OXFORD UNIVERSITY PRESS | Briefings In Bioinformatics

Reducing healthcare disparities using multiple multiethnic data distributions with fine-tuning of transfer learning

Journal:	*Briefings in Bioinformatics*
Manuscript ID	BIB-21-1509.R2
Manuscript Type:	Case Study
Date Submitted by the Author:	20-Jan-2022
Complete List of Authors:	Toseef, Muhammad; City University of Hong Kong, Department of Computer Science Li, Xiangtao; Jilin University, School of Artificial Intelligence Wong, Ka-Chun; City University of Hong Kong, Computer Science
Keywords:	ethnic disparities, transfer learning, domain adaptation, deep learning
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
BIB-21-1509.R2.tex	

Briefings in Bioinformatics, 2021, pp. 1–10

doi: DOI HERE
Advance Access Publication Date: Day Month Year

Reducing healthcare disparities using multiple multiethnic data distributions with fine-tuning of transfer learning

Muhammad Toseef,[1] Xiangtao Li[2*] and Ka-Chun Wong[1,3*]

[1]Department of Computer Science, City University of Hong Kong, Hong Kong SAR, [2]School of Artificial Intelligence, Jilin University, Jilin, China and [3]Hong Kong Institute for Data Science, City University of Hong Kong, Hong Kong SAR
[*]Corresponding authors.
Email: kc.w@cityu.edu.hk and lixt314@jlu.edu.cn

Abstract

Healthcare disparities in multiethnic medical data is a major challenge; the main reason lies in the unequal data distribution of ethnic groups among many data cohorts. Biomedical data collected from different cancer genome research projects may consist of mainly one ethnic group, such as people with European ancestry. In contrast, the data distribution of other ethnic races such as African, Asian, Hispanic, and Native Americans can be less visible than the counterpart. Data inequality in the biomedical field is an important research problem, resulting in the diverse performance of machine learning models while creating healthcare disparities. Previous researches have reduced the healthcare disparities only using limited data distributions. In our study, we work on fine-tuning of deep learning and transfer learning models with different multiethnic data distributions for the prognosis of 33 cancer types. In previous studies, to reduce the healthcare disparities, only a single ethnic cohort was used as the target domain with one major source domain. In contrast, we focused on multiple ethnic cohorts as the target domain in transfer learning using the TCGA and MMRF CoMMpass study datasets. After performance comparison for experiments with new data distributions, our proposed model shows better performance for transfer learning schemes compared to the baseline approach for old and new data distributation experiments.

Key words: ethnic disparities, transfer learning, domain adaptation, deep learning

Introduction

After the arrival of the pandemic in 2019, the death ratio records among different ethnic races showed that people from some specific groups were more prone to infection and expiry than some other races. For example, in the United Kingdom, men from the African race have four times more chances to die as compared with other races [1]. Similarly, in New York, USA the African and Hispanic children lost their guardians and parents, twice as compared with American or Asian children [1].

With the technological developments associated with the coronavirus pandemic, the need for bioinformatics has become prominent. Hence, the role of personalized medicine data management can alleviate this gap [2, 3, 4]. As data grows day by day, medical data inequality has become a challenging problem. According to a survey, the current battles over income inequality can turn into struggles over data inequality in the next ten years [5, 6]. Data inequality can be defined as where the data distribution is unbalanced amongst the different members. The possible reasons for medical data inequality in different data cohorts can be poverty and healthcare access, biological, or sociocultural differences [7]. Data inequality problem is the main concern in many fields such as legal systems, advertisements [8, 5] and for many machine learning problems such as natural language processing, and computer vision [9, 10, 11]. Despite the breakthroughs in bioinformatics with mature computational methods and big data, health data inequality among ethnic groups will create a global health problem for artificial intelligence models. The cancer omics data from different cohorts allows the bioscientists and data engineers to work on complex cancer problems [12, 13, 14, 15]. The problem associated with healthcare disparities lies in the uneven data distribution among ethnic groups, leading to the low performance of machine learning models [16, 17].

Data from different cancer genome research projects, including The Cancer Genome Atlas Program [18, 19], OncoArray [20], The Therapeutically Applicable Research to Generate Effective Treatments initiative [21], and other studies are mostly collected from specific ethnic groups. Data distribution among different groups have been shown in Fig. 1.

This data discrepancy in biomedical data creates the availability of insufficient data for training machine learning models with low-performance accuracy. The data inequality in ethnic groups can lead to healthcare disparities; for instance, the majority of biomedical data comes from European ancestry. Therefore, it is not surprising that the machine learning models can be more accurate for a specific group of people than other ethnic races [22, 23]. To address this challenge, a recent attempt was proposed to solve the data inequality problem [24]. In this study, the authors developed a deep transfer learning models to reduce the health care disparities among ethnic groups. This approach followed the multiethnic learning schemes, where authors used the domain adaptation task of transfer learning. The proposed mixture learning, independent learning, and transfer learning experiments were based on source and target domains training and testing data distributions. The authors observed that data distribution inequality and discrepancy is the main reason for performance disparities. To overcome such performance gap, they reduced biomedical data inequality among ethnic groups using transfer learning.

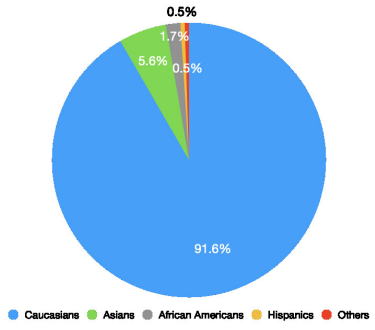


Fig. 1. Data inequality among different ethnic races: these data statistics represents collected samples from TARGET, OncoArray, TCGA, and 416 other cancer projects

Intelligent systems can learn from data and make predictions and patterns [25], especially with the availability of abundant training data [26, 27]. Transfer learning [28, 29] can be defined as where the distribution *P1* learned from one situation could be generalized to another distribution *P2* [30]. However, if the collected data mainly consist of only one domain, the system cannot generalize well. One example is a situation where the electronic health record data have been collected primarily for old patients, while the data for young patients is very limited. Now if the system predicts outcomes for a small data group (i.e., young patients) based on the past data from a large data group (i.e., old patients), the results can be biased; for instance, the system might indicate serious health risk factors for a young patient with high blood pressure. Transfer learning is helpful to train the deep learning model efficiently from the biased data, where domain adaptation is a sub-task of transfer learning [31, 32].

Among the previous approaches to reduce the health care disparities [24], although the proposed model achieved good results, the performance with the combinations of ethnic cohorts' compositions, cancer types, feature types and clinical endpoint outcomes were not fully explored. In this work, we performed experiments with new feature types with baseline model settings and observed the low-performance transfer learning experiments. After that, we fine-tuned the DNN models for transfer learning with different model settings and again performed the new and baseline experiments (Supplementary Table 1). We performed the experiments with fine-tuning of transfer learning model with 3-folds and 4-folds stratified cross-validation. In the baseline paper, authors performed 224 experiments [1], and results showed that this work has the potential to improve the results with fine-tuning of the current approach. Further, we analyzed those experiments with new factors from the feature space and found significant decrease in model performance.

To the best of our knowledge, the current methods developed for reducing the healthcare parities associated with data inequality distribution of ethnic races were not evaluated with different settings of the ethnic composition of the cohorts, omics data type, feature types, cancer types, and clinical endpoint outcomes. Those features with different arrangements may affect downstream model performance. We analyzed the model performance with new data distributation of ethnic cohorts and with unseen factors from the available feature space.

The main objective of this study is to improve the performance of transfer learning experiments with fine-tuning of deep learning models . We found that, with new combinations of features (Supplementary Table. 1), the transfer learning model performance was not good with low AUROC score. We fine-tuned and applied our domain adaptation model to these low-performance experiments. The results show slight to significant improvements for domain adaptation model.

The rest of the paper is organized as follows. In section 2, we have described the main algorithms used in our proposed approach with the dataset description. Section 3 discusses the experiments settings and performance evaluation of the deep learning models. In section 4, we discussed different experiments with source and target data distributions. Section 5 provides the conclusion of our work with possible future research directions.

Materials and methods

In this section, we discussed the datasets used in this study and main methods and deep learning models. We used the preprocess data by [24] for all experiments, a preprocessed version is provided on figshare. We also described the domain adaptation and transfer learning with deep learning model architecture.

Dataset

The datasets we adopted in model training and testing come from three different sources, including TCGA [19, 33], TARGET [19], and MMRF CoMMpass [34]. The entity distribution of data in all three datasets is listed as follows: European 80.5%, African 9.2%, East Asian 6.1%, Native Americans 3.6%, and others 0.7%. The Cancer Genome Atlas [35] (TCGAA) project was started in 2005. The data

[1] https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-020-18918-3/MediaObjects/41467_2020_18918_MOESM4_ESM.xlsx

distribution for seven main cancer types has been shown in Supplementary Figure S2. As shown in the figure, most data belong to the European ancestry, while the data distribution gap for other ethnic races is distantly large.

To ensure a fair comparison, we followed and adopted the dataset preprocessed by [24] and other publicly available data[2], we used the version 6 of dataset given by the authors. For all cancer types, EA ethnicity contributes almost 90% of the whole data. Such data inequality is the biggest challenge to reduce healthcare disparities and improve the accuracy of machine learning models for all ethnic groups. As the number of patients in the TCGA cohort for African people is less than 10 percent, like other ethnic cohorts with too few cases, transfer learning and domain adaptation can help generalize the model performance.

For both datasets, missing values were handled by removing the sample with missing values more than 20% in 189 protein features and 17176 mRNA feature expression. The clinical endpoint outcome time such as target year and genetic ancestry information were also considered and those samples were also removed with missing information.

TCGA Dataset

The Cancer Genome Atlas Program dataset was downloaded from Genome Data Commons (GDC) of the National Cancer Institute[3] for all cancer types. For the TCGA dataset, 189 protein features and 17176 mRNA feature expressions were used. For ethnic group distribution of the dataset, each ancestry was determined based on The Cancer Genetic Ancestry [33].

MMRF CoMMpass Dataset

This dataset contains the patient information with incurable blood cancer Multiple myeloma [36]. According to the ethnic composition of the CoMMpass studies, Caucasians are the major ethnic race with 77% of the whole data cohort, while other ethnic groups are: Africans 16%, Asians 2%, and others 5%. The Multiple Myeloma CoMMpass Study dataset was downloaded from Genomic Data Commons Data Portal [4]. For all machine learning experiments with mRNA as feature type, 600 mRNA features from MMRF CoMMpass dataset were selected using highest mean absolute deviation.

Domain Adaptation

For a successful machine learning model, it is assumed that we have enough training data have the same data and have the same feature space. However, in real-world datasets it is very common to have data across different domains creating data bias. Dataset bias usually happens when we have data mainly form one domain, such as training dataset is primarily form one demographic or from one culture [37]. Domain adaptation is a transfer learning task which is helpful when we have biased dataset. Domain adaptation is a type of transductive transfer learning where source and target task is same and source and target domain are different but related to each other [31].

In a domain adaptation task, the source domain or feature space is denoted by χ which consists of a specific dataset $X = X_1, X_2, \ldots, X_n$ and the label space Y, where $Y = Y_1, Y_2, \ldots, Y_n$, and a marginal probability distribution on dataset $P(X)$ [18, 31]. The predictive function f is used to learn from label pairs and can be expressed as $P(Y|X)$. A domain can be defined as $D = \chi, P(X)$, in a domain feature space and the marginal probability distributions on the dataset are the main components [38]. Domain adaptation can be defined as where the marginal probabilities between source and target domains are different, as shown Eq. in 1, where $P(X_S)$ and $P(X_T)$ are marginal probabilities for source and target, respectively.

$$P(X_S) \neq P(X_T) \quad (1)$$

In this study, as we have the same task on the hand but data different data distributions, so domain adaptation is the best suitable to learn data distributation from source data such as EA cohort. Furthermore, we outlined the transfer learning and definitions in supplementary methods section 1A.

Framework Overview

Our main contributions of this work is to use different ethnic cohorts data as target domain for fine-tuning of transfer learning and deep learning models. We divided our experiments on the basis of ethnic cohorts data and then performed extensive sets of experiments with different features (Supplementary Figure S2) with fine-tuning of machine learning experiments. We worked on data distribution of ethnic cohorts which were not fully explored in the previous studies. Asian Americans (ASA) and Native Americans (NAT-A) cohorts data is selected as target domain with European American cohort (EA) as source domain data for all multiethnic experiments. The flowchart of performed experiments has been shown in Fig. 2.

In the first phase of experiments, we worked with baseline model settings and data distributation, and performed experiments with new features (Supplementary Figure S2). All new features experiments were performed with baseline model settings and with a rigorous sets of fine-tuning models. We performed these experiments with different hyperparameters for deep learning and domain adaptation models.

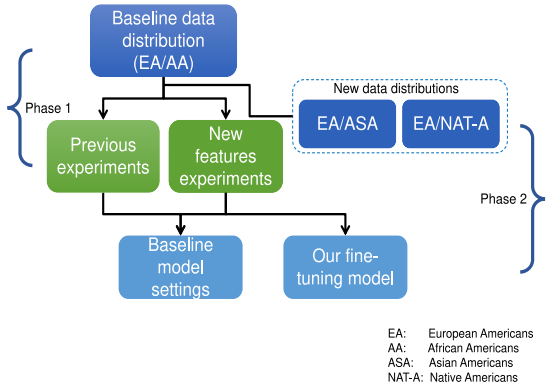


Fig. 2. Flowchart of performed experiments with baseline data distribution and using new data cohorts as target domain data. Phase 1 is consists of all experiments performed with data distribution using baseline approach with AA data as the target domain. Phase 2 has two directions: first experiments were performed using ASA data as target domain and then NAT-A data as the target domain

In the second phase of experiments, first we selected ASA cohort data as target domain and performed phase 1

[2] https://figshare.com/articles/dataset/Gao_Y_Cui_Y_2020_/12811574

[3] https://gdc.cancer.gov

[4] https://portal.gdc.cancer.gov/

experiments. After that, we used the same new features and then performed experiments using NAT-A cohort data. In phase 2, we performed all experiments with baseline model our fine-tuning models. Overall architecture of our approach has been shown in Fig. 3.

Deep neural network model

We used Lasagne[5] and Theano[6] libraries to train the deep neural networks. For each machine learning task (Table 1), we follow the same multiethnic scheme proposed by [24](Gao Cui, 2020). The baseline experiments were performed with the 20 cancer types with protein and mRNA features and 4 clinical endpoint outcomes with reasonably good performance for mixture learning and European cohort independent learning scheme.

The overall structure for the deep learning model was adapted from pyramid architecture [39], where the input layer of the model contains the data features of protein and mRNA. For protein features, the input layer has 189 nodes; with mRNA features, the input layer has 200 nodes [24]. Followed by a fully connected layer with 128 nodes, a dropout layer, followed by another fully connected layer with 64 nodes, followed by a dropout layer. In each machine learning experiment, two multivariate logistic regression models were applied to calculate the regression parameters of the source and target domain of the domain adaptation model, as shown below as $Y^{African}$ and $Y^{European}$.

$$Y^{African} = \frac{1}{(1 + e^{-\beta^{African}.X^{African}})} \quad (2)$$

$$Y^{European} = \frac{1}{(1 + e^{-\beta^{European}.X^{European}})} \quad (3)$$

We fine-tuned the deep learning models with different values of hyperparameters for all sets of experiments. We used stochastic gradient descent (SGD) optimizer for baseline model settings, and RMSprop [40] optimizer in our model to minimize the cross-entropy loss function with the learning rate ranges from 0.01 to 0.005. For all layers in deep learning models, the rectified linear unit (ReLU) activation function was used to avoid gradient vanishing problem, it can be expressed as $f(x) = max(0, x)$, and dropout value was set to 50% to 70%). The batch size for different experiments was set differently according to the number of total cases available.

For the independent learning schemes, where the cases were small, we set the batch size to 8, 16. and 24 for different experiments. For the European independent learning scheme, the batch size was set to 32 and 64 in our experiments because the dataset size was relatively large. For transfer learning scheme, the batch size was tuned with 16 and 24. The maximum iteration for each experiment was set to 50, 100, and 200 To avoid premature stopping, the Nesterov momentum was applied where momentum=0.9 [41].

For model training and testing, we used the 4-folds stratified cross-validation to split the data [42]. We omit the experiments for combinations where the cases for both source and target domains were less than five to ensure the reliable model performance. For each multiethnic experiment, the data was stratified differently; for mixture learning schemes (1,2 and 3) and for both independent learning schemes, the data was stratified using clinical outcome endpoint. For transfer learning experiments, first A sample data were stratified using clinical outcome endpoints, then initially model was trained using E samples. After that, the domain adaptation model was trained using the A. Finally, the domain adaptation model was then tested using A samples data. The deep learning model was fine-tuned for the generalization of the domain adaptation model [43]. For all six types of multiethnic experiments, training and testing distribution are tabulated in Table 1.

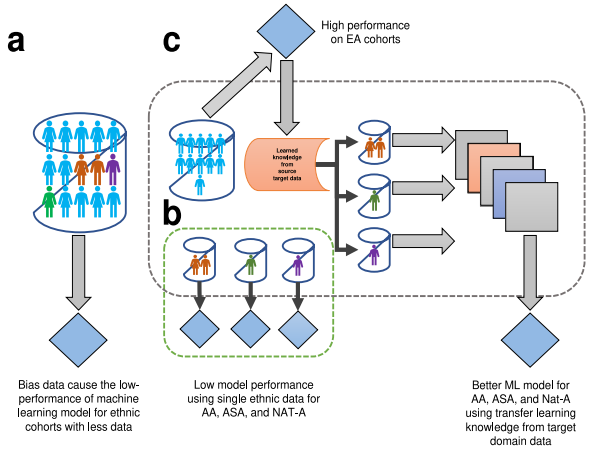


Fig. 3. Overall flowchart of the multiethnic experiment schemes. (a) The performance of ML model is bias because of data discrepancy with high performance for EA cohort and low performance for other cohorts; (b) Poor data distributions bring the low performance for AA, ASA, NAT-A and other cohorts; (c) Transfer learning helps to improve the performance of AA, ASA, and NAT-A cohorts after using the learned knowledge from EA cohorts

Experiments and result analysis

In this section, we have presented the experiments settings and result analysis for fine-tuning models. The experiments were implemented in Python 3.6.12 using TensorFlow 1.13.1 and sklearn package (Pedregosa et al., 2011). We performed the model training and testing using NVIDIA GeForce RTX2080 Ti with 64 GB memory.

In each experiment, we trained deep learning models to classify the two prognosis outcomes for two ethnic cohorts, where experiment was run for 20 times with 3-folds and 4-fold stratified cross-validation [42], for baseline approach and our model, respectively. After performing the extensive experiments, we generalized the set of hyperparameters with best performance across different model architectures for fine-tuning and domain adaptation models.

Performance Evaluation

To evaluate the model performance, we used the area under receiver operating characteristic (AUROC) [44] method. It measures the total area under the receiver operating curve [45] for two parameters, true positive rate (TPR) and false positive rate (FPR) for a classification model, as shown in Eq. 4 and 5, respectively. A To ensure the fair comparison for all experiments and for reproducible results, we performed each multiethnic experiment for AUROC score The final reported results for each experiment was weighted for 20 runs.

[5] https://lasagne.readthedocs.io/en/latest/
[6] http://deeplearning.net/software/theano/

Table 1. Multiethnic machine learning schemes based on data distributation: experiments are performed on the basis of data distribution for fine-tuning and domain adaptation models

Multiethnic experiment schemes	Baseline data distribution		New data distributions			
	Training data	Testing data	Training data	Testing data	Training data	Testing data
Mixture learning 1	EA +AA	EA +AA	EA + ASA	EA + ASA	EA + NAT-A	EA + NAT-A
Mixture learning 2	EA +AA	EA	EA + ASA	EA	EA + NAT-A	EA
Mixture learning 3	EA +AA	AA	EA + ASA	ASA	EA + NAT-A	NAT-A
Independent learning 1	EA	EA	EA	EA	EA	EA
Independent learning 2	AA	AA	ASA	ASA	NAT-A	NAT-A
Transfer learning	Source: EA / Target: AA	AA	Source: EA / Target: ASA	ASA	Source: EA / Target: NAT-A	NAT-A

$$TPR = \frac{TruePositive}{TruePositive + FalseNegative} \quad (4)$$

$$FPR = \frac{FalsePositive}{TrueNegative + FalsePositive} \quad (5)$$

Baseline Approach

This section described the multiethnic baseline approach experiments and models settings [24]. In baseline paper, the authors performed all multiethnic experiments using European ancestry data as source domain and African ancestry data as target domain, this data distributation has been shown in Table 1. Two deep learning models and one domain adaptation model were used with 3-folds stratified cross-validation. The authors performed all experiments for 40 cancer types, protein and mRNA as feature types, and 4 clinical endpoint outcomes with five different machine learning schemes.

The five multiethnic schemes are mixture learning 1, mixture learning 2, mixture learning 3, Independent learning 1, independent learning 2, and domain adaptation. For mixture learning schemes 1,2, and 3, EA and AA data were used to train the deep learning models, and also for model testing in mixture learning. For mixture learning scheme 2 and 3, the testing data was EA and AA, respectively. For two independent learning schemes, they used EA and AA data separately to train and test the deep learning models. For transfer learning scheme, source data was EA with AA as target data to train the model, while model testing was done only using AA data. The performance for three mixture learning and independent learning schemes for EA cohort was good using the European ancestry data as source and target domain, while it had low performance for all other ethnic cohorts because of inadequate data.

Experiments

This section discuss the experiment with different distribution of ethic cohorts in our study. We divided our experiments in three categories, first we performed experiments using EA as source domain data and AA as target domain data across different features combinations. We performed these experiments with fine-tuning of deep learning models with 4-folds stratified cross-validation. Secondly, we performed the experiments with new data distribution, where we used Asian Americans and Native Americans cohorts data instead of African American cohort, with EA cohort data as source domain for multiethnic experiments. Data distributions for all experiments has been shown in Table 1. The overall structure of the performed experiments has been shown in Fig. 2. For fair comparison in both set of experiments, we used the same hyperparameter values for fine-tuning models. We designed and performed the experiments with the following guidelines:

1. First, we analysed the low-performance experiments with the new features combinations with baseline approach [24] (Supplementary Fig. S1 and Supplementary Table 1).
2. Then we tried the low-performance experiments (1.) with new architectures of fine-tuning models and with different settings of stratified cross-validation.
3. After that, we analyzed model architecture with best performance over previously performed same experiments, specifically for the transfer learning experiments.
4. Then we performed experiments with new data distributation of ethnic cohorts

Each experiment has four attributes alongside with ethnic data composition. These four attributes are given below.

(a) 33 cancer types, further divided in 8 main categories (Supplementary Table S1)
(b) two feature types
(c) four clinical endpoint outcomes
(d) 1,2,3,4,5 years threshold time to an event

The two features types defined in (a) are mRNA and Protein expressions. Four endpoint outcomes defined in (b) are: disease-specific survival (DSS), the overall survival of the patient (OS), progression-free interval (PFI), and disease-free interval (DFI) [46]. The flowchart of a single experiment has been shown in Supplementary Fig. S3.

To ensure the machine learning model reliability, we made sure that at least five cases are present for each prognosis class (dead or alive) for EA → AA, EA → ASA, and EA → NAT-A distributions. We described the performed experiments on the basis of data distributions in the next sections.

Experiments based on EA/AA data distribution

We started the experiments from point (1) in section 3.3 to identify the low performance experiments specially for transfer learning task. In this first phase of experiments, we mainly used EA as source domain while AA as target domain, same data distributation followed in the baseline approach. We found the gap in the performance of many experiments for transfer learning and independent learning tasks were negligible, showing the low performance of domain adaptation task. After identifying low performance experiments, we fine-tuned the transfer learning experiments with extensive sets of experiments and generalized the best set of hyperparameters. After analyzing the experiments with lower performance for the previous approaches, we performed the same experiments for

Table 2. AUROC comparison of transfer learning scheme of baseline approach and our fine-tuned models from each data distributation

Sr. No.	Experiment settings	Data distributions	AUROC for transfer learning					
			(Gao & Cui, 2020)			Our model		
			min	max	avg	min	max	avg
1.	SARC-mRNA-OS-2YR	EA → AA	0.409	0.62	0.535	0.52	0.84	0.676
2.	SARC-mRNA-DSS-3YR	EA → AA	0.51	0.84	0.63	0.64	0.767	0.68
3.	UCEC-Protein-DFI-3YR	EA → AA	0.43	0.729	0.593	0.49	0.79	0.629
4.	BLCA-Protein-PFI-4YR	EA → AA	0.6	0.91	0.787	0.55	0.95	0.857
5.	BRCA-mRNA-PFI-3YR	EA → ASA	0.5	0.66	0.6	0.575	0.73	0.639
6.	STAD-mRNA-OS-1YR	EA → ASA	0.701	0.729	0.602	0.49	0.79	0.64
7.	PanSCCs-mRNA-OS-2YR	EA → ASA	0.47	0.729	0.649	0.579	0.565	0.629
8.	UCEC-mRNA-DSS-4YR	EA → ESA	0.328	0.656	0.524	0.406	0.687	0.55
9.	CESE-mRNA-OS-2YR	EA → NAT-A	0.459	0.704	0.563	0.445	0.717	0.612
10.	CESE-mRNA-DSS-1YR	EA → NAT-A	0.2	0.628	0.439	0.342	0.726	0.509
11.	PanGyn-mRNA-OS-2YR	EA → NAT-A	0.295	0.5	0.42	0.436	0.613	0.512
12.	CESE-mRNA-DSS-2YR	EA → NAT-A	0.394	0.67	0.544	0.462	0.663	0.6

transfer learning domain adaptation with new model settings, as describes in section 2.4.

The results for all performed experiments are shown in Supplementary Table 1. AUROC results for transfer learning scheme for selected experiments are presented in Table 2, weighted for 20 runs for each experiment. Two experiments for SARC cancer type with mRNA feature, and UCEC cancer type with protein feature are presented. It can be seen that our model outperformed the baseline approach with 0.535 to 0.676, 0.63 to 0.68, and 0.593 to 0.629 for reported experiments.

Furthermore, the results comparison for EA/AA data distributation has been shown in Figures 4 and 5. In Fig. 5, comparisons of five multiethnic experiments for baseline approach and our model has been shown for UCEC cancer type. It can be observed that results are good for mixture learning 2 and independent learning, as EA data cohort is used for model teasing. The performance of transfer learning with baseline approach LUAD, LUSC, SARC, and UECE cancer types has been in Fig. 4. It can be seen that our model has outperformed the transfer learning experiments performed with baseline approach.

Experiments based on EA/ASA data distribution

In the first step of second phase, we performed all new experiments with ASA cohort as target domain data. We only selected cancer types with enough data available (Supplementary Figure S2). Furthermore, we filtered out the feature combinations from experiments lists where five cases were not present for all four prognosis classes (Supplementary Table 2). The selected cancer types with this data distributation are BLCA, BRCA, COADREAD, ESCA, PanGI, STAD, and STES.

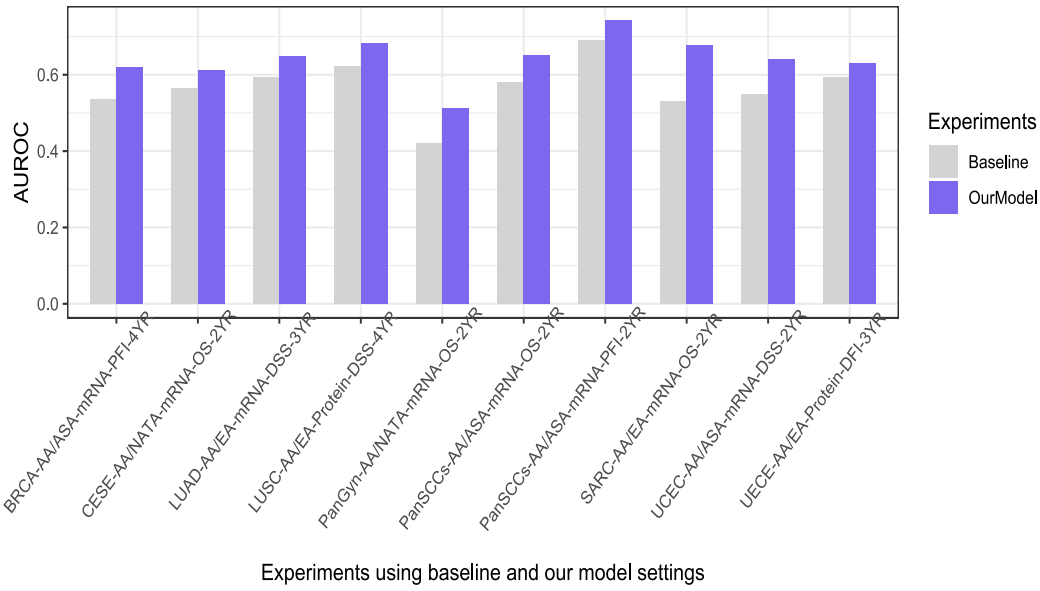


Fig. 4. Comparison of baseline approach and our fine tuning model

The results of this distribution is presented in Fig. 6. As in the previous experiments, mixture learning 2 and independent learning 1 results are better because model testing is done using EA data, while mixture learning 3 and independent learning 2 performance is low because model is tested using Asian Americans data. The comparison of transfer learning schemes has been shown in Fig. 4 for BRCA, PanSCCS, and UCEC cancers.

Experiments based on EA/NAT-A data distribution

In another data distributation in the second step of phase 2, we used Native Americans cohort as target domain data. As the Native American cases are very small in many cancer types, so mainly we cancer categories GBMLGG, PanGybn, PanSCCs, and CESE cancer type. As same in the previous experiments, first we filtered out the experiment combinations with cases less than 5 in all prognosis classes (Supplementary Table 2).

Fig. 7 represents the results using Native Americans as target domain for CESE cancer type, where AUROC for transfer learning schemes is improved from 0.56 to 0.62 for our model, also it is reflected in Table 2.

We have observed that some features combinations in experiments plays significant role than others, such as clinical endpoint outcome and target year combine together. While checking for at least 5 cases for four prognosis classes, as shown in Supplementary Table 2, we noticed that overall survival and disease specific survival clinical endpoint plays a significant role with 4 and 5 target year. Most of available data in major cancer categories from Table S2 has the progression-free interval (PFI) and overall survival rate (OS) with either 4 or 5 target year.

Conclusion

In this study, we worked on the fine tuning of transfer learning and domain adaptation models to reduce the healthcare disparities for data bias in ethnic cohorts of TCGA and MMRF CoMMpass datasets. We performed experiments in two phases, where we used multiple cohorts data as target domains. In the first phase, we used the baseline model data distributions with our fine-tuning models. In the second phase, we used Asian Americans and Native Americans data as target domain. In both phases, we used European American cohort data as source domain, because availability of enough data for transfer learning experiments. To the best of our knowledge, this is the first attempt to use multiple multiethnic data distributions to reduce the healthcare disparities. For both phases, extensive sets of experiments were performed using baseline and our fine-tuning models for comparisons. Results showed that our model improved the transfer learning scheme of multiethnic experiments with better AUROC. In future, we plan to leverage the proposed domain adaptation for COVID-19 multiethnic data.

Key Points

- We proposed the fine-tuning of deep learning and transfer learning models to reduce the healthcare disparities for machine learning models.
- Extensive sets of experiments were performed in two phases based on different ethnic data distributions for transfer learning tasks.
- 33 cancer types from TCGA study are used for European Americans data as source domain and African Americans, Asian Americans, and Native Americans data as traget domains.

Funding

This research was substantially sponsored by the research projects (Grant No. 32170654 and Grant No. 32000464) supported by the National Natural Science Foundation of China and was substantially supported by the Shenzhen Research Institute, City University of Hong Kong. The work described in this paper was substantially supported by the grant from the Health and Medical Research Fund, the Food and Health Bureau, The Government of the Hong Kong Special Administrative Region [07181426]. The work described in this paper was partially supported by the grants from City University of Hong Kong (CityU 11202219, CityU 11203520, CityU 11203221).

Availability

Supplementary data is available here and source code is available at this GitHub repository.

References

1. A lack of data on race hampers efforts to tackle inequalities. https://www.economist.com/leaders/2020/11/21/a-lack-of-data-on-race-hampers-efforts-to-tackle-inequalities.
2. Yue Cao, Thomas Andrew Geddes, Jean Yee Hwa Yang, and Pengyi Yang. Ensemble deep learning in bioinformatics. *Nature Machine Intelligence*, 2(9):500–508, 2020.
3. Binhua Tang, Zixiang Pan, Kang Yin, and Asif Khateeb. Recent advances of deep learning in bioinformatics and computational biology. *Frontiers in genetics*, 10:214, 2019.
4. Haoyang Li, Shuye Tian, Yu Li, Qiming Fang, Renbo Tan, Yijie Pan, Chao Huang, Ying Xu, and Xin Gao. Modern deep learning in bioinformatics. *Journal of molecular cell biology*, 2020.
5. Isaac E Kim Jr and Indra Neil Sarkar. Racial representation disparity of population-level genomic sequencing efforts. In *MedInfo*, pages 974–978.
6. Courtney R Lyles, Mitchell R Lunn, Juno Obedin-Maliver, and Kirsten Bibbins-Domingo. The new era of precision population health: insights for the all of us research program and beyond. *Journal of translational medicine*, 16(1):1–4, 2018.
7. Michael C Gibbons. A historical overview of health disparities and the potential of ehealth solutions. *Journal of medical Internet research*, 7(5):e50, 2005.
8. Santiago Guerrero, Andrés López-Cortés, Alberto Indacochea, Jennyfer M García-Cárdenas, Ana Karina Zambrano, Alejandro Cabrera-Andrade, Patricia Guevara-Ramírez, Diana Abigail González, Paola E Leone, and

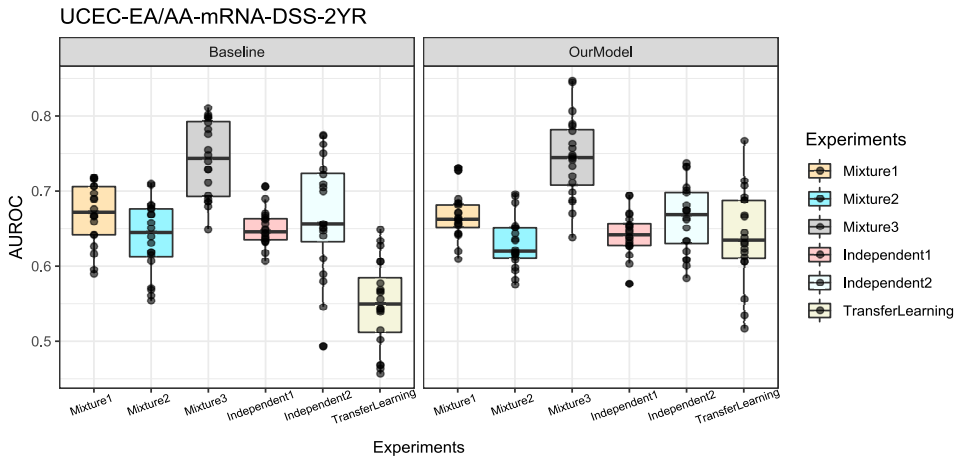


Fig. 5. Multiethnic schemes comparison with baseline model setting using African American cohort as target domain with European American as source domain for UCEC cancer type

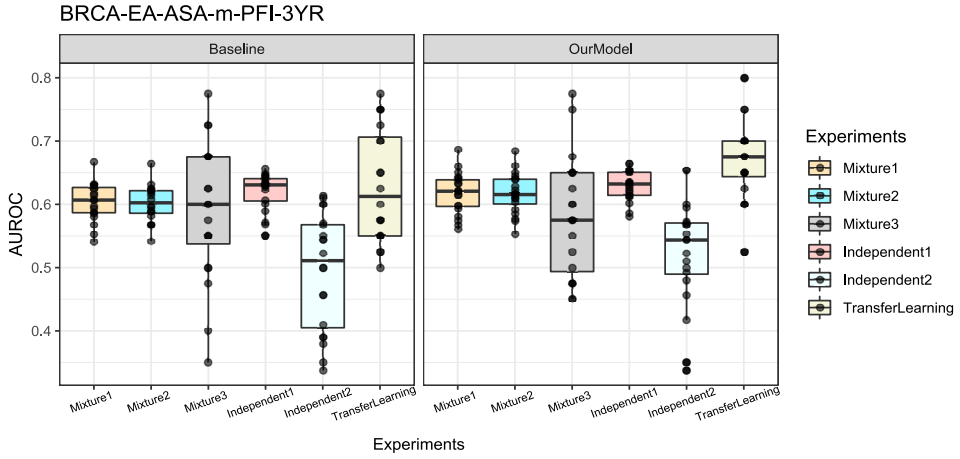


Fig. 6. Multiethnic schemes comparison with baseline model setting using Asian Americans cohort as target domain with European American as source domain for BRCA cancer category

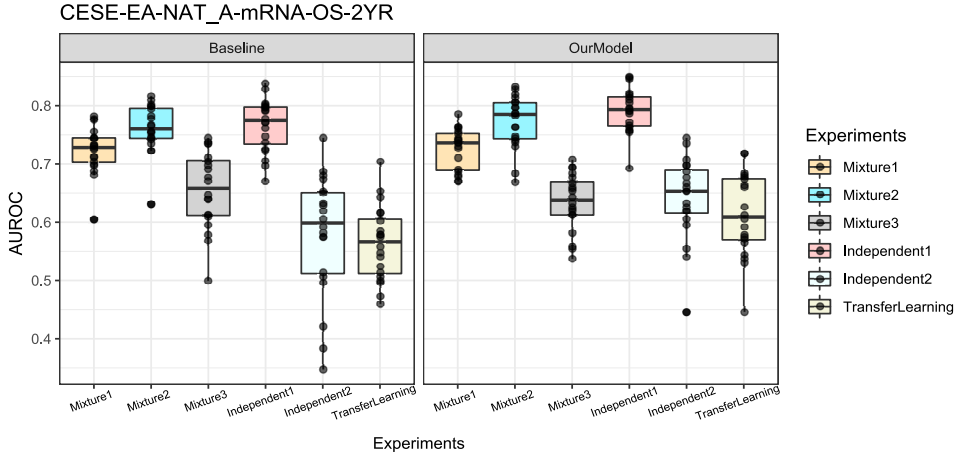


Fig. 7. Multiethnic schemes comparison with baseline model setting using Native Americans cohort as target domain with European American as source domain for CESE cancer category

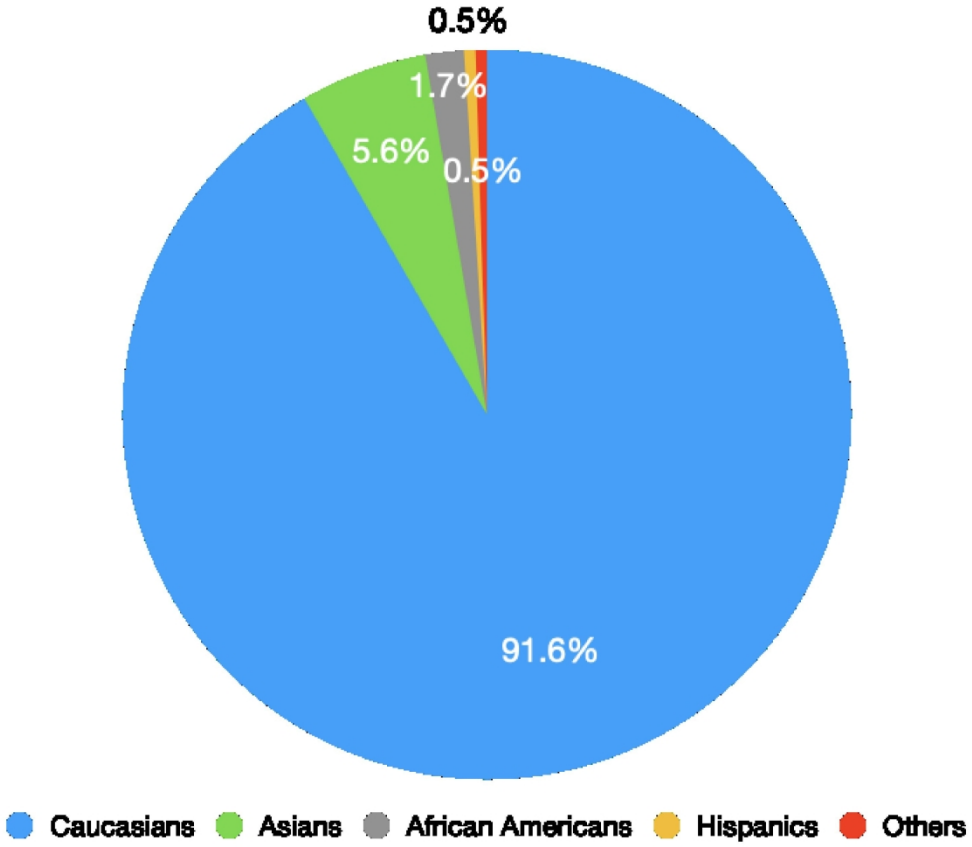
César Paz-y Miño. Analysis of racial/ethnic representation in select basic and applied cancer research studies. *Scientific reports*, 8(1):1–8, 2018.
9. Cathy O'neil. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishing Group, 2016.
10. Safiya Umoja Noble. *Algorithms of oppression: How search engines reinforce racism*. NYU Press, 2018.
11. Virginia Eubanks. *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press, 2018.
12. Carolyn Hutter and Jean Claude Zenklusen. The cancer genome atlas: creating lasting value beyond its data. *Cell*, 173(2):283–285, 2018.
13. Katherine A Hoadley, Christina Yau, Toshinori Hinoue, Denise M Wolf, Alexander J Lazar, Esther Drill, Ronglai Shen, Alison M Taylor, Andrew D Cherniack, and Vésteinn Thorsson. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell*, 173(2):291–304. e6, 2018.
14. Mathias Uhlen, Cheng Zhang, Sunjae Lee, Evelina Sjöstedt, Linn Fagerberg, Gholamreza Bidkhori, Rui Benfeitas, Muhammad Arif, Zhengtao Liu, and Fredrik Edfors. A pathology atlas of the human cancer transcriptome. *Science*, 357(6352), 2017.
15. Tathiane M Malta, Artem Sokolov, Andrew J Gentles, Tomasz Burzykowski, Laila Poisson, John N Weinstein, Bożena Kamińska, Joerg Huelsken, Larsson Omberg, and Olivier Gevaert. Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell*, 173(2):338–354. e15, 2018.
16. Robin M Weinick, Elizabeth A Jacobs, Lisa Cacari Stone, Alexander N Ortega, and Helen Burstin. Hispanic healthcare disparities: challenging the myth of a monolithic hispanic population. *Medical care*, pages 313–320, 2004.
17. Rumi Chunara, Yuan Zhao, Ji Chen, Katharine Lawrence, Paul A Testa, Oded Nov, and Devin M Mann. Telemedicine and healthcare disparities: a cohort study in a large healthcare system in new york city during covid-19. *Journal of the American Medical Informatics Association*, 28(1):33–41, 2021.
18. Jianfang Liu, Tara Lichtenberg, Katherine A Hoadley, Laila M Poisson, Alexander J Lazar, Andrew D Cherniack, Albert J Kovatich, Christopher C Benz, Douglas A Levine, and Adrian V Lee. An integrated tcga pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell*, 173(2):400–416. e11, 2018.
19. NIH. The cancer genome atlas program. https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga, 2006.
20. Christopher I Amos, Joe Dennis, Zhaoming Wang, Jinyoung Byun, Fredrick R Schumacher, Simon A Gayther, Graham Casey, David J Hunter, Thomas A Sellers, and Stephen B Gruber. The oncoarray consortium: a network for understanding the genetic architecture of common cancers. *Cancer Epidemiology and Prevention Biomarkers*, 26(1):126–135, 2017.
21. NCI. Target: Therapeutically applicable research to generate effective treatments. https://ocg.cancer.gov/programs/target.
22. Alicia R Martin, Masahiro Kanai, Yoichiro Kamatani, Yukinori Okada, Benjamin M Neale, and Mark J Daly. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature genetics*, 51(4):584–591, 2019.
23. Alvin Rajkomar, Michaela Hardt, Michael D Howell, Greg Corrado, and Marshall H Chin. Ensuring fairness in machine learning to advance health equity. *Annals of internal medicine*, 169(12):866–872, 2018.
24. Yan Gao and Yan Cui. Deep transfer learning for reducing health care disparities arising from biomedical data inequality. *Nature Communications*, 11(1):1–1, 2020.
25. Jie Lu, Vahid Behbood, Peng Hao, Hua Zuo, Shan Xue, and Guangquan Zhang. Transfer learning using computational intelligence: A survey. *Knowledge-Based Systems*, 80:14–23, 2015.
26. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034.
27. Michael I Jordan and Tom M Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015.
28. Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big data*, 3(1):1–40, 2016.
29. Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, and Chunfang Liu. A survey on deep transfer learning. In *International conference on artificial neural networks*, pages 270–279. Springer.
30. Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.
31. Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.
32. Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*, 79(1-2):151–175, 2010.
33. Jiao Yuan, Zhongyi Hu, Brandon A Mahal, Sihai D Zhao, Kevin H Kensler, Jingjiang Pi, Xiaowen Hu, Youyou Zhang, Yueying Wang, and Junjie Jiang. Integrated analysis of genetic ancestry and genomic alterations across cancers. *Cancer cell*, 34(4):549–560. e9, 2018.
34. MMRF. The relating clinical outcomes in multiple myeloma to personal assessment of genetic profile. https://themmrf.org/we-are-curing-multiplemyeloma/mmrf-commpass-study/.
35. TCGA. The cancer genetic ancestry atlas. http://52.25.87.215/TCGAA.
36. Scott R Goldsmith, Mark A Fiala, James Dukeman, Armin Ghobadi, Keith Stockerl-Goldstein, Mark A Schroeder, Michael Tomasson, Tanya M Wildes, and Ravi Vij. Next generation sequencing-based validation of the revised international staging system for multiple myeloma: an analysis of the mmrf commpass study. *Clinical Lymphoma Myeloma and Leukemia*, 19(5):285–289, 2019.
37. Facial recognition is accurate, if you're a white guy. `https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html`.
38. Wen Xu, Jing He, and Yanfeng Shu. *Transfer Learning and Deep Domain Adaptation*. IntechOpen, 2020.
39. Son Lam Phung and Abdesselam Bouzerdoum. A pyramidal neural network for visual pattern recognition. *IEEE transactions on neural networks*, 18(2):329–343, 2007.

40. Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.
41. Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. On the importance of initialization and momentum in deep learning. In *International conference on machine learning*, pages 1139–1147. PMLR.
42. Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and regression trees*. CRC press, 1984.
43. Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? *arXiv preprint arXiv:1411.1792*, 2014.
44. Tom Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874, 2006.
45. Kelly H Zou, A James O'Malley, and Laura Mauri. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation*, 115(5):654–657, 2007.
46. Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, and Vincent Dubourg. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

Muhammad Toseef is currently a PhD student in the Department of Computer science, City University of Hong Kong, Hong Kong SAR. His research interests include bioinformatics and computational biology.

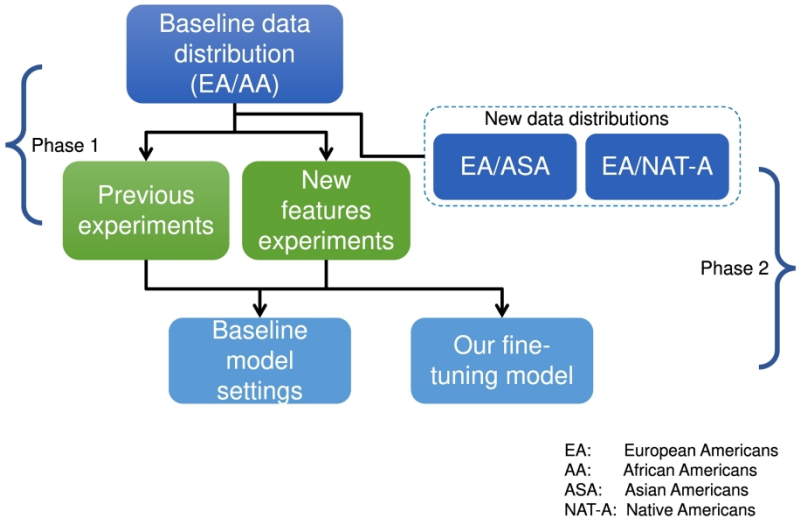
Xiangtao Li is a professor in the School of Artificial Intelligence, Jilin University, Jilin, China. His research interests include bioinformatics, computational biology and evolutionary data mining.

Ka-Chun Wong assumed his duty as an associate professor at City University of Hong Kong, Hong Kong SAR. His research interests include bioinformatics, computational biology, evolutionary computation, data mining, machine learning and interdisciplinary research. He is merited as the associate editor of Bio Data Mining in 2016. In addition, he is on the editorial board of Applied Soft Computing since 2016. Remarkably, he has solely edited two books published by Springer and CRC Press, attracting 30 peer-reviewed book chapters around the world.



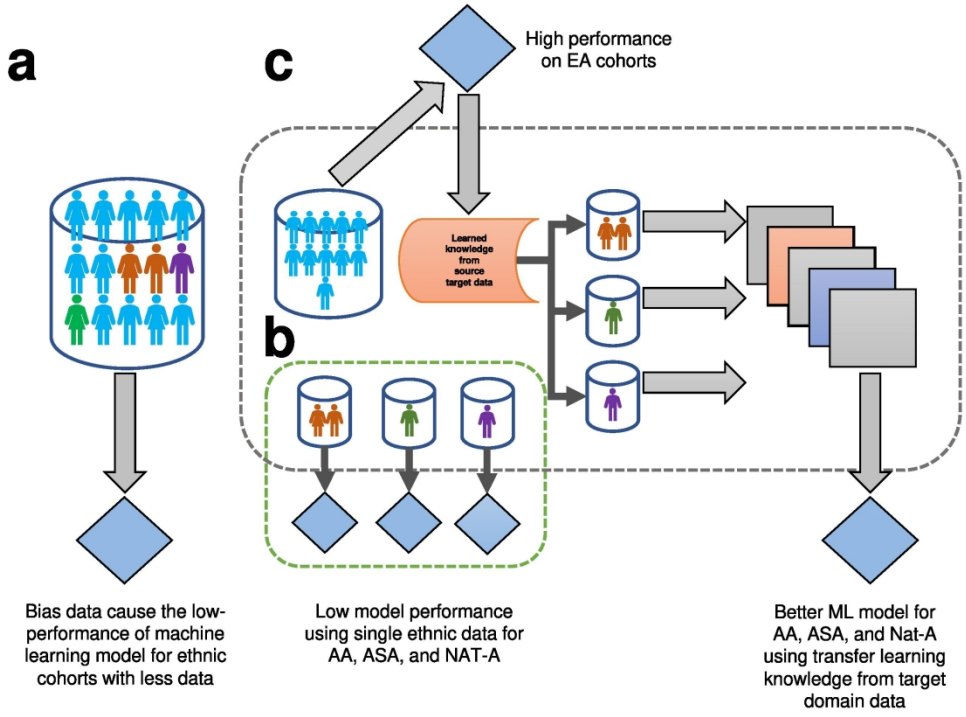
Data inequality among different ethnic races: these data statistics represent collected samples from TARGET, OncoArray, TCGA, and 416 other cancer projects

127x113mm (600 x 600 DPI)



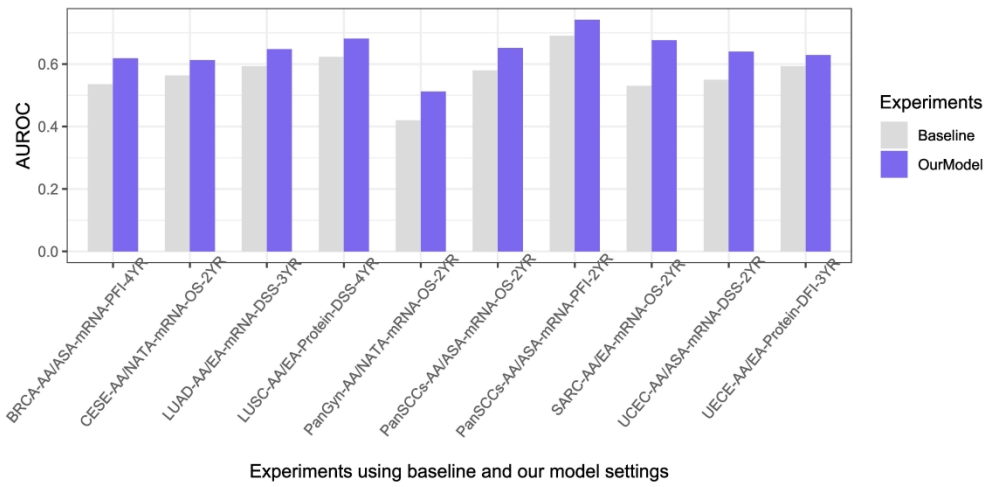
Flowchart of performed experiments with baseline data distribution and using new data cohorts as target domain data. Phase 1 is consists of all experiments performed with data distribution using baseline approach with AA data as the target domain. Phase 2 has two directions: first experiments were performed using ASA data as target domain and then NAT-A data as the target domain

338x190mm (300 x 300 DPI)



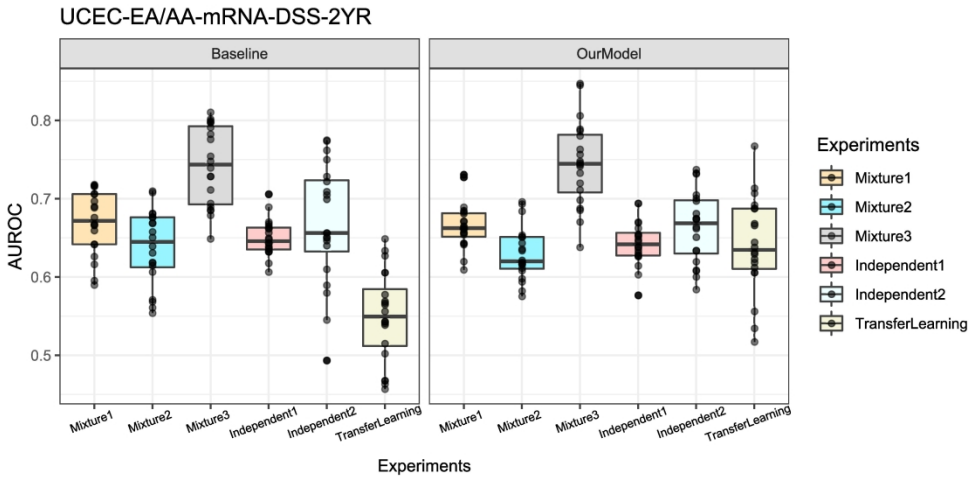
Overall flowchart of the multiethnic experiment schemes. (a) The performance of ML model is bias because of data discrepancy with high performance for EA cohort and low performance for other cohorts; (b) Poor data distributions bring the low performance for AA, ASA, NAT-A, and other cohorts; (c) Transfer learning helps to improve the performance of AA, ASA, and NAT-A cohorts after using the learned knowledge from EA cohorts

115x83mm (600 x 600 DPI)



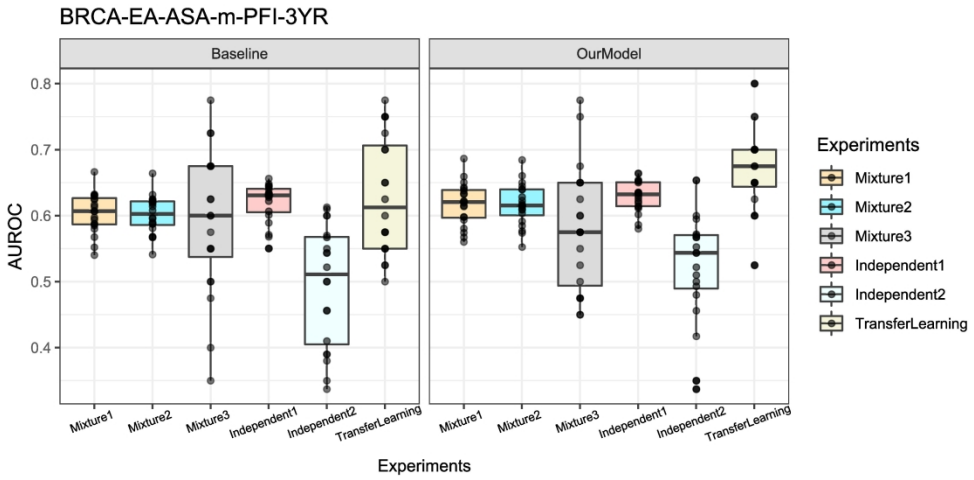
Comparison of baseline approach and our fine-tuning model

209x103mm (600 x 600 DPI)



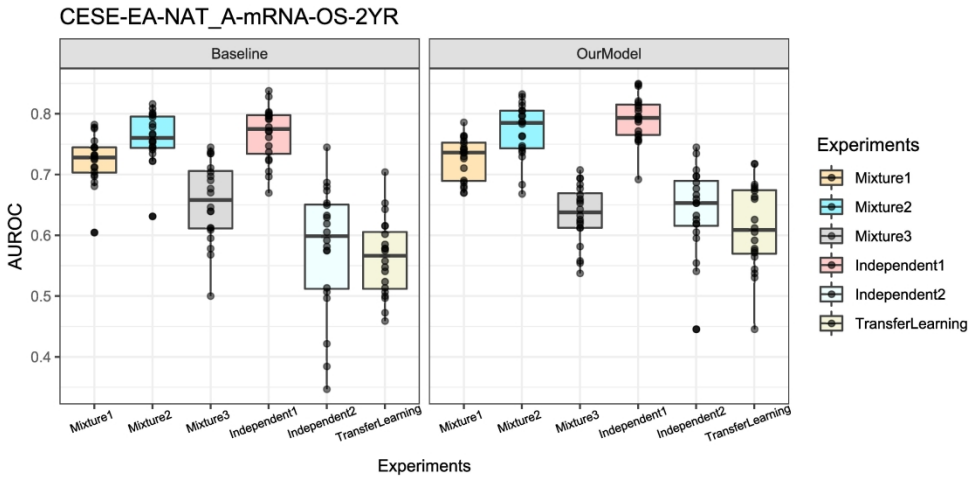
Multiethnic schemes comparison with baseline model setting using African American cohort as target domain with European American as source domain for UCEC cancer type

209x102mm (600 x 600 DPI)



Multiethnic schemes comparison with baseline model setting using Asian Americans cohort as target domain with European American as source domain for BRCA cancer category

209x102mm (600 x 600 DPI)



Multiethnic schemes comparison with baseline model setting using Native Americans cohort as target domain with European American as source domain for CESE cancer category

209x102mm (600 x 600 DPI)

Ethnic disparity reduction with domain adaptation and fine-tuning for healthcare inequality

Muhammad Toseef, Xiangtao Li*, and Ka-Chun Wong*

1. INTRODUCTION

A. Transfer learning

In traditional machine learning, to perform a task *T*, an algorithm learns using the experience *E* with available data [1]. This single machine learning task is isolated and no learned knowledge is retained for later task to learn. On the other hand, transfer learning is helpful to learn the knowledge transferred from the previous task. A pictorial definition of difference between traditional machine learning and transfer learning has been shown in Fig. S1. In traditional machine learning, tow tasks *T1* and *T2* are trained on isolated ML models using datasets A and B, respectively. In case of ML model B, it may not generalize well for Task *T1* and brings the performance degradation because of inadequate and bias data or domain shift in data. Transfer learning helps to learned knowledge in form of weights and features from Task *T1*, where data to train machine learning model is well enough.

In every task of transfer learning, there is a source domain $\mathcal{D}_s$ and target domain $\mathcal{D}_T$ with source and target tasks $\mathcal{T}_S$ and $\mathcal{T}_T$, respectively [2]. A domain $\mathcal{D}$ can be defined as in Eq. S1

$$\mathcal{D} = \{\mathcal{X}, P(X)\} \quad \text{(S1)}$$

where, $\mathcal{X}$ is a feature space, $P(X)$is marginal probability, and *X* is a sample data point [2]. So, every domain in transfer learning consist of two components described above, feature space and marginal probability.

$$P(X) = \{x_1, \ldots, x_n\}, \qquad x_i \in \mathcal{X} \quad \text{(S2)}$$

A domain task $\mathcal{T}$ can be defined as:

$$\mathcal{T} = \{\mathcal{Y}, P(X \mid Y)\} = \{\mathcal{Y}, \eta\} \qquad \mathcal{Y} = \{y_1, \ldots, y_n\}, \quad y_i \in \mathcal{Y} \quad \text{(S3)}$$

where, $\mathcal{Y}$ is label space, η is a predictive function learned from $\{x_i, y_i\}$.

We can define transfer learning where from a source domain $\mathcal{D}_s$ with a source task T_s, the objective is to learn the knowledge from the target conditional distribution $P(Y_T|X_T)$ in target domain $\mathcal{D}_T$. In transfer learning task, $\mathcal{D}_S \neq \mathcal{D}_T$ or $\mathcal{T}_S \neq \mathcal{T}_T$. Domain adaptation is a sub-task of transfer learning where source and domain tasks remain the same but domain shift happens from $\mathcal{D}_s$to $\mathcal{D}_T$, such as, $\mathcal{X}_S \neq \mathcal{X}_T$ [1,2]. Domain adaptation is quite helpful where data shift happens in source and target domains, such in the case of ethnic disparities in the cancer datasets.

In the proposed study, for all experiments, we used the European American (EA) cohort as $\mathcal{D}_s$, while African Americans (AA), Asian Americans (ASA), and Native Americans (NAT-A) cohorts as $\mathcal{D}_T$ in different set of experiments.

B. Cancer Types

The total reported cancer types in this study are 33, but for better data distribution for source domain DS, many single cancer types were composed into main 8 cancer categories, as shown in Supplementary Table S1. In all TCGA cohorts' datasets used in this study, the EA contributes is 80.5%, 9.2% ASA, 6.1% AA, 3.6% NAT-A, and 0.7% others. Based on this distribution for 8 main cancer categories in TCGA cohorts, has been shown in Supplementary Fig. S1.

2. SUPPLEMENTARY FIGURES

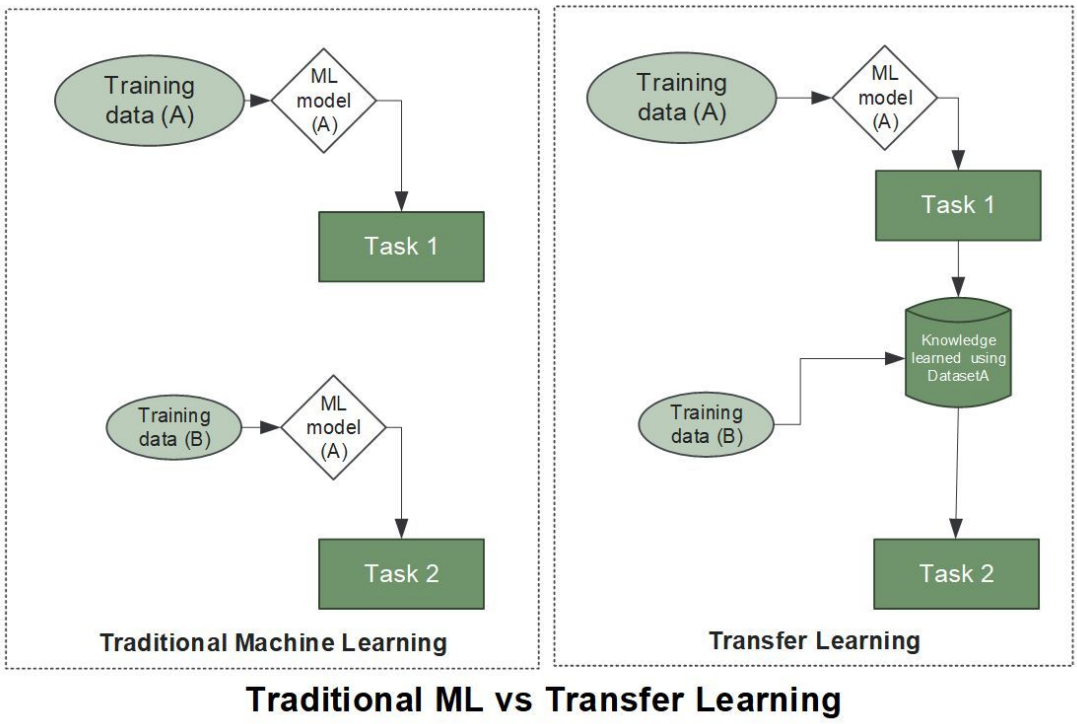


Fig. S1. The comparison between traditional machine learning and transfer learning

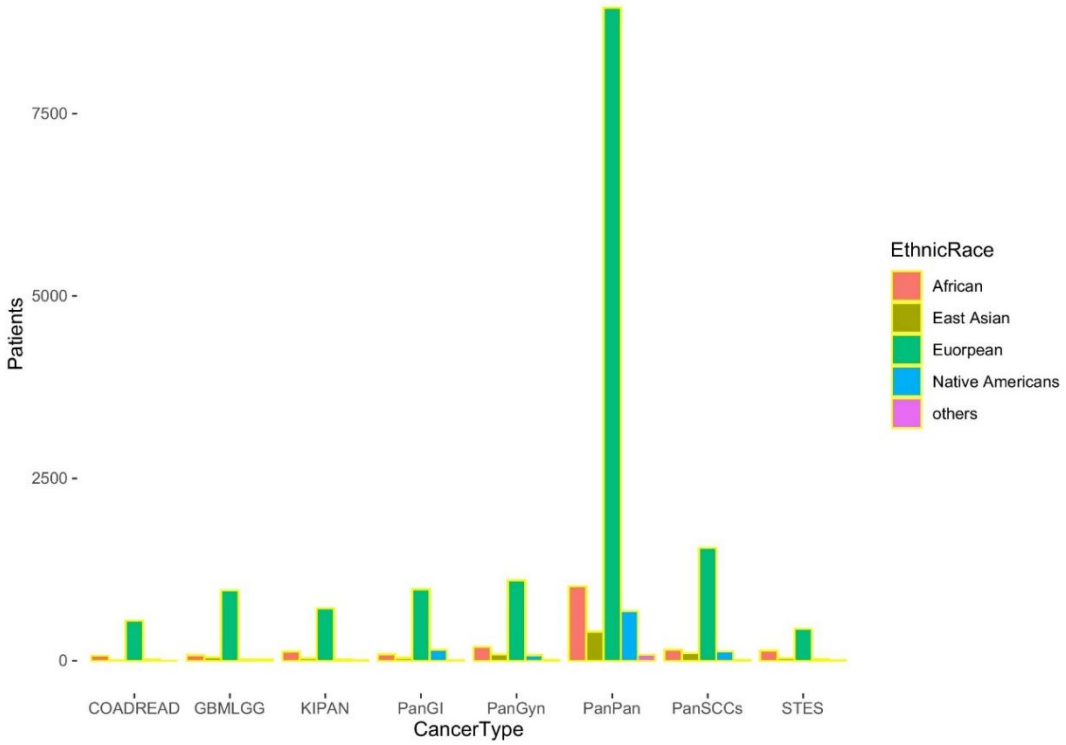


Fig. S2. Major cancer types data distribution for EA, AA, ASA, NAT-A, and others

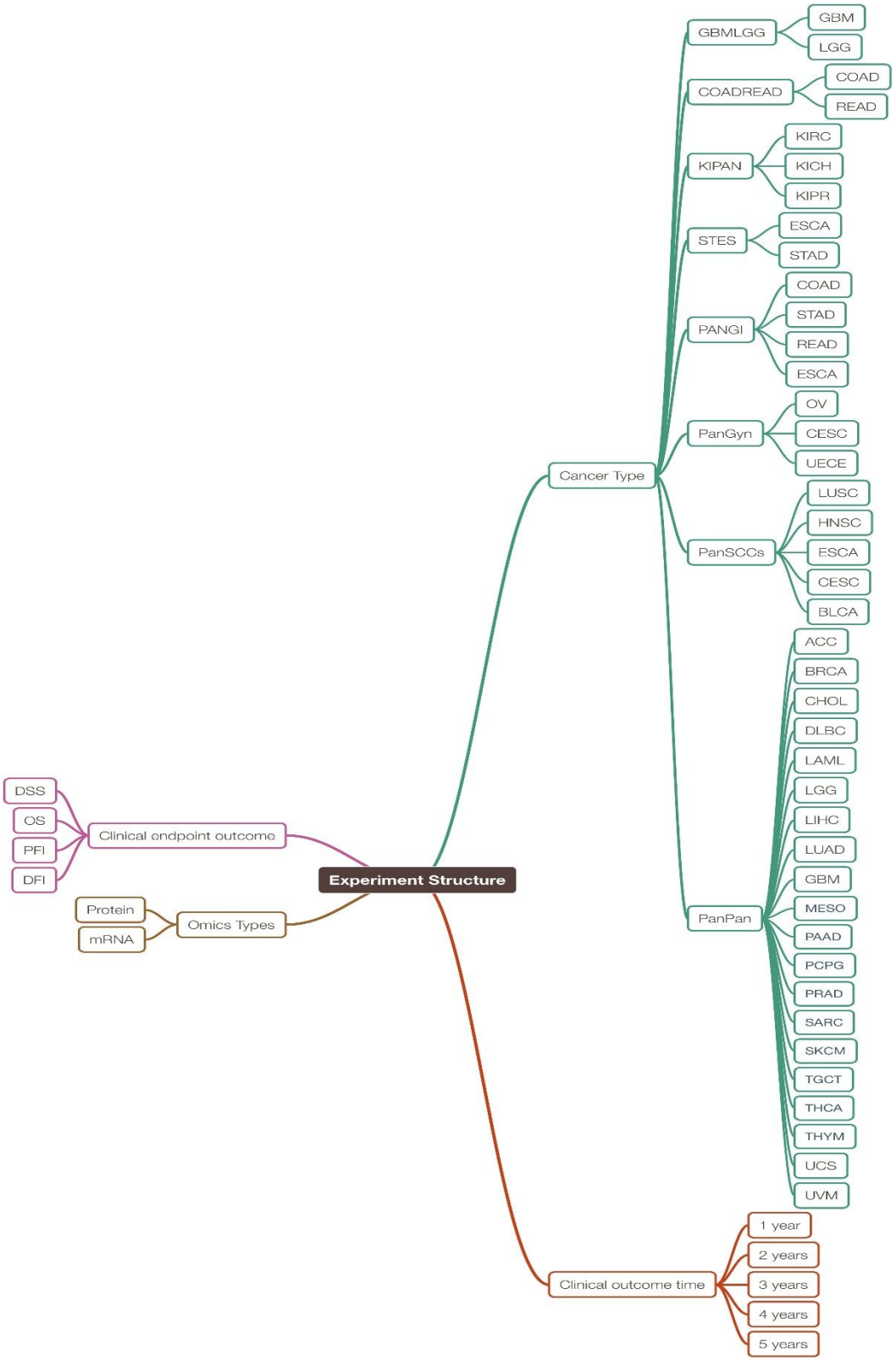


Fig. S3. A mind map representation of experiment features with all possible combinations. Each experiment is consists of 4 features: i) cancer type: 8 cancer categories representing 33 different types of cancers, ii) omics types: represents either mRNA or protein feature, iii) clinical endpoint outcome: this represents the prognosis class with possible outcome of disease, iv) clinical outcome time: this feature shows how many years are associated with clinical end point.

2. SUPPLEMENTARY TABELS

Table S1. 8 Cancer categories for 33 cancer types used in experiments

Sr.	Cancer category	Cancer types
1.	GBMLGG	GBM, LGG
2.	COADREAD	COAD, READ
3.	KIPAN	KIRC, KICH, KIRP
4.	STES	ESCA, STAD
5.	PanGI	COAD, STAD, READ, ESCA
6.	PanGyn	OV, CESC, USC, UCEC
7.	PanSCCs	LUSC, HNSC, ESCA, CESC, BLCA
8.	PanPan	ACC, BLCA, BRCA, CESC, COAD, DLBC, ESCA, GBM, HNSC, KICH, KIRC, KIRP, LAML, LGG, LIHC, LUAD, LUSC, MESO, OV, PAAD, PCPG, PRAD, READ, SARC, SKCM, STAD, TGCT, THCA, THYM, UCEC, UCS, UVM

Table S2. Hyperparameter values for our fine-tuning algorithm

Hyperparameter	Value
Learning rate	0.00001
Batch size	16
Dropout	0.5
Train epoch	100
Epochs	20
L1 reg	0.0001
L2 reg	0.0001

We have performed new data distributions experiments with baseline features combinations with at least five cases presents for prognosis classes. There are two prognosis classes in the baseline paper: either alive or dead, with clinical endpoint outcome feature. As the authors mentioned in the baseline paper, they only selected experiments with at least 5 cases for each prognosis class for source and target domains. So, for fair comparisons with the baseline paper, first, we checked if both source and target domain meet this condition. We only worked on the experiments where at least 5 cases were present in all classes. We categorize these experiments as ‘completed’ or ‘not-completed’ as shown in Table S3**.** Completed and not-completed experiments with baseline features combinations with new data distributions.

Table S3. Completed and not-completed experiments with baseline features combinations with new data distributions

Multi-ethnic data distributions		Completed	Not-completed	Total
Source domain	Target domain			
EA	ASA	46	178	224
EA	NAT-A	50	174	224

In our fine-tuning models for baseline and new data distributions, the AUROC score has improved significantly for diseases (Figure S2) where enough target domain data was available. In Table 4, we can see that the AUROC score has improved from 0.787 to 0.857 for the BLCA-Protein-PFI-4YR experiment using Native Americans as target domain data with the baseline model and our model. In another experiment, using Native Americans data as target domain, our model performed better as compared to baseline model with this new data distribution. For experiment using 'CESE-mRNA-DSS-2YR' features combination, AUROC score improves from 0.544 to 0.60. Furthermore, experiments with new and old data distributions with baseline and our model settings has been given in Supplementary Table 1. We have showed the complete 20 runs of UCEC-mRNA-DSS-2 experiment using EA→AA data distribution for baseline and our fine-tuning model, as shown in Table S3 and Table 4, respectively.

Table S4. AUROC score for UCEC-mRNA-DSS-2 with 3-folds (baseline model)

Sr. No.	folds	Mixture learning 1	Mixture learning 2	Mixture learning 3	Independent learning 1	Independent learning 2	Transfer learning
1	3	0.666581	0.614612	0.782328	0.606393	0.650862	0.605603
2	3	0.59541	0.553881	0.689655	0.661644	0.721983	0.456897
3	3	0.715446	0.680594	0.790948	0.639269	0.65625	0.538793
4	3	0.616168	0.571005	0.728448	0.632877	0.579741	0.633621
5	3	0.65936	0.639269	0.693966	0.705708	0.761853	0.540948
6	3	0.589866	0.568493	0.648707	0.662557	0.699353	0.467672
7	3	0.641955	0.606393	0.711207	0.649543	0.493534	0.564655
8	3	0.666065	0.650457	0.678879	0.656621	0.728448	0.556034
9	3	0.675993	0.618037	0.797414	0.635845	0.640086	0.568966
10	3	0.688499	0.657763	0.799569	0.63242	0.773707	0.56681
11	3	0.69688	0.676941	0.728448	0.648174	0.646552	0.577586
12	3	0.667612	0.630822	0.75431	0.643379	0.649784	0.515086
13	3	0.689917	0.668493	0.747845	0.67032	0.704741	0.627155
14	3	0.626225	0.560731	0.775862	0.61758	0.774784	0.467672
15	3	0.706163	0.66895	0.801724	0.636986	0.58944	0.648707
16	3	0.718025	0.676027	0.810345	0.667808	0.709052	0.543103
17	3	0.705905	0.707763	0.685345	0.632192	0.65625	0.463362
18	3	0.64131	0.618721	0.685345	0.665068	0.545259	0.605603
19	3	0.716735	0.681507	0.797414	0.689269	0.75	0.502155
20	3	0.716864	0.709817	0.739224	0.643151	0.609914	0.543103

Table S5. AUROC score for UCEC-mRNA-DSS-2 with 4-folds (our fine-tuning model)

Sr. No.	folds	Mixture learning 1	Mixture learning 2	Mixture learning 3	Independent learning 1	Independent learning 2	Transfer learning
1	4	0.680892	0.651142	0.756466	0.647717	0.633621	0.6875
2	4	0.609206	0.575342	0.6875	0.634475	0.732759	0.631466
3	4	0.643244	0.617123	0.698276	0.654566	0.737069	0.62931
4	4	0.68476	0.635616	0.788793	0.638356	0.704741	0.69181
5	4	0.664776	0.61895	0.74569	0.645205	0.633621	0.637931
6	4	0.641181	0.582192	0.786638	0.669178	0.600216	0.605603
7	4	0.663486	0.616438	0.762931	0.651826	0.681034	0.668103
8	4	0.670835	0.63379	0.741379	0.576712	0.674569	0.713362
9	4	0.727308	0.684018	0.846983	0.653425	0.700431	0.767241
10	4	0.619778	0.597945	0.670259	0.626256	0.731681	0.612069
11	4	0.642728	0.608219	0.719828	0.67032	0.697198	0.556034
12	4	0.656782	0.621233	0.747845	0.627854	0.650862	0.534483
13	4	0.671351	0.651142	0.711207	0.662329	0.661638	0.706897
14	4	0.661423	0.593836	0.844828	0.626484	0.674569	0.517241
15	4	0.73066	0.696119	0.806034	0.638128	0.663793	0.665948
16	4	0.683084	0.653425	0.732759	0.630594	0.584052	0.605603
17	4	0.689144	0.693151	0.637931	0.614384	0.673491	0.644397
18	4	0.656266	0.643379	0.685345	0.693836	0.619612	0.618534
19	4	0.654203	0.611416	0.743534	0.694064	0.608836	0.622845
20	4	0.661294	0.616438	0.780172	0.602968	0.607759	0.6875

REFERENCES

1. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning (MIT Press, 2016). http://www.deeplearningbook.org.

2. S. J. Pan and Q. Yang, “A survey on transfer learning,” IEEE Transactions on knowledge dataengineering22, 1345–1359 (2009).

Experiment setting	Peformed experiment	Soure/Target (data)	Model settings	Algorithm	Folds	Epochs	AUROC for Mixture 0	AUROC for Mixture 1	AUROC for Mixture 2	AUROC for Independent 1	AUROC for Independent 2	AUROC for Transfer Learning	Min of TL	Max of TL	Average of TL
BLCA-mRNA-PFI-1YR		EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.652207921	0.647265536	0.701234568	0.651846688	0.725617284	0.778395062	0.62963	0.950617	0.778395062
			Our Fine-tuning model		4	20	0.638851485	0.632557478	0.719753086	0.644070112	0.70617284	0.82345679	0.703704	0.938272	0.82345679
BLCA-mRNA-PFI-3YR		EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.614263441	0.594553571	0.945	0.586095238	0.7525	0.7675	0.55	1	0.7675
			Our Fine-tuning model		4	20	0.615645161	0.596309524	0.9275	0.589827381	0.1025	0.9375	0.75	1	0.9375
BLCA-mRNA-PFI-4YR		EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.610882353	0.592533508	0.915	0.592671911	0.7525	0.7875	0.6	1	0.7875
			Our Fine-tuning model		4	20	0.609418301	0.594383741	0.84	0.58608683	0.1025	0.8575	0.55	1	0.8575
BLCA-mRNA-PFI-5YR		EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.506749538	0.493843537	0.705	0.499285714	0.7	0.22	0	0.4	0.22
			Our Fine-tuning model		4	20	0.506749538	0.493843537	0.705	0.499285714	0.7	0.29	0	0.6	0.29
BLCA-Protein-PFI-3YR		EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.614263441	0.594553571	0.945	0.586095238	0.7525	0.7675	0.55	1	0.7675
			Our Fine-tuning model		4	20	0.615645161	0.596309524	0.9275	0.589827381	0.1025	0.9375	0.75	1	0.9375
BLCA-Protein-PFI-4YR		EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.610882353	0.592533508	0.915	0.592671911	0.7525	0.7875	0.6	1	0.7875
			Our Fine-tuning model		4	20	0.609418301	0.594383741	0.84	0.58608683	0.1025	0.8575	0.55	1	0.8575
BLC-Protein-PFI-5YR		EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.506749538	0.493843537	0.705	0.499285714	0.7	0.22	0.1	0.4	0.22
			Our Fine-tuning model		4	20	0.506749538	0.493843537	0.705	0.499285714	0.7	0.29	0.18	0.6	0.29
GBMLGG-mRNA-DSS-2YR		EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.890030719	0.891120584	0.885227273	0.891375037	0.835606061	0.931818182	0.863636	0.969697	0.931818182
			Our Fine-tuning model		4	20	0.892392627	0.893830343	0.879924242	0.892663825	0.852462121	0.931439394	0.893939	0.969697	0.931439394
UCEC-mRNA-DSS-2YR	Old	EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.670048994	0.638013699	0.742349138	0.649840183	0.667079741	0.549676724	0.456897	0.648707	0.549676724
			Our Fine-tuning model		4	20	0.665620165	0.630045662	0.744719828	0.64293379	0.663577586	0.640193966	0.517241	0.767241	0.640193966
UCEC-mRNA-DSS-3YR	Old	EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.740126812	0.721621315	0.768581081	0.704342404	0.735754505	0.697072072	0.5	0.813063	0.697072072
			Our Fine-tuning model		4	20	0.75938794	0.739818594	0.786261261	0.72521542	0.735247748	0.667905405	0.583333	0.77027	0.667905405
UCEC-mRNA-DSS-4YR	Old	EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.721733649	0.705170455	0.730133333	0.707855114	0.720933333	0.676	0.568	0.757333	0.676
			Our Fine-tuning model		4	20	0.721323877	0.709190341	0.7156	0.709602273	0.7208	0.6476	0.522667	0.768	0.6476
UCEC-mRNA-DSS-4YR	New	EA/ASA	Baseline	Fine-tuning algorithm 1	3	20	0.710967813	0.715525568	0.671875	0.709105114	0.360546875	0.52421875	0.328125	0.65625	0.52421875
			Our Fine-tuning model		4	20	0.715575397	0.719559659	0.67890625	0.7103125	0.36015625	0.55078125	0.40625	0.6875	0.55078125
UCEC-mRNA-DSS-5YR	Old	EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.703229424	0.708712121	0.662171053	0.712794613	0.712417763	0.58125	0.480263	0.694079	0.58125
			Our Fine-tuning model		4	20	0.707423189	0.71233165	0.670559211	0.697685185	0.733881579	0.585690789	0.516447	0.720395	0.585690789
UCEC-mRNA-DSS-5YR	New	EA/ASA	Baseline	Fine-tuning algorithm 1	3	20	0.728178747	0.738215488	0.59921875	0.713257576	0.360546875	0.52578125	0.40625	0.625	0.52578125
			Our Fine-tuning model		4	20	0.709014128	0.716435185	0.60234375	0.695159933	0.36015625	0.54375	0.421875	0.625	0.54375
UCEC-mRNA-OS-1YR	New	EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.596075795	0.559200841	0.70673913	0.614437434	0.702173913	0.517934783	0.421739	0.684783	0.517934783
			Our Fine-tuning model		4	20	0.588037897	0.549001052	0.705978261	0.581871714	0.702173913	0.662717391	0.55	0.767391	0.662717391
UCEC-mRNA-OS-2YR	Old	EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.62443902	0.609166667	0.688314176	0.635461433	0.711637931	0.475383142	0.37931	0.614943	0.475383142
			Our Fine-tuning model		4	20	0.611523638	0.586239669	0.711111111	0.635750689	0.719683908	0.494061303	0.346743	0.672414	0.494061303
UCEC-mRNA-OS-3YR	Old	EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.642925022	0.618764077	0.714189189	0.613914696	0.704826255	0.629054054	0.505792	0.745174	0.629054054
			Our Fine-tuning model		4	20	0.637092415	0.612753378	0.720559846	0.620432151	0.683832046	0.627220077	0.559846	0.716216	0.627220077
UCEC-mRNA-OS-3YR	New	EA/ASA	Baseline	Fine-tuning algorithm 1	3	20	0.629481793	0.622846284	0.65	0.616420327	0.503333333	0.651666667	0.483333	0.75	0.651666667
			Our Fine-tuning model		3	20	0.629481793	0.622846284	0.65	0.616420327	0.503333333	0.534166667	0.383333	0.716667	0.534166667
UCEC-mRNA-OS-4YR	Old	EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.631550265	0.610611888	0.682777778	0.617097902	0.661555556	0.690666667	0.582222	0.782222	0.690666667
			Our Fine-tuning model		4	20	0.637534392	0.620087413	0.677888889	0.622159091	0.661833333	0.623666667	0.531111	0.704444	0.623666667
UCEC-mRNA-OS-5YR	Old	EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.673754328	0.677559156	0.657202216	0.683088992	0.672645429	0.619113573	0.473684	0.734072	0.619113573
			Our Fine-tuning model		4	20	0.683222941	0.690432099	0.647645429	0.686432613	0.669390582	0.641412742	0.506925	0.772853	0.641412742
UCEC-mRNA-PFI-1YR	New	EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.523956152	0.492894345	0.652485795	0.467107781	0.59350142	0.536221591	0.321023	0.663352	0.536221591
			Our Fine-tuning model		4	20	0.501662304	0.478295068	0.589985795	0.459210247	0.576100852	0.556747159	0.420455	0.774148	0.556747159
UCEC-mRNA-PFI-2YR	New	EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.594528689	0.57860409	0.651	0.556838211	0.680944444	0.637555556	0.576101	0.741111	0.637555556
			Our Fine-tuning model		4	20	0.59901127	0.5844571	0.6525	0.577548221	0.670444444	0.521277778	0.364444	0.665556	0.521277778
UCEC-mRNA-PFI-3YR	Old	EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.657686285	0.637719392	0.706041667	0.622790026	0.663576389	0.708125	0.597222	0.801389	0.708125
			Our Fine-tuning model		4	20	0.663483623	0.639415668	0.727847222	0.623169949	0.668055556	0.7075	0.631944	0.815278	0.7075
UCEC-mRNA-PFI-4YR	Old	EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.657468437	0.631768069	0.718622449	0.636634	0.694855442	0.725170068	0.676871	0.802721	0.725170068
			Our Fine-tuning model		4	20	0.665436957	0.63956717	0.732057823	0.635177624	0.68082483	0.674744898	0.54932	0.823129	0.674744898
UCEC-mRNA-PFI-5YR	Old	EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.632593795	0.633135417	0.622794118	0.624947917	0.585556723	0.650105042	0.556723	0.758403	0.650105042
			Our Fine-tuning model		4	20	0.632864358	0.631989583	0.627836134	0.639875	0.607195378	0.626995798	0.523109	0.72479	0.626995798
UCEC-Protein-DSS-1YR	New	EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.752166172	0.749968072	0.727467105	0.704789272	0.413980263	0.720394737	0.720395	0.720395	0.720394737
			Our Fine-tuning model		4	20	0.716780415	0.688473819	0.733223684	0.680970626	0.507648026	0.6625	0.657895	0.664474	0.6625
KIPAN-mRNA-DSS-1YR	Old	EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.753435807	0.750193211	0.773333333	0.749794324	0.609404762	0.855952381	0.792063	0.919048	0.855952381
			Our Fine-tuning model		4	20	0.767749877	0.760985997	0.805396825	0.739736984	0.624166667	0.829603175	0.687302	0.920635	0.829603175
KIPAN-mRNA-DSS-2YR	Old	EA/AA	Baseline	Fine-tuning algorithm 1	3	20	0.78543461	0.785742097	0.790816327	0.766879886	0.736791383	0.759637188	0.707483	0.823129	0.759637188
			Our Fine-tuning model		4	20									
CESE-mRNA-OS-2YR	New	EA/NAT_A	Baseline	Fine-tuning algorithm 1	3	20	0.723329193	0.760764944	0.655102041	0.766514691	0.578401361	0.563435374	0.459184	0.704082	0.563435374
			Our Fine-tuning model		4	20	0.726111801	0.771453901	0.63452381	0.79154002	0.640816327	0.612755102	0.445578	0.717687	0.612755102
CESE-mRNA-OS-3YR	New	EA/NAT_A	Baseline	Fine-tuning algorithm 1	3	20	0.665817469	0.707333333	0.585561497	0.726128205	0.629411765	0.52644385	0.358289	0.695187	0.52644385
			Our Fine-tuning model		4	20	0.679269317	0.724871795	0.578609626	0.739307692	0.639572193	0.542379679	0.331551	0.690374	0.542379679
CESE-mRNA-DSS-1YR	New	EA/NAT_A	CESE-mRNA-DSS-1YR	Fine-tuning algorithm 1	3	20	0.592503925	0.676309524	0.499107143	0.681428571	0.637946429	0.439434524	0	0.627976	0.439434524
			Our Fine-tuning model		4	20	0.632437206	0.672797619	0.572470238	0.665595238	0.65922619	0.50952381	0.342262	0.72619	0.50952381
CESE-mRNA-DSS-2YR	New	EA/NAT_A	Baseline	Fine-tuning algorithm 1	3	20	0.719285714	0.777266484	0.635544218	0.742135989	0.578571429	0.544897959	0.394558	0.670068	0.544897959
			Our Fine-tuning model		4	20	0.741026786	0.807177198	0.640646259	0.780940934	0.640646259	0.595748299	0.462585	0.663265	0.595748299
CESE-mRNA-DSS-3YR	New	EA/NAT_A	Baseline	Fine-tuning algorithm 1	3	20	0.669965812	0.709544837	0.588636364	0.707302989	0.653977273	0.571590909	0.426136	0.698864	0.571590909
			Our Fine-tuning model		4	20	0.692136752	0.733967391	0.636647727	0.747826087	0.664488636	0.6	0.431818	0.721591	0.6
BRCA-mRNA-PFI-3YR	New	EA/ASA	Baseline	Fine-tuning algorithm 1	3	20	0.603760046	0.601586022	0.59125	0.620305474	0.5	0.63	0.5	0.775	0.63
			Our Fine-tuning model		4	20	0.617413671	0.616666667	0.59	0.629068915	0.5	0.67	0.525	0.8	0.67
BRCA-mRNA-PFI-4YR	New	EA/ASA	BRCA-mRNA-PFI-4YR	Fine-tuning algorithm 1	3	20	0.599155612	0.594019608	0.621428571	0.599414566	0.5	0.535714286	0.371429	0.657143	0.535714286
			Our Fine-tuning model		4	20	0.611489796	0.607789916	0.635714286	0.594515406	0.5	0.618571429	0.542857	0.771429	0.618571429
STAD-mRNA-OS-1YR	New	EA/ASA	Baseline	Fine-tuning algorithm 1	3	20	0.592549864	0.612457429	0.530363636	0.628437633	0.52	0.602363636	0.5	0.660909	0.602363636
			Our Fine-tuning model		4	20	0.587584091	0.609780758	0.519363636	0.627697957	0.517909091	0.639363636	0.575455	0.73	0.639363636
PanGyn-mRNA-OS-2YR	New	EA/NAT_A	PanGyn-mRNA-OS-2YR	Fine-tuning algorithm 1	3	20	0.56407185	0.57721902	0.480842912	0.571641341	0.474042146	0.420210728	0.295019	0.505747	0.420210728
			Our Fine-tuning model		4	20	0.547211602	0.559207493	0.457758621	0.569409222	0.534003831	0.512164751	0.436782	0.613027	0.512164751
PanGyn-mRNA-OS-3YR	New	EA/NAT_A	Baseline	Fine-tuning algorithm 1	3	20	0.633509301	0.647446998	0.492105263	0.657872633	0.516873065	0.405727554	0.291022	0.513932	0.405727554
			Our Fine-tuning model		4	20	0.638691973	0.651365917	0.506656347	0.663770837	0.53126935	0.431733746	0.337461	0.575851	0.431733746
PanGyn-mRNA-OS-4YR	New	EA/NAT_A	Baseline	Fine-tuning algorithm 1	3	20	0.655004915	0.665431672	0.517763158	0.672044045	0.602631579	0.379276316	0.256579	0.532895	0.379276316
			Our Fine-tuning model		4	20	0.66604855	0.676760498	0.5	0.677329274	0.579276316	0.402631579	0.256579	0.506579	0.402631579

ESCA-mRNA-OS-1YR	New	EA/ASA	Baseline	Fine-tuning algorithm 1	3	20	0.509510582	0.544960861	0.335714286	0.570572407	0.382857143	0.46244898	0.306122	0.697959	0.46244898
			Our Fine-tuning model		4	20	0.509656085	0.546722114	0.324693878	0.568909002	0.360408163	0.506530612	0.346939	0.787755	0.506530612
PanSCCs-mRNA-OS-2YR	New	EA/ASA	Baseline	Fine-tuning algorithm 1	3	20	0.623710155	0.625462048	0.670636364	0.618727354	0.577045455	0.579727273	0.469091	0.649091	0.579727273
			Our Fine-tuning model		4	20	0.616569681	0.617813462	0.679636364	0.61453652	0.581181818	0.651727273	0.565455	0.734545	0.651727273
PanSCCs-mRNA-PFI-2YR	New	EA/ASA	Baseline	Fine-tuning algorithm 1	3	20	0.635427677	0.627396524	0.73445122	0.624303819	0.686699695	0.69070122	0.596037	0.810976	0.69070122
			Our Fine-tuning model		4	20	0.642937779	0.634240975	0.755868902	0.624402454	0.690853659	0.741768293	0.63872	0.833841	0.741768293
PanSCCs-mRNA-PFI-3YR	New	EA/ASA	Baseline	Fine-tuning algorithm 1	3	20	0.628119626	0.617995653	0.81765625	0.609878066	0.598359375	0.8275	0.675	0.91875	0.8275
			Our Fine-tuning model		4	20	0.625868865	0.614160109	0.833125	0.59466121	0.665390625	0.86625	0.809375	0.928125	0.86625
PanSCCs-mRNA-OS-3YR	New	EA/ASA	Baseline	Fine-tuning algorithm 1	3	20	0.667223319	0.671348159	0.553532609	0.671203427	0.363949275	0.45317029	0.338768	0.547101	0.45317029
			Our Fine-tuning model		4	20	0.667427496	0.669996719	0.584601449	0.665146555	0.367980072	0.458423913	0.351449	0.577899	0.458423913
PanGyn-mRNA-DSS-4YR	New	EA/NAT_A	Baseline	Fine-tuning algorithm 1	3	20	0.561745999	0.577661417	0.466374269	0.598068635	0.487768031	0.401364522	0.28655	0.502924	0.401364522
			Our Fine-tuning model		4	20	0.553402711	0.569182745	0.448635478	0.591366254	0.518079922	0.435185185	0.274854	0.551657	0.435185185

EA	European Americans
AA	African Americans
ASA	Asian Americans
NAT-A	Native Americans
OTHER	Other ethnic races
ACC	Adrenocortical carcinoma
BLCA	Bladder urothelial carcinoma
BRCA	Breast invasive carcinoma
CESC	Cervical and endocervical cancers
CHOL	Cholangiocarcinoma
COAD	Colon adenocarcinoma
DLBC	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma
ESCA	Esophageal carcinoma
GBM	Glioblastoma multiforme
HNSC	Head and Neck squamous cell carcinoma
KICH	Kidney Chromophobe
KIRC	Kidney renal clear cell carcinoma
KIRP	Kidney renal papillary cell carcinoma
LAML	Acute Myeloid Leukemia
LGG	Brain Lower Grade Glioma
LIHC	Liver hepatocellular carcinoma
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
MESO	Mesothelioma
OV	Ovarian serous cystadenocarcinoma
PAAD	Pancreatic adenocarcinoma
PCPG	Pheochromocytoma and Paraganglioma
PRAD	Prostate adenocarcinoma
READ	Rectum adenocarcinoma
SARC	Sarcoma
SKCM	Skin Cutaneous Melanoma
STAD	Stomach adenocarcinoma
TGCT	Testicular Germ Cell Tumors
THCA	Thyroid carcinoma
THYM	Thymoma
UCEC	Uterine Corpus Endometrial Carcinoma
UCS	Uterine Carcinosarcoma
UVM	Uveal Melanoma
COADREAD	Colorectal adenocarcinoma (COAD+READ)
GBMLGG	Glioma (GBM+LGG)
STES	Stomach and Esophageal carcinoma
PanGyn	Pan-gynecological cancers (OV+CESC+USC+UCEC)
PanGI	Pan-gastrointestinal cancers (COAD+READ+STAD+ESCA)
PanSCC	Pan-squamous cancers (LUSC+HNSC+ESCA+CESC+BLCA)
KIPAN	Pan-kidney cancers (KICH+KIRC+KIRP)
MM	Multiple Myeloma

Experiment	Source/Target	Status	White alive	White dead	Nat-A alive	Native dead	Total White	Total Nat-A	Total alive	Total dead	Total cases
CESC Protein OS 1	EA/NAT_A	Not complete	91	3	13	1	94	14	104	4	108
CESC Protein OS 2	EA/NAT_A	Not complete	61	5	5	1	66	6	66	6	72
CESC Protein OS 3	EA/NAT_A	Not complete	37	2	12	2	39	14	49	4	53
CESC Protein OS 4	EA/NAT_A	Not complete	27	2	14	1	29	15	41	3	44
CESC Protein OS 5	EA/NAT_A	Not complete	24	14	0	2	38	2	24	16	40
CESC Protein DSS 1	EA/NAT_A	Not complete	89	0	13	1	89	14	102	1	103
CESC Protein DSS 2	EA/NAT_A	Not complete	59	2	5	1	61	6	64	3	67
CESC Protein DSS 3	EA/NAT_A	Not complete	36	8	2	2	44	4	38	10	48
CESC Protein DSS 4	EA/NAT_A	Not complete	26	9	1	2	35	3	27	11	38
CESC Protein DSS 5	EA/NAT_A	Not complete	23	9	0	2	32	2	23	11	34
CESC Protein PFI 1	EA/NAT_A	Not complete	87	5	13	1	92	14	100	6	106
CESC Protein PFI 2	EA/NAT_A	Not complete	53	13	5	1	66	6	58	14	72
CESC Protein PFI 3	EA/NAT_A	Not complete	32	16	2	2	48	4	34	18	52
CESC Protein PFI 4	EA/NAT_A	Not complete	24	16	1	2	40	3	25	18	43
CESC Protein PFI 5	EA/NAT_A	Not complete	21	16	0	2	37	2	21	18	39
CESC Protein DFI 1	EA/NAT_A	Not complete	72	2	11	1	74	12	83	3	86
CESC Protein DFI 2	EA/NAT_A	Not complete	44	9	3	1	53	4	47	10	57
CESC Protein DFI 3	EA/NAT_A	Not complete	27	11	1	1	38	2	28	12	40
CESC Protein DFI 4	EA/NAT_A	Not complete	19	11	1	1	30	2	20	12	32
CESC Protein DFI 5	EA/NAT_A	Not complete	16	11	1	0	27	1	17	11	28
CESC mRNA OS 1	EA/NAT_A	Complete	143	9	42	8	152	50	185	17	202
CESC mRNA OS 2	EA/NAT_A	Complete	94	21	21	14	115	35	115	35	150
CESC mRNA OS 3	EA/NAT_A	Complete	65	30	11	17	95	28	76	47	123
CESC mRNA OS 4	EA/NAT_A	Not complete	66	35	4	17	101	21	70	52	122
CESC mRNA OS 5	EA/NAT_A	Not complete	33	35	1	17	68	18	34	52	86
CESC mRNA DSS 1	EA/NAT_A	Complete	140	6	42	8	146	50	182	14	196
CESC mRNA DSS 2	EA/NAT_A	Complete	91	16	21	14	107	35	112	30	142
CESC mRNA DSS 3	EA/NAT_A	Complete	64	23	11	16	87	27	75	39	114
CESC mRNA DSS 4	EA/NAT_A	Not complete	45	25	4	16	70	20	49	41	90
CESC mRNA DSS 5	EA/NAT_A	Not complete	32	25	1	16	57	17	33	41	74
PanGyn mRNA OS 1	EA/NAT_A	Complete	694	55	58	9	749	67	752	64	816
PanGyn mRNA OS 2	EA/NAT_A	Complete	492	126	32	16	618	48	524	142	666
PanGyn mRNA OS 3	EA/NAT_A	Complete	334	185	17	19	519	36	351	204	555
PanGyn mRNA OS 4	EA/NAT_A	Complete	239	224	8	19	232	27	243	243	486
PanGyn mRNA OS 5	EA/NAT_A	Not complete	159	247	3	21	250	24	268	268	536
PanGyn mRNA PFI 1	EA/NAT_A	Complete	603	132	55	12	187	67	144	144	288
PanGyn mRNA PFI 2	EA/NAT_A	Complete	361	255	27	22	282	49	277	277	554

PanGyn mRNA PFI 3	EA/NAT_A	Complete	364	479	8	40	843	48	372	519	891
PanGyn mRNA PFI 4	EA/NAT_A										
PanGyn mRNA PFI 5	EA/NAT_A										

Experiment	Source/Target	Status	White alive	White dead	Asian alive	Asian dead	Total White	Total Asians	Total alive	Total dead	Total cases
BRCA Protein OS 1	EA/ASA	Not complete	573	13	42	0	586	42	615	13	628
BRCA Protein OS 2	EA/ASA	Not complete	391	24	11	1	415	12	402	25	427
BRCA Protein OS 3	EA/ASA	Not complete	288	45	8	3	333	11	296	48	344
BRCA Protein OS 4	EA/ASA	Not complete	216	57	7	3	273	10	223	60	283
BRCA Protein OS 5	EA/ASA	Not complete	163	69	4	4	232	8	167	73	240
BRCA Protein DSS 1	EA/ASA	Not complete	555	8	42	0	563	42	597	8	605
BRCA Protein DSS 2	EA/ASA	Not complete	375	15	11	1	390	12	386	16	402
BRCA Protein DSS 3	EA/ASA	Not complete	273	28	8	3	301	11	281	31	312
BRCA Protein DSS 4	EA/ASA	Not complete	204	32	7	3	236	10	211	35	246
BRCA Protein DSS 5	EA/ASA	Not complete	155	37	4	4	192	8	159	41	200
BRCA Protein PFI 1	EA/ASA	Not complete	562	19	40	2	581	42	602	21	623
BRCA Protein PFI 2	EA/ASA	Not complete	377	37	10	3	414	13	387	40	427
BRCA Protein PFI 3	EA/ASA	Complete	273	56	7	5	329	12	280	61	341
BRCA Protein PFI 4	EA/ASA	Complete	199	66	6	5	265	11	205	71	276
BRCA Protein PFI 5	EA/ASA	Not complete	144	77	3	6	221	9	147	83	230
BRCA Protein DFI 1	EA/ASA	Not complete	486	11	35	2	497	37	521	13	534
BRCA Protein DFI 2	EA/ASA	Not complete	315	23	7	2	338	9	322	25	347
BRCA Protein DFI 3	EA/ASA	Not complete	233	31	5	3	264	8	238	34	272
BRCA Protein DFI 4	EA/ASA	Not complete	170	37	4	3	207	7	174	40	214
BRCA Protein DFI 5	EA/ASA	Not complete	122	44	3	2	166	5	125	46	171
BRCA mRNA OS 1	EA/ASA	Not complete	690	20	43	0	710	43	733	20	753
BRCA mRNA OS 2	EA/ASA	Not complete	474	28	12	1	502	13	486	29	515
BRCA mRNA OS 3	EA/ASA	Not complete	346	50	9	3	396	12	355	53	408
BRCA mRNA OS 4	EA/ASA	Not complete	258	62	8	3	320	11	266	65	331
BRCA mRNA OS 5	EA/ASA	Not complete	198	76	5	4	274	9	203	80	283
BRCA mRNA DSS 1	EA/ASA	Not complete	672	13	43	0	685	43	715	13	728
BRCA mRNA DSS 2	EA/ASA	Not complete	458	17	12	1	475	13	470	18	488
BRCA mRNA DSS 3	EA/ASA	Not complete	331	31	9	3	362	12	340	34	374
BRCA mRNA DSS 4	EA/ASA	Not complete	246	35	8	3	281	11	254	38	292
BRCA mRNA DSS 5	EA/ASA	Not complete	189	41	5	4	230	9	194	45	239
BRCA mRNA PFI 1	EA/ASA	Not complete	678	41	25	2	719	27	703	43	746
BRCA mRNA PFI 2	EA/ASA	Not complete	459	41	11	3	500	14	470	44	514
BRCA mRNA PFI 3	EA/ASA	Complete	330	62	8	5	392	13	338	67	405
BRCA mRNA PFI 4	EA/ASA	Complete	238	75	7	5	313	12	245	80	325
BRCA mRNA PFI 5	EA/ASA	Not complete	175	89	4	6	264	10	179	95	274
BRCA mRNA DFI 1	EA/ASA	Not complete	594	16	36	2	610	38	630	18	648
BRCA mRNA DFI 2	EA/ASA	Not complete	392	25	8	2	417	10	400	27	427
BRCA mRNA DFI 3	EA/ASA	Not complete	288	33	6	3	321	9	294	36	330
BRCA mRNA DFI 4	EA/ASA	Not complete	207	41	5	3	248	8	212	44	256
BRCA mRNA DFI 5	EA/ASA	Not complete	152	50	3	3	202	6	155	53	208
STAD mRNA OS 1	EA/ASA	Complete	174	54	55	20	228	75	229	74	303
STAD mRNA OS 2	EA/ASA	Complete	90	77	18	24	167	42	108	101	209

STAD mRNA OS 3	EA/ASA	Not complete	104	36	25	2	140	27	129	38	167
STAD mRNA OS 4	EA/ASA	Not complete	18	106	0	25	124	25	18	131	149
STAD mRNA OS 5	EA/ASA	Not complete	11	108	0	25	119	25	11	133	144
COADREAD Protein OS 1	EA/ASA	Not complete	317	42	11	1	359	12	328	43	371
COADREAD Protein OS 2	EA/ASA	Not complete	196	60	1	1	256	2	197	61	258
COADREAD Protein OS 3	EA/ASA	Not complete	106	69	1	1	175	2	107	70	177
COADREAD Protein OS 4	EA/ASA	Not complete	54	82	1	1	136	2	55	83	138
COADREAD Protein OS 5	EA/ASA	Not complete	43	88	1	1	131	2	44	89	133
COADREAD Protein DSS 1	EA/ASA	Not complete	301	30	11	1	331	12	312	31	343
COADREAD Protein DSS 2	EA/ASA	Not complete	183	42	1	1	225	2	184	43	227
COADREAD Protein DSS 3	EA/ASA	Not complete	97	45	1	1	142	2	98	46	144
COADREAD Protein DSS 4	EA/ASA	Not complete	48	49	1	1	97	2	49	50	99
COADREAD Protein DSS 5	EA/ASA	Not complete	31	51	1	1	82	2	32	52	84
COADREAD Protein PFI 1	EA/ASA	Not complete	293	56	10	2	349	12	303	58	361
COADREAD Protein PFI 2	EA/ASA	Not complete	171	84	1	2	255	3	172	86	258
COADREAD Protein PFI 3	EA/ASA	Not complete	86	97	1	2	183	3	87	99	186
COADREAD Protein PFI 4	EA/ASA	Not complete	44	102	1	2	146	3	45	104	149
COADREAD Protein PFI 5	EA/ASA	Not complete	30	102	0	3	132	3	30	105	135
COADREAD Protein DFI 1	EA/ASA	Not complete	120	8	10	1	128	11	130	9	139
COADREAD Protein DFI 2	EA/ASA	Not complete	81	12	1	1	93	2	82	13	95
COADREAD Protein DFI 3	EA/ASA	Not complete	38	14	1	1	52	2	39	15	54
COADREAD Protein DFI 4	EA/ASA	Not complete	15	15	1	1	30	2	16	16	32
COADREAD Protein DFI 5	EA/ASA	Not complete	9	15	0	2	24	2	9	17	26
COADREAD mRNA OS 1	EA/ASA	Not complete	396	52	12	1	448	13	408	53	461
COADREAD mRNA OS 2	EA/ASA	Not complete	252	74	2	1	326	3	254	75	329
COADREAD mRNA OS 3	EA/ASA	Not complete	138	82	1	1	220	2	139	83	222
COADREAD mRNA OS 4	EA/ASA	Not complete	67	96	1	1	163	2	68	97	165
COADREAD mRNA OS 5	EA/ASA	Not complete	44	102	1	1	146	2	45	103	148
COADREAD mRNA DSS 1	EA/ASA	Not complete	379	38	12	1	417	13	391	39	430
COADREAD mRNA DSS 2	EA/ASA	Not complete	238	53	2	1	291	3	240	54	294
COADREAD mRNA DSS 3	EA/ASA	Not complete	129	54	1	1	183	2	130	55	185
COADREAD mRNA DSS 4	EA/ASA	Not complete	61	60	1	1	121	2	62	61	123
COADREAD mRNA DSS 5	EA/ASA	Not complete	41	61	1	1	102	2	42	62	104
COADREAD mRNA PFI 1	EA/ASA	Not complete	367	69	11	2	436	13	378	71	449
COADREAD mRNA PFI 2	EA/ASA	Not complete	218	106	2	2	324	4	220	108	328
COADREAD mRNA PFI 3	EA/ASA	Not complete	110	119	1	2	229	3	111	121	232
COADREAD mRNA PFI 4	EA/ASA	Not complete	56	125	1	2	181	3	57	127	184
COADREAD mRNA PFI 5	EA/ASA	Not complete	37	125	0	3	162	3	37	128	165
COADREAD mRNA DFI 1	EA/ASA	Not complete	166	9	11	1	175	12	177	10	187
COADREAD mRNA DFI 2	EA/ASA	Not complete	114	17	2	1	131	3	116	18	134
COADREAD mRNA DFI 3	EA/ASA	Not complete	57	18	1	1	75	2	58	19	77
COADREAD mRNA DFI 4	EA/ASA	Not complete	24	20	1	1	44	2	25	21	46
COADREAD mRNA DFI 5	EA/ASA	Not complete	15	20	0	2	35	2	15	22	37

ESCA mRNA OS 1	EA/ASA	Complete	73	28	35	7	101	42	108	35	143	
ESCA mRNA OS 2	EA/ASA	Not complete	35	45	2	8	80	10	37	53	90	
ESCA mRNA OS 3	EA/ASA	Not complete	15	51	1	8	66	9	16	59	75	
ESCA mRNA OS 4	EA/ASA	Not complete	9	55	0	9	64	9	9	64	73	
ESCA mRNA OS 5	EA/ASA	Not complete	3	57	0	9	60	9	3	66	69	
ESCA mRNA DSS 1	EA/ASA	Not complete	71	16	35	4	87	39	106	20	126	
ESCA mRNA DSS 2	EA/ASA	Not complete	34	29	2	4	63	6	36	33	69	
ESCA mRNA DSS 3	EA/ASA	Not complete	15	34	1	4	49	5	16	38	54	
ESCA mRNA DSS 4	EA/ASA	Not complete	9	36	0	5	45	5	9	41	50	
ESCA mRNA DSS 5	EA/ASA	Not complete	3	38	0	5	41	5	3	43	46	
ESCA mRNA PFI 1	EA/ASA	Complete	59	34	20	19	93	39	79	53	132	B
ESCA mRNA PFI 2	EA/ASA	Not complete	30	47	0	22	77	22	30	69	99	
ESCA mRNA PFI 3	EA/ASA	Not complete	12	53	0	22	65	22	12	75	87	
ESCA mRNA PFI 4	EA/ASA	Not complete	7	56	0	22	63	22	7	78	85	
ESCA mRNA PFI 5	EA/ASA	Not complete	3	56	3	0	59	3	6	56	62	
ESCA mRNA DFI 1	EA/ASA	Not complete	29	1	18	17	30	35	47	18	65	
ESCA mRNA DFI 2	EA/ASA	Not complete	14	4	0	19	18	19	14	23	37	
ESCA mRNA DFI 3	EA/ASA	Not complete	4	4	0	19	8	19	4	23	27	
ESCA mRNA DFI 4	EA/ASA	Not complete	3	4	0	19	7	19	3	23	26	
ESCA mRNA DFI 5	EA/ASA	Not complete	2	4	0	19	6	19	2	23	25	
ESCA Protein OS 1	EA/ASA	Complete	48	18	34	6	66	40	82	24	106	B
ESCA Protein OS 2	EA/ASA	Not complete	24	26	2	7	50	9	26	33	59	
ESCA Protein OS 3	EA/ASA	Not complete	9	30	1	0	39	1	10	30	40	
ESCA Protein OS 4	EA/ASA	Not complete	6	33	0	8	39	8	6	41	47	
ESCA Protein OS 5	EA/ASA	Not complete	2	34	0	8	36	8	2	42	44	
ESCA Protein DSS 1	EA/ASA	Not complete	46	8	34	3	54	37	80	11	91	
ESCA Protein DSS 2	EA/ASA	Not complete	23	13	2	3	36	5	25	16	41	
ESCA Protein DSS 3	EA/ASA	Not complete	9	16	1	3	25	4	10	19	29	
ESCA Protein DSS 4	EA/ASA	Not complete	6	18	0	4	24	4	6	22	28	
ESCA Protein DSS 5	EA/ASA	Not complete	2	19	0	4	21	4	2	23	25	
ESCA Protein PFI 1	EA/ASA	Complete	39	18	19	18	57	37	58	36	94	
ESCA Protein PFI 2	EA/ASA	Not complete	21	25	0	20	46	20	21	45	66	
ESCA Protein PFI 3	EA/ASA	Not complete	28	8	0	20	36	20	28	28	56	
ESCA Protein PFI 4	EA/ASA	Not complete	5	30	0	20	35	20	5	50	55	
ESCA Protein PFI 5	EA/ASA	Not complete	2	30	0	20	32	20	2	50	52	
ESCA Protein DFI 1	EA/ASA	Not complete	20	1	17	17	21	34	37	18	55	
ESCA Protein DFI 2	EA/ASA	Not complete	11	2	0	18	13	18	11	20	31	
ESCA Protein DFI 3	EA/ASA	Not complete	3	1	0	18	4	18	3	19	22	
ESCA Protein DFI 4	EA/ASA	Not complete	3	1	0	18	4	18	3	19	22	
ESCA Protein DFI 5	EA/ASA	Not complete	2	1	0	18	3	18	2	19	21	
PanSCCs mRNA OS 1	EA/ASA	Complete	1115	233	93	15	1348	108	1208	248	1456	B
PanSCCs mRNA OS 2	EA/ASA	Complete	655	420	22	25	1075	47	677	445	1122	
PanSCCs mRNA OS 3	EA/ASA	Complete	418	502	10	25	920	35	428	527	955	B

PanSCCs mRNA OS 4	EA/ASA	Complete	293	537	8	26	830	34	301	563	864	B
PanSCCs mRNA OS 5	EA/ASA	Complete	197	556	6	27	753	33	203	583	786	B
PanSCCs mRNA PFI 1	EA/ASA						0	0	0	0	0	
PanSCCs mRNA PFI 2	EA/ASA						0	0	0	0	0	
PanSCCs mRNA PFI 3	EA/ASA						0	0	0	0	0	
PanSCCs mRNA PFI 4	EA/ASA		257	496	7	40	753	47	264	536	800	B
PanSCCs mRNA PFI 5	EA/ASA		161	517	4	41	678	45	165	558	723	

Experiment	Source/Target	Status	White alive	White dead	NAT/ASA alive	NAT/ASA dead	Total White	Total NAT/ASA	Total alive	Total dead	Total cases
GBMLGG-Protein-OS-3YR	EA/NAT-A	Not complete	138	201	3	4	339	7	141	205	346
	EA/ASA-A	Complete	170	163	12	7	333	19	182	170	352
PRAD-mRNA-PFI-3YR	EA/NAT-A	Not complete	139	59	0	3	198	3	139	62	201
	EA/ASA-A	Not complete	139	59	3	3	198	6	142	62	204
KIPAN-Protein-DSS-3YR	EA/NAT-A	Not complete	325	81	19	1	406	20	344	82	426
	EA/ASA-A	Not complete	325	81	4	2	406	6	329	83	412
PanGyn-mRNA-DFI-5YR	EA/NAT-A	Not complete	88	142	2	7	230	9	90	149	239
	EA/ASA-A	Complete	88	142	17	18	230	35	105	160	265
GBMLGG-Protein-DSS-3YR	EA/NAT-A						0	0	0	0	0
	EA/ASA-A						0	0	0	0	0
KIPAN-Protein-PFI-3YR	EA/NAT-A	Not complete	281	144	18	2	425	20	299	146	445
	EA/ASA-A	Not complete	281	144	4	3	425	7	285	147	432
KIPAN--Protein-OS-3YR	EA/NAT-A	Not complete	329	117	19	1	446	20	348	118	466
	EA/ASA-A	Not complete	329	117	4	2	446	6	333	119	452
BRCA-Protein-OS-4YR	EA/NAT-A	Not complete	47	125	4	3	172	7	51	128	179
	EA/ASA-A	Not complete	216	57	7	3	273	10	223	60	283
BLCA-mRNA-OS-1YR	EA/NAT-A	Not complete	231	68	11	2	299	13	242	70	312
	EA/ASA-A	Not complete	231	68	32	3	299	35	263	71	334
BLCA-mRNA-PFI-1YR	EA/NAT-A	Not complete	191	92	10	2	283	12	201	94	295
	EA/ASA-A	Complete	191	92	26	7	283	33	217	99	316
BRCA-mRNA-DSS-2YR	EA/NAT-A	Not complete	458	17	16	1	475	17	474	18	492
	EA/ASA-A	Not complete	458	17	12	1	475	13	470	18	488
BRCA-mRNA-DSS-3YR	EA/NAT-A	Not complete	331	31	9	2	362	11	340	33	373
	EA/ASA-A	Not complete	331	31	9	3	362	12	340	34	374
BRCA-mRNA-DSS-4YR	EA/NAT-A	Not complete	246	35	8	2	281	10	254	37	291
	EA/ASA-A	Not complete	246	35	8	3	281	11	254	38	292
BRCA-mRNA-DSS-5YR	EA/NAT-A	Not complete	189	41	6	2	230	8	195	43	238
	EA/ASA-A	Not complete	189	41	5	4	230	9	194	45	239
BRCA-mRNA-OS-2YR	EA/NAT-A	Not complete	474	28	16	1	502	17	490	29	519
	EA/ASA-A	Not complete	474	28	12	1	502	13	486	29	515
BRCA-mRNA-PFI-2YR	EA/NAT-A	Not complete	459	41	13	4	500	17	472	45	517
	EA/ASA-A	Not complete	459	41	11	3	500	14	470	44	514
BRCA-Protein-DFI-2YR	EA/NAT-A	Not complete	315	23	8	3	338	11	323	26	349
	EA/ASA-A	Not complete	315	23	7	2	338	9	322	25	347
BRCA-Protein-DSS-3YR	EA/NAT-A	Not complete	273	28	6	2	301	8	279	30	309
	EA/ASA-A	Not complete	273	28	8	3	301	11	281	31	312
BRCA-Protein-DSS-4YR	EA/NAT-A	Not complete	204	32	5	2	236	7	209	34	243
	EA/ASA-A	Not complete	204	32	7	3	236	10	211	35	246
BRCA-Protein-DSS-5YR	EA/NAT-A	Not complete	155	37	5	2	192	7	160	39	199
	EA/ASA-A	Not complete	155	37	4	4	192	8	159	41	200

BRCA-Protein-PFI-1YR	EA/NAT-A	Not complete	562	19	15	2	581	17	577	21	598
	EA/ASA-A	Not complete	562	19	40	2	581	42	602	21	623
BRCA-Protein-PFI-2YR	EA/NAT-A	Not complete	377	37	8	3	414	11	385	40	425
	EA/ASA-A	Not complete	377	37	10	3	414	13	387	40	427
BRCA-Protein-PFI-4YR	EA/NAT-A	Not complete	199	66	5	3	265	8	204	69	273
	EA/ASA-A	Complete	199	66	6	5	265	11	205	71	276
CESC-mRNA-DSS-4YR	EA/NAT-A	Not complete	45	25	4	16	70	20	49	41	90
	EA/ASA-A	Not complete	45	25	2	1	70	3	47	26	73
CESC-mRNA-DSS-5YR	EA/NAT-A	Not complete	33	35	1	17	68	18	34	52	86
	EA/ASA-A	Not complete	32	25	2	1	57	3	34	26	60
CESC-mRNA-OS-4YR	EA/NAT-A	Not complete	46	35	4	17	81	21	50	52	102
	EA/ASA-A	Not complete	46	35	2	2	81	4	48	37	85
CESC-mRNA-OS-5YR	EA/NAT-A	Not complete	33	35	1	17	68	18	34	52	86
	EA/ASA-A	Not complete	33	35	2	2	68	4	35	37	72
CESC-mRNA-PFI-2YR	EA/NAT-A	Complete	83	30	18	18	113	36	101	48	149
	EA/ASA-A	Not complete	83	30	4	5	113	9	87	35	122
CESC-mRNA-PFI-3YR	EA/NAT-A	Complete	58	34	10	20	92	30	68	54	122
	EA/ASA-A	Not complete	58	34	3	4	92	7	61	38	99
CESC-mRNA-PFI-4YR	EA/NAT-A	Not complete	42	35	4	20	77	24	46	55	101
	EA/ASA-A	Not complete	42	35	2	4	77	6	44	39	83
CESC-mRNA-PFI-5YR	EA/NAT-A	Not complete	29	35	1	20	64	21	30	55	85
	EA/ASA-A	Not complete	29	35	1	5	64	6	30	40	70
COAD-mRNA-DSS-4YR	EA/NAT-A						0	0	0	0	0
	EA/ASA-A	Not complete	50	48	1	1	98	2	51	49	100
COAD-mRNA-DSS-5YR	EA/NAT-A						0	0	0	0	0
	EA/ASA-A	Not complete	36	49	1	1	85	2	37	50	87
COADREAD-mRNA-DSS-2YR	EA/NAT-A	Not complete	41	61	0	1	102	1	41	62	103
	EA/ASA-A	Not complete	238	53	2	1	291	3	240	54	294
COADREAD-mRNA-DSS-4YR	EA/NAT-A	Not complete	61	60	0	1	121	1	61	61	122
	EA/ASA-A	Not complete	61	60	1	1	121	2	62	61	123
COADREAD-mRNA-DSS-5YR	EA/NAT-A	Not complete	41	61	0	1	102	1	41	62	103
	EA/ASA-A	Not complete	41	61	1	1	102	2	42	62	104
GBMLGG-mRNA-DSS-1YR	EA/NAT-A	Not complete	403	68	18	2	471	20	421	70	491
	EA/ASA-A	Not complete	403	68	8	0	471	8	411	68	479
GBMLGG-mRNA-DSS-2YR	EA/NAT-A	Not complete	231	131	9	3	362	12	240	134	374
	EA/ASA-A	Not complete	231	131	1	1	362	2	232	132	364
GBMLGG-mRNA-OS-1YR	EA/NAT-A	Not complete	415	77	18	2	492	20	433	79	512
	EA/ASA-A	Not complete	415	77	8	0	492	8	423	77	500
GBMLGG-mRNA-OS-2YR	EA/NAT-A	Not complete	239	145	9	3	384	12	248	148	396
	EA/ASA-A	Not complete	239	145	1	1	384	2	240	146	386
GBMLGG-mRNA-PFI-1YR	EA/NAT-A	Complete	342	150	16	5	492	21	358	155	513

[illegible]	EA/ASA-A	Not complete	342	150	7	1	492	8	349	151	500
GBMLGG-mRNA-PFI-2YR	EA/NAT-A	Complete	179	223	8	6	402	14	187	229	416
	EA/ASA-A	Not complete	179	223	1	2	402	3	180	225	405
GBMLGG-Protein-DSS-1YR	EA/NAT-A	Not complete	371	77	18	1	448	19	389	78	467
	EA/ASA-A	Not complete	371	77	6	0	448	6	377	77	454
GBMLGG-Protein-DSS-2YR	EA/NAT-A	Not complete	202	153	8	3	355	11	210	156	366
	EA/ASA-A	Not complete	202	153	1	0	355	1	203	153	356
GBMLGG-Protein-OS-1YR	EA/NAT-A	Not complete	380	91	18	1	471	19	398	92	490
	EA/ASA-A	Not complete	380	91	6	0	471	6	386	91	477
GBMLGG-Protein-OS-2YR	EA/NAT-A	Not complete	209	170	8	3	379	11	217	173	390
	EA/ASA-A	Not complete	209	170	1	0	379	1	210	170	380
GBMLGG-Protein-PFI-1YR	EA/NAT-A	Complete	293	178	16	5	471	21	309	183	492
	EA/ASA-A	Not complete	293	178	6	0	471	6	299	178	477
GBMLGG-Protein-PFI-2YR	EA/NAT-A	Complete	150	245	7	7	395	14	157	252	409
	EA/ASA-A	Not complete	150	245	1	1	395	2	151	246	397
HNSC-mRNA-DSS-2YR	EA/NAT-A	Complete	198	81	9	5	279	14	207	86	293
	EA/ASA-A	Not complete	198	81	0	2	279	2	198	83	281
HNSC-mRNA-DSS-3YR	EA/NAT-A	Complete	123	93	8	5	216	13	131	98	229
	EA/ASA-A	Not complete	123	93	0	2	216	2	123	95	218
KIPAN-mRNA-DFI-2YR	EA/NAT-A	Not complete	156	17	13	1	173	14	169	18	187
	EA/ASA-A	Not complete	156	17	0	2	173	2	156	19	175
KIPAN-mRNA-DFI-3YR	EA/NAT-A	Not complete	118	21	11	1	139	12	129	22	151
	EA/ASA-A	Not complete	118	21	0	2	139	2	118	23	141
KIPAN-mRNA-DFI-4YR	EA/NAT-A	Not complete	98	26	7	2	124	9	105	28	133
	EA/ASA-A	Not complete	98	26	0	2	124	2	98	28	126
KIPAN-mRNA-DFI-5YR	EA/NAT-A	Not complete	75	27	6	2	102	8	81	29	110
	EA/ASA-A	Not complete	75	27	0	2	102	2	75	29	104
KIPAN-mRNA-DSS-1YR	EA/NAT-A	Not complete	587	41	30	0	628	30	617	41	658
	EA/ASA-A	Not complete	587	41	11	1	628	12	598	42	640
KIPAN-mRNA-DSS-2YR	EA/NAT-A	Not complete	469	69	27	1	538	28	496	70	566
	EA/ASA-A	Not complete	469	69	6	1	538	7	475	70	545
KIPAN-mRNA-DSS-3YR	EA/NAT-A	Not complete	371	89	24	1	460	25	395	90	485
	EA/ASA-A	Not complete	371	89	4	2	460	6	375	91	466
KIPAN-mRNA-DSS-4YR	EA/NAT-A	Not complete	290	104	18	1	394	19	308	105	413
	EA/ASA-A	Not complete	290	104	1	3	394	4	291	107	398
KIPAN-mRNA-DSS-5YR	EA/NAT-A	Not complete	206	116	15	1	322	16	221	117	338
	EA/ASA-A	Not complete	206	116	0	3	322	3	206	119	325
KIPAN-mRNA-OS-1YR	EA/NAT-A	Not complete	599	56	30	0	655	30	629	56	685
	EA/ASA-A	Not complete	599	66	11	1	665	12	610	67	677
KIPAN-mRNA-OS-2YR	EA/NAT-A	Not complete	479	96	27	1	575	28	506	97	603
	EA/ASA-A	Not complete	479	96	6	1	575	7	485	97	582

KIPAN-mRNA-OS-3YR	EA/NAT-A	Not complete	377	128	24	1	505	25	401	129	530
	EA/ASA-A	Not complete	377	128	4	2	505	6	381	130	511
KIPAN-mRNA-OS-4YR	EA/NAT-A	Not complete	296	152	18	1	448	19	314	153	467
	EA/ASA-A	Not complete	296	152	1	3	448	4	297	155	452
KIPAN-mRNA-OS-5YR	EA/NAT-A	Not complete	212	174	15	1	386	16	227	175	402
	EA/ASA-A	Not complete	212	174	0	3	386	3	212	177	389
KIPAN-mRNA-PFI-1YR	EA/NAT-A	Not complete	540	98	29	1	638	30	569	99	668
	EA/ASA-A	Not complete	540	98	10	3	638	13	550	101	651
KIPAN-mRNA-PFI-2YR	EA/NAT-A	Not complete	417	135	26	2	552	28	443	137	580
	EA/ASA-A	Not complete	417	135	5	3	552	8	422	138	560
KIPAN-mRNA-PFI-3YR	EA/NAT-A	Not complete	327	153	23	2	480	25	350	155	505
	EA/ASA-A	Not complete	327	153	4	3	480	7	331	156	487
KIPAN-mRNA-PFI-4YR	EA/NAT-A	Not complete	254	166	17	3	420	20	271	169	440
	EA/ASA-A	Not complete	254	166	1	4	420	5	255	170	425
KIPAN-mRNA-PFI-5YR	EA/NAT-A	Complete	181	174	12	5	355	17	193	179	372
	EA/ASA-A	Not complete	181	174	0	4	355	4	181	178	359
KIPAN-Protein-DSS-1YR	EA/NAT-A	Not complete	502	39	22	0	541	22	524	39	563
	EA/ASA-A	Not complete	502	39	11	1	541	12	513	40	553
KIPAN-Protein-DSS-2YR	EA/NAT-A	Not complete	408	63	20	1	471	21	428	64	492
	EA/ASA-A	Not complete	408	63	6	1	471	7	414	64	478
KIPAN-Protein-DSS-4YR	EA/NAT-A	Not complete	255	96	15	1	351	16	270	97	367
	EA/ASA-A	Not complete	255	96	1	3	351	4	256	99	355
KIPAN-Protein-DSS-5YR	EA/NAT-A	Not complete	180	107	13	1	287	14	193	108	301
	EA/ASA-A	Not complete	180	107	0	3	287	3	180	110	290
KIPAN-Protein-OS-1YR	EA/NAT-A	Not complete	511	54	22	0	565	22	533	54	587
	EA/ASA-A	Not complete	511	54	11	1	565	12	522	55	577
KIPAN-Protein-OS-2YR	EA/NAT-A	Not complete	416	87	20	1	503	21	436	88	524
	EA/ASA-A	Not complete	416	87	6	1	503	7	422	88	510
KIPAN-Protein-OS-4YR	EA/NAT-A	Not complete	259	140	15	1	399	16	274	141	415
	EA/ASA-A	Not complete	259	140	3	1	399	4	262	141	403
KIPAN-Protein-OS-5YR	EA/NAT-A	Not complete	184	160	13	1	344	14	197	161	358
	EA/ASA-A	Not complete	184	160	0	3	344	3	184	163	347
KIPAN-Protein-PFI-1YR	EA/NAT-A	Not complete	458	91	21	1	549	22	479	92	571
	EA/ASA-A	Not complete	458	91	10	3	549	13	468	94	562
KIPAN-Protein-PFI-2YR	EA/NAT-A	Not complete	356	128	19	2	484	21	375	130	505
	EA/ASA-A	Not complete	356	128	5	3	484	8	361	131	492
KIPAN-Protein-PFI-4YR	EA/NAT-A	Not complete	218	156	14	3	374	17	232	159	391
	EA/ASA-A	Not complete	218	156	1	4	374	5	219	160	379
KIPAN-Protein-PFI-5YR	EA/NAT-A	Not complete	153	164	11	4	317	15	164	168	332
	EA/ASA-A	Not complete	153	164	0	4	317	4	153	168	321
KIRC-mRNA-DSS-5YR	EA/NAT-A	Not complete	130	91	11	0	221	11	141	91	232

[illegible]	EA/ASA-A	Not complete	130	91	0	1	221	1	130	92	222
KIRC-mRNA-OS-2YR	EA/NAT-A	Not complete	313	76	17	0	389	17	330	76	406
	EA/ASA-A	Not complete	313	76	5	0	389	5	318	76	394
KIRC-mRNA-OS-3YR	EA/NAT-A	Not complete	255	103	16	0	358	16	271	103	374
	EA/ASA-A	Not complete	255	103	4	0	358	4	259	103	362
KIRC-mRNA-OS-4YR	EA/NAT-A	Not complete	193	122	13	0	315	13	206	122	328
	EA/ASA-A	Not complete	193	122	1	1	315	2	194	123	317
KIRC-mRNA-OS-5YR	EA/NAT-A	Not complete	133	138	11	0	271	11	144	138	282
	EA/ASA-A	Not complete	133	138	0	1	271	1	133	139	272
KIRC-mRNA-PFI-1YR	EA/NAT-A	Not complete	330	73	18	0	403	18	348	73	421
	EA/ASA-A	Not complete	330	73	7	0	403	7	337	73	410
KIRC-mRNA-PFI-2YR	EA/NAT-A	Not complete	266	101	16	1	367	17	282	102	384
	EA/ASA-A	Not complete	266	101	5	0	367	5	271	101	372
KIRC-mRNA-PFI-3YR	EA/NAT-A	Not complete	219	113	15	1	332	16	234	114	348
	EA/ASA-A	Not complete	219	113	4	0	332	4	223	113	336
KIRC-mRNA-PFI-4YR	EA/NAT-A	Not complete	162	123	12	2	285	14	174	125	299
	EA/ASA-A	Not complete	162	123	1	1	285	2	163	124	287
KIRC-mRNA-PFI-5YR	EA/NAT-A	Not complete	109	129	8	4	238	12	117	133	250
	EA/ASA-A	Not complete	109	129	1	0	238	1	110	129	239
KIRC-Protein-DSS-5YR	EA/NAT-A	Not complete	113	88	9	0	201	9	122	88	210
	EA/ASA-A	Not complete	113	88	0	1	201	1	113	89	202
KIRC-Protein-OS-2YR	EA/NAT-A	Not complete	282	73	13	0	355	13	295	73	368
	EA/ASA-A	Not complete	282	73	5	0	355	5	287	73	360
KIRC-Protein-OS-3YR	EA/NAT-A	Not complete	227	98	12	0	325	12	239	98	337
	EA/ASA-A	Not complete	227	98	4	0	325	4	231	98	329
KIRC-Protein-OS-4YR	EA/NAT-A	Not complete	172	117	10	0	289	10	182	117	299
	EA/ASA-A	Not complete	172	117	1	1	289	2	173	118	291
KIRC-Protein-OS-5YR	EA/NAT-A	Not complete	116	132	9	0	248	9	125	132	257
	EA/ASA-A	Not complete	116	132	0	1	248	1	116	133	249
KIRC-Protein-PFI-1YR	EA/NAT-A	Not complete	297	71	14	0	368	14	311	71	382
	EA/ASA-A	Not complete	297	71	7	0	368	7	304	71	375
KIRC-Protein-PFI-2YR	EA/NAT-A	Not complete	236	100	12	1	336	13	248	101	349
	EA/ASA-A	Not complete	236	100	5	0	336	5	241	100	341
KIRC-Protein-PFI-3YR	EA/NAT-A	Not complete	191	111	11	1	302	12	202	112	314
	EA/ASA-A	Not complete	191	111	4	0	302	4	195	111	306
KIRC-Protein-PFI-4YR	EA/NAT-A	Not complete	142	121	9	2	263	11	151	123	274
	EA/ASA-A	Not complete	142	121	1	1	263	2	143	122	265
KIRC-Protein-PFI-5YR	EA/NAT-A	Not complete	92	127	7	3	219	10	99	130	229
	EA/ASA-A	Not complete	92	127	0	1	219	1	92	128	220
KIRP-mRNA-DSS-2YR	EA/NAT-A	Not complete	115	15	7	1	130	8	122	16	138
	EA/ASA-A	Not complete	115	15	0	1	130	1	115	16	131

KIRP-mRNA-DSS-3YR	EA/NAT-A	Not complete	79	15	5	1	94	6	84	16	100
	EA/ASA-A	Not complete	79	15	1	0	94	1	80	15	95
KIRP-mRNA-DSS-4YR	EA/NAT-A	Not complete	62	18	3	1	80	4	65	19	84
	EA/ASA-A	Not complete	62	18	0	1	80	1	62	19	81
KIRP-mRNA-DSS-5YR	EA/NAT-A	Not complete	43	20	2	1	63	3	45	21	66
	EA/ASA-A	Not complete	43	20	0	1	63	1	43	21	64
KIRP-mRNA-OS-2YR	EA/NAT-A	Not complete	119	17	7	1	136	8	126	18	144
	EA/ASA-A	Not complete	119	17	0	1	136	1	119	18	137
KIRP-mRNA-OS-3YR	EA/NAT-A	Not complete	82	20	5	1	102	6	87	21	108
	EA/ASA-A	Not complete	82	20	0	1	102	1	82	21	103
KIRP-mRNA-OS-4YR	EA/NAT-A	Not complete	65	24	3	1	89	4	68	25	93
	EA/ASA-A	Not complete	65	24	0	1	89	1	65	25	90
KIRP-mRNA-OS-5YR	EA/NAT-A	Not complete	46	29	2	1	75	3	48	30	78
	EA/ASA-A	Not complete	46	29	0	1	75	1	46	30	76
KIRP-mRNA-PFI-1YR	EA/NAT-A	Not complete	163	21	8	1	184	9	171	22	193
	EA/ASA-A	Not complete	163	21	2	2	184	4	165	23	188
KIRP-mRNA-PFI-2YR	EA/NAT-A	Not complete	107	29	7	1	136	8	114	30	144
	EA/ASA-A	Not complete	107	29	0	2	136	2	107	31	138
KIRP-mRNA-PFI-3YR	EA/NAT-A	Not complete	70	33	5	1	103	6	75	34	109
	EA/ASA-A	Not complete	70	33	0	2	103	2	70	35	105
KIRP-mRNA-PFI-4YR	EA/NAT-A	Not complete	55	36	3	1	91	4	58	37	95
	EA/ASA-A	Not complete	55	36	0	2	91	2	55	38	93
KIRP-mRNA-PFI-5YR	EA/NAT-A	Not complete	39	38	2	1	77	3	41	39	80
	EA/ASA-A	Not complete	39	39	0	2	78	2	39	41	80
KIRP-Protein-DSS-2YR	EA/NAT-A	Not complete	88	10	4	1	98	5	92	11	103
	EA/ASA-A	Not complete	88	10	0	1	98	1	88	11	99
KIRP-Protein-DSS-3YR	EA/NAT-A	Not complete	63	10	4	1	73	5	67	11	78
	EA/ASA-A	Not complete	63	10	0	1	73	1	63	11	74
KIRP-Protein-DSS-4YR	EA/NAT-A	Not complete	49	13	3	1	62	4	52	14	66
	EA/ASA-A	Not complete	49	13	0	1	62	1	49	14	63
KIRP-Protein-DSS-5YR	EA/NAT-A	Not complete	34	15	2	1	49	3	36	16	52
	EA/ASA-A	Not complete	34	15	0	1	49	1	34	16	50
KIRP-Protein-OS-2YR	EA/NAT-A	Not complete	90	11	4	1	101	5	94	12	106
	EA/ASA-A	Not complete	90	11	0	1	101	1	90	12	102
KIRP-Protein-PFI-1YR	EA/NAT-A	Not complete	117	16	4	1	133	5	121	17	138
	EA/ASA-A	Not complete	117	16	2	2	133	4	119	18	137
KIRP-Protein-PFI-2YR	EA/NAT-A	Not complete	79	23	4	1	102	5	83	24	107
	EA/ASA-A	Not complete	79	23	0	2	102	2	79	25	104
KIRP-Protein-PFI-3YR	EA/NAT-A	Not complete	53	27	4	1	80	5	57	28	85
	EA/ASA-A	Not complete	53	27	0	2	80	2	53	29	82
LGG-mRNA-PFI-2YR	EA/NAT-A	Not complete	168	116	8	4	284	12	176	120	296

[illegible]	EA/ASA-A	Not complete	168	116	1	1	284	2	169	117	286
LGG-Protein-PFI-2YR	EA/NAT-A	Not complete	134	92	7	4	226	11	141	96	237
	EA/ASA-A	Not complete	134	92	0	1	226	1	134	93	227
LUAD-mRNA-DSS-1YR	EA/NAT-A	Not complete	323	26	3	1	349	4	326	27	353
	EA/ASA-A	Not complete	323	26	5	1	349	6	328	27	355
LUSC-mRNA-DSS-2YR	EA/NAT-A	Not complete	187	47	1	0	234	1	188	47	235
	EA/ASA-A	Not complete	187	47	4	0	234	4	191	47	238
LUSC-Protein-DSS-4YR	EA/NAT-A						0	0	0	0	0
	EA/ASA-A	Not complete	64	48	0	1	112	1	64	49	113
LUSC-Protein-OS-4YR	EA/NAT-A						0	0	0	0	0
	EA/ASA-A	Not complete	73	98	1	1	171	2	74	99	173
OV-mRNA-DSS-2YR	EA/NAT-A	Not complete	147	40	2	0	187	2	149	40	189
	EA/ASA-A	Not complete	147	40	5	2	187	7	152	42	194
OV-mRNA-OS-2YR	EA/NAT-A	Not complete	159	52	2	0	211	2	161	52	213
	EA/ASA-A	Not complete	159	52	5	2	211	7	164	54	218
OV-Protein-DSS-3YR	EA/NAT-A	Not complete	163	90	3	1	253	4	166	91	257
	EA/ASA-A	Not complete	163	90	2	1	253	3	165	91	256
OV-Protein-OS-3YR	EA/NAT-A	Not complete	176	103	3	2	279	5	179	105	284
	EA/ASA-A	Not complete	176	103	2	1	279	3	178	104	282
OV-Protein-OS-4YR	EA/NAT-A	Not complete	119	144	3	2	263	5	122	146	268
	EA/ASA-A	Not complete	119	114	2	1	233	3	121	115	236
PanGI-mRNA-DSS-2YR	EA/NAT-A	Complete	348	131	11	15	479	26	359	146	505
	EA/ASA-A	Complete	348	131	22	22	479	44	370	153	523
PanGI-mRNA-DSS-3YR	EA/NAT-A	Complete	180	144	10	16	324	26	190	160	350
	EA/ASA-A	Not complete	180	144	4	23	324	27	184	167	351
PanGI-mRNA-DSS-4YR	EA/NAT-A	Not complete	88	153	4	18	241	22	92	171	263
	EA/ASA-A	Not complete	88	153	1	24	241	25	89	177	266
PanGI-mRNA-DSS-5YR	EA/NAT-A	Not complete	55	157	2	19	212	21	57	176	233
	EA/ASA-A	Not complete	55	157	1	24	212	25	56	181	237
PanGI-mRNA-OS-2YR	EA/NAT-A	Complete	364	209	11	19	573	30	375	228	603
	EA/ASA-A	Complete	364	209	22	33	573	55	386	242	628
PanGI-mRNA-OS-3YR	EA/NAT-A	Complete	189	237	10	20	426	30	199	257	456
	EA/ASA-A	Not complete	189	237	4	34	426	38	193	271	464
PanGI-mRNA-OS-4YR	EA/NAT-A	Not complete	94	257	4	22	351	26	98	279	377
	EA/ASA-A	Not complete	94	257	1	35	351	36	95	292	387
PanGI-mRNA-OS-5YR	EA/NAT-A	Not complete	58	267	2	23	325	25	60	290	350
	EA/ASA-A	Not complete	58	267	1	35	325	36	59	302	361
PanGI-mRNA-PFI-1YR	EA/NAT-A	Complete	578	152	21	10	730	31	599	162	761
	EA/ASA-A	Complete	578	152	78	42	730	120	656	194	850
PanGI-mRNA-PFI-3YR	EA/NAT-A	Complete	152	254	6	21	406	27	158	275	433
	EA/ASA-A	Complete	152	254	3	51	406	54	155	305	460

PanGI-mRNA-PFI-4YR	EA/NAT-A	Not complete	78	265	4	21	343	25	82	286	368
	EA/ASA-A	Not complete	78	265	1	0	343	1	79	265	344
PanGI-mRNA-PFI-5YR	EA/NAT-A	Not complete	51	266	2	22	317	24	53	288	341
	EA/ASA-A	Not complete	51	266	0	52	317	52	51	318	369
PanGI-Protein-DSS-4YR	EA/NAT-A	Not complete	70	114	1	12	184	13	71	126	197
	EA/ASA-A	Not complete	70	114	1	19	184	20	71	133	204
PanGI-Protein-OS-2YR	EA/NAT-A	Complete	291	169	5	13	460	18	296	182	478
	EA/ASA-A	Complete	291	169	14	28	460	42	305	197	502
PanGI-Protein-OS-3YR	EA/NAT-A	Complete	146	195	5	13	341	18	151	208	359
	EA/ASA-A	Not complete	146	195	4	29	341	33	150	224	374
PanGI-Protein-OS-4YR	EA/NAT-A	Not complete	76	213	1	15	289	16	77	228	305
	EA/ASA-A	Not complete	76	213	1	30	289	31	77	243	320
PanGyn-mRNA-DFI-2YR	EA/NAT-A	Complete	295	101	19	7	396	26	314	108	422
	EA/ASA-A	Complete	295	101	29	6	396	35	324	107	431
PanGyn-mRNA-DFI-3YR	EA/NAT-A	Complete	196	127	10	7	323	17	206	134	340
	EA/ASA-A	Complete	196	127	22	6	323	28	218	133	351
PanGyn-mRNA-DFI-4YR	EA/NAT-A	Not complete	140	136	4	7	276	11	144	143	287
	EA/ASA-A	Complete	140	136	18	7	276	25	158	143	301
PanGyn-mRNA-DSS-1YR	EA/NAT-A	Complete	671	37	57	9	708	66	728	46	774
	EA/ASA-A	Not complete	671	37	51	3	708	54	722	40	762
PanGyn-mRNA-DSS-2YR	EA/NAT-A	Complete	475	95	31	16	570	47	506	111	617
	EA/ASA-A	Complete	475	95	38	5	570	43	513	100	613
PanGyn-mRNA-DSS-3YR	EA/NAT-A	Complete	322	144	16	18	466	34	338	162	500
	EA/ASA-A	Complete	322	144	24	7	466	31	346	151	497
PanGyn-mRNA-DSS-4YR	EA/NAT-A	Complete	234	172	7	18	406	25	241	190	431
	EA/ASA-A	Complete	234	172	19	8	406	27	253	180	433
PanGyn-mRNA-DSS-5YR	EA/NAT-A	Not complete	155	191	3	19	346	22	158	210	368
	EA/ASA-A	Complete	155	191	19	8	346	27	174	199	373
PanGyn-mRNA-OS-2YR	EA/NAT-A	Complete	492	126	32	16	618	48	524	142	666
	EA/ASA-A	Complete	492	126	38	7	618	45	530	133	663
PanGyn-mRNA-OS-3YR	EA/NAT-A	Complete	334	185	17	19	519	36	351	204	555
	EA/ASA-A	Complete	334	185	24	9	519	33	358	194	552
PanGyn-mRNA-OS-4YR	EA/NAT-A	Complete	239	224	8	19	463	27	247	243	490
	EA/ASA-A	Complete	239	224	19	10	463	29	258	234	492
PanGyn-mRNA-OS-5YR	EA/NAT-A	Not complete	159	247	3	21	406	24	162	268	430
	EA/ASA-A	Complete	159	247	19	10	406	29	178	257	435
PanGyn-mRNA-PFI-1YR	EA/NAT-A	Complete	603	132	55	12	735	67	658	144	802
	EA/ASA-A	Complete	603	132	49	6	735	55	652	138	790
PanGyn-mRNA-PFI-2YR	EA/NAT-A	Complete	361	255	27	22	616	49	388	277	665
	EA/ASA-A	Complete	361	255	30	14	616	44	391	269	660
PanGyn-mRNA-PFI-3YR	EA/NAT-A	Complete	234	295	13	25	529	38	247	320	567

PanGyn-mRNA-PFI-4YR	EA/ASA-A	Complete	234	295	23	14	529	37	257	309	566
	EA/NAT-A	Complete	165	311	6	25	476	31	171	336	507
PanGyn-mRNA-PFI-5YR	EA/ASA-A	Complete	165	311	18	15	476	33	183	326	509
	EA/NAT-A	Not complete	106	319	2	25	425	27	108	344	452
PanGyn-Protein-DFI-1YR	EA/ASA-A	Complete	106	319	17	16	425	33	123	335	458
	EA/NAT-A	Not complete	435	46	25	2	481	27	460	48	508
PanGyn-Protein-DFI-2YR	EA/ASA-A	Not complete	435	46	32	2	481	34	467	48	515
	EA/NAT-A	Not complete	269	126	11	4	395	15	280	130	410
PanGyn-Protein-DFI-3YR	EA/ASA-A	Complete	269	126	20	6	395	26	289	132	421
	EA/NAT-A	Not complete	170	163	4	4	333	8	174	167	341
PanGyn-Protein-DFI-4YR	EA/ASA-A	Complete	170	163	12	7	333	19	182	170	352
	EA/NAT-A	Not complete	120	173	2	4	293	6	122	177	299
PanGyn-Protein-DFI-5YR	EA/ASA-A	Complete	120	173	10	8	293	18	130	181	311
	EA/NAT-A	Not complete	80	180	1	4	260	5	81	184	265
PanGyn-Protein-DSS-1YR	EA/ASA-A	Complete	80	180	10	8	260	18	90	188	278
	EA/NAT-A	Not complete	651	32	31	2	683	33	682	34	716
PanGyn-Protein-DSS-2YR	EA/ASA-A	Not complete	651	32	41	1	683	42	692	33	725
	EA/NAT-A	Not complete	468	91	17	4	559	21	485	95	580
PanGyn-Protein-DSS-3YR	EA/ASA-A	Not complete	468	91	30	2	559	32	498	93	591
	EA/NAT-A	Complete	325	147	8	5	472	13	333	152	485
PanGyn-Protein-DSS-4YR	EA/ASA-A	Not complete	325	147	15	4	472	19	340	151	491
	EA/NAT-A	Complete	234	183	5	5	417	10	239	188	427
PanGyn-Protein-DSS-5YR	EA/ASA-A	Complete	234	183	12	5	417	17	246	188	434
	EA/NAT-A	Not complete	163	210	2	7	373	9	165	217	382
PanGyn-Protein-OS-2YR	EA/ASA-A	Complete	163	210	11	6	373	17	174	216	390
	EA/NAT-A	Not complete	487	117	19	4	604	23	506	121	627
PanGyn-Protein-OS-3YR	EA/ASA-A	Not complete	487	117	30	4	604	34	517	121	638
	EA/NAT-A	Complete	341	179	9	6	520	15	350	185	535
PanGyn-Protein-OS-4YR	EA/ASA-A	Complete	341	179	15	6	520	21	356	185	541
	EA/NAT-A	Complete	241	226	6	6	467	12	247	232	479
PanGyn-Protein-OS-5YR	EA/ASA-A	Complete	241	226	12	7	467	19	253	233	486
	EA/NAT-A	Not complete	169	258	2	9	427	11	171	267	438
PanGyn-Protein-PFI-1YR	EA/ASA-A	Complete	169	258	11	8	427	19	180	266	446
	EA/NAT-A	Not complete	566	145	31	4	711	35	597	149	746
PanGyn-Protein-PFI-2YR	EA/ASA-A	Not complete	566	145	39	4	711	43	605	149	754
	EA/NAT-A	Complete	326	284	14	9	610	23	340	293	633
PanGyn-Protein-PFI-3YR	EA/ASA-A	Complete	326	284	21	22	610	43	347	306	653
	EA/NAT-A	Complete	199	340	5	11	539	16	204	351	555
PanGyn-Protein-PFI-4YR	EA/ASA-A	Complete	199	340	13	13	539	26	212	353	565
	EA/NAT-A	Not complete	140	355	3	11	495	14	143	366	509
	EA/ASA-A	Complete	140	355	10	14	495	24	150	369	519

PanGyn-Protein-PFI-5YR	EA/NAT-A	Not complete	94	366	1	11	460	12	95	377	472
	EA/ASA-A	Complete	94	366	10	14	460	24	104	380	484
PanSCC-mRNA-DSS-1YR	EA/NAT-A	Complete	671	37	57	9	708	66	728	46	774
	EA/ASA-A	Complete	1055	130	92	6	1185	98	1147	136	1283
PanSCC-mRNA-DSS-2YR	EA/NAT-A	Complete	475	95	31	16	570	47	506	111	617
	EA/ASA-A	Complete	627	259	22	12	886	34	649	271	920
PanSCC-mRNA-DSS-3YR	EA/NAT-A	Complete	322	144	16	18	466	34	338	162	500
	EA/ASA-A	Complete	391	310	10	11	701	21	401	321	722
PanSCC-mRNA-DSS-4YR	EA/NAT-A	Complete	234	172	7	18	406	25	241	190	431
	EA/ASA-A	Complete	272	326	8	12	598	20	280	338	618
PanSCC-mRNA-DSS-5YR	EA/NAT-A	Not complete	155	191	3	19	346	22	158	210	368
	EA/ASA-A	Complete	182	336	6	12	518	18	188	348	536
PanSCC-mRNA-PFI-1YR	EA/NAT-A	Complete	979	290	70	18	1269	88	1049	308	1357
	EA/ASA-A	Complete	979	290	68	32	1269	100	1047	322	1369
PanSCC-Protein-DSS-3YR	EA/NAT-A	Complete	258	201	11	8	459	19	269	209	478
	EA/ASA-A	Complete	258	201	6	8	459	14	264	209	473
PanSCC-Protein-DSS-4YR	EA/NAT-A	Complete	193	213	6	8	406	14	199	221	420
	EA/ASA-A	Complete	193	213	5	9	406	14	198	222	420
PanSCC-Protein-OS-4YR	EA/NAT-A	Complete	206	356	6	13	562	19	212	369	581
	EA/ASA-A	Complete	206	356	5	18	562	23	211	374	585
PanSCC-Protein-PFI-3YR	EA/NAT-A	Complete	238	303	11	11	541	22	249	314	563
	EA/ASA-A	Not complete	238	303	4	36	541	40	242	339	581
PRAD-mRNA-PFI-4YR	EA/NAT-A	Complete	257	496	14	36	753	50	271	532	803
	EA/ASA-A	Not complete	74	69	2	3	143	5	76	72	148
PRAD-mRNA-PFI-5YR	EA/NAT-A	Complete	161	517	8	37	678	45	169	554	723
	EA/ASA-A	Not complete	47	73	1	3	120	4	48	76	124
SARC-mRNA-DSS-3YR	EA/NAT-A	Not complete	95	49	4	3	144	7	99	52	151
	EA/ASA-A	Not complete	95	49	0	1	144	1	95	50	145
SARC-mRNA-DSS-4YR	EA/NAT-A	Not complete	67	56	3	3	123	6	70	59	129
	EA/ASA-A	Not complete	67	56	0	1	123	1	67	57	124
SARC-mRNA-OS-2YR	EA/NAT-A	Not complete	135	35	4	3	170	7	139	38	177
	EA/ASA-A	Not complete	138	43	0	1	181	1	138	44	182
SARC-mRNA-OS-3YR	EA/NAT-A	Not complete	95	49	4	3	144	7	99	52	151
	EA/ASA-A	Not complete	98	57	0	1	155	1	98	58	156
SARC-mRNA-OS-4YR	EA/NAT-A	Not complete	67	56	3	3	123	6	70	59	129
	EA/ASA-A	Not complete	69	66	0	1	135	1	69	67	136
SARC-Protein-OS-2YR	EA/NAT-A	Not complete	122	40	4	3	162	7	126	43	169
	EA/ASA-A	Not complete	122	40	0	1	162	1	122	41	163
STAD-mRNA-DSS-2YR	EA/NAT-A	Complete	76	49	8	11	125	19	84	60	144
	EA/ASA-A	Complete	76	49	18	17	125	35	94	66	160
STAD-mRNA-PFI-2YR	EA/NAT-A	Complete	65	78	7	13	143	20	72	91	163

[illegible]	EA/ASA-A	Complete	65	78	16	26	143	42	81	104	185
UCEC-mRNA-DSS-2YR	EA/NAT-A	Complete	219	20	8	2	239	10	227	22	249
	EA/ASA-A	Not complete	219	20	25	1	239	26	244	21	265
UCEC-mRNA-DSS-3YR	EA/NAT-A	Not complete	147	30	3	2	177	5	150	32	182
	EA/ASA-A	Not complete	147	30	20	3	177	23	167	33	200
UCEC-mRNA-DSS-4YR	EA/NAT-A	Not complete	110	32	1	2	142	3	111	34	145
	EA/ASA-A	Not complete	110	32	16	4	142	20	126	36	162
UCEC-mRNA-DSS-5YR	EA/NAT-A	Not complete	72	33	2	1	105	3	74	34	108
	EA/ASA-A	Not complete	72	33	16	4	105	20	88	37	125
UCEC-mRNA-OS-2YR	EA/NAT-A	Not complete	220	33	8	2	253	10	228	35	263
	EA/ASA-A	Not complete	220	33	25	1	253	26	245	34	279
UCEC-mRNA-OS-3YR	EA/NAT-A	Not complete	148	48	3	2	196	5	151	50	201
	EA/ASA-A	Not complete	148	48	20	3	196	23	168	51	219
UCEC-mRNA-OS-4YR	EA/NAT-A	Not complete	110	52	1	2	162	3	111	54	165
	EA/ASA-A	Not complete	110	52	16	4	162	20	126	56	182
UCEC-mRNA-OS-5YR	EA/NAT-A	Not complete	72	54	1	2	126	3	73	56	129
	EA/ASA-A	Not complete	72	54	16	4	126	20	88	58	146
UCEC-mRNA-PFI-3YR	EA/NAT-A	Not complete	128	73	3	2	201	5	131	75	206
	EA/ASA-A	Not complete	128	73	19	4	201	23	147	77	224
UCEC-mRNA-PFI-4YR	EA/NAT-A	Not complete	93	79	1	3	172	4	94	82	176
	EA/ASA-A	Not complete	93	79	16	4	172	20	109	83	192
UCEC-mRNA-PFI-5YR	EA/NAT-A	Not complete	60	80	1	3	140	4	61	83	144
	EA/ASA-A	Not complete	60	80	16	4	140	20	76	84	160
UCEC-Protein-DSS-2YR	EA/NAT-A	Not complete	178	18	8	2	196	10	186	20	206
	EA/ASA-A	Not complete	178	18	18	0	196	18	196	18	214
UCEC-Protein-DSS-3YR	EA/NAT-A	Not complete	122	26	3	2	148	5	125	28	153
	EA/ASA-A	Not complete	122	26	13	2	148	15	135	28	163
UCEC-Protein-DSS-4YR	EA/NAT-A	Not complete	90	28	1	2	118	3	91	30	121
	EA/ASA-A	Not complete	90	28	10	3	118	13	100	31	131
UCEC-Protein-OS-2YR	EA/NAT-A	Not complete	179	30	8	2	209	10	187	32	219
	EA/ASA-A	Not complete	179	30	18	0	209	18	197	30	227
UCEC-Protein-OS-4YR	EA/NAT-A	Not complete	90	45	1	2	135	3	91	47	138
	EA/ASA-A	Not complete	90	45	10	3	135	13	100	48	148
UCEC-Protein-OS-5YR	EA/NAT-A	Not complete	59	47	1	2	106	3	60	49	109
	EA/ASA-A	Not complete	59	47	10	3	106	13	69	50	119
UCEC-Protein-PFI-3YR	EA/NAT-A	Not complete	105	62	2	3	167	5	107	65	172
	EA/ASA-A	Not complete	105	62	12	3	167	15	117	65	182
UCEC-Protein-PFI-4YR	EA/NAT-A	Not complete	76	65	1	3	141	4	77	68	145
	EA/ASA-A	Not complete	76	65	3	10	141	13	79	75	154

Point-by-Point Responses to the Reviewers' Comments (BIB-21-1509-R1)

Authors' Point-by-Point Responses to the Reviewers' comments on "Reducing healthcare disparities using multiple multiethnic data distributions with fine-tuning of transfer learning" by Toseef, Li, and Wong (BIB-21-1509-R1) submitted to Briefings in Bioinformatics.

Editor's Comments to Author

Reviewers feel the authors have addressed majority of the questions, and only minor revisions are required before its publication in BIB.

Our response: Thank you very much for your time. We have provided the response to the reviewer's comments and provided the publicly available GitHub repository with user instructions.

Reviewer #3 (Reviewers' Comments to Author):

Our response: Thank you very much for your valuable comments and suggestion. We have provided the response to your comments as highlighted points.

The authors provided only source code and datasets. That is not enough. They should provide a package or website for user to analyze their own data. And they should provide a detail instruction to teach user how to use their package or website. Since most of users are not familiar with deep learning and transfer learning..

Our response: Thank you very much for your insightful suggestion. We agree with you and provided a public GitHub repository with user instructions to use the software. All mRNA and protein features datasets are also available. Users can run multiple multiethnic transfer learning experiments by changing the source and target data distributions in EA-ASA and EA-NAT folders. Furthermore, we are working on ethnic cohorts for COVID-19 datasets and we will integrate this project with this research project hosted on a web server for better user experience.

Source code with user instruction, supplementary data, and supplementary tables are provided on the following links.

Supplementary data is available at
https://www.dropbox.com/sh/0a2vh55oulzemze/AADjGcde6z5FzEGIf_FSuDQaa?dl=0

Source code with user instructions is available at
https://github.com/mtoseef99/multiple-multiethnic-disparities-reduction

General response: Thank you very much for your time and valuable suggestions. We would like to say thanks to three reviewers for their time and insightful comments to improve the manuscript.