Quantifying Intratumor Heterogeneity by Key Genes Selected Using Concrete Autoencoder

Raihanul Bari Tanvir
Knight Foundation School of Computing and Information Sciences
Florida International University
Miami, FL, USA
rtanv003@fiu.edu

Masrur Sobhan
Knight Foundation School of Computing and Information Sciences
Florida International University
Miami, FL, USA
msobh002@fiu.edu

Abdullah Al Mamun
Knight Foundation School of Computing and Information Sciences
Florida International University
Miami, FL, USA
mmamu009@fiu.edu

Ananda Mohan Mondal*
Knight Foundation School of Computing and Information Sciences
Florida International University
Miami, FL, USA
amondal@fiu.edu

***Abstract*— The tumor cell population in a cancer tissue has distinct molecular characteristics and exhibits different phenotypes, thus, resulting in different subpopulations. This phenomenon is known as Intratumor Heterogeneity (ITH), which is a major contributor in drug resistance, poor prognosis, etc. Therefore, quantifying the levels of ITH in cancer patients is essential and there are many algorithms which do so in different ways, using different types of omics data. DEPTH (Deviating gene Expression Profiling Tumor Heterogeneity) is the latest algorithm that uses transcriptomic data to evaluate the ITH score. It shows promising performance, has strong similarity with six other algorithms, and has advantage over two algorithms that uses same type of data (tITH, sITH). However, it has a major drawback that it uses expression values of all the genes (~20K genes) in quantifying ITH levels. We hypothesize that a subset of key genes is sufficient to quantify the ITH level for a tumor. To prove our hypothesis, we developed a deep learning-based computational framework using unsupervised Concrete Autoencoder (CAE) to select a set of cancer-specific key genes that can be used to evaluate the ITH score. For experiment, we used gene expression profile data of tumor cohorts of breast, kidney and lung cancer from TCGA repository. We selected three sets of key genes, each set related to breast, kidney, and lung tumor cohorts, using multi-run CAE. For the three cancers stated and three molecular subtypes of lung cancer, we calculated ITH level using all genes and key genes selected by CAE and performed a side-by-side comparison. It was found that similar conclusions can be reached for survival and prognostic outcomes based on ITH scores derived from all genes and the sets of key genes. Additionally, for subtypes of lung cancer, the comparative distribution of ITH scores derived from all genes and key genes remains similar. Based on these observations, it can be stated that, a subset of key genes, instead of all genes, is sufficient for ITH quantification. Our results also showed that many of the key genes are prognostically significant, which can be used as possible therapeutic targets.**

Keywords*— *Concrete Autoencoder, Deep Learning, Gene Expression, Intratumor Heterogeneity, ITH

I. Introduction

Intratumor Heterogeneity (ITH) refers to different types of tumor cell subpopulations within a tumor [1]. Even though these cell subpopulations have the same origin (tumor tissue, patient), they exhibit different phenotypes and possess different molecular characteristics. ITH is one of the main challenges for targeted cancer therapy, as the difference in tumor cells and their microenvironments makes it harder for targeted cancer therapy to eradicate cancer cells [2][3]. Therefore, an accurate assessment of ITH is essential to understand the tumor dynamics and development of effective and durable therapeutic strategies.

The causes of ITH can be many depending on different levels, such as genome, epigenome, transcriptome, etc. [4]. For example, reduced DNA damage mechanisms, microenvironmental factors (hypoxia, acidosis, etc.) [5], subclonal evolution [2], etc. are contributors of ITH in genomic level. Methylation of tumor suppressor genes is an example of ITH at epigenomic level [6]. Different patterns of gene expression contribute to ITH at transcriptome level, which is observed to mirror ITH in genomic or epigenomic level or both [5][7]. This makes transcriptomic data suitable for quantifying ITH.

There are different algorithms for quantifying ITH, such as ABSOLUTE [8], MATH [9], EXPANDS [10], PhyloWGS [11]. These algorithms use genomic data, such as – copy number alterations (CNA), somatic mutation profiles, etc. There are also algorithms that take advantage of transcriptome profile that mirrors ITH in genomic and/or epigenomic level, such as- tITH [12], sITH [13], and DEPTH [14]. Among the algorithms that quantify ITH, DEPTH showed strong similarity in calculating a degree of ITH with other genome-based and transcriptome-based methods. DEPTH has certain advantages over its two other transcriptome-based counterparts, tITH and sITH. First, unlike the other two, DEPTH can calculate ITH score without normal samples. Second, it calculates ITH score using only gene

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

expression data. tITH requires protein-protein interaction (PPI) network along with gene expression data.

DEPTH algorithm also has certain drawbacks. While it uses only gene expression data, it uses expression values of all genes (~20K genes) in calculating ITH score. While calculating ITH, not all the genes may contribute to ITH. Additionally, not all genes may have alterations in gene expression patterns due to the effect of ITH in genomic and/or epigenomic level. There may be a subset of genes which is sufficient in calculating ITH score at transcriptome level. In this study, we propose a deep learning-based computational framework based on unsupervised concrete autoencoder (CAE) [15] to identify a set of key genes that can be used to quantify ITH. CAE can select the most informative subset of features from a large set of features which also have a minimum reconstruction error. First, we select a set of key genes from breast (BRCA), kidney (KIRC) and lung cancer (LUAD) using the expression profile data from TCGA repository. Then we calculate the ITH score for the tumor patients in three cancer types using all genes and the key genes employing DEPTH algorithm. We also calculate DEPTH score using all genes and key genes for different molecular subtype of LUAD. In both cases, we show that a set of key genes (k=100) can achieve similar results as using all genes (k=20,000+). Our contribution is the development of a computational framework to identify key genes that can be used in evaluating the ITH score. In other words, we do not need all the genes used in the previous work in evaluating the ITH score.

II. Materials and Methods

A. Dataset Collection

We collected gene expression dataset of TCGA Breast Carcinoma (BRCA), Lung Adenocarcinoma (LUAD) and Kidney Renal Carcinoma (KIRC) from USCS Xena Browser database [16]. Each dataset contains expression profiles of 20,530 mRNAs. The distribution of tumor and normal samples for three cancers is given in Table I.

TABLE I. Summary of gene expression data of BRCA, LUAD and KIRC cohorts from TCGA.

Cancer Type	# Tumor	# Normal	# Genes
BRCA	1097	114	20530
KIRC	533	72	20530
LUAD	515	59	20530

B. Concrete Autoencoder to Select Cancer-specific Key Genes

Concrete Autoencoder (CAE) [15], a unsupervised deep learning approach, is used to identify cancer-specific key genes. CAE identifies features that are most informative for a given dataset [15], [17].

CAE Architecture and Working Principle: Figure 1 shows the general architecture of CAE. CAE differs from the standard Autoencoder in the encoder part, where CAE employs a concrete selector layer. This selector layer is based on Concrete distribution [18] or Gumbel Softmax distribution [19], which is a relaxed variant of discrete distribution. The selector layer is used to incorporate discrete distribution into deep learning algorithms, and as an example, CAE uses it to learn a subset of features that are most informative and produces minimum reconstruction error. In the learning phase, the selector layer learns a subset of features, which depends on a hyperparameter called Temperature (T), which is gradually lowered during training phase to a low value using a simple annealing scheduling. This gradual decrease in temperature helps the concrete distribution to learn and select a definite subset of features [15]. Unlike the encoder part of the CAE, the decoder part resembles closely with the standard Autoencoder.

In the selector layer, each unit selects a unique feature with the highest probability from the original feature space. Thus, CAE selects the most informative subset of features and the reconstruction of original feature space using the selected subset of features produces minimum reconstruction error. In the original Autoencoder, the features learned at the encoder part are latent features, whereas the features learned at CAE are actual features.

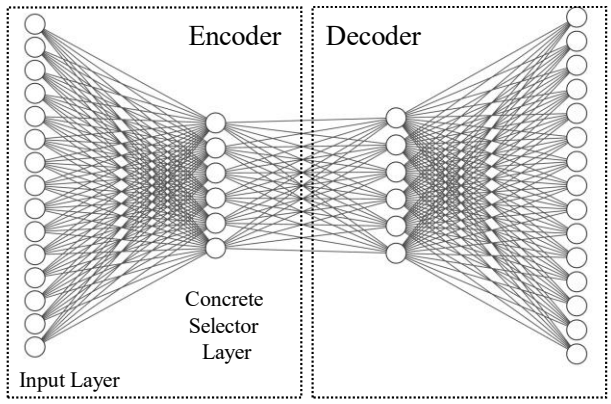

Fig. 1. The general architecture of CAE. The encoder part contains the input layer and the concrete selector layer. The decoder in this figure is a 2-layer neural network, with the final layer having the same number of nodes as the input layer.

CAE was trained on the each of the gene expression data of BRCA, KIRC and LUAD and 100 features were selected in each run. While training CAE, the dataset was divided randomly into 80/20 split for training and testing.

C. Selection of Hyperparameter and Tuning

We kept three of the parameters the same as used in the original CAE developed by Abid et al. [15]. These parameters are (i) leaky ReLU as activation function with a threshold value of 0.1, (ii) 10% dropout rate, and (iii) temperature T reduced using a simple annealing schedule from 10 to 0.1, For decoder part of CAE, we used a neural network with two hidden layers, each layer containing 300 units as used in our previous work with expression profile data [17].

Number of epochs and learning rate was tuned using random search [20]. For *tuning learning rate*, we varied the number of epochs from 2000 to 10000. The number of epochs were set to high since the number of samples are low (ranges between 515 for KIRC to 1097 for BRCA) compared to the feature space of ~20K genes. Another reason for using large number of epochs is to reach the high mean-max probability. Mean-max probability is the average of maximum probability value for each feature selected by a node in the concrete selector layer.

Each node in concrete selector layer selects the feature with the highest probability. Therefore, one of the main goals in training of CAE is to make sure that the mean-max probability is high (>= 0.98). The learning rate was varied from 0.001 to 0.01 with a step size of 0.001 and it was found that 0.002 was optimum in reaching the mean-max probability >= 0.98. The number of epochs set for BRCA, KIRC and LUAD were 3500, 9000, and 8000, respectively. The average number of epochs to reach the threshold of mean-max probability >= 0.98 were 3066, 8191 and 7250 for BRCA, KIRC, and LUAD datasets, respectively.

D. Training CAE

Figure 2 shows the characteristics curve for CAE or an instance of the training behaviors of CAE for LUAD dataset. The hyperparameter, Temperature (T) was reduced using a simple annealing schedule from 10 to 0.1 from the start epoch to the last. The reconstruction errors (loss) for training set and validation set are plotted using blue and red curve respectively. It shows that both errors were relatively high during the early phase of training as expected and both reaches to a minimum plateau at the end. Also, the mean-max probability finally approaches 1.0 (yellow curve).

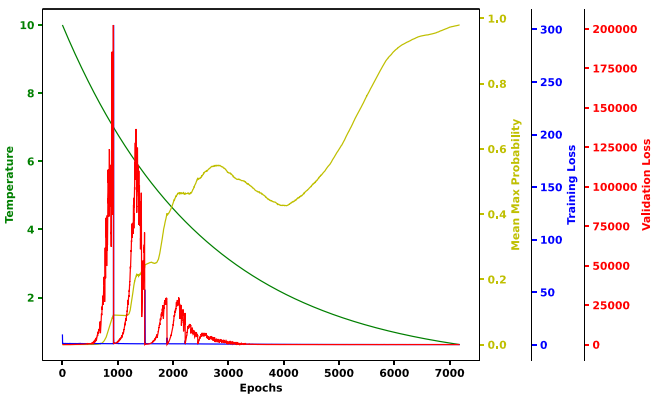

Fig. 2. The characteristics curve of CAE. Reconstruction errors with training data (blue) and validation data (red) were plotted. The green and yellow curve represent the values of Temperature (T) and mean-max probability, respectively.

Implementation: The CAE was implemented using Keras (https://keras.io/). Experiments were conducted in parallel manner on high performance cluster with NVIDIA Quatro K620 GPU with 384 cores and 2GB memory devices.

E. ITH scoring method

To calculate the score of Intratumor Heterogeneity (ITH), we used a scoring method named- Deviating Gene Expression Profiling Tumor Heterogeneity, or DEPTH in short [14]. This method is mRNA-based, and it only uses the gene expression values of tumor samples and normal samples. The DEPTH score of a tumor sample TS is defined as –

$$\frac{1}{m-1}\sqrt{\sum_{i=1}^{m}\left(\left(ex(G_i,\,TS)-\frac{1}{n}\sum_{j=1}^{n}ex(G_i,NS_j)\right)^2-\frac{1}{m}\sum_{j=1}^{m}\left(ex(G_i,TS)-\frac{1}{m}\sum_{i=1}^{m}ex(G_i,NS_j)\right)^2\right)^2}$$

Where, TS is the tumor sample for which the score is being calculated. G_i is the i-th Gene and m is the number of genes. n is the number of normal patients. $ex(G_i,\,S)$ expression of gene in sample S. It assigns a score to each patient. It is based on standard deviations of the gene expression values variations in tumors from mean value of normal samples. We calculated ITH score for each of the cancer patients of BRCA, KIRC and LUAD employing DEPTH using two sets of genes. One score using all the genes as used in [14] and the other using only the key genes selected by multi-run CAE.

F. Survival Analysis

Survival Analysis was performed to check whether two groups of patients based on high and low ITH scores are significantly distinguishable in prognosis. In our analysis, the event of interest is the death of cancer patients.

Survival Analysis Based on ITH Scores: Samples were sorted in descending order of ITH score, and then the top one-third and the bottom one-third of the total samples were taken as two groups. This analysis was done to compare the prognostic capability of ITH scores derived using all genes [14] and key genes (our study) (see Section III-C, III-D).

Survival Analysis Based on Each of Key Genes: The cohort was divided into two groups based on median of the expression values of the gene in question. This survival analysis was done to check how many of key genes are prognostic.

In both cases, after forming two groups, Kaplan-Meier curve was plotted, and Log-rank test was performed to check the statistical significance of difference in survival function.

III. Results and Discussion

A. Selecting Key Genes – Multi-Run CAE

Due to the stochastic nature of CAE, the model was trained 10 times, and in each run, 100 features were selected for each cancer cohort - BRCA, KIRC and LUAD. Figure 3(a) shows the stochastic nature of CAE since only 16 features are common between three single-run CAE. In case of 10-run CAE, top 100 features were selected from the combined list sorted in descending order based on the frequency of appearance of a feature in 10 runs. It is clear from Figure 3(b) that there are 53 genes common between three batches of 10-runs. Thus, multi-run approach was adopted to select the robust set of features. To select the key features, the top 100 frequent features were chosen based on the assumption that the more frequent a feature in different runs, the more informative the feature is. The summary of 10-run CAE for the three cancer cohorts are shown in Table II. The combined lists of features derived from 10-run CAE consist of 469, 527, and 435 genes for BRCA, KIRC, and LUAD, respectively. The range of frequency of top 100 features is 3 to 10 for each cancer cohort, which means that the most frequent features appeared in all 10 runs and the least frequent one appeared in 3 runs.

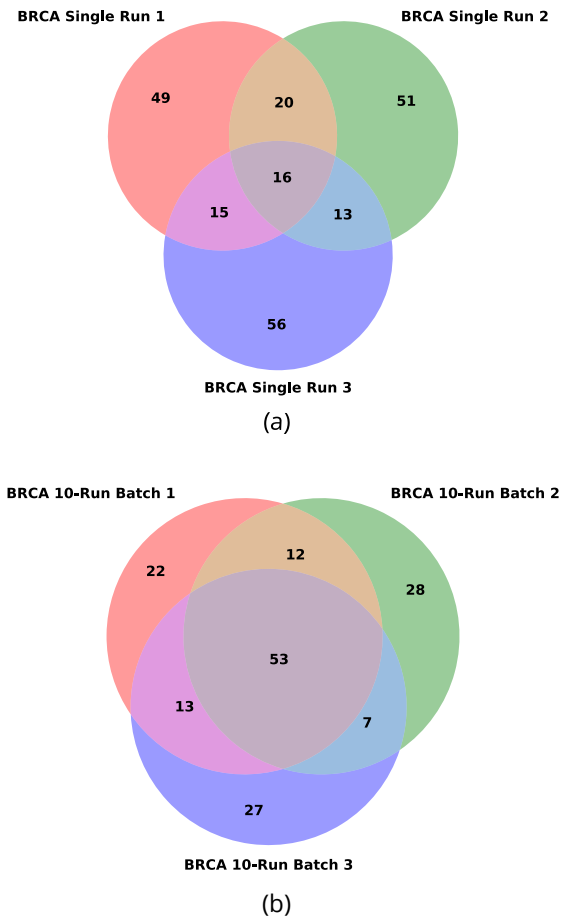

Fig. 3. Selecting robust set of features. (a) Venn diagram of three sets of 100 genes from three single-run CAE; CAE produces only 16 features common between three single runs. (b) three sets of top 100 features from 10-run CAE. 10-run CAE produces more features (53 genes) common between three batches of runs. Thus, multi-run CAE produces robust set of features.

TABLE II. SUMMARY OF MULTI-RUN CAE RUNS FOR BRCA, KIRC, LUAD COHORT.

Cancer Type	# Genes in Multi-run	Frequency Range in Top 100
BRCA	469	[3-10]
KIRC	527	[3-10]
LUAD	435	[3-10]

B. *Commonality Check*

We investigated the commonality of the key gene sets of three cancers, which is shown by the Venn diagram in Figure 4. It shows that there is no common gene between three gene sets. But there are a few genes common between each pairs of gene sets: 5 between BRCA and LUAD, 3 between KIRC and BRCA, and 3 between KIRC and LUAD. Since the size of each set is 100 and there are only few genes common between two sets and none between the three sets, thus, the key gene sets are cancer-specific.

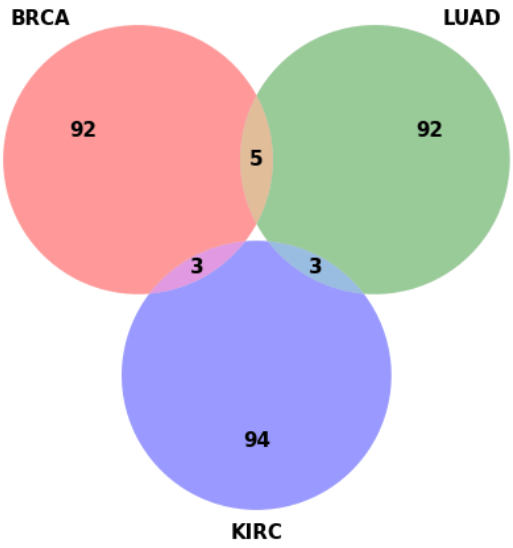

Fig. 4. Venn diagram of three sets of Key genes for BRCA, KIRC and LUAD cohort.

C. *All Genes vs. Key Genes in ITH scoring: Whole Cancer Cohort*

In this section, we compare the ITH scores calculated for BRCA, KIRC and LUAD cohorts using two different sets of genes: (i) DEPTH score calculated using all genes [14] and (ii) DEPTH score calculated using only the key genes selected by the multi-run CAE system (our work). Survival analysis is used to compare the two ITH scores. Figure 5 presents the results of survival analyses for three cancer cohorts based on ITH scores derived from all genes and key gene set.

In Figure 5, Kaplan Meier plots are shown for BRCA (a-b), LUAD (c-d) and KIRC (e-f) cohorts. In the plots in the top three, the groups are formed based on DEPTH scores calculated using all the genes. For the bottom three, groups were formed using DEPTH scores calculated using the key genes. In all the six Kaplan Meier plots, it is evident that high DEPTH score is related to poor prognosis and low DEPTH scores have higher chance of survival. By comparing plots (a) with (b), (c) with (d), and (e) with (f), it is clear that the similar results can be achieved for each cancer cohort by using only the 100 key genes. Thus, we do not need all 20K genes to evaluate the ITH scores as done in [14]. The hazard ratio and the log-rank P-values are slightly different between two corresponding plots. For instance, for BRCA, P-value of 0.0067 and Hazard Ratio of 1.75 were found in survival analysis using all genes; whereas key genes produced a P-value of 0.0369 and Hazard Ratio of 1.52. Similar pattern can be observed for KIRC and LUAD as well. But unlike BRCA, KIRC and LUAD shows better outcome for key genes than all genes as they produced reduced P-values and increased Hazard Ratios. It is evident that, only the key genes can be used to quantify ITH in whole cohort of each cancer type.

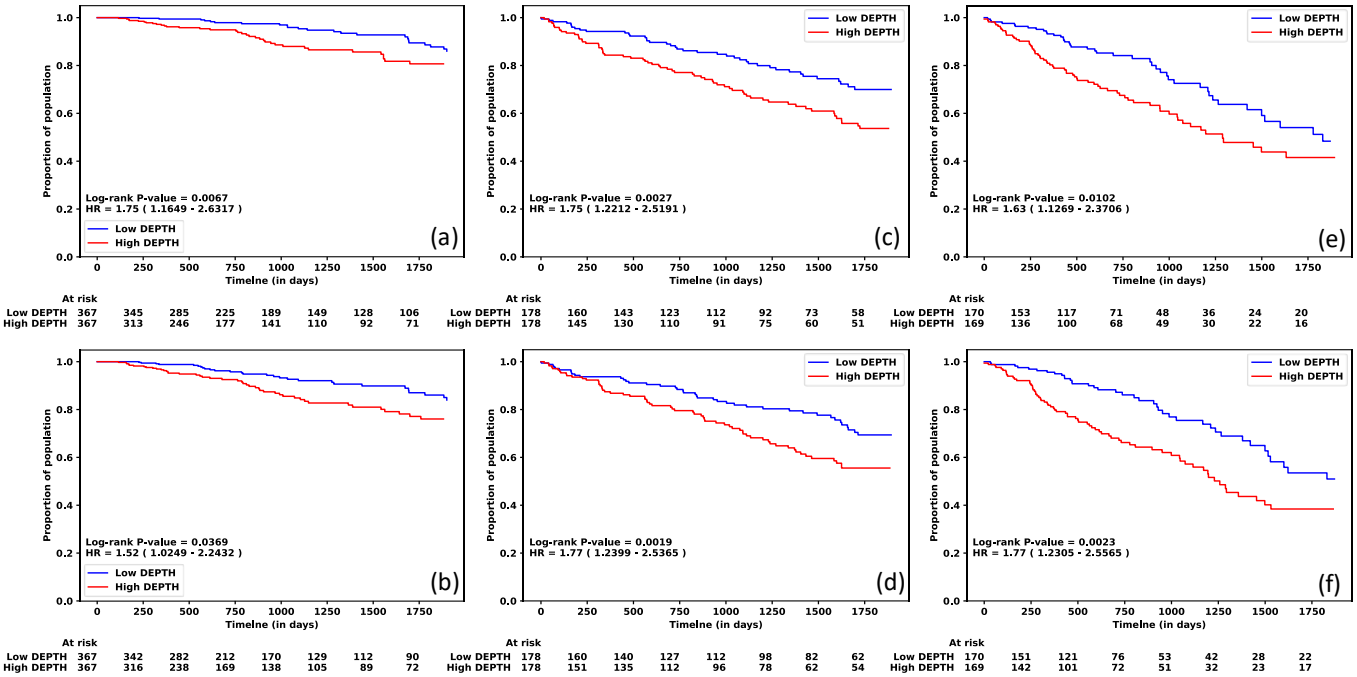

Fig. 5. Prognostic analysis using DEPTH score. Kaplan Meier plots for BRCA(a)(b), KIRC(c)(d) and LUAD(e)(f). Top-row KM plots: groupings were done based on DEPTH scores calculated using all genes ((a),(c),(e)). Bottom-row plots: groupings were done based on DEPTH scores calculated using key genes ((b),(d),(f)).

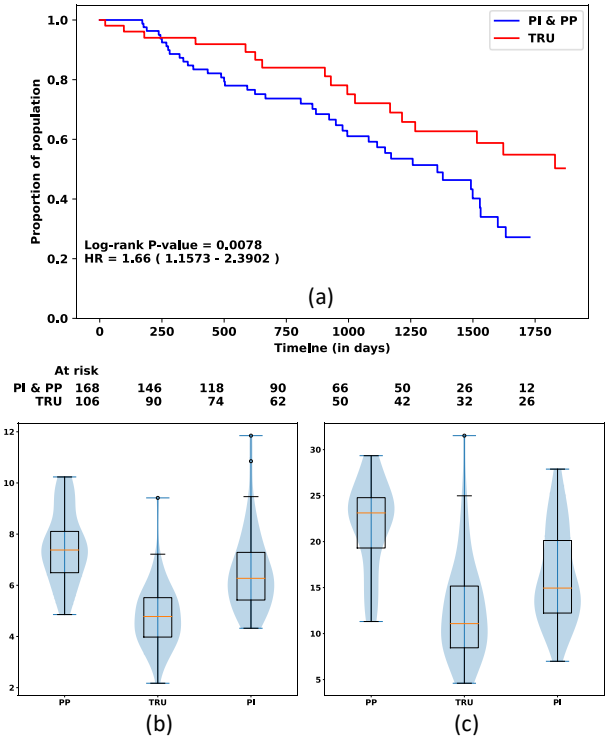

Fig. 6. Comparison of ITH score using all genes and key genes in molecular Subtypes of LUAD. (a) Kaplan-Meier Curve comparing the survial of TRU subtype vs PI and PP subtypes combined. (b) Distribution of ITH score (DEPTH) in three molecular subtypes of LUAD, calculated using all genes in terms of boxplot and violinplot. (c) Distribution of ITH score (DEPTH) in three molecular subtypes of LUAD, calculated using key genes of LUAD.

D. All Genes vs. Key Genes in ITH scoring: LUAD subtypes

In this section, we show the comparison of ITH scores (DEPTH) for LUAD subtypes calculated using all genes versus the key genes. Of 435 LUAD patients, 55, 34 and 54 are labeled as Terminal Respiratory Unit (TRU), Proximal Proliferation (PP) and Proximal Inflammation (PI). This molecular subtyping is done in [21]. We compare ITH score for two cases: all genes vs. key genes. Figure 6(a) shows the prognostic difference between TRU and PI and PP combined, from where it is evident that, TRU subtype is prognostically favorable and it has higher chance of survival than other two subtypes combined. Figure 6(b) and 6(c) show the ITH score distribution for three subtypes, using all genes and key genes. In both plots, it is seen that TRU has comparatively lower values in ITH score than the other subtypes. Although, the set of genes used to calculate the ITH score are different in 6(b) and 6(c) and the values are changed, the overall comparative distribution remained the same. Based on this observation, it can be concluded that to calculate the ITH score, only the key genes will suffice.

E. Survival Analysis of Key Genes

Survival analysis was performed on each of the key genes from their respective cancer cohort to identify whether they possess prognostic capabilities. Figure 7 shows the forest plot of the prognostically significant genes, along with the summary of survival analyses in terms of Logrank P-value and Hazard Ratio with 95% confidence interval. The thresholds for prognostically significant genes are Logrank P-value ≤ 0.05 and Hazard Ratio, HR $\neq 1$. Of 100 key genes for BRCA, 15 are prognostically significant. Similarly, for KIRC and LUAD, 30 and 61 genes are prognostically significant, as shown in Table III.

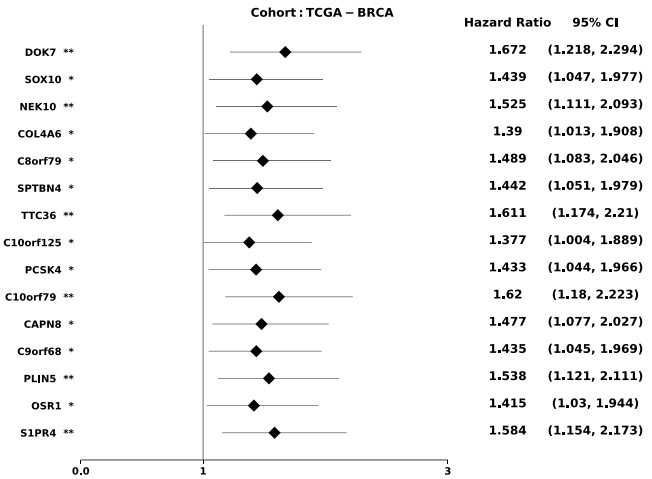

Fig. 7. Forest plot of key genes for BRCA. Summary of survival analyses of key genes that are prognostically significant. It shows the Hazard Ratio of the value (diamond shape) and its 95% confidence interval (straight line crossing the diamond). The logrank P-value is represented by the number of stars alongside the gene (* - P ⩽ 0.05, ** - P ⩽ 0.01, *** - P ⩽ 0.001, **** - P ⩽ 0.0001).

TABLE III. NUMBER OF PROGNOSTIC GENES FOR EACH TYPE OF CANCERS.

Cancer Type	# Prognostic Genes
BRCA	15
KIRC	30
LUAD	61

IV. CONCLUSION

In this study, we hypothesize that a subset of key genes, instead of all genes (~20K genes), is sufficient to evaluate the ITH scores of individual tumors. To prove our hypothesis, we developed a computational framework based on multi-run concrete autoencoder that selects a set of key genes for a tumor cohort that can be used to evaluate the ITH score for individual tumor. It was shown that, instead of all genes used in the existing state-of-the-art method, a subset of key genes is sufficient to evaluate the ITH score using gene expression profile data. We showed that ITH scores derived using the subset of key genes and all genes produced the similar survival and prognostic outcomes for three different cancers (BRCA, KIRC, and LUAD) as well as molecular subtypes of LUAD. Based on this study, we conclude that a subset of key genes is sufficient to quantify the ITH at transcriptome level. We also showed that many of these key genes are prognostically significant, which can be investigated farther as possible therapeutic targets.

ACKNOWLEDGMENT

The work has been partially supported by the NSF CAREER Award (#1651917 transferred to #1901628) to AMM.

REFERENCES

[1] M. Jamal-Hanjani, S. A. Quezada, J. Larkin, and C. Swanton, “Translational implications of tumor heterogeneity,” *Clinical Cancer Research*. 2015, doi: 10.1158/1078-0432.CCR-14-1429.

[2] M. A. Qazi *et al.*, “Intratumoral heterogeneity: Pathways to treatment resistance and relapse in human glioblastoma,” *Ann. Oncol.*, 2017, doi: 10.1093/annonc/mdx169.

[3] R. Reinartz *et al.*, “Functional Subclone profiling for prediction of treatment-induced intratumor population shifts and discovery of rational drug combinations in human glioblastoma,” *Clin. Cancer Res.*, 2017, doi: 10.1158/1078-0432.CCR-15-2089.

[4] T. M. Grzywa, W. Paskal, and P. K. Włodarski, “Intratumor and Intertumor Heterogeneity in Melanoma,” *Transl. Oncol.*, vol. 10, no. 6, pp. 956–975, Dec. 2017, doi: 10.1016/j.tranon.2017.09.007.

[5] R. J. Gillies, D. Verduzco, and R. A. Gatenby, “Evolutionary dynamics of carcinogenesis and why targeted therapy does not work,” *Nature Reviews Cancer*. 2012, doi: 10.1038/nrc3298.

[6] J. Takamizawa *et al.*, “Reduced Expression of the let-7 MicroRNAs in Human Lung Cancers in Association with Shortened Postoperative Survival,” *Cancer Res.*, vol. 64, no. 11, pp. 3753–3756, 2004, doi: 10.1158/0008-5472.CAN-04-0637.

[7] L. Sigalotti *et al.*, “Intratumor heterogeneity of cancer/testis antigens expression in human cutaneous melanoma is methylation-regulated and functionally reverted by 5-Aza-2′-deoxycytidine,” *Cancer Res.*, 2004, doi: 10.1158/0008-5472.CAN-04-1442.

[8] S. L. Carter *et al.*, “Absolute quantification of somatic DNA alterations in human cancer,” *Nat. Biotechnol.*, vol. 30, no. 5, pp. 413–421, May 2012, doi: 10.1038/nbt.2203.

[9] E. A. Mroz and J. W. Rocco, “MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma,” *Oral Oncol.*, vol. 49, no. 3, pp. 211–215, Mar. 2013, doi: 10.1016/j.oraloncology.2012.09.007.

[10] N. Andor, J. V. Harness, S. Müller, H. W. Mewes, and C. Petritsch, “EXPANDS: expanding ploidy and allele frequency on nested subpopulations,” *Bioinformatics*, vol. 30, no. 1, pp. 50–60, Jan. 2014, doi: 10.1093/bioinformatics/btt622.

[11] A. G. Deshwar, S. Vembu, C. K. Yung, G. H. Jang, L. Stein, and Q. Morris, “PhyloWGS: Reconstructing subclonal composition and evolution from whole-genome sequencing of tumors,” *Genome Biol. 2015 161*, vol. 16, no. 1, pp. 1–20, Feb. 2015, doi: 10.1186/S13059-015-0602-8.

[12] Y. Park, S. Lim, J.-W. Nam, and S. Kim, “Measuring intratumor heterogeneity by network entropy using RNA-seq data,” *Sci. Reports 2016 61*, vol. 6, no. 1, pp. 1–12, Nov. 2016, doi: 10.1038/srep37767.

[13] M. Kim, S. Lee, S. Lim, and S. K. Id, “SpliceHetero: An information theoretic approach for measuring spliceomic intratumor heterogeneity from bulk tumor RNA-seq,” *PLoS One*, 2019, doi: 10.1371/journal.pone.0223520.

[14] M. Li, Z. Zhang, L. Li, and X. Wang, “An algorithm to quantify intratumor heterogeneity based on alterations

of gene expression profiles," *Commun. Biol.*, vol. 3, no. 1, pp. 1–19, Dec. 2020, doi: 10.1038/s42003-020-01230-7.

[15] A. Abid, M. F. Balin, and J. Zou, "Concrete autoencoders: Differentiable feature selection and reconstruction," *36th Int. Conf. Mach. Learn. ICML 2019*, vol. 2019-June, pp. 694–711, 2019.

[16] M. Goldman *et al.*, "The UCSC Xena platform for public and private cancer genomics data visualization and interpretation," *bioRxiv*, p. 326470, Sep. 2018, doi: 10.1101/326470.

[17] A. Al Mamun *et al.*, "Multi-run Concrete Autoencoder to Identify Prognostic lncRNAs for 12 Cancers," *bioRxiv*, p. 2021.08.01.454691, Aug. 2021, doi: 10.1101/2021.08.01.454691.

[18] C. J. Maddison, A. Mnih, and Y. W. Teh, "The Concrete Distribution: A Continuous Relaxation of Discrete Random Variables," *5th Int. Conf. Learn. Represent. ICLR 2017 - Conf. Track Proc.*, Nov. 2016.

[19] E. Jang, S. Gu, and B. Poole, "Categorical reparameterization with gumbel-softmax," 2017.

[20] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, 2012.

[21] "Comprehensive molecular profiling of lung adenocarcinoma," *Nature*, 2014, doi: 10.1038/nature13385.