**Supplementary File S1: Description of normalization methods used**

**Data normalization**

Four normalization approaches were used on the training, test and confirmation datasets:

* “**Unit scale normalization**”: For a given sample vector $v=(v\_{1},…,v\_{n})$:

$$v\_{normed}=v.\frac{1}{\left‖v\right‖\_{2}^{2}}$$

With $\left‖v\right‖\_{2}^{2}$ the $l\_{2}$norm of *v.* The purpose of this normalization is to obtain samples with a norm scaled to 1.

* “**Median scale normalization**”: The median absolute deviation (*mad*) is a robust estimator of the variability of a univariate sample. For a given feature vector $x=(x\_{1},…,x\_{n})$:

$$mad\left(x\right)=median(\left\{\left|x\_{i}-median(x)\right|, x\_{i}\in x\right\})$$

 The “median scale normalization” is defined by:

$$x\_{scaled}=\left(x-median\left(x\right)\right).\frac{1}{mad(x)}$$

* “**Robust scale normalization**”: For each feature, the mean and the standard deviation is computed using values between the first and the last quantile from the training set (in the case of an ordered array of values, values used begin after the first 25% until reaching 75% of the total values).

$$x\_{whitened}= \left\{\frac{x\_{i}-mean\_{25-75}(x\_{i})}{std\_{25-75}(x\_{i})}, x\_{i}\in x\right\}$$

This normalization presents the advantage to compute robust mean and standard deviation estimates without being influenced by possible outliers.

* “**Rank normalization**”: For a given sample vector $v=(v\_{1},…,v\_{n})$, the values $v\_{i}$ are ordered and the function $rank(v\_{i})$ returns the rank of the value $v\_{i}$. (1 if it is the lowest and *n* if it is the highest value).

$$v\_{rank}=\frac{1}{n}.(rank(v\_{1}),…,rank(v\_{n}))$$

This normalization aims to compare samples using the relative ranks of their features without considering their values.

We used the 3 pre-processed TCGA HCC omics data sets as the input for the autoencoders framework. We stacked the 3 matrices that are then scaled using the unit scale normalization. Autoencoders were previously showed to be good frameworks to fuse heterogeneous features (1). To classify new samples, we first selected the common set of features between the new samples and the training set (the TCGA dataset) to create a new training matrix. If the samples have multi-omic features, we processed each type of omic independently.

We conducted 2-step (MAD+scale) normalization. First, we used the median scale (MAD) normalization on both the training set and the new test samples. A previous normalization approach based on mad estimator were already used to normalize samples from RNA-seq data (2). Then, for mRNA and DNA methylation data, we applied the robust scale normalization on the training data set, and then scaled the new confirmation samples using the means and the standard deviation computed from the training set. For the miRNA data, we used unit scale normalization for both the miRNA training and the confirmation matrix.

As an alternative procedure to MAD+scale approach, we also used the rank normalization to normalize samples from all the training, test and confirmation sets before the classification. This procedure can be used to classify a unique new sample which doesn’t belong to any cohort. We additionally tested two other alternatives of preprocessing, using just robust scale or median scale normalizations alone. These three alternative normalization procedures showed overall decent performances in term of C-index and p-value for the different test and confirmation sets (Table S2).

**References**

1. Liu F, Li H, Ren C, Bo X, Shu W. PEDLA: predicting enhancers with a deep learning-based algorithmic framework. Sci Rep 2016;6:28517.

2. Zhang B, Wang J, Wang X, Zhu J, Liu Q, Shi Z*, et al.* Proteogenomic characterization of human colon and rectal cancer. Nature 2014;513(7518):382-7.