Article

Examination of Independent Prognostic Power of Gene Expressions and Histopathological Imaging Features in Cancer

Tingyan Zhong [1], Mengyun Wu [2],* and Shuangge Ma [3],*

[1] SJTU-Yale Joint Center for Biostatistics, Department of Bioinformatics and Biostatistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China; tyzhong@sjtu.edu.cn

[2] School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China

[3] Department of Biostatistics, Yale University, New Haven, CT 06520, USA

* Correspondence: wu.mengyun@mail.shufe.edu.cn (M.W.); shuangge.ma@yale.edu (S.M.)

check for updates

Abstract: Cancer prognosis is of essential interest, and extensive research has been conducted searching for biomarkers with prognostic power. Recent studies have shown that both omics profiles and histopathological imaging features have prognostic power. There are also studies exploring integrating the two types of measurements for prognosis modeling. However, there is a lack of study rigorously examining whether omics measurements have independent prognostic power conditional on histopathological imaging features, and vice versa. In this article, we adopt a rigorous statistical testing framework and test whether an individual gene expression measurement can improve prognosis modeling conditional on high-dimensional imaging features, and a parallel analysis is conducted reversing the roles of gene expressions and imaging features. In the analysis of The Cancer Genome Atlas (TCGA) lung adenocarcinoma and liver hepatocellular carcinoma data, it is found that multiple individual genes, conditional on imaging features, can lead to significant improvement in prognosis modeling; however, individual imaging features, conditional on gene expressions, only offer limited prognostic power. Being among the first to examine the independent prognostic power, this study may assist better understanding the "connectedness" between omics profiles and histopathological imaging features and provide important insights for data integration in cancer modeling.

Keywords: cancer prognosis; independent prognostic power; omics profiles; histopathological imaging features

1. Introduction

In cancer research and practice, prognosis is of essential interest. Extensive statistical modeling has been conducted, and yet there is still much room for additional research [1,2]. Multiple families of biomarkers have been used in cancer prognosis modeling. In the past decades, with the development of profiling techniques, omics data have been extensively used and shown to be effective. For example, in [3], a signature composed of 21 gene expressions is used for modeling breast cancer prognosis. In [4], hsa-mir-155 and hsa-let-7a-2 microRNAs are found as prognostic for lung cancer. In [5], the prognostic power of methylated RASSF1A and/or APC serum DNA for breast cancer is identified. Such findings are biologically highly sensible as cancer is a genetic disease and the development and progression of cancer are strongly affected by molecular changes.

Imaging techniques have been extensively used in cancer practice. In particular, for definitive diagnosis, biopsies are usually ordered, and pathologists review the slides of representative sections

of tissues. The histopathological imaging data generated in this process directly reflect the histological organization and morphological characteristics of tumor cells and their surrounding tumor microenvironment. More recently, beyond diagnosis, histopathological imaging data have also been used for modeling other cancer outcomes/phenotypes. For example, Harpole et al. [6] showed that the levels of tumor cell dedifferentiation are associated with survival outcomes. Automated histopathological imaging analysis systems have been shown to be effective in the prognostic determination of various malignancies, including breast cancer [7], neuroblastoma [8], lymphoma [9], nonsmall cell lung cancer [10], precancerous lesions in the esophagus [11], and others.

Omics profiles and histopathological images are biologically connected and contain overlapping information. In particular, properties of tumors and their microenvironment, as described in histopathological images, are highly regulated by molecular changes. Multiple statistical modelings have been conducted on their interconnections. For example, Yu et al. [10] found the correlation of quantitative histopathological features with TP53 mutation in lung adenocarcinoma. Cooper et al. [12] demonstrated that PDGFRA, EGFR, and MDM2 amplifications are associated with imaging features such as greater minor axis length, area, and circularity in glioblastoma.

On the other hand, it is also possible that omics profiles and histopathological images may contain independent information. Tissues used to generate omics measurements are usually heterogeneous and mixtures of different components, making the observed omics measurements represent an aggregation of distinct cells [12]. Histopathological images, with extremely high spatial resolutions, also contain information on tissue relationships and characteristics of the spatial organizations of tumor cells. Features of histopathological images, beyond regulated by molecular changes, may also be affected by personal immune system, microenvironment, and other factors. There are a few recent studies integrating omics measurements and imaging features and showing that such a data integration can lead to improved prognosis modeling for breast cancer [13], brain tumor [14], and nonsmall cell lung cancer [15]. Such studies seem to suggest that there is independent information in omics and histopathological imaging data.

Our literature review suggests that there is a lack of study rigorously quantifying whether omics profiles contain independent information, conditional on histopathological images, on cancer prognosis, and vice versa. It is noted that the aforementioned and other data integration studies do not suggest which type of measurement has independent information. In addition, they are mostly estimation-based and do not have a rigorous statistical inference framework. Some studies rely on prescreening to accommodate high data dimensionality and can be statistically limited. This study is conducted to directly tackle these problems. It can advance from the existing literature in multiple aspects. Specifically, it delivers an analysis pipeline which has a rigorous statistical inference framework and can show whether individuals of a type of measurement have additional prognostic power conditional on the other type of measurement. It can provide solid evidence on whether data integration is needed in modeling and clinical practice. In addition, data analysis may also provide further insights into two deadly cancers.

2. Data

The Cancer Genome Atlas (TCGA) is one of the largest and most comprehensive cancer projects, and is organized by the National Cancer Institute (NCI) and National Human Genome Research Institute (NHGRI) [16]. It has published high quality omics data and histopathological images for 33 types of cancer. In this study, we analyze data on lung adenocarcinoma (LUAD) and liver hepatocellular carcinoma (LIHC), both of which have high mortality rates and pose increasing public concerns. They also have relatively larger sample sizes, which is critical for making reliable findings. The proposed analysis can be extended to other cancer types straightforwardly. For omics measurements, we consider gene expressions, which have a central role in cancer prognosis modeling. The response of interest is overall survival time, which is right censored. It is noted that other types of omics measurements,

for example protein expressions and microRNAs, and other prognosis outcomes can be analyzed in a similar manner.

2.1. mRNA Gene Expression Measurements

We download the mRNA gene expression data from cBioPortal (http://cbioportal.org), which have been measured using the Illumina Hiseq2000 RNA Sequencing Version 2 analysis platform, and processed and normalized using the RSEM software. We refer to the literature [17,18] and TCGA website for more detailed information on data collection and processing. Data are available on 25,031 gene expression measurements. As the number of cancer-related genes is not expected to be large, we conduct a simple prescreening to reduce the dimensionality and also improve the stability of estimation. Specifically, the top 5000 genes with the largest variances are selected for downstream analysis.

2.2. Histopathological Imaging Features

We download the whole-slide histopathology images directly from the TCGA website (https://portal.gdc.cancer.gov), which are in the svs format. These tissue slides are formalin-fixed, paraffin-embedded slides with which the cell morphology is well-preserved and thus appropriate for image feature recognition. They are captured at 20× or 40× magnification by the Aperio medical scanner. To extract imaging features, we conduct the following three-step preprocessing. The overall flowchart is provided in Figure 1.
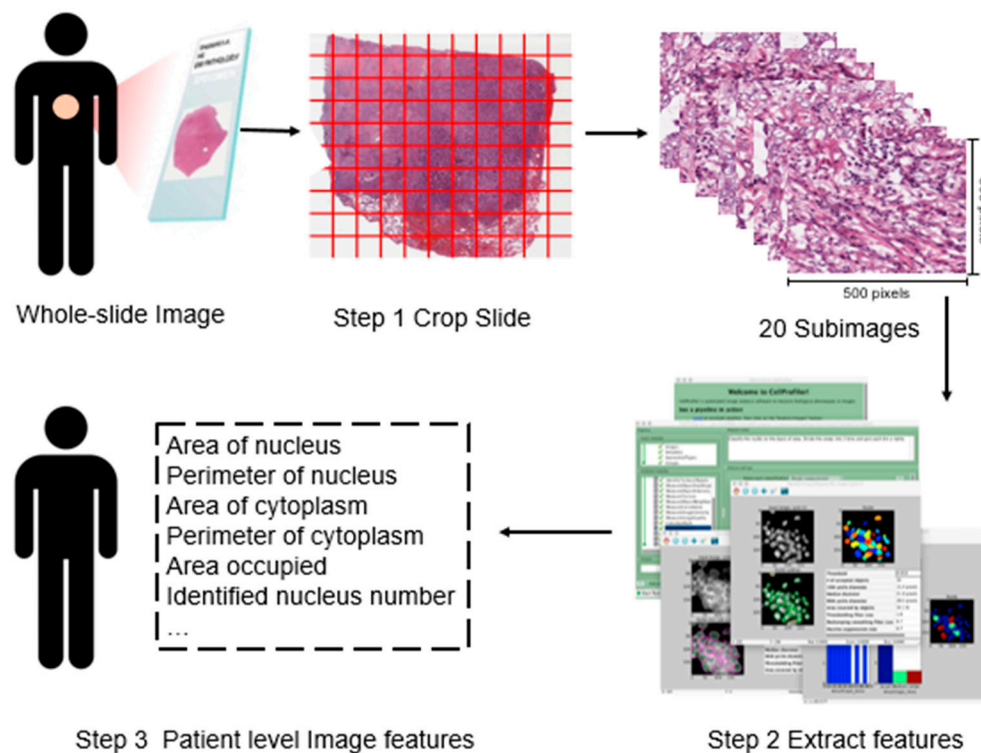

Figure 1. Flowchart of extracting imaging features. Step 1: whole-slide histopathology images are cropped into small subimages of 500 × 500 pixels, and 20 subimages are then randomly selected. Step 2: Imaging features are extracted using CellProfiler [19] for each subimage. Step 3: For each patient, features are averaged.

In the first step, we process the downloaded svs files using the Openslide Python library [20]. Specifically, to make data in an appropriate format for morphological feature extraction, each svs image is cropped into small subimages of 500 × 500 pixels, and each subimage is then saved in the tiff format. Among them, subimages that contain mostly (>50%) background white space are filtered out.

From the remaining subimages, twenty representative ones are randomly selected for each svs image to avoid potential subjective bias and reduce computational cost [21].

In the second step, we use CellProfiler [19] to extract features from each subimage. CellProfiler is an open source cell image analysis platform developed by the Broad Institute. Specifically, image colors are separated based on hematoxylin staining and eosin staining, and converted to grayscale to extract regional features, such as cell and tissue texture and granularity. These are "classic" image processing features, which have been examined in a large number of imaging studies. It is noted that they are not expected to be specific histopathological features, and cannot be recognized by pathologist. More specifically, texture describes a set of metrics calculated in CellProfiler to quantify the perceived texture of histopathological images, and includes information on the spatial arrangement of color or intensities in images. Granularity describes the size of how well the structure element fits in images. Then, cell nuclei are identified and segmented to collect the cell-level features, such as cell size, shape, distribution, and texture of nuclei. Other cell features such as regional occupation, area fraction, and neighboring architecture are also captured. The above process generates 832 features for each subimage. A further screening is conducted to exclude irrelevant features such as file size and execution information. Finally, for each subimage, a total of 772 features are available for analysis.

In the third step, for each slide, the average feature values over twenty representative subimages are computed. When a subject has multiple slides, the average values over multiple slides are further computed for downstream analysis.

The above preprocessing is applied to both LUAD and LIHC data. There are missing values in imaging feature data. Subjects with more than 25% missing imaging features are excluded. The remaining missing values are imputed using sample medians.

Remark 1. *We adopt CellProfiler to extract imaging features, which is a popular choice in recent literature. The feature names are automatically provided by CellProfiler. The extracted features represent objective attributes of histopathological images, including the area and perimeter of nucleus and cytoplasm, mean and standard deviation of these measures, and other general image attributes. For each patient, multiple slides and subimages may be available. To extract as much information as possible, we consider multiple slides and subimages simultaneously, and the average values are used to summarize information. A closer examination suggests that they are not explicitly associated with specific histopathological findings, such as cellularity, atypia, anaplasia, nuclear pleomorphism, and some others. However, these attributes have been shown to be associated with pathological stages [7,10] and prognosis [21], and also been used in multiple existing cancer modeling studies [13,15]. For example, Yu et al. [22] show that some of these features are associated with specific subtypes of lung cancer. In this study, our goal is to rigorously quantify independent information in omics profiles and histopathological images for cancer prognosis, as opposed to focusing on specific interpretation of imaging features. As such, these features are sensible for our analysis, although they may not have simple, explicit biological interpretations.*

2.3. Available Data

After data matching, we obtain records on 316 and 358 subjects for LUAD and LIHC, respectively, with 5000 gene expression measurements, 772 imaging features, and survival time. For LUAD, there are 103 deaths during follow-up, with survival times ranging from 0 to 214.77 months (median 6.03 months). For LIHC, there are 121 deaths during follow-up, with survival times ranging from 0 to 120.73 months (median 19.25 months). Summary information of the analyzed subjects is provided in Table 1.

Table 1. Sample characteristics.

Characteristic	LUAD	LIHC
Sample size	316	358
Age at diagnosis: median (range)	66 (39–88)	61 (16–90)
Follow-up: median (range)	6.03 (0–214.77)	19.25 (0–120.73)
Vital status: *n* (%)		
Alive	213 (67.4%)	233 (65.0%)
Deceased	103 (32.6%)	125 (35.0%)
Sex: *n* (%)		
Male	144 (45.6%)	242 (67.6%)
Female	172 (54.4%)	116 (32.4%)
Cancer stage: *n* (%)		
I	180 (57.0%)	166 (46.4%)
II	69 (18.7%)	82 (22.9%)
III	41 (13%)	83 (23.2%)
IV	21 (6.6%)	5 (1.4%)
NA	5 (1.6%)	22 (6.1%)

3. Methods

For describing cancer survival, we use the accelerated failure time (AFT) model. This model has been extensively adopted in cancer studies with high-dimensional variables because of its lucid interpretations and low computational cost [23]. For each gene expression, we consider the prognosis model with its effect along with all imaging features. A statistically rigorous test is then conducted on the gene expression's effect, which can suggest whether this particular gene has independent information for prognosis conditional on the imaging features. Then a parallel analysis is conducted, reversing the roles of gene expressions and imaging features. With a special emphasis on omics and imaging features, clinical factors are not included in the prognosis models.

Consider n independent subjects. For the ith subject, denote $x_i = (x_{i1}, x_{x2}, \ldots, x_{ip})$ and $z_i = (z_{i1}, z_{i2}, \ldots, z_{iq})$ as the p- and q-dimensional vectors of gene expressions and imaging features, T_i and C_i as the logarithm of the event and censoring times. With right censoring, we observe $(y_i = \min(T_i, C_i), \delta_i = \text{I}(T_i \leq C_i), x_i, z_i)$.

First consider the analysis with one gene expression and all imaging features. The analysis with one imaging feature and all gene expressions can be conducted in a similar manner and will not be reiterated. For the jth gene expression, consider the following AFT model:

$$T_i = \alpha + x_{ij}\beta_j + z_i\theta_j + \varepsilon_i \tag{1}$$

where α is the intercept, β_j and $\theta_j = (\theta_{j1}, \ldots, \theta_{jq})'$ are the unknown coefficients for the jth gene expression and all imaging features, and ε_i is the random error. A statistical test

$$\text{H}_0: \ \beta_j = 0 \ \text{vs.} \ \text{H}_1: \ \beta_j \neq 0$$

can reveal whether x_{ij} is independently associated with T_i given z_i. Here, loosely speaking, a smaller p value can indicate a stronger association/more prognostic power. The analysis is challenging with the high dimensionality of imaging features, which makes the "standard" estimation and inference techniques not applicable. To tackle this problem, we consider a high-dimensional inference approach [24] recently developed under a related but simpler context. Specifically, for estimation, the

weighted least squares approach is adopted. Assume that data have been sorted according to y_i's from the smallest to the largest. Then we have the Kaplan–Meier weights defined as

$$w_1 = \frac{\delta_1}{n}, \ w_i = \frac{y_i}{n-i+1} \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)^{\delta_j}, i = 2, \ldots, n,$$

where a further normalization is conducted to make $\sum_{i=1}^n w_i = n$. Let $W = \text{diag}\{w_1, w_2, w_3, \ldots, w_n\}$, $x_{\cdot j}$, $z_{\cdot k}$, and y denote the vectors composed of x_{ij}'s, z_{ik}'s, and y_i's, which are weighted normalized such that $\mathbf{1}'W x_{\cdot j} = 0$, $\mathbf{1}'W z_{\cdot k} = 0$, and $\mathbf{1}'W y = 0$, and $Z = (z_{\cdot 1}, \ldots, z_{\cdot q})$. With the high data dimensionality, regularized estimation is needed. In addition, not all imaging features are expected to be associated with survival, posing a variable selection problem. Consider the weighted penalized estimation:

$$\begin{cases} \left(\widetilde{\beta}_j^*, \widetilde{\theta}_j \right) = \text{argmin}_{\beta_j, \theta \in R^{1+q}} \left\{ \frac{1}{2n} \| W^{\frac{1}{2}} \left(y - x_{\cdot j} \beta_j - Z\theta_j \right) \|_2^2 + \lambda_0 \sum_{k=1}^q |\theta_{jk}| \right\} \\ \widetilde{b}_j = \text{argmin}_{b_j \in R^q} \left\{ \frac{1}{2n} \| W^{\frac{1}{2}} \left(x_{\cdot j} - Z b_j \right) \|_2^2 + \lambda_1 \sum_{k=1}^q |b_{jk}| \right\} \end{cases}, \quad (2)$$

where $\|\cdot\|_2$ denotes the L$_2$ norm, $b_j = (b_{j1}, \ldots, b_{jq})'$ is the unknown coefficient vector, and $\lambda_0, \lambda_1 > 0$ are data-dependent tuning parameters for Lasso penalties.

With $\widetilde{\theta}_j$ and $\widetilde{b}_j$, the final estimate $\widetilde{\beta}_j$ of β_j is defined as

$$\widetilde{\beta}_j = \left(\widetilde{x}_{\cdot j}' W x_{\cdot j} \right)^{-1} \widetilde{x}_{\cdot j}' W \left(y - Z \widetilde{\theta}_j \right), \quad (3)$$

where $\widetilde{x}_{\cdot j} = x_{\cdot j} - Z \widetilde{b}_j$. It has been shown in [24] that $\sqrt{n} \left(\widetilde{\beta}_j - \beta_j \right)$ is asymptotically normal with variance defined as

$$\widetilde{\Sigma} \left(\widetilde{\beta}_j \right) = \frac{1}{n} \left(\widetilde{x}_{\cdot j}' W x_{\cdot j} \right)^{-1} \widetilde{\Sigma_1} \left(x_{\cdot j}' W \widetilde{x}_{\cdot j} \right)^{-1},$$

where $\widetilde{\Sigma_1}$ is the sample variance based on $\widetilde{x}_{ij}(y_i - x_{ij}\widetilde{\beta}_j + z_i\widetilde{\theta}_j)$. With this asymptotic normal distribution, the test statistic for H$_0$: $\beta_j = 0$ can be defined as

$$\widetilde{\beta}_j / \sqrt{\widetilde{\Sigma} \left(\widetilde{\beta}_j \right) / n},$$

which follows Student's T distribution. The unadjusted *p* value can then be obtained. When all gene expressions are considered together, to accommodate multiple comparisons, *p* values are adjusted using the voxel-level false discovery rate approach [25].

　　An advantage of the above analysis is that it can be realized via simple coding. The most challenging step is the estimation in (2), which can be realized using the R function *glmnet*. The tuning parameters λ_0 and λ_1 are selected using the EBIC approach [26]. To facilitate data analysis in and beyond this study, we have developed R code implementing the proposed approach. To illustrate its usage, we have also provided an example R file with the LUAD data. The code and data are publicly available at http://www.github.com/shuanggema/TestLDHD as well as in Supplementary Materials.

Remark 2. *The effectiveness of the AFT model for cancer prognosis modeling has been well tested. Penalization has been shown effective for screening out irrelevant variables and accommodating high dimensionality. As shown in [24], the estimation (2) can effectively "single out" the effect of the one gene expression. It is noted that, as the gene expression effect is of particular interest, its coefficient is not subject to penalization in estimation. A "byproduct" of penalized estimation is that imaging features associated with prognosis, conditional on the one gene expression effect, are identified, which may assist understanding the associations between imaging features and prognosis as well as the associations between imaging features and gene expressions. The statistical distribution result for the test statistic has been rigorously proved in [24]. The T distribution makes inference very easy.*

4. Results

4.1. Identification of Gene Expressions with Independent Prognostic Power Conditional on Imaging Features

We first apply the analysis approach described above to identify individual gene expressions that have independent prognostic power conditional on the high dimensional histopathological imaging features. With significance level 0.05 as cutoff, 85 genes are identified as significantly associated with prognosis for LUAD. Detailed estimation results are provided in the Table S1. A quick literature search of PubMed suggests that many of the findings have sound biological basis. For top ten representative genes, we provide brief information and references on their associations with lung cancer prognosis in Table 2. For LIHC, we identify 386 genes as significantly associated with prognosis conditional on imaging features. Again, it is found that many of those genes have established evidences of being associated with liver cancer prognosis. Brief information on the representative genes is provided in Table 3. The solid evidences from existing studies provide support to the validity of analysis. To more comprehensively comprehend the identified genes, we examine the genes' functional and biological connections by conducting Gene Ontology (GO) and KEGG pathway enrichment analysis. The processes with the smallest *p* values are presented in Figure 2. It is observed that both LUAD and LIHC have enriched processes associated with cell adhesion. This may result from high-dimensional imaging features capturing more information on cell interaction. The identified genes are also enriched in traditional cancer hallmarks, such as the positive regulation of mitotic cell cycle, proteoglycans in cancer, and extracellular matrix regulation. It is interesting that some immune response-related pathways, such as MHC assembly, are also identified, considering that there are promising developments in immunotherapy for cancer treatment recently.
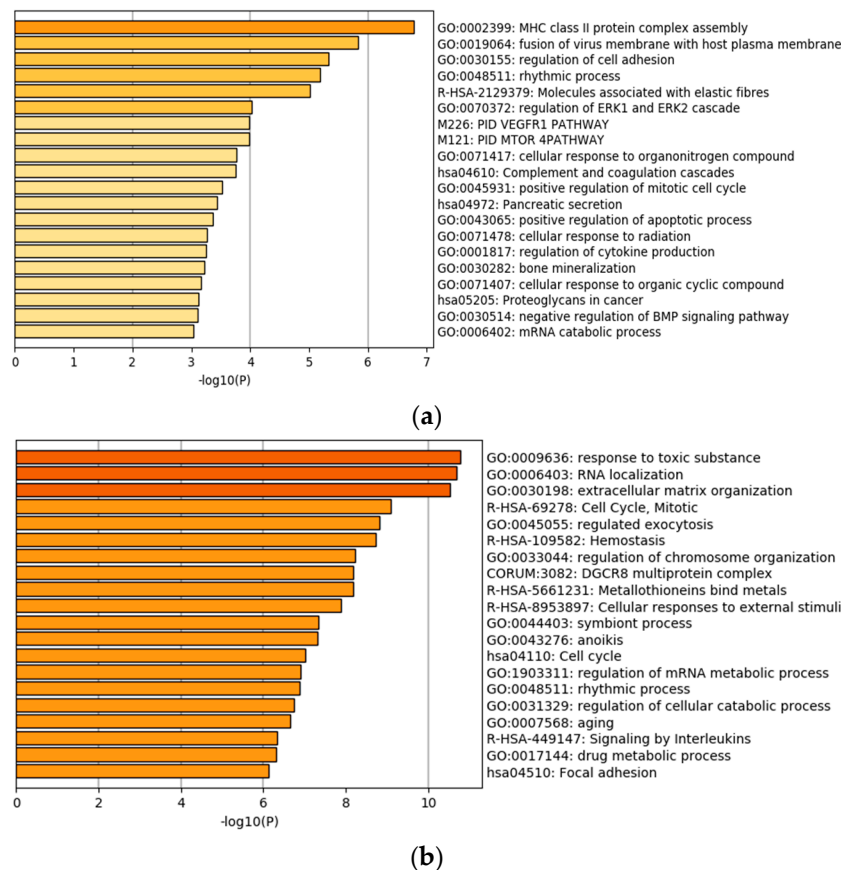

(**a**)

(**b**)

Figure 2. Gene ontology (GO) and pathway enrichment analysis of the identified genes. (**a**) lung adenocarcinoma (LUAD), (**b**) liver hepatocellular carcinoma (LIHC).

It is also noted that the proposed analysis takes an angle different from the literature. As such, there are also new findings. For example, the top ranked genes (with the smallest *p* values) also include DCAF6 and LITAF for LUAD and IARS and LRPPRC for LIHC. Gene DCAF6 is a transcriptional cofactor that enhances androgen receptor (AR)-mediated transcriptional activity. Chen et al. [27] have found that the expression of DCAF6 is upregulated in prostate cancer patient and may be a candidate tumor promoter for prostate cancer, indicating its important role in cancer etiology. For gene LITAF, Zhou et al. [28] have established the regulatory axis of AMPK–LITAF–TNFSF15, where AMPK activation upregulates the transcription of LITAF and consequently upregulates the expression of TNFSF15. TNFSF15 inhibits the growth of prostate cancer cells and bovine aortic endothelial cells in vitro, supporting that LITAF may function as a tumor suppressor. Gene IARS is the coding gene of isoleucyl-tRNA synthetase (ARSs). ARSs has been shown to catalyze the amino acylation of tRNAs by their cognate amino acid, linking amino acids with the correct nucleotide triplets and ensuring the correct transformation of the genetic code to the protein level. Kopajtich et al. [29] have reported that biallelic IARS mutations can cause infantile hepatopathy, suggesting its potential association with liver function. LRPPRC is another interesting gene, which regulates the expression of all mitochondrial DNA-encoded mRNAs, and thus has important contributions to mitochondrial function. Tian et al. [30] have examined the expression of LRPPRC in six types of cancer and observed that LRPPRC plays an important role in tumorigenesis through the resistance to apoptosis and high invasive activity. Other newly identified genes, such as FLS353 [31], IPO7 [32], PDS5A [33], and MPP-2 [34], have also been demonstrated to be related to tumorigenesis. This brief literature search suggests that the implications of these new findings in lung and liver cancer prognosis have not been well established, however, they have been observed to have important contributions to cancer etiology. The new findings are also not surprising. It is conceptually sensible that the strongest "signals" are perhaps reflected in both omics profiles and histopathological images. The proposed analysis seeks for additional "signals" in gene expressions that are not reflected in images. As such, the findings may complement those in the literature.

Table 2. Analysis of LUAD data: representative identified genes.

Gene	Evidence	PMID
HYAL2	Real-time PCR studies showed that HYAL2 genes were down regulated in non-small cell lung cancer [35].	19140316
MAPK1IP1L	MAPK1IP1L gene was found to be related with acquired resistance to MET inhibitors in lung cancer cells [36].	28396363
HLA-DRA	Lack of surface class II expression was found to be associated with a specific defect in HLA-DRA induction in non-small cell lung carcinoma cells [37].	8786310
HNRNPK	Higher levels of hnRNP mRNAs were found in SCLC as compared to NSCLC. hnRNP K protein localization varied with cellular confluence [38].	12871776
GPNMB	Osteoactivin (GPNMB) ectodomain protein was shown to promote growth and invasive behavior of human lung cancer cells [39].	26883195
BMP2	Positive correlation was found between gene expressions of two angiogenic factors, VEGF and BMP-2, in lung cancer patients [40].	19324447
COMMD6	COMMD9 was demonstrated to promote TFDP1/E2F1 transcriptional activity via interaction with TFDP1 in non-small cell lung cancer [41].	27871936
HLA-DRB1	Lung cancer patients in Japan showed an increased frequency of HLA-DRB1*0901 and a decreased frequency of HLA-DRB1*1302 and DRB1*14-related alleles when compared to the other subjects [42].	9808426
LARP1	LARP1 post-transcriptionally regulates mTOR and contributes to cancer progression [43].	25531318
ZAK	ZAK inhibits human lung cancer cell growth via ERK and JNK activation in an AP-1-dependent manner [44].	20331627

* The star is a sign indicating the location of the mutated allel.

Table 3. Analysis of LIHC data: representative identified genes.

Gene	Evidence	PMID
LAPTM4B	LAPTM4B is a potential proto-oncogene, whose overexpression is involved in carcinogenesis and progression of HCC [45].	12902989
CAPZA1	CAPZA1 expression levels were negatively correlated with the biological characteristics of primary HCC and patient prognosis [46].	28093067
PLOD2	PLOD2 expression was identified as a significant, independent factor of poor prognosis for HCC patients [47].	22098155
STIP1	STIP1 was upregulated in HCC and associated with poor clinical prognosis [48].	28887036
IGF1	Inhibition of IGF-1R tyrosine kinase (IGF-1R-TK) by NVP-AEW541 induces growth inhibition, apoptosis and cell cycle arrest in human HCC cell lines without accompanying cytotoxicity [49].	16530734
HTATIP2	HepG2 cells that expressed transgenic HTATIP2 formed more invasive tumors in mice following administration of sorafenib. Sorafenib therapy prolonged recurrence-free survival in patients who expressed lower levels of HTATIP2 compared with higher levels [50].	22922424
GNAI3	GNAI3 inhibits tumor cell migration and invasion and is post-transcriptionally regulated by miR-222 in hepatocellular carcinoma [51].	25444921
XPO1	Exportin-1 (XPO1, CRM1) mediates the nuclear export of several key growth regulatory and tumor suppressor proteins [52].	25030088
PLVAP	PLVAP was identified as a gene specifically expressed in vascular endothelial cells of HCC but not in non-tumorous liver tissues [53].	25376302
EPAS1	HIF-2alpha/EPAS1 expression may play an important role in tumor progression and prognosis of HCC [54].	17589895

Analysis is further conducted to better comprehend the additional prognostic information in gene expressions. Specifically, for each identified gene expression, two AFT models are considered. The first model, referred to as A1, contains the single gene expression as well as selected imaging features; in contrast, the second model, referred to as A2, contains only the selected imaging features. Comparing the two models can straightforwardly describe the contribution of the gene expression to prognosis. To facilitate the comparison, subjects are classified into low and high risk groups with equal sizes based on the survival times predicted using A1 and A2, and the log rank test is conducted to compare survival of the two groups. p values from the log rank tests are provided in the Table S1. For LUAD, it is observed that 56 out of the 85 A1 type models can effectively separate the high and low risk groups (with the adjusted log rank test p values < 0.05); In contrast, 47 of the A2 type models have significant adjusted log rank p values. In addition, 51 A1 type models have p values smaller than their A2 counterparts. In the analysis of LIHC data, the numbers of significant adjusted p values for the A1 and A2 type of models are 237 and 66, respectively. In addition, 286 genes have p values smaller under the A1 type models than their A2 counterparts. These results suggest that gene expressions can significantly improve prognosis modeling beyond histopathological imaging features. For a more intuitive presentation, we present the Kaplan–Meier curves for two representative genes in Figure 3. Specifically, we show the survival functions of the low and high risk groups under the A1 and A2 models. The A1 models have a much clearer and more significant separation. Such results suggest that, for clinical practice, it may be necessary to integrate gene expression data beyond histopathological imaging data.
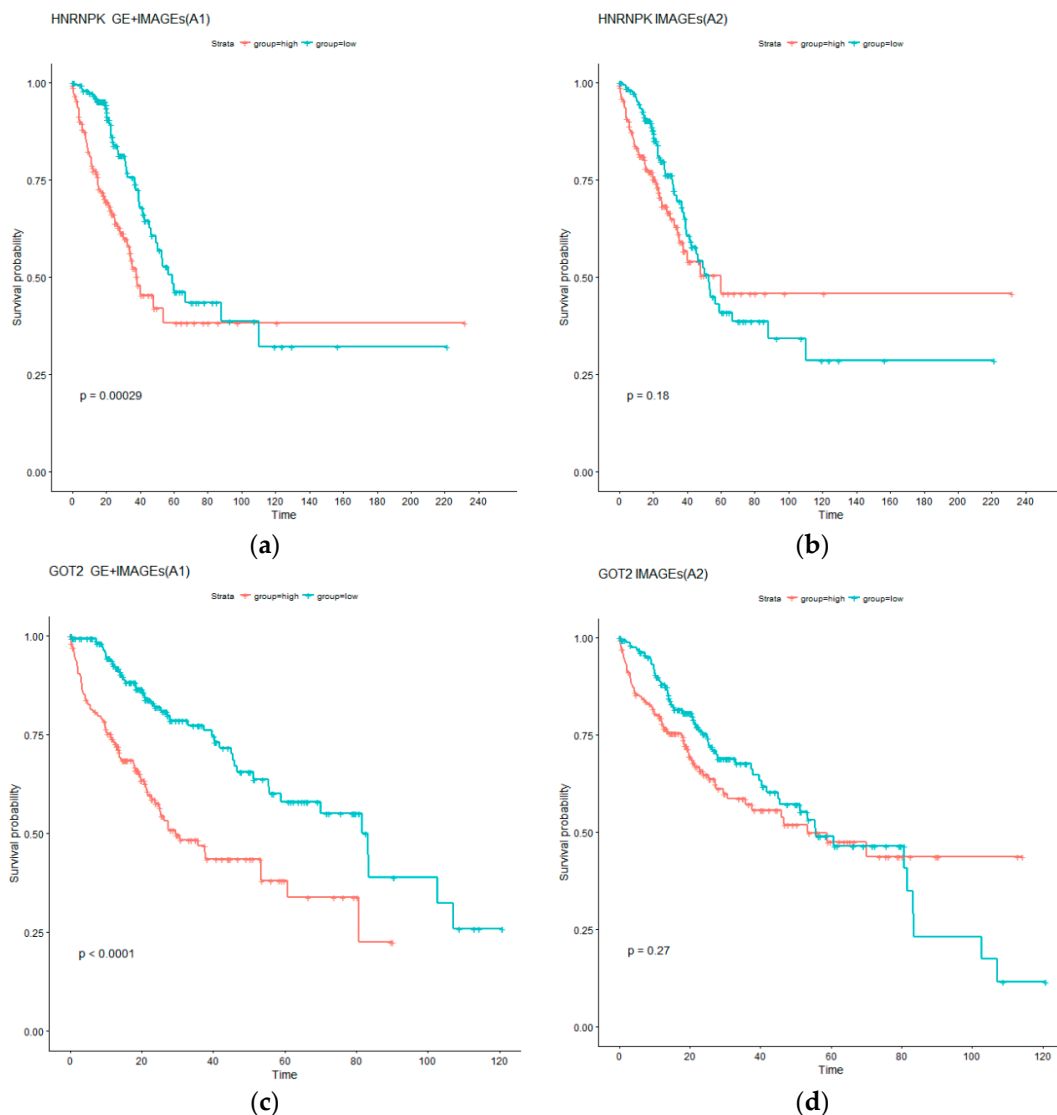
Figure 3. Kaplan–Meier (KM) curves for low (blue) and high (red) risk groups under models A1 and A2. (**a**,**b**) for LUAD: Gene HNRNPK as well as selected imaging features (A1), and only selected imaging features (A2). (**c**,**d**) for LIHC: Gene GOT2 as well as selected imaging features (A1), and only selected imaging features (A2). *p* values are computed from log rank tests.

4.2. Identification of Imaging Features with Independent Prognostic Power Conditional on Gene Expressions

Analysis parallel to the above is conducted, reversing the roles of gene expressions and imaging features. With significance level 0.05, the identified imaging features along with their *p* values are shown in Table 4. Detailed estimation results are provided in the Table S1. In the LUAD data analysis, 11 imaging features are found as significantly associated with prognosis conditional on the selected gene expressions. Among them, six are texture related, which is consistent with the literature [21]. In the analysis of LIHC data, nine imaging features are found as significantly associated with prognosis conditional on gene expressions, among which eight belong to the morphological category of nuclei texture. There is one notable difference between gene expression and histopathological imaging measurements. With extensive functional studies accumulated over years, the biological functions of many genes are known; In contrast, the biological interpretations of imaging features mostly remain unclear. As such, interpretations as in Table 2; Table 3 cannot be pursued.

Table 4. Identified histopathological features.

LUAD		LIHC	
Feature Name	**Adjusted p Value**	**Feature Name**	**Adjusted p Value**
Mean_Identifyeosinprimarycytoplasm_Texture_Correlation_maskosingray_3_00	1.16×10^{-4}	StDev_Identifyhemasub2_Texture_DifferenceEntropy_ImageAfterMath_3_02	9.46×10^{-6}
Median_Identifyeosinprimarycytoplasm_Texture_Correlation_maskosingray_3_00	1.59×10^{-4}	StDev_Identifyhemasub2_Texture_SumEntropy_ImageAfterMath_3_00	9.46×10^{-6}
StDev_Identifyeosinprimarycytoplasm_Texture_Correlation_maskosingray_3_00	1.59×10^{-4}	StDev_Identifyhemasub2_Texture_SumEntropy_ImageAfterMath_3_02	1.85×10^{-5}
StDev_Identifyhemasub2_AreaShape_Orientation	1.59×10^{-4}	StDev_Identifyhemasub2_Texture_DifferenceEntropy_ImageAfterMath_3_00	2.92×10^{-5}
StDev_Identifyhemasub2_AreaShape_Zernike_6_6	1.05×10^{-4}	StDev_Identifyhemasub2_Texture_SumEntropy_ImageAfterMath_3_01	3.64×10^{-5}
StDev_Identifyhemasub2_AreaShape_Zernike_9_1	1.59×10^{-4}	StDev_Identifyhemasub2_Texture_DifferenceEntropy_ImageAfterMath_3_01	4.08×10^{-5}
StDev_Identifyhemasub2_AreaShape_Zernike_9_9	1.59×10^{-4}	StDev_Identifyhemasub2_Texture_DifferenceEntropy_ImageAfterMath_3_03	4.82×10^{-5}
StDev_Identifyhemasub2_Texture_DifferenceEntropy_ImageAfterMath_3_03	1.64×10^{-4}	StDev_Identifyhemasub2_Texture_SumEntropy_ImageAfterMath_3_03	9.24×10^{-5}
StDev_Identifyhemasub2_Texture_SumEntropy_ImageAfterMath_3_01	1.64×10^{-4}	Granularity_2_ImageAfterMath	1.07×10^{-4}
Texture_Correlation_maskosingray_3_00	1.59×10^{-4}		
Granularity_15_ImageAfterMath	7.66×10^{-4}		

As in the above section, additional analysis is conducted. Specifically, two types of AFT models are considered. The first type, referred to as B1, is based on an individual imaging feature and selected gene expressions; whereas the second type, referred to as B2, is only based on selected gene expressions. Detailed results are provided in the Table S1. For LUAD, it is observed that ten out of the 11 B1 type models can effectively separate the high and low risk groups, whereas 11 of the B2 type models have significant adjusted log rank p values. For the identified imaging feature Granularity_15_ImageAfterMath, the B1 model is not significant, however, the corresponding B2 model is significant. This can be explained by the larger degrees of freedom of the B1 model and possible correlation between the two types of measurements. It is also found that six imaging features have p values smaller under the B1 type models than their B2 counterparts. In the LIHC data analysis, the overall findings are similar. Specifically, all tests under both B1 and B2 types of models have adjusted p values smaller than 0.05. All tests under B1 have p values larger than those under B2. Again, this can be possibly explained by the larger numbers of parameters under B1 and correlations. The KM curves for two representative imaging features are examined in Figure 4. The differences between the left and right panels are much less distinct compared to Figure 3. Overall, the analysis suggests that, conditional on the presence of gene expression measurements, histopathological imaging features have limited independent prognostic power.
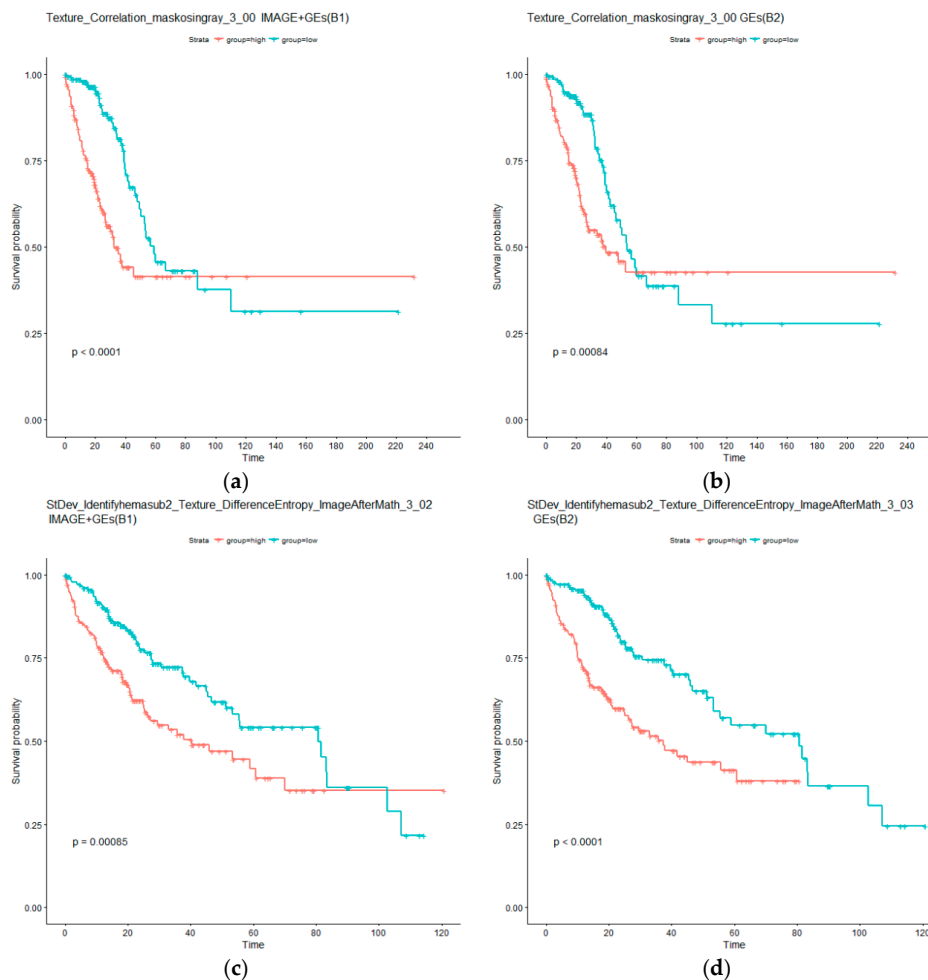

Figure 4. KM curves for low (blue) and high (red) risk groups under models B1 and B2. (**a**,**b**) for LUAD: Imaging feature Texture_Correlation_maskosingray_3_00 as well as selected gene expressions (B1), and only selected gene expressions (B2). (**c**,**d**) for LIHC: Imaging feature StDev_Identifyhemasub2_Texture_DifferenceEntropy_ImageAfterMath_3_00 as well as selected gene expressions (B1), and only selected gene expressions (B2). p values are computed from log rank tests.

5. Discussion

In the literature, the separate and integrated analyses of omics and histopathological imaging data have been conducted. This study has taken a different perspective and tried to answer the fundamental question of whether a type of measurement has independent prognostic power conditional on the collective effect of the other type. A rigorous statistical testing approach, developed under simpler settings, has been adopted and extended. In the analysis of TCGA data on two cancer types, it has been found that, conditional on imaging features, individual gene expressions may still have significant prognostic power; however, conditional on gene expressions, individual imaging features may have limited prognostic power. As such, at least for the analyzed datasets, gene expressions and histopathological imaging features have an irreversible independent relationship in modeling prognosis. Such findings, to the best of our knowledge, are the first in the literature. They may have important implications for cancer practice. Specifically, in cancer clinical practice, gene expression profiling is becoming routine. Our analysis suggests that, for modeling prognosis, when gene expression data are available, clinicians may not need to order histopathological imaging. However, this is not true the other way around. For a more accurate prognosis modeling, with imaging data, clinicians may still want to order gene expression profiling, and use integrated analysis techniques (developed in published and future literature) to integrate gene expression and imaging data. The biological implications of the findings, pertinent to the associations between gene expressions and imaging features, are worth further investigation.

This study can be extended in multiple directions. Specifically, other types of omics measurements, such as DNA mutations, miRNA expressions, methylation, and copy number variations, can be analyzed similarly. It is noted that the findings are not necessarily the same as in this article. In our analysis, we have considered high dimensional imaging features, which may include more information but do not have simple biological interpretations. It is possible to conduct a similar analysis using low dimensional imaging features such as vascular invasion and lymphocyte cells—this is postponed to future research. It is noted that in the literature [55], techniques have already been developed to identify cancer regions, extract such low dimensional imaging features, and use them in analysis. The analysis has been focused on gene expression and imaging data. We acknowledge the importance of clinical, environmental, and other factors in cancer prognosis. These factors are not included in analysis with the following concerns. First, the most important objective of this study is to evaluate the overlapping/independent information in omics and histopathological imaging data. As such, we focus on these two types of measurements. Second, to the best of our knowledge, techniques for making inference with the "one gene expression + low dimensional clinical factors + high dimensional imaging features" model are not available. This analysis can be more complex than that in this article with the significant differences between clinical factors and imaging features. Table 1 has been provided so that researchers can comprehend properties of the TCGA cohorts. Extending the proposed analysis and accommodating clinical factors will be postponed to future research. It is also of interest to conduct similar analysis for other types of cancer outcomes. For a continuous cancer marker, the described analysis technique can be directly applied. New developments will be needed for other types of outcomes/phenotypes. It is also noted that the detailed results can be data and model dependent. It is impossible to conduct analysis with all cancer types and models. However, the "spirit" of the proposed analysis will be broadly applicable. Although with certain limitations, being the first of its kind, this study may still provide important insights into cancer modeling and characteristics of cancer as reflected in omics profiles and histopathological images.

6. Conclusions

Omics and histopathological imaging data have been co-collected in cancer practice and analyzed in recent integrated analysis. In this study, we have presented a pipeline for analyzing their independent prognostic power conditional on the other type of data. The adopted statistical inference technique has a solid ground and can be broadly applied. The "asymmetric" finding is interesting, has

not been observed in the literature, and has sound interpretations. It is reasonable to expect that the proposed analysis and its downstream integrated analysis will gain popularity fast and have a deep impact on cancer practice.

References

1. Mallett, S.; Royston, P.; Dutton, S.; Waters, R.; Altman, D.G. Reporting methods in studies developing prognostic models in cancer: A review. *BMC Med.* **2010**, *8*, 20. [CrossRef] [PubMed]

2. Kourou, K.; Exarchos, T.P.; Exarchos, K.P.; Karamouzis, M.V.; Fotiadis, D.I. Machine learning applications in cancer prognosis and prediction. *Comput. Struct. Biotechnol. J.* **2015**, *13*, 8–17. [CrossRef] [PubMed]

3. Rath, M.G.; Uhlmann, L.; Fiedler, M.; Heil, J.; Golatta, M.; Dinkic, C.; Hennigs, A.; Schott, S.; Ernst, V.; Koch, T.; et al. Oncotype DX((R)) in breast cancer patients: Clinical experience, outcome and follow-up-a case-control study. *Arch. Gynecol. Obstet.* **2018**, *297*, 443–447. [CrossRef] [PubMed]

4. Yanaihara, N.; Caplen, N.; Bowman, E.; Seike, M.; Kumamoto, K.; Yi, M.; Stephens, R.M.; Okamoto, A.; Yokota, J.; Tanaka, T.; et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* **2006**, *9*, 189–198. [CrossRef] [PubMed]

5. Muller, H.M.; Widschwendter, A.; Fiegl, H.; Ivarsson, L.; Goebel, G.; Perkmann, E.; Marth, C.; Widschwendter, M. DNA methylation in serum of breast cancer patients: An independent prognostic marker. *Cancer Res.* **2003**, *63*, 7641–7645. [PubMed]

6. Harpole, D.H., Jr.; Herndon, J.E., 2nd; Wolfe, W.G.; Iglehart, J.D.; Marks, J.R. A prognostic model of recurrence and death in stage I non-small cell lung cancer utilizing presentation, histopathology, and oncoprotein expression. *Cancer Res.* **1995**, *55*, 51–56. [PubMed]

7. Romo-Bucheli, D.; Janowczyk, A.; Gilmore, H.; Romero, E.; Madabhushi, A. Automated tubulenuclei quantification and correlation with Oncotype DX risk categories in ER+ breast cancer whole slide Images. *Sci. Rep.* **2016**, *6*, 32706. [CrossRef] [PubMed]

8. Sertel, O.; Kong, J.; Shimada, H.; Catalyurek, U.V.; Saltz, J.H.; Gurcan, M.N. Computer-aided prognosis of neuroblastoma on whole-slide images: Classification of stromal development. *Pattern Recognit.* **2009**, *42*, 1093–1103. [CrossRef]

9. Sertel, O.; Kong, J.; Catalyurek, U.V.; Lozanski, G.; Saltz, J.H.; Gurcan, M.N. Histopathological Image Analysis Using Model-Based Intermediate Representations and Color Texture: Follicular Lymphoma Grading. *J. Signal Process. Syst.* **2009**, *55*, 169–183. [CrossRef]

10. Yu, K.H.; Berry, G.J.; Rubin, D.L.; Re, C.; Altman, R.B.; Snyder, M. Association of omics features with histopathology patterns in lung adenocarcinoma. *Cell Syst.* **2017**, *5*, 620–627.e3. [CrossRef]

11. Sabo, E.; Beck, A.H.; Montgomery, E.A.; Bhattacharya, B.; Meitner, P.; Wang, J.Y.; Resnick, M.B. Computerized morphometry as an aid in determining the grade of dysplasia and progression to adenocarcinoma in Barrett's esophagus. *Lab. Investig.* **2006**, *86*, 1261–1271. [CrossRef]

12. Cooper, L.A.; Kong, J.; Gutman, D.A.; Dunn, W.D.; Nalisnik, M.; Brat, D.J. Novel genotype-phenotype associations in human cancers enabled by advanced molecular platforms and computational analysis of whole slide images. *Lab. Investig.* **2015**, *95*, 366–376. [CrossRef] [PubMed]

13. Sun, D.; Li, A.; Tang, B.; Wang, M. Integrating genomic data and pathological images to effectively predict breast cancer clinical outcome. *Comput. Methods Programs Biomed.* **2018**, *161*, 45–53. [CrossRef] [PubMed]

14. Mobadersany, P.; Yousefi, S.; Amgad, M.; Gutman, D.A.; Barnholtz-Sloan, J.S.; Vega, J.V.; Brat, D.J.; Cooper, L.A. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E2970–E2979. [CrossRef] [PubMed]

15. Zhu, X.L.; Yao, J.W.; Luo, X.; Xiao, G.H.; Xie, Y.; Gazdar, A.; Huang, J.Z. Lung cancer survival prediction from pathological images and genetic data—An integration study. In Proceedings of the 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), Prague, Czech Republic, 13–16 April 2016; pp. 1173–1176.

16. Hutter, C.; Zenklusen, J.C. The Cancer Genome Atlas: Creating Lasting Value beyond Its Data. *Cell* **2018**, *173*, 283–285. [CrossRef] [PubMed]

17. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B.E.; Sumer, S.O.; Aksoy, B.A.; Jacobsen, A.; Byrne, C.J.; Heuer, M.L.; Larsson, E.; et al. The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2012**, *2*, 401–404. [CrossRef] [PubMed]

18. Gao, J.; Aksoy, B.A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S.O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E.; et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **2013**, *6*, pl1. [CrossRef] [PubMed]

19. Soliman, K. CellProfiler: Novel Automated Image Segmentation Procedure for Super-Resolution Microscopy. *Biol. Proced. Online* **2015**, *17*, 11. [CrossRef]

20. Goode, A.; Gilbert, B.; Harkes, J.; Jukic, D.; Satyanarayanan, M. OpenSlide: A vendor-neutral software foundation for digital pathology. *J. Pathol. Inform.* **2013**, *4*, 27. [CrossRef] [PubMed]

21. Luo, X.; Zang, X.; Yang, L.; Huang, J.; Liang, F.; Rodriguez-Canales, J.; Wistuba, I.I.; Gazdar, A.; Xie, Y.; Xiao, G. Comprehensive Computational Pathological Image Analysis Predicts Lung Cancer Prognosis. *J. Thorac. Oncol.* **2017**, *12*, 501–509. [CrossRef] [PubMed]

22. Yu, K.H.; Zhang, C.; Berry, G.J.; Altman, R.B.; Re, C.; Rubin, D.L.; Snyder, M. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat. Commun.* **2016**, *7*, 12474. [CrossRef] [PubMed]

23. Buckley, J.; James, I. Linear regression with censored data. *Biometrika* **1979**, *66*, 429–436. [CrossRef]

24. Chai, H.; Zhang, Q.Z.; Huang, J.; Ma, S.G. Inference for low-dimensional covariates in a high-dimensional accelerated failure time model. *Stat. Sin.* **2019**. [CrossRef]

25. Benjamini, Y.; Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **2001**, *29*, 1165–1188.

26. Chen, J.H.; Chen, Z.H. Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* **2008**, *95*, 759–771. [CrossRef]

27. Chen, H.H.; Fan, P.; Chang, S.W.; Tsao, Y.P.; Huang, H.P.; Chen, S.L. NRIP/DCAF6 stabilizes the androgen receptor protein by displacing DDB2 from the CUL4A-DDB1 E3 ligase complex in prostate cancer. *Oncotarget* **2017**, *8*, 21501–21515. [CrossRef] [PubMed]

28. Zhou, J.; Yang, Z.; Tsuji, T.; Gong, J.; Xie, J.; Chen, C.; Li, W.; Amar, S.; Luo, Z. LITAF and TNFSF15, two downstream targets of AMPK, exert inhibitory effects on tumor growth. *Oncogene* **2011**, *30*, 1892–1900. [CrossRef]

29. Kopajtich, R.; Murayama, K.; Janecke, A.R.; Haack, T.B.; Breuer, M.; Knisely, A.S.; Harting, I.; Ohashi, T.; Okazaki, Y.; Watanabe, D.; et al. Biallelic IARS Mutations Cause Growth Retardation with Prenatal Onset, Intellectual Disability, Muscular Hypotonia, and Infantile Hepatopathy. *Am. J. Hum. Genet.* **2016**, *99*, 414–422. [CrossRef]

30. Tian, T.; Ikeda, J.; Wang, Y.; Mamat, S.; Luo, W.; Aozasa, K.; Morii, E. Role of leucine-rich pentatricopeptide repeat motif-containing protein (LRPPRC) for anti-apoptosis and tumourigenesis in cancers. *Eur. J. Cancer* **2012**, *48*, 2462–2473. [CrossRef]

31. Van Beijnum, J.R.; Moerkerk, P.T.M.; Gerbers, A.J.; de Brune, A.P.; Arends, J.W.; Hoogenboom, H.R.; Hufton, S.E. Target validation for genomics using peptide-specific phage antibodies: A study of five gene products overexpressed in colorectal cancer. *Int. J. Cancer* **2002**, *101*, 118–127. [CrossRef]

32. Szczyrba, J.; Nolte, E.; Hart, M.; Doll, C.; Wach, S.; Taubert, H.; Keck, B.; Kremmer, E.; Stohr, R.; Hartmann, A.; et al. Identification of ZNF217, hnRNP-K, VEGF-A and IPO7 as targets for microRNAs that are downregulated in prostate carcinoma. *Int. J. Cancer* **2013**, *132*, 775–784. [CrossRef] [PubMed]

33. Losada, A. Cohesin in cancer: Chromosome segregation and beyond. *Nat. Rev. Cancer* **2014**, *14*, 389–393. [CrossRef] [PubMed]

34. Sivula, A.; Talvensaari-Mattila, A.; Lundin, J.; Joensuu, H.; Haglund, C.; Ristimaki, A.; Turpeenniemi-Hujanen, T. Association of cyclooxygenase-2 and matrix metalloproteinase-2 expression in human breast cancer. *Breast Cancer Res. Treat.* **2005**, *89*, 215–220. [CrossRef] [PubMed]

35. Anedchenko, E.A.; Dmitriev, A.A.; Krasnov, G.S.; Kondrat'eva, T.T.; Kopantsev, E.P.; Vinogradova, T.V.; Zinov'eva, M.V.; Zborovskaia, I.B.; Polotskii, B.E.; Sakharova, O.V.; et al. Down-regulation of RBSP3/CTDSPL, NPRL2/G21, RASSF1A, ITGA9, HYAL1 and HYAL2 genes in non-small cell lung cancer. *Mol. Biol. (Mosk)* **2008**, *42*, 965–976. [CrossRef] [PubMed]

36. Gimenez-Xavier, P.; Pros, E.; Bonastre, E.; Moran, S.; Aza, A.; Grana, O.; Gomez-Lopez, G.; Derdak, S.; Dabad, M.; Esteve-Codina, A.; et al. Genomic and Molecular Screenings Identify Different Mechanisms for Acquired Resistance to MET Inhibitors in Lung Cancer Cells. *Mol. Cancer Ther.* **2017**, *16*, 1366–1376. [CrossRef] [PubMed]

37. Lu, Y.; Boss, J.M.; Hu, S.X.; Xu, H.J.; Blanck, G. Apoptosis-independent retinoblastoma protein rescue of HLA class II messenger RNA IFN-gamma inducibility in non-small cell lung carcinoma cells. Lack of surface class II expression associated with a specific defect in HLA-DRA induction. *J. Immunol.* **1996**, *156*, 2495–2502. [PubMed]

38. Pino, I.; Pio, R.; Toledo, G.; Zabalegui, N.; Vicent, S.; Rey, N.; Lozano, M.D.; Torre, W.; Garcia-Foncillas, J.; Montuenga, L.M. Altered patterns of expression of members of the heterogeneous nuclear ribonucleoprotein (hnRNP) family in lung cancer. *Lung Cancer* **2003**, *41*, 131–143. [CrossRef]

39. Oyewumi, M.O.; Manickavasagam, D.; Novak, K.; Wehrung, D.; Paulic, N.; Moussa, F.M.; Sondag, G.R.; Safadi, F.F. Osteoactivin (GPNMB) ectodomain protein promotes growth and invasive behavior of human lung cancer cells. *Oncotarget* **2016**, *7*, 13932–13944. [CrossRef] [PubMed]

40. Bieniasz, M.; Oszajca, K.; Eusebio, M.; Kordiak, J.; Bartkowiak, J.; Szemraj, J. The positive correlation between gene expression of the two angiogenic factors: VEGF and BMP-2 in lung cancer patients. *Lung Cancer* **2009**, *66*, 319–326. [CrossRef]

41. Zhan, W.; Wang, W.; Han, T.; Xie, C.; Zhang, T.; Gan, M.; Wang, J.B. COMMD9 promotes TFDP1/E2F1 transcriptional activity via interaction with TFDP1 in non-small cell lung cancer. *Cell Signal.* **2017**, *30*, 59–66. [CrossRef] [PubMed]

42. Tokumoto, H. Analysis of HLA-DRB1-related alleles in Japanese patients with lung cancer—relationship to genetic susceptibility and resistance to lung cancer. *J. Cancer Res. Clin. Oncol.* **1998**, *124*, 511–516. [CrossRef] [PubMed]

43. Mura, M.; Hopkins, T.G.; Michael, T.; Abd-Latip, N.; Weir, J.; Aboagye, E.; Mauri, F.; Jameson, C.; Sturge, J.; Gabra, H.; et al. LARP1 post-transcriptionally regulates mTOR and contributes to cancer progression. *Oncogene* **2015**, *34*, 5025–5036. [CrossRef] [PubMed]

44. Yang, J.J.; Lee, Y.J.; Hung, H.H.; Tseng, W.P.; Tu, C.C.; Lee, H.; Wu, W.J. ZAK inhibits human lung cancer cell growth via ERK and JNK activation in an AP-1-dependent manner. *Cancer Sci.* **2010**, *101*, 1374–1381. [CrossRef] [PubMed]

45. Shao, G.Z.; Zhou, R.L.; Zhang, Q.Y.; Zhang, Y.; Liu, J.J.; Rui, J.A.; Wei, X.; Ye, D.X. Molecular cloning and characterization of LAPTM4B, a novel gene upregulated in hepatocellular carcinoma. *Oncogene* **2003**, *22*, 5060–5069. [CrossRef]

46. Huang, D.; Cao, L.; Zheng, S. CAPZA1 modulates EMT by regulating actin cytoskeleton remodelling in hepatocellular carcinoma. *J. Exp. Clin. Cancer Res.* **2017**, *36*, 13. [CrossRef] [PubMed]

47. Noda, T.; Yamamoto, H.; Takemasa, I.; Yamada, D.; Uemura, M.; Wada, H.; Kobayashi, S.; Marubashi, S.; Eguchi, H.; Tanemura, M.; et al. PLOD2 induced under hypoxia is a novel prognostic factor for hepatocellular carcinoma after curative resection. *Liver Int.* **2012**, *32*, 110–118. [CrossRef] [PubMed]

48. Chen, Z.; Xu, L.; Su, T.; Ke, Z.; Peng, Z.; Zhang, N.; Peng, S.; Zhang, Q.; Liu, G.; Wei, G.; et al. Autocrine STIP1 signaling promotes tumor growth and is associated with disease outcome in hepatocellular carcinoma. *Biochem. Biophys. Res. Commun.* **2017**, *493*, 365–372. [CrossRef] [PubMed]

49. Hopfner, M.; Huether, A.; Sutter, A.P.; Baradari, V.; Schuppan, D.; Scherubl, H. Blockade of IGF-1 receptor tyrosine kinase has antineoplastic effects in hepatocellular carcinoma cells. *Biochem. Pharmacol.* **2006**, *71*, 1435–1448. [CrossRef] [PubMed]

50. Zhang, W.; Sun, H.C.; Wang, W.Q.; Zhang, Q.B.; Zhuang, P.Y.; Xiong, Y.Q.; Zhu, X.D.; Xu, H.X.; Kong, L.Q.; Wu, W.Z.; et al. Sorafenib down-regulates expression of HTATIP2 to promote invasiveness and metastasis of orthotopic hepatocellular carcinoma tumors in mice. *Gastroenterology* **2012**, *143*, 1641–1649. [CrossRef]

51. Zhang, Y.; Yao, J.; Huan, L.; Lian, J.; Bao, C.; Li, Y.; Ge, C.; Li, J.; Yao, M.; Liang, L.; et al. GNAI3 inhibits tumor cell migration and invasion and is post-transcriptionally regulated by miR-222 in hepatocellular carcinoma. *Cancer Lett.* **2015**, *356*, 978–984. [CrossRef]

52. Zheng, Y.; Gery, S.; Sun, H.; Shacham, S.; Kauffman, M.; Koeffler, H.P. KPT-330 inhibitor of XPO1-mediated nuclear export has anti-proliferative activity in hepatocellular carcinoma. *Cancer Chemother. Pharmacol.* **2014**, *74*, 487–495. [CrossRef] [PubMed]

53. Wang, Y.H.; Cheng, T.Y.; Chen, T.Y.; Chang, K.M.; Chuang, V.P.; Kao, K.J. Plasmalemmal Vesicle Associated Protein (PLVAP) as a therapeutic target for treatment of hepatocellular carcinoma. *BMC Cancer* **2014**, *14*, 815. [CrossRef] [PubMed]

54. Bangoura, G.; Liu, Z.S.; Qian, Q.; Jiang, C.Q.; Yang, G.F.; Jing, S. Prognostic significance of HIF-2alpha/EPAS1 expression in hepatocellular carcinoma. *World J. Gastroenterol.* **2007**, *13*, 3176–3182. [CrossRef] [PubMed]

55. Bejnordi, B.E.; Veta, M.; van Diest, P.J.; van Ginneken, B.; Karssemeijer, N.; Litjens, G.; van der Laak, J.; the CAMELYON16 Consortium; Hermsen, M.; Manson, Q.F.; et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA* **2017**, *318*, 2199–2210. [CrossRef] [PubMed]