1.1. *Metrics of performance for each model*

Below we discuss the metrics utilized to evaluate performance of each model. As described above, multiple variations of models from each data modality were constructed, and the ones that performed best, within a single data level, were chosen for performance comparison across data levels. The Rscripts for model evaluation are included in the "*Rscripts_regression_analysis.zip*" file included in the **supplementary material**.

1.1.1. *Proteogenomics*

To measure the efficacy of each model, we calculate the measures of accuracy, precision, recall and F-score (using function "*prediction*" from the R-package "*ROCR*"[1]). Each subset of the proteogenomics datasets is modeled using multinomial log-linear regression and cross-validated (70% training) over 100 iterations. The measures listed previously are computed across all the labels within a cohort (label versus NOT label) and the results report the average. Traditional definitions of all four measures are utilized, and are additionally weighed by the "support" (i.e the fraction of the corresponding label in the test set).

1.1.2. *Histology Images*

In order to effectively compare the results of the each of the models from CNNs to proteogenomic regression models, we calculate the measures of precision, recall and F-score for each clinical cohort (averaged for all labels). The input tiles were divided to training, "dev" (subset of data utilized by the CNN to iteratively measure improvement and subsequently update feature weights to finalize the model) and test sets to calculate these performance metrics, while ensuring that different tiles from the same patient are not contained in both the testing and training sets. The "*sklearn.metrics.classification_report*"[2] python module was used for evaluating the image classification models.

References

1. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. *Bioinformatics*. 2005;21(20):7881. http://rocr.bioinf.mpi-sb.mpg.de.
2. Bowles M. Machine learning in Python. *Igarss 2014*. 2014;(1):1-5. doi:10.1007/s13398-014-0173-7.2.