfitting, we added an *L1* regularization penalty α_w on the weight vector W_i, and an *L2* regularization penalty α_a on the nodes activities: $F_{1 \rightarrow k}(x)$. Thus, the objective function above becomes:

$$L(x, x') = logloss(x, x') + \sum_{i=1}^{k} \left(\alpha_w ||W_i||_1 + \alpha_a ||F_{1 \rightarrow i}(x)||_2^2 \right)$$

We implemented an autoencoder with three hidden layers (500, 100, and 500 nodes, respectively) using the Python Keras library (https://github.com/fchollet/keras). We used the bottleneck layer of the autoencoder to produce new features from

the omics data. The values α_a and α_w were set to 0.0001 and 0.001. Finally, to train the autoencoder, we used the gradient descent algorithm with 10 epochs and 50% dropout. Here, epoch means the iteration of the learning algorithm (stochastic gradient descent) through the entire training dataset. During one epoch, the learning algorithm processes each instance of training data once.

Transformed feature selection and K-means clustering

The autoencoder reduced the initial number of features to 100 new features obtained from the bottleneck layer. Next, for each of these transformed features produced by the autoencoder, we built a univariate Cox proportional hazards (Cox-PH) model and selected features from which a significant Cox-PH model is obtained (log-rank $P < 0.05$). We then used these reduced new features to cluster the samples using the K-means clustering algorithm. We determined the optimal number of clusters with two metrics: Silhouette index (37) and Calinski–Harabasz criterion (38). We used the *scikit-learn* package for K-means implementation (39).

Data partitioning and robustness assessment

We used a cross-validation (CV)–like procedure to partition the TCGA dataset as follows: We used a 60%/40% split (training/test sets) of the TCGA data to have sufficient number of test samples that generate evaluation metrics. We first randomly split the 360 samples from TCGA into 5 folds. We then used 2 of the 5 folds as the test set, and the remaining 3 folds as the training set. With this approach, we obtained 10 new combinations (folds). For each of these 10 new folds, we constructed a model using the 60% samples (training set) and predicted the labels in test set (held-out). This data partitioning was only used to assess the robustness of the model. For each training fold, a distinct autoencoder and a classifier (see below) were built to predict the labels of the test fold. The labels of the TCGA samples are finally inferred using an autoencoder built with all the samples, and these labels were used for the prediction of the confirmation datasets.

Supervised classification

After obtaining the labels from K-means clustering, we built a supervised classification model(s) using the SVM algorithm. We normalized each omics layer in the training set and then selected the top N features that are most correlated with the cluster labels (obtained from K-means) based on ANOVA F values. We set default N values as 100 for mRNAs, 50 for methylation, and 50 for miRNAs.

To predict on TCGA 3-omics held-out test data, we built an SVM classifier from a combination of the top 100 mRNAs, 50 methylation, and 50 miRNA features selected by ANOVA. To predict each of the other five confirmation cohorts used in this study, we built an SVM classifier on each omics type, using the corresponding top 100 mRNAs, 50 methylation, or 50 miRNA features selected by ANOVA, respectively. For a confirmation cohort from a specific omic layer, we first selected common features (mRNAs, genes with CpG sites, or miRNAs) between this cohort and the corresponding omic layer in the TCGA training set. Specifically, the common features between the five cohorts and the TCGA training dataset are 14,634 for the LIRI-JP cohort; 9,311 for the NCI cohort; 174 for the Chinese cohort; 10,550 for the E-TABM-36 cohort; and 19,883 for the Hawaiian cohort.

We then applied two scaling steps on both the training set and confirmation cohort samples. We first used a median scaling on

both the training set and the new test samples, where each feature is rescaled according to its median and absolute median deviation. This approach was used to normalize samples from RNA-Seq data previously (40). For mRNA and DNA methylation data, we then applied a robust scaling on the training set and confirmation samples using the means and the standard deviations of the training set (41). For miRNA confirmation data, we applied the unit-scale normalization for both the miRNA training and the confirmation cohort. When predicting a single sample, an alternative rank normalization (rather than robust or unit-scale normalization) can be applied to both the new sample and samples from the training sets (see more details in Supplementary File S1).

We used the *scikit-learn* package to perform grid search to find the best hyperparameters of the SVM model(s) using 5-fold CV and built SVM models.

Evaluation metrics for models

The metrics used closely reflect the accuracy of survival prediction in the subgroups identified. Three sets of evaluation metrics were used.

Concordance index. The concordance index (C-index) can be seen as the fraction of all pairs of individuals whose predicted survival times are correctly ordered (42) and is based on Harrell C statistics (43). A C-index score around 0.70 indicates a good model, whereas a score around 0.50 means random background. To compute the C-index, we first built a Cox-PH model using the training set (cluster labels and survival data) and then predicted survival using the labels of the test/confirmation set. We then calculated the C-index using function concordance.index in the R *survcomp* package (44). To compute the C-index using the multiple clinical features, we built a Cox-PH using the *glmnet* package instead (45). We opted to perform penalization through ridge regression rather than the default Lasso penalization. Before building the Cox-PH model, we performed a 10-fold CV to find the best lambda.

Log-rank P value of Cox-PH regression. We plotted the Kaplan–Meier survival curves of the two risk groups and calculated the log-rank P value of the survival difference between them. We used the Cox-PH model for survival analysis (46), similar to described before (47, 48), using the R *survival* package (49).

Brier score. It is another score function that measures the accuracy of probabilistic prediction (50). In survival analysis, the Brier score measures the mean of the difference between the observed and the estimated survival beyond a certain time (51). The score ranges between 0 and 1, and a larger score indicates higher inaccuracy. We used the implementation of the Brier score from the R *survcomp* package.

Alternative approaches to the DL framework

We compared the performances of the DL framework with two alternative approaches. In the first approach, we performed principal component analysis (PCA) and used the same number (100) of principal components as those features in the bottleneck layer of Fig. 1. We then identified the subset (13) of PCA features significantly associated with survival using univariate Cox-PH models, using the same procedure as the Cox-PH step in Fig. 1. In the second approach, we selected the top 37 features among
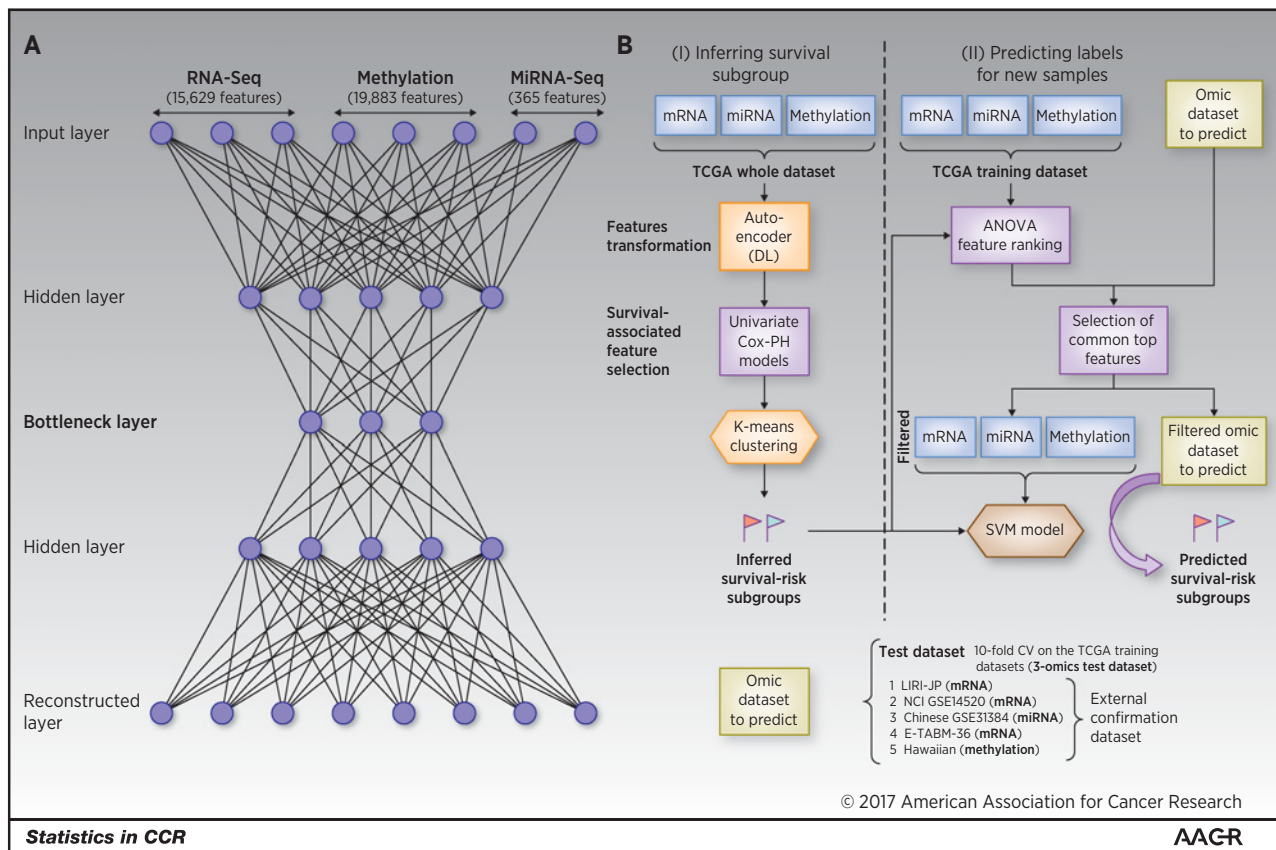
Figure 1.
Overall workflow. **A,** Autoencoder architecture used to integrate 3 omics of HCC data. **B,** Workflow combining DL and machine learning techniques to predict HCC survival subgroups. The workflow includes two steps: step 1, inferring survival subgroups; step 2, predicting risk labels for new samples. In step 1, mRNA, DNA methylation, and miRNA features from the TCGA HCC cohort are stacked up as input features for the autoencoder, a DL method; then each of the new, transformed features in the bottleneck layer of the autoencoder is then subjected to single variate Cox-PH models to select the features associated with survival; then K-mean clustering is applied to samples represented by these features to identify survival-risk groups. In step 2, mRNA, methylation, and miRNA input features are ranked by ANOVA test *F* values, those features that are in common with the predicting dataset are selected, then the top features are used to build an SVM model(s) to predict the survival-risk labels of new datasets.

all 3 omics features using single-variant Cox-PH models based on the C-index scores. We clustered the samples using the same K-means procedure as in Fig. 1.

Functional analysis

A number of functional analyses were performed to understand the characteristics of two survival-risk subtypes of TCGA HCC samples.

TP53 mutation analysis. We analyzed the somatic mutation frequency distributions in the survival subtypes for the *TP53* gene among the TCGA and LIRI-JP cohorts. The TCGA and LIRI-JP cohorts have exome sequencing and whole genome sequencing data for 186 and 230 samples with survival data, respectively. We performed Fisher test on *TP53* mutation between two survival-risk groups.

Clinical covariate analysis. We tested the associations of our identified subtypes with other clinical factors, including gender, race, grade, stage, and risk factors, using Fisher exact tests. To test whether the two survival-risk subtypes have prognostic values in

addition to clinical characteristics, we built a combined Cox-PH model with survival-risk classification and clinical data, and compared it to the one with only clinical data (stage, grade, race, gender, age, and risk factor).

Differential expression. To identify the differentially expressed genes between the two survival-risk subtypes, we performed the differential expression analysis for the mRNA, miRNA expression, and methylation genes. We used the *DESeq2* package (52) to identify the differential gene and miRNA expression between the two subtypes (false discovery rate, or FDR, < 0.05). In addition, we used $\log_2$ fold change greater than 1 as filtering for mRNA/miRNA. For methylation data, we transformed the beta values into *M* values as elsewhere (53, 54) using the *lumi* package in R (55). We fit the linear model for each gene using the *lmFit* function followed by the empirical Bayes method, using the *limma* package in R (56). It uses moderate *t* tests to determine significant difference in methylation for each gene between S1 and S2 subtypes (Benjamini–Hochberg corrected $P < 0.05$). In addition, we used averaged *M* value differences greater than 1 as filtering. We used volcano plot to show the differentially methylated genes in two subtypes.

Enriched pathway analysis. We used upregulated and down-regulated genes for the KEGG pathway analysis, using the functional annotation tool from the online DAVID interface (57, 58). We used the modified Fisher exact test *P* value (EASE score provided by DAVID) threshold of 0.10 to consider a pathway significant. We plotted the gene pathway network using Gephi (59).

Results

Two differential survival subtypes are identified in TCGA multi-omics HCC data

From the TCGA HCC project, we obtained 360 tumor samples that had coupled RNA-Seq, miRNA-Seq, and DNA methylation data. For these 360 samples, we preprocessed the data as described in the "Materials and Methods" section and obtained 15,629 genes from RNA-Seq, 365 miRNAs from miRNA-Seq, and 19,883 genes from DNA methylation data as input features. These three types of omics features were stacked together using an autoencoder, a DL framework (36). The architecture of the autoencoder is shown

in Fig. 1A. We used the activity of the 100 nodes from the bottleneck hidden layer as new features. We then conducted univariate Cox-PH regression on each of the 100 features and identified 37 features significantly (log-rank $P < 0.05$) associated with survival. These 37 features were subjective to K-means clustering, with cluster number K ranging from 2 to 6. Using silhouette index and the Calinski–Harabasz criterion, we found that K = 2 was the optimum with the best scores for both metrics (Supplementary Fig. S1A). Furthermore, the survival analysis on the full TCGA HCC data shows that the survivals in the two subclusters are drastically different (log-rank $P = 7.13e-6$; Fig. 2A). Moreover, K = 2 to 6 yielded Kaplan–Meier survival curves that essentially represent two significantly different survival groups (Supplementary Fig. S1B). Thus, we determined that K = 2 was the optimal number of classes for the subsequent supervised machine learning processes.

We next used the two classes determined above as the labels to build a classification model using the SVM algorithm with CV (Fig. 1B). We split the 360 TCGA samples into 10 folds using a 60/40 ratio for training and test data. We chose a 60/40 split rather
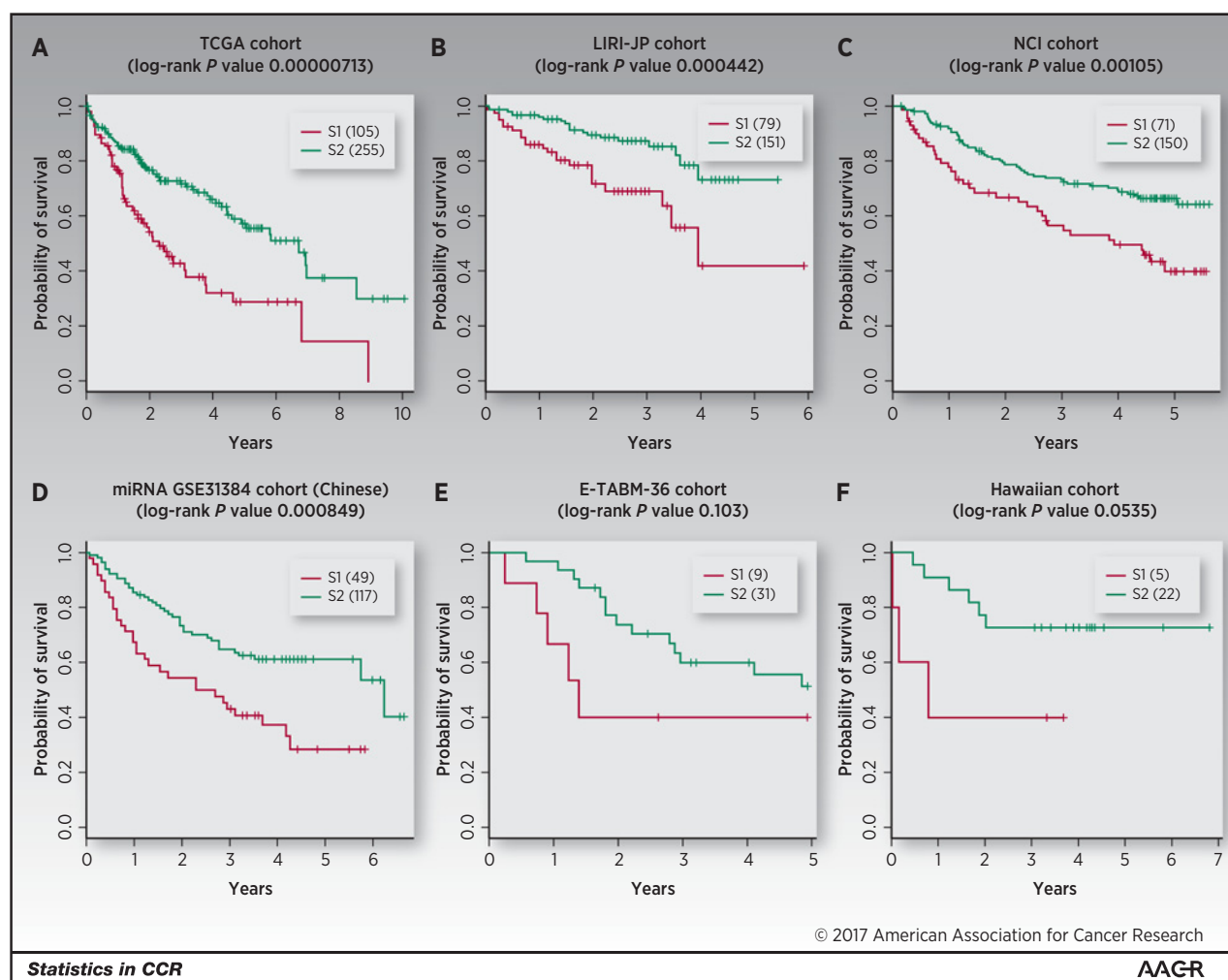

Figure 2.
Significant survival differences for the TCGA and external confirmation cohorts: TCGA cohort (**A**), LIRI-JP cohort (**B**), NCI cohort (**C**), Chinese cohort (**D**), E-TABM-36 cohort (**E**), and Hawaiian cohort (**F**).

Table 1. CV-based performance robustness of the SVM classifier on training and test set in TCGA cohort

Dataset	10-fold CV	C-index	Brier score	Log-rank *P* (geo. mean)
Training	3-omics training (60%)	0.70 (±0.04)	0.19 (±0.01)	0.001
Test	3-omics test (40%)	0.69 (±0.08)	0.20 (±0.02)	0.005
	RNA only	0.68 (±0.07)	0.20 (±0.02)	0.01
	miRNA only	0.69 (±0.07)	0.20 (±0.02)	0.003
	Methylation only	0.66 (±0.07)	0.20 (±0.02)	0.031

Abbreviation: geo., geometric.

than a conventional 90/10 split to have sufficient test samples for sensible log-rank *P* values in the survival analysis (see Materials and Methods). In addition, we assessed the accuracy of the survival subtype predictions using C-index, which measures the fraction of all pairs of individuals whose predicted survival times are ordered correctly (42). We also calculated the error of the model fitting on survival data using Brier score (50). On average, the training data generated high C-index (0.70 ± 0.04), low Brier score (0.19 ± 0.01), and significant average log-rank *P* value (0.001) on survival difference (Table 1). A similar trend was observed for the 3-omics held-out test data, with C-index = 0.69 ± 0.08, Brier score = 0.20 ± 0.02, and average survival *P* = 0.005 (Table 1). When tested on each single omic layer of data, this multi-omics model also has decent performances in terms of C-index, low Brier scores, and log-rank *P* (Table 1). These results demonstrate that the classification model using cluster labels is robust to predict survival-specific clusters. Supplementary Table S2 enlists the top K features for 3-omics selected by ANOVA for the SVM-based classification in the full TCGA cohort.

The survival subtypes are robustly validated in five independent cohorts

To demonstrate the robustness of the classification model at predicting survival outcomes, we validated the model on a variety of five independent cohorts, each of which had only mRNA, or miRNA or methylation omics data (Table 2; Figs. 2B–F). The common top features selected by ANOVA prior to SVM classification (between TCGA and five cohorts) are as follows: LIRI-JP (94%), NCI (74%), Chinese-GSE31384 (58%), E-TABM-36 (82%), and Hawaiian (100%). The LIRI-JP dataset is the RNA-Seq dataset with the most patients (*n* = 230); we achieved a good C-index of 0.75, a low Brier error rate of 0.16, and a log-rank *P* value of 4.4e−4 between the two subtypes. For the second largest (*n* = 221) cohort (NCI GSE14520), the two subgroups have a decent C-index of 0.67 and low Brier error rate of 0.18, with a log-rank *P* value of 1.05e−3 (Table 2). For the Chinese cohort (GSE31384), the miRNA array data with 166 samples, the two subgroups have a C-index of 0.69, a low Brier error rate of 0.21, and a log-rank *P* value of 8.49e−4 (Table 2). Impressively, the C-indices for the two smallest cohorts, E-TABM-36 (40 samples) and Hawaiian (27 samples), are very good, with values of 0.77 and 0.82, respectively. The *P* values obtained for the small cohorts are

not significant due to small sample size, with values of 0.103 and 0.0535, respectively (Fig. 2E and F).

The DL-based methodology outperforms alternative approaches

We compared the performance of the model described in Fig. 1B to two alternative approaches (Supplementary Fig. S2). In the first approach, we replaced the autoencoder with the conventional dimension reduction method principal component analysis (PCA). Similar to the 100 features from hidden nodes in the autoencoder, we obtained the top 100 principal components, which were then subjective to univariate Cox-PH. As a result, 13 principal component features remained. However, this approach failed to give significant log-rank *P* value (*P* = 0.14) in detecting survival subgroups (Supplementary Fig. S2A). It also yielded significantly lower C-index for test data (0.62; Supplementary Table S3) as compared with the model using the autoencoder. In the second approach, we bypassed the autoencoder step, performed univariate Cox-PH analysis on each input feature in the 3 omics data types, and kept the top 37 features based on the C-index scores (Supplementary Fig. S2B). This model gave a *P* value of 3.0e−8, still much less significant than the DL method (6.0e−9; Supplementary Fig. S2C). More importantly, these alternative approaches failed overall to find significant subgroups in the majority of the confirmation sets. The only significance is in the LIRI-JP dataset using the Cox-PH approach (Supplementary Table S3).

Worth noticing, the 3-omics–based DL model gives better prediction metrics in CV when compared with single-omics–based DL models (Supplementary Table S4), suggesting that multi-omics data are, indeed, better than single-omics data for model building. Finally, autoencoders fitted with a high number of epochs, with more than three hidden layers, or with a high number of hidden nodes presented significant decreases of the performances. However, only one hidden layer or too few hidden nodes appeared less efficient (Supplementary Table S3).

Adding clinical information does not improve DL-based multi-omics model

It remains to be seen whether the DL-based multi-omics model will improve the predictability by adding clinical information. Therefore, we assessed the performance of alternative models with clinical variables as the features, either alone or in combination with previous DL-based multi-omics model (Table 3). When

Table 2. Performance of classifier for the five external confirmation cohorts

Confirmation cohort	Omics data type	Reference	Samples (*N*)	C-index	Brier score	Log-rank *P*
LIRI-JP	RNA-Seq	(31)	230	0.75	0.16	4.4e−4
NCI	mRNA microarray	(32)	221	0.67	0.18	1.05e−3
Chinese	miRNA array	(33)	166	0.69	0.21	8.49e−4
E-TABM-36	mRNA microarray	(16)	40	0.77	0.19	0.103
Hawaiian	DNA methylation	(34)	27	0.82	0.19	5.35e−2

Table 3. Performance of the model using clinical features on confirmation cohorts

Confirmation cohort	C-index (clinic only)	C-index (combined[a])	Brier score	Log-rank P
LIRI-JP	0.55	0.74	0.16	0
NCI	0.45	0.65	0.19	0.007
E-TABM-36	0.50	0.75	0.19	0.056
Hawaiian	0.70	0.87	0.19	0.003

[a]Combined, clinical + DL-based class labels.

clinical features were used as the sole feature set for survival prediction, the models' performances were much poorer (Table 3) when compared with the DL-based genomic model (Table 2). Then we combined the clinical features with the 3 omics layers before the K-means clustering step in Fig. 1B. Surprisingly, the C-indices of the combined model were not better on the confirmation cohorts with larger sample sizes (LIRI-JP and NCI cohorts) compared with those of the DL-based multi-omics model. C-index and P value were only slightly but not statistically significantly better for the Hawaiian cohort, which has only 27 samples. We thus conclude that the DL-based multi-omics model performs sufficiently well, even without clinical features. We speculate that the reason is due to the unique advantage of the DL neural network, which can capture the redundant contributions of clinical features through their correlated genomic features.

Associations of survival subgroups with clinical covariates

We performed the Fisher exact test between the two survival subgroups and the clinical variables from the TCGA cohort, and found that only grade ($P = 0.0004$) and stage ($P = 0.002$) were significantly associated with survival, as expected. As HCC is aggravated by multiple risk factors, including HBV, HCV, and alcohol, we also tested our model within subpopulations stratified by individual risk factors (Table 4). Impressively, our model performed very well on all the risk factor categories, with C-indices ranging from 0.69 to 0.79 and Brier scores between 0.19 and 0.20. Log-rank P values were significant in HBV-infected patients ($P = 0.04$), alcohol consumers ($P = 0.005$), and others category ($P = 0.0035$). The only nonsignificant P value ($P = 0.20$) was obtained from the HCV-infected patients, probably attributed to the small group size ($n = 31$).

$TP53$ is one of the most frequently mutated genes in HCC, and its inactivation mutations have been reported to be associated with poor survival in HCC (60). Between the two survival subgroups S1 and S2 in the TCGA samples, $TP53$ is more frequently mutated in the aggressive subtype S1 (Fisher test $P = 0.042$). Furthermore, $TP53$ inactivation mutations are associated with the aggressive subtype S1 in the LIRI-JP cohort, where whole genome sequencing data are available ($P = 0.024$).

Functional analysis of the survival subgroups in TCGA HCC samples

We used the $DESeq2$ package (52) for differential gene expression between the two identified subtypes. After applying the filter

Table 4. Full model performance within each subpopulation stratified by the clinical confounders in the TCGA cohort

Confounder	Samples (N)	C-index	Brier score	Log-rank P
HBV	74	0.74	0.20	0.04
HCV	31	0.69	0.19	0.20
Alcohol	67	0.79	0.20	0.005
Others	59	0.77	0.19	0.0035

of $\log_2$ fold change >1 and FDR <0.05, we obtained 820 upregulated and 530 downregulated genes in the aggressive subcluster S1. Figure 3 shows the comparative expression profile of these 1,350 genes after normalization. The upregulated genes in the S1 cluster include the stemness marker genes $EPCAM$ ($P = 5.7e-6$) and $KRT19$ ($P = 6.7e-15$) and tumor marker gene $BIRC5$ ($P = 1.2e-13$), which were also reported earlier to be associated with aggressive HCC subtype (61–63). In addition, 18 genes ($ADH1B$, $ALDOA$, $APOC3$, $CYP4F12$, $EPHX2$, KHK, $PFKFB3$, $PKLR$, PLG, RGN, $RGS2$, $RNASE4$, $SERPINC1$, $SLC22A7$, $SLC2A2$, $SPHK1$, $SULT2A1$, and $TM4SF1$) differentially expressed in the two subtypes have similar trends of expression as in the previous study, where a panel of 65-gene signature was reported to be associated with HCC survival (64).

Using the differentially expressed genes above, we conducted KEGG pathway analysis to pinpoint the pathways enriched in the two subtypes. As we have used EASE score in DAVID as the enrichment method, these results should be interpretive only (65). These subtypes have different and (almost) disjoint active pathways, confirming that they are distinct subgroups at the pathway level (Fig. 4). Aggressive subtype S1 is enriched with cancer related pathways, Wnt signaling pathway, PI3K–Akt signaling pathway, and so on (Fig. 4A). The Wnt signaling pathway was reported to be associated with aggressive HCC previously (66). In contrast, the moderate subtype S2 has activated metabolism-related pathways including drug metabolism, amino acid and fatty acid metabolism, and so on (Fig. 4B). Further biological functional studies are needed to confirm the signaling pathway (for S1) versus metabolic pathway (for S2) preferences between the two survival groups. We performed similar differential analysis for miRNA expression and methylation data, and detected 23 miRNAs and 55 genes' methylation statistically different between the two subgroups (Supplementary Fig. S3; Supplementary File S2).

Discussion

Heterogeneity is one of the bottlenecks for understanding HCC etiology. Although there are many studies for subtype identification of HCC patients, embedding survival outcome of the patients as part of the procedure of identified subtypes has not been reported before. Moreover, most reported HCC subtype models have either no or very few external confirmation cohorts. This calls for better strategies where the identified subtypes could reflect the phenotypic outcome of the patients, that is, the survival directly. Current work includes the integration of the multi-omics data from the same patients, giving an edge by exploiting the improved signal-to-noise ratio. To our knowledge, we are the first to use the DL framework to integrate multi-omics information in HCC. It propels DL to develop a risk-stratification model, not only for prognostication but also instrumental for improvising risk-adapted therapy in HCC.

We have identified two subtypes from the molecular level. This model is robust and perhaps more superior than other

Pathway	Genes (N)	%	EASE score
Pathways in cancer	27	0.024	4.10E-03
PI3K–Akt signaling pathway	24	0.021	6.28E-03
Focal adhesion	20	0.018	3.10E-04
Proteoglycans in cancer	19	0.017	5.99E-04
Hippo signaling pathway	15	0.013	1.81E-03
Regulation of actin cytoskeleton	15	0.013	3.08E-02
ECM-receptor interaction	14	0.012	2.24E-05
Axon guidance	13	0.012	3.24E-03
Wnt signaling pathway	12	0.011	1.62E-02
Protein digestion and absorption	11	0.010	1.82E-03

Pathway	Genes (N)	%	EASE score
Metabolic pathways	123	0.190	7.98E-27
Chemical carcinogenesis	27	0.042	7.33E-18
Biosynthesis of antibiotics	27	0.042	1.57E-07
Retinol metabolism	24	0.037	7.29E-17
Drug metabolism—cytochrome P450	22	0.034	4.17E-14
Metabolism of xenobiotics by cytochrome P450	22	0.034	2.72E-13
Steroid hormone biosynthesis	18	0.028	3.08E-11
Bile secretion	18	0.028	6.32E-10
PPAR signaling pathway	17	0.026	3.36E-09
Peroxisome	17	0.026	8.81E-08
Carbon metabolism	17	0.026	6.59E-06
Complement and coagulation cascades	15	0.023	2.96E-07
Drug metabolism—other enzymes	14	0.022	1.14E-08
Glycolysis/gluconeogenesis	13	0.020	8.39E-06
Fatty acid degradation	12	0.019	6.20E-07
Glycine, serine, and threonine metabolism	11	0.017	1.58E-06
Tryptophan metabolism	11	0.017	2.04E-06

Statistics in CCR

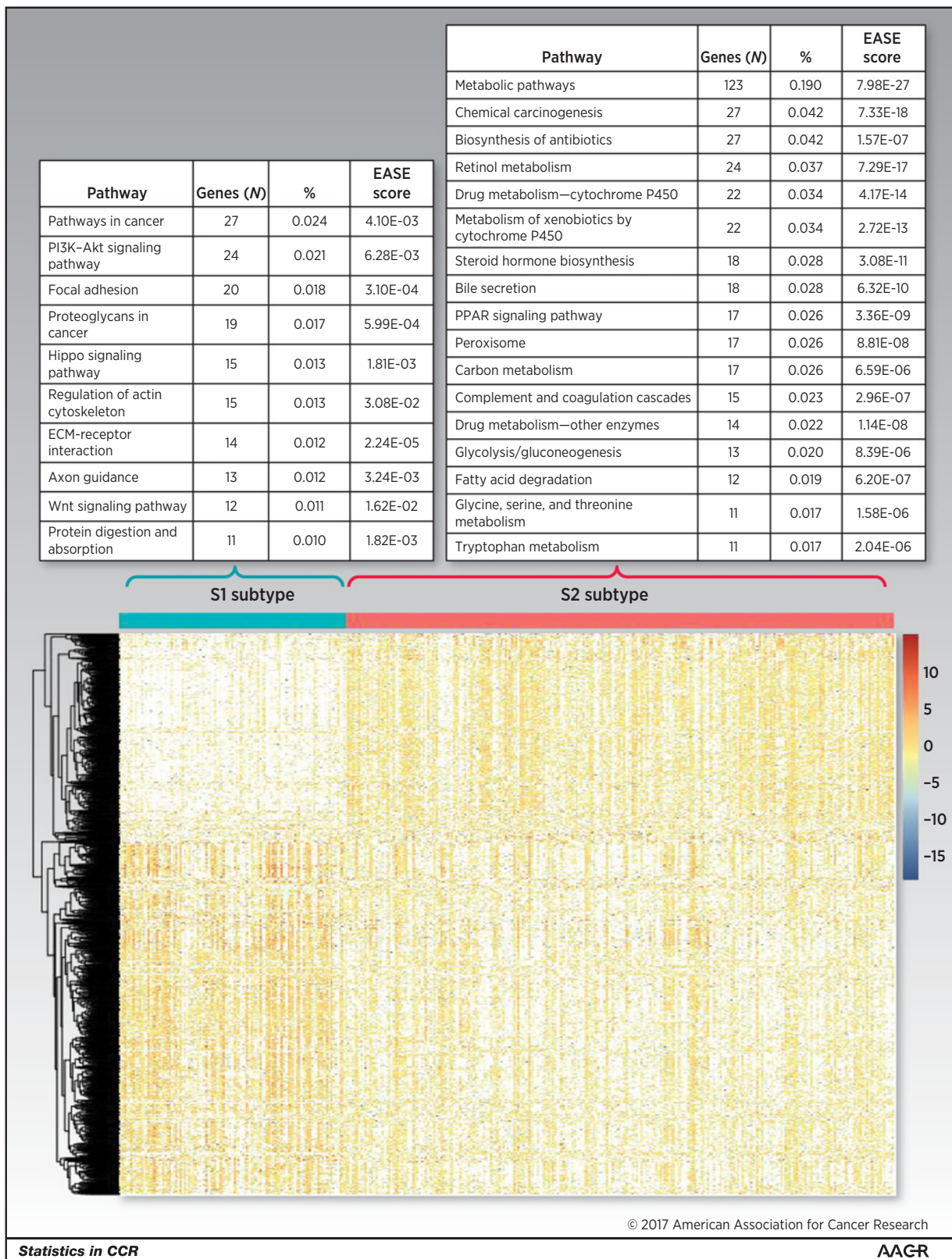© 2017 American Association for Cancer Research

AACR

Figure 3.
Differentially expressed genes and their enriched pathways in the two subtypes from the TCGA cohort. S1: aggressive (higher risk survival) subtype; S2: moderate (lower risk survival) subtype.
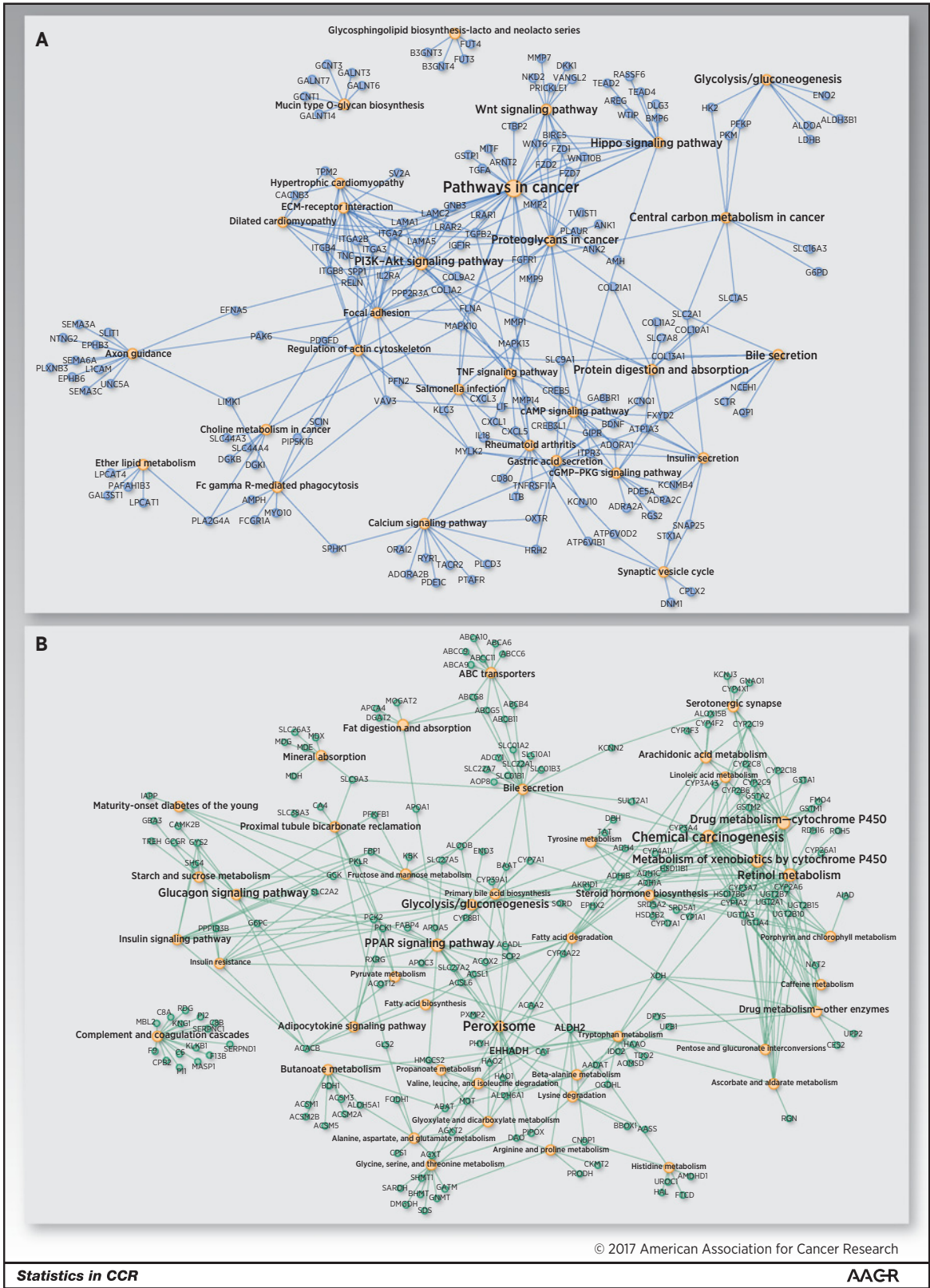
Figure 4.
Bipartite graph for significantly enriched KEGG pathways and upregulated genes in the two subtypes. Enriched pathway gene analysis for upregulated genes in the S1 aggressive tumor subgroup (**A**) and less aggressive S2 subgroup (**B**).

approaches, manifested in several levels. First, CV results gave the consistent performance in TCGA HCC test samples, implying the reliability and robustness of the model. Second, the DL technique used in the model has captured sufficient variations due to potential clinical risk factors, such that it performs as accurately or even better than, having additional clinical features in the model. Third, the autoencoder framework shows much more efficiency to identify features linked to survival compared with PCA or individual Cox-PH–based models. Finally and most importantly, this model is repetitively validated in five additional cohorts, ranging from RNA-Seq, mRNA microarray, miRNA array, and DNA methylation platforms.

In association with clinical characteristics, the more aggressive subtype (S1) has consistent trends of association with higher *TP53* inactivation mutation frequencies in the TCGA and LIRI-JP cohorts, which is in concordance with the previous study (60). Association of stemness markers (*KRT19* and *EPCAM*) with the S1 subtype is also in congruence with the literature (61, 62). Moreover, the S1 subtype is enriched with activated Wnt signaling pathway (66). Despite our effort, the one-to-one comparison with previous studies is not feasible due to the absence of cluster label information in original reports and lack of survival data in some cases. Fortunately, we were able to identify five external confirmation cohorts encompassing different omic datasets and succeeded in validating the subtypes among them. These results gave enough confidence that the two survival subtype–specific model proposed in this report is of direct clinical importance and may be useful to improving the survival of patients with HCC.

Some caveats are worth discussion below. First, we used the whole TCGA dataset in step 1 (Fig. 1B) to learn the class label of the TCGA samples in an unsupervised way. Therefore, when we build an SVM model using the TCGA training dataset and apply it on TCGA testing data, the C-statistics may be inflated; however, when we apply the SVM model to the other external datasets, these datasets give more unbiased C-statistics, as they are not part of the SVM model construction process. Also, our current model is trained on the TCGA HCC data, and it has been reported earlier that TCGA samples are impure (67). Liver tumor samples (LIHC) was reported to have better than average purity among 21 tumor types—higher than breast cancers in TCGA. Also, to obtain only HCC samples, we have procured the data from the TCGA website with their clinical annotation for liver hepatocellular carcinoma under the LIHC flag. The purity issue, along with the heterogeneous nature of HCC due to various risk factor, may explain why we do not have a C-index better than 0.80 in the TCGA training data. To

further examine the effects of risk factors on the model, we built submodels for samples with only HBV, HCV, and alcohol risk factors. We obtained C-indices of 0.90, 0.92, and 0.83 on HBV, HCV, and alcohol-affected TCGA subpopulations. Thus, the heterogeneity of the population does affect the model performance. However, issues exist to test these models on external datasets, as the submodels were built on small training data, thus they could suffer from over-fitting in confirmation cohorts. Moreover, sample risk factors are not always known for public cohorts, restricting our confirmation effort. Albeit these issues, the current TCGA-based model has an average C-index of 0.74 on five external confirmations, indicating that the model is generally predictive. In addition, we used log-rank *P* value and Brier score as other performance metrics to assess our pipeline. In the future, we plan to collaborate with clinicians to prospective cohorts and improve the model over time.

Disclosure of Potential Conflicts of Interest

K. Chaudhary, O.B. Poirion, and L.X. Garmire are listed as inventors on a provisional patent for deep learning–based survival subpopulation prediction held by the University of Hawaii. No potential conflicts of interest were disclosed by the other author.

Authors' Contributions

Conception and design: O.B. Poirion, L.X. Garmire
Development of methodology: K. Chaudhary, O.B. Poirion, L.X. Garmire
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): O.B. Poirion
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): K. Chaudhary, O.B. Poirion L. Lu, L.X. Garmire
Writing, review, and/or revision of the manuscript: K. Chaudhary, O.B. Poirion, L. Lu, L.X. Garmire
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): K. Chaudhary, O.B. Poirion
Study supervision: L.X. Garmire

Acknowledgments

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015;65:87–108.
2. American Cancer Society. Key statistics about liver cancer. Available from: https://www.cancer.org/cancer/liver-cancer/about/what-is-key-statistics.html.
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin 2016;66:7–30.
4. Trevisani F, Magini G, Santi V, Morselli-Labate AM, Cantarini MC, Di Nolfo MA, et al. Impact of etiology of cirrhosis on the survival of patients diagnosed with hepatocellular carcinoma during surveillance. Am J Gastroenterol 2007;102:1022–31.
5. Nguyen VT, Law MG, Dore GJ. Hepatitis B-related hepatocellular carcinoma: epidemiological characteristics and disease burden. J Viral Hepat 2009;16:453–63.
6. Chen CH, Huang GT, Yang PM, Chen PJ, Lai MY, Chen DS, et al. Hepatitis B- and C-related hepatocellular carcinomas yield different clinical features and prognosis. Eur J Cancer 2006;42:2524–9.
7. Chen CH, Su WW, Yang SS, Chang TT, Cheng KS, Lin HH, et al. Long-term trends and geographic variations in the survival of patients with hepatocellular carcinoma: analysis of 11,312 patients in Taiwan. J Gastroenterol Hepatol 2006;21:1561–6.
8. American Cancer Society. Liver cancer survival rates. Available from: https://www.cancer.org/cancer/liver-cancer/detection-diagnosis-staging/survival-rates.html.
9. El-Serag HB. Hepatocellular carcinoma. N Engl J Med 2011;365:1118–27.
10. Marrero JA, Kudo M, Bronowicki JP. The challenge of prognosis and staging for hepatocellular carcinoma. Oncologist 2010;15 Suppl 4:23–33.

11. Colagrande S, Inghilesi AL, Aburas S, Taliani GG, Nardi C, Marra F. Challenges of advanced hepatocellular carcinoma. World J Gastroenterol 2016;22:7645–59.
12. Llovet JM, Ricci S, Mazzaferro V, Hilgard P, Gane E, Blanc JF, et al. Sorafenib in advanced hepatocellular carcinoma. N Engl J Med 2008;359:378–90.
13. Chen X, Cheung ST, So S, Fan ST, Barry C, Higgins J, et al. Gene expression patterns in human liver cancers. Mol Biol Cell 2002;13:1929–39.
14. Lee JS, Chu IS, Heo J, Calvisi DF, Sun Z, Roskams T, et al. Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. Hepatology 2004;40:667–76.
15. Breuhahn K, Vreden S, Haddad R, Beckebaum S, Stippel D, Flemming P, et al. Molecular profiling of human hepatocellular carcinoma defines mutually exclusive interferon regulation and insulin-like growth factor II overexpression. Cancer Res 2004;64:6058–64.
16. Boyault S, Rickman DS, de Reynies A, Balabaud C, Rebouissou S, Jeannot E, et al. Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. Hepatology 2007;45:42–52.
17. Chiang DY, Villanueva A, Hoshida Y, Peix J, Newell P, Minguez B, et al. Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. Cancer Res 2008;68:6779–88.
18. Hoshida Y, Nijman SM, Kobayashi M, Chan JA, Brunet JP, Chiang DY, et al. Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. Cancer Res 2009;69:7385–92.
19. Goossens N, Sun X, Hoshida Y. Molecular classification of hepatocellular carcinoma: potential therapeutic implications. Hepat Oncol 2015;2:371–9.
20. Liu G, Dong C, Liu L. Integrated multiple "-omics" data reveal subtypes of hepatocellular carcinoma. PLoS One 2016;11:e0165457.
21. Huang S, Chaudhary K, Garmire LX. More is better: recent progress in multi-omics data integration methods. Front Genet 2017;8:84.
22. Tan J, Ung M, Cheng C, Greene CS. Unsupervised feature construction and knowledge extraction from genome-wide assays of breast cancer with denoising autoencoders. Pac Symp Biocomput 2015;20:132–43.
23. Chen L, Cai C, Chen V, Lu X. Learning a hierarchical representation of the yeast transcriptomic machinery using an autoencoder model. BMC Bioinformatics 2016;17 Suppl 1:9.
24. Khalili M, Alavi Majd H, Khodakarim S, Ahadi B, Hamidpour M. Prediction of the thromboembolic syndrome: an application of artificial neural networks in gene expression data analysis. J Paramed Sci 2016;7:8.
25. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep 2016;6:26094.
26. Chen Q, Song X, Yamada H, Shibasaki R. Learning deep representation from big and heterogeneous data for traffic accident inference. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence; 2016 Feb 12–17; Phoenix, AZ. Palo Alto (CA): AAAI Press; 2016. p. 338–44.
27. Tan J, Hammond JH, Hogan DA, Greene CS. ADAGE-Based integration of publicly available *Pseudomonas aeruginosa* gene expression data with denoising autoencoders illuminates microbe–host interactions. mSystems 2016;1:e00025–15.
28. Zhu Y, Qiu P, Ji Y. TCGA-assembler: open-source software for retrieving and processing TCGA data. Nat Methods 2014;11:599–600.
29. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, et al. Similarity network fusion for aggregating data types on a genomic scale. Nat Methods 2014;11:333–7.
30. Xiang Q, Dai X, Deng Y, He C, Wang J, Feng J, et al. Missing value imputation for microarray gene expression data using histone acetylation information. BMC Bioinformatics 2008;9:252.
31. Fujimoto A, Furuta M, Totoki Y, Tsunoda T, Kato M, Shiraishi Y, et al. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. Nat Genet 2016;48:500–9.
32. Roessler S, Jia HL, Budhu A, Forgues M, Ye QH, Lee JS, et al. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. Cancer Res 2010;70:10202–12.
33. Wei R, Huang GL, Zhang MY, Li BK, Zhang HZ, Shi M, et al. Clinical significance and prognostic value of microRNA expression signatures in hepatocellular carcinoma. Clin Cancer Res 2013;19:4780–91.
34. Song MA, Tiirikainen M, Kwee S, Okimoto G, Yu H, Wong LL. Elucidating the landscape of aberrant DNA methylation in hepatocellular carcinoma. PLoS One 2013;8:e55761.
35. Liu F, Li H, Ren C, Bo X, Shu W. PEDLA: predicting enhancers with a deep learning-based algorithmic framework. Sci Rep 2016;6:28517.
36. Bengio Y. Learning deep architectures for AI. Foundations and Trends in Machine Learning 2009;2:1–127.
37. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987;20:53–65.
38. Caliński T, Harabasz J. A dendrite method for cluster analysis. Communications in Statistics 1974;3:1–27.
39. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825–30.
40. Zhang B, Wang J, Wang X, Zhu J, Liu Q, Shi Z, et al. Proteogenomic characterization of human colon and rectal cancer. Nature 2014;513:382–7.
41. Angermueller C, Parnamaa T, Parts L, Stegle O. Deep learning for computational biology. Mol Syst Biol 2016;12:878.
42. Steck H, Krishnapuram B, Dehing-oberije C, Lambin P, Raykar VC. On ranking in survival analysis: bounds on the concordance index. In: JC Platt, D Koller, Y Singer, ST Roweis, editors. Advances in neural information processing systems 20. Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems (NIPS 2007); 2007 Dec 3–6; Vancouver, BC, Canada. La Jolla (CA): Neural Information Processing Systems Foundation; 2008. p. 1209–16.
43. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361–87.
44. Schröder MS, Culhane AC, Quackenbush J, Haibe-Kains B. survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. Bioinformatics 2011;27:3206–8.
45. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33:1.
46. Cox DR. Regression models and life-tables. In: Kotz S, Johnson NL, editors. Breakthroughs in statistics: methodology and distribution. vol. 2. New York: Springer; 1992. p 527–41.
47. Wei R, De Vivo I, Huang S, Zhu X, Risch H, Moore JH, et al. Meta-dimensional data integration identifies critical pathways for susceptibility, tumorigenesis and progression of endometrial cancer. Oncotarget 2016;7:55249–63.
48. Huang S, Yee C, Ching T, Yu H, Garmire LX. A novel model to combine clinical and pathway-based transcriptomic information for the prognosis prediction of breast cancer. PLoS Comput Biol 2014;10:e1003851.
49. Therneau T. A package for survival analysis in S. R package version 2.38. 2015. Available from: http://CRAN.R-project.org/package=survival.
50. Brier GW. Verification of forecasts expressed in terms of probability. Monthly Weather Review 1950;78:1–3.
51. Zhang X, Li Y, Akinyemiju T, Ojesina A, Buckhaults P, Liu N, et al. Pathway-structured predictive model for cancer survival prediction: a two-stage approach. Genetics 2017;205:89–100.
52. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
53. Ching T, Ha J, Song MA, Tiirikainen M, Molnar J, Berry MJ, et al. Genome-scale hypomethylation in the cord blood DNAs associated with early onset preeclampsia. Clin Epigenetics 2015;7:21.
54. Ching T, Song MA, Tiirikainen M, Molnar J, Berry M, Towner D, et al. Genome-wide hypermethylation coupled with promoter hypomethylation in the chorioamniotic membranes of early onset pre-eclampsia. Mol Hum Reprod 2014;20:885–904.
55. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics 2008;24:1547–8.
56. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.
57. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009;37:1–13.
58. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44–57.
59. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. Icwsm 2009;8:361–2. Available from: http://gephi.org.
60. Villanueva A, Hoshida Y. Depicting the role of TP53 in hepatocellular carcinoma progression. J Hepatol 2011;55:724–5.

61. Yamashita T, Forgues M, Wang W, Kim JW, Ye Q, Jia H, et al. EpCAM and alpha-fetoprotein expression defines novel prognostic subtypes of hepatocellular carcinoma. Cancer Res 2008;68:1451–61.

62. Andersen JB, Loi R, Perra A, Factor VM, Ledda-Columbano GM, Columbano A, et al. Progenitor-derived hepatocellular carcinoma model in the rat. Hepatology 2010;51:1401–9.

63. Cao L, Li C, Shen S, Yan Y, Ji W, Wang J, et al. OCT4 increases BIRC5 and CCND1 expression and promotes cancer progression in hepatocellular carcinoma. BMC Cancer 2013;13:82.

64. Kim SM, Leem SH, Chu IS, Park YY, Kim SC, Kim SB, et al. Sixty-five gene-based risk score classifier predicts overall survival in hepatocellular carcinoma. Hepatology 2012;55:1443–52.

65. Goeman JJ, Buhlmann P. Analyzing gene expression data in terms of gene sets: methodological issues. Bioinformatics 2007;23:980–7.

66. White BD, Chien AJ, Dawson DW. Dysregulation of Wnt/beta-catenin signaling in gastrointestinal cancers. Gastroenterology 2012;142:219–32.

67. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. Nat Commun 2015;6:8971.