

Prognostic power of molecular and clinical data across cancer types

Xinsen Xu (✉ xuxinsen@gmail.com)

Shanghai Jiao Tong University School of Medicine Affiliated Renji Hospital <https://orcid.org/0000-0002-3339-9324>

Wei Wang

Shanghai Jiao Tong University School of Medicine Affiliated Renji Hospital

Min He

Shanghai Jiao Tong University School of Medicine Affiliated Renji Hospital

Linhua Yang

Shanghai Jiao Tong University School of Medicine Affiliated Renji Hospital

Wei Chen

Shanghai Jiao Tong University School of Medicine Affiliated Renji Hospital

Jian Wang

Shanghai Jiao Tong University School of Medicine Affiliated Renji Hospital

Research article

Keywords: cancer, prognosis, genome, molecular, clinical

Posted Date: January 17th, 2020

DOI: <https://doi.org/10.21203/rs.2.21114/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Precision medicine holds promise in prognostication of human cancer. By analyzing the Cancer Genome Atlas (TCGA) data, we evaluated the prognostic power of molecular and clinical data across 33 cancer types.

Methods

The clinical and molecular data of more than 11,000 patients were obtained from the TCGA database. Top features associated with overall survival were identified. Concordance index of each data type was calculated to investigate the prognostic power. The performance differences among clinical data, molecular data and combination data (integration of molecular data with clinical data) were evaluated.

Results

The prognostic power of combination data was significantly higher than the molecular data in 108 of 163 comparisons. However, it was only significantly higher than the clinical data in 27 of 163 comparisons. The clinical data seemed to be the most informative prognostic variable in almost half cancer types (14/33). Deeper insights into the low grade glioma models showed that integration of clinical data with molecular data yielded better prognostic modelling than either data used alone. From the pan-cancer level, the combination data was shown to be the most informative prognostic predictor when the sample size was large. In addition, mutation data also showed significant prognostic value.

Conclusions

Molecular markers complement the traditional diagnostic approaches in the pursuit of precision medicine. The combination of reliable clinical data, multidimensional genomic measurements and mature bioinformatics algorithms may confer more robust prognostic value that will inform clinical decision making.

Background

The new era of precision medicine holds promise for the personalized prognostic prediction. As the currently most widely used system, the TNM staging by the American Joint Committee on Cancer (AJCC) is still far from satisfactory. Fortunately, technological advances have greatly increased our understanding of the molecular basis of tumor development. Molecular markers complement the traditional system in the pursuit of precision medicine[1].

For instance, Roth et al. reported that including molecular markers with independent prognostic clinical factors might result in more accurate prognostic prediction than the TNM staging alone[2]. Yuan et al. and Zhao et al. investigated the prediction benefits of combining molecular and clinical data in six cancer

types, and demonstrated that despite limited quantitative gains, some improvement was shown[3, 4]. However, with respect to cancer prognosis, few studies systematically analyze the predictive power of genome data from the pan-cancer point of view.

The Cancer Genome Atlas (TCGA) project has depicted multi-dimensional maps of genomic changes in more than 11,000 patients. As the new milestone of the TCGA project, the Pan-Cancer Atlas analyzed molecular aberrations across cancer types and reclassified multi-cancer groups with potential clinical utility, that therapies effective in one cancer type might be extended to others with similar genomic background[5–7].

Thus, by analyzing the TCGA Pan-Cancer gene expression, methylation, mutation, copy number variation, miRNA and protein expression data, we depicted the prognostic atlas and evaluated the predictive power of molecular and clinical data across 33 cancer types.

Methods

Data set compilation and data processing

Clinical and molecular data were acquired from the TCGA GDAC Firehose System (<http://gdac.broadinstitute.org/>) and the Pan-Cancer Atlas repository (<https://gdc.cancer.gov/node/905/>). Patients with complete clinical and molecular data were screened for further analysis. In order to keep data consistency across cancer types, we selected age and tumor stage as clinical variables, and included gene expression, methylation, mutation, copy number variation, miRNA and protein expression as molecular data. If tumor stage information was missing in some cancer type, it was substituted by specific clinical score or category. The core sample set was defined as the samples with complete data in all platforms (except mutation data) and clinical data. The mutation analysis was performed separately, since the intersection between patients with mutation data and patients with other platform data was relatively small. Similar to the previous reports, for the core data set in each cancer type, patients were randomly split as the training group (80%) and testing group (20%) for 100 times to calculate the concordance index (c-index)[3]. The molecular data obtained from the database were processed as shown in Figure S1.

Performance evaluation of the clinical data alone, molecular data alone and combination data

For the clinical data alone or molecular data alone, top features associated with patient survival were first identified by the univariate cox analysis, and then converged with LASSO in the training group (R package “glmnet”) to select top features. The top features were then applied in the testing group for performance evaluation. After 100 randomizations, the mean concordance index (c-index) was calculated (R package “survcomp”). With respect to the combination data, clinical features that were significantly correlated with

patient survival were first identified as the baseline to build the cox model. Then the molecular variables that better fit the model were included by a feature-selection step against the residuals. After 100 randomizations, the concordance index (c-index) was calculated (R package “survcomp”).

Statistical analysis

The c-index heatmap was constructed using Python module “Matplotlib.pyplot”. To compare the performance differences (c-index) among clinical data, molecular data and combination data, the wilcoxon signed rank test was applied to calculate the P value, with a two-tailed $P < 0.05$ considered significant. The survival curves were constructed by the Kaplan-Meier method and compared by the log-rank test, which were stratified by different prognostic scores. Oncoplot of the mutation data was constructed using the R package “maftools”. Forest plot was constructed to show the association between mutated genes and survival outcomes.

Results

Patient characteristics

We included 33 tumor types in the core sample set, namely adrenocortical carcinoma (ACC), bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), cervical and endocervical cancers (CESC), cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), lymphoid neoplasm diffuse large B-cell lymphoma (DLBC), esophageal carcinoma (ESCA), glioblastoma multiforme (GBM), head and neck squamous cell carcinoma (HNSC), kidney chromophobe (KICH), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), acute myeloid leukemia (LAML), brain lower grade glioma (LGG), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), mesothelioma (MESO), ovarian serous cystadenocarcinoma (OV), pancreatic adenocarcinoma (PAAD), pheochromocytoma and paraganglioma (PCPG), prostate adenocarcinoma (PRAD), rectum adenocarcinoma (READ), sarcoma (SARC), skin cutaneous melanoma (SKCM), stomach adenocarcinoma (STAD), testicular germ cell tumors (TGCT), thyroid carcinoma (THCA), thymoma (THYM), uterine corpus endometrial carcinoma (UCEC), uterine carcinosarcoma (UCS), and uveal melanoma (UVM).

The original platform information of each tumor was shown in Table S1. The sample size, death number, tumor stage distribution and mean age of the core sample set and the mutation sample set in each tumor were summarized in Table S2.

Prognostic power of diverse clinical data, molecular data and combination data in each cancer type

The significant amplifications or deletions calculated from the GISTIC and significant mutated genes calculated from the MutSig were shown in Table S3. The features of miRNA, mRNA, methylation and

RPPA data were obtained directly from the level 3 data (Figure S1). After univariate screen and LASSO for all patients in each cancer type, the significant prognostic factors identified by the cox regression analysis were shown in Table S4-S5.

Patients were randomly split as the training group (80%) and testing group (20%) for 100 times to calculate the concordance index (c-index). The mean c-index calculated from each platform data in each cancer type was shown in Fig. 1, with the darker color blocks representing higher c-index values. The prediction performance of clinical variables varied across cancer types, with the c-index from as high as 0.86 for THCA to as low as 0.40 for CESC (Fig. 1). Among all molecular data, the mRNA seems to be the most informative prognostic variable, with the highest mean c-index of 0.58 (Fig. 1)[4].

There were 163 comparisons between the molecular data and combination data in total, and 163 comparisons between the clinical data and combination data in total (5 molecular data platforms in each of the 33 cancer types, except that the protein data was missing in LAML and UVM, namely, $5 \times 33 - 2 = 163$). Additional predictive power was observed in 108 of the total 163 comparisons between the combination data and molecular data in all cancer types, with the c-index of combination data significantly higher than that of the molecular data alone ($P < 0.05$) (Table S6).

However, on the other hand, in almost half of the cancer types (14/33), the clinical data still appeared to be the most informative index for cancer prediction, with the highest c-index value among the clinical data, molecular data and combination data (Fig. 1). Incorporating molecular data into the clinical model can only boost the prediction accuracy in 27 out of 163 comparisons ($P < 0.05$) (Table S6).

Deeper insights from top-performing prognostic models

To get deeper insights, we further demonstrated the practical applicability of the prognostic model in LGG, since the c-index of LGG was the highest among all the 33 cancer types (Fig. 1).

As the TNM staging system was not applicable in TCGA LGG, we included the age and Karnofsky Performance Status Scale (KPS) as the clinical variables. We divided the patients with different KPS score into high risk (score = 20, 30, 40), high-intermediate risk (score = 50, 60), low-intermediate risk (score = 70, 80), and low risk (score = 90, 100). Using the clinical data alone, the survival outcomes were significantly distinguished by these risk groups (log-rank $P < 0.01$, c-index = 0.75, Fig. 2A). When combining the clinical data with molecular data, the survival outcomes were also significantly different among these subgroups defined by the integrative models involving CNV (Fig. 2B, log-rank $P < 0.01$, c-index = 0.80), methylation (Fig. 2C, log-rank $P < 0.01$, c-index = 0.90), miRNA (Fig. 2D, log-rank $P < 0.01$, c-index = 0.85), mRNA (Fig. 2E, log-rank $P < 0.01$, c-index = 0.72) and RPPA (Fig. 2F, log-rank $P < 0.01$, c-index = 0.92). The detailed models were described in Table S7. On the other hand, the c-index values based on the different classification methods showed that three integrative models (methylation, miRNA and RPPA) outperformed the clinical KPS score model in terms of prognostic power ($P < 0.05$) (Fig. 2).

Pan-cancer analysis of adenocarcinoma, squamous cell carcinoma, neuronal tumors and kidney tumors

As the performance of prognostic prediction depends on the sample size involved, we tried to evaluate the prognostic power of different models from a higher level[8]. Recently, Malta et al. sorted cancer types into different groups by the stemness indices obtained from transcriptomic and epigenetic features[5]. According to their findings, we combined the BRCA, CHOL, COAD, ESCA (adenocarcinoma), LUAD, OV, PAAD, PRAD, READ, STAD and UCEC into the adenocarcinoma, combined the CESC, ESCA (squamous carcinoma), HNSC and LUSC into the squamous cell carcinoma, combined the GBM, LGG, PCPG, SKCM and UVM into the neuronal tumors, and combined the KICH, KIRC and KIRP into the kidney tumors. The mean c-index calculated from each platform data was shown in Fig. 3.

In adenocarcinoma and neuronal tumors, additional predictive power was observed in all combination data, compared with molecular data alone ($P < 0.05$) or clinical data alone ($P < 0.05$) (Fig. 3A, C). In the squamous cell carcinoma and kidney tumors, incorporating clinical data into the molecular model could also boost the prediction accuracy in all the 5 different platforms ($P < 0.05$, Fig. 3B, D). However, when compared with the clinical data alone, additional predictive power was missed in several integrative models involving CNV and methylation in squamous cell carcinoma, and methylation and miRNA in kidney tumors ($P > 0.05$, Fig. 3B, D).

Survival analysis based on mutation data

Under most circumstances in the 100 randomizations, few features passed the univariate cox screen and LASSO, since the number of patients with positive gene mutation was much less than that of the wild type patients. Instead, we only performed the cox regression analysis to explore the significant prognostic mutated genes in each cancer type, which was shown in Table S8.

At the pan-cancer level, we explored the association of driver mutations with prognosis[9]. In adenocarcinoma, squamous cell carcinoma, neuronal tumors and kidney tumors, we incorporated top mutated driver genes with high mutation frequencies ($> 10\%$) to correlate with survival outcomes. As shown in Fig. 4A-B, the hazard ratio (HR) of mutant TP53 versus wild type TP53 was 2.05 [95% confidence interval (CI): 1.81–2.31] in adenocarcinoma and 1.44 (95% CI: 1.19–1.74) in squamous cell carcinoma, respectively. Compared with PTEN wild type population, PTEN mutation carriers had a reduced incidence of mortality in adenocarcinoma (HR: 0.53, 95%CI: 0.42–0.66) (Fig. 4A). However, the mutant PTEN conferred an increased mortality in neuronal tumors (HR: 2.43, 95% CI: 2.02–2.94) (Fig. 4C). The other significant prognostic mutated genes included in PIK3CA, KRAS, and ARID1A in adenocarcinoma (Fig. 4A), NOTCH1 in squamous cell carcinoma, IDH1, ATRX, BRAF, APOB, SPTA1, NF1, EGFR and CIC in neuronal tumors (Fig. 4C), PBRM1 and SETD2 in kidney tumors (Fig. 4D).

Discussion

Till now, our study was probably the most comprehensive analysis evaluating the prognostic power of molecular data and clinical data across cancer types.

In our study, the different tumor biological characteristics, mortality rates and sample sizes in different cancer types resulted in the varied predictive power. Among the 5 types of molecular data, the mRNA seems to be the most informative prognostic variable (mean c-index: 0.58), which was consistent with the previous study reported by Zhao et al[4]. Under most circumstances, the combination data outperformed the molecular data alone. However, the clinical data still seemed to be the most informative prognostic variable in almost half cancer types.

On the other hand, in some occasions, the integrative models were already able to boost the prediction accuracy of the molecular data or clinical data alone[3, 4]. In our results, the c-index values in LGG data were the highest among all the 33 cancer types. Thus, we built several prognostic models of LGG to take a deeper insight. According to previous studies, although there were several prognostic models reported to be useful for LGG, they were only based on the molecular markers[10, 11]. Our integrative models included both molecular data and clinical data, which were demonstrated to outperform the clinical models or molecular models.

As Zhu et. al demonstrated that the performance of prognostic prediction was improved when the sample size increased, we further compared the prognostic power from the pan-cancer level[8]. In kidney tumors (941 patients) and squamous cell carcinoma (1436 patients), although the combination data seemed to confer the highest prognostic power in most cases, several integrative models, which involved methylation and miRNA in kidney tumors, and CNV and methylation in squamous cell carcinoma, did not beat the clinical model of the tumor. However, in neuronal tumors (1841 patients) and adenocarcinoma (4523 patients), which included more patients, the combination data showed the highest prognostic power. Although the different biological characteristics and stemness might partially explain the difference, we conclude that the advantage of the integrative models would be more robust when the sample size increased[6]. So far, the sample size of the TCGA dataset was yet not large enough to fully demonstrate the performance of prognostic prediction in each tumor type.

The accumulation of gene mutations would impair cell division checkpoints and lead to tumorigenesis[12]. According to our results, the most recurrent mutations in each cancer type might not necessarily confer the prognostic value. They were just frequently mutated due to certain cellular processes that were disrupted in certain cancer type[12]. While at the pan-cancer level, we observed that the proportion of prognostic genes were much higher in recurrent driver mutation genes than the normal mutated genes. As driver mutations would result in cancer initiation and progression, they were more likely to be the oncogenes and prognostic genes. We identified several top recurrent mutated genes with prognostic value, such as TP53, PTEN, PIK3CA, KRAS, and so on. Interestingly, we showed the opposite prognostic effects of PTEN mutation on adenocarcinoma and neuronal tumors. We suppose the distinct effects of nonsense mutation and missense mutation of PTEN might lead to the opposite mutational effects of PTEN in different types of cancer[13–15].

Conclusions

Molecular markers complement the traditional diagnostic approaches in the pursuit of precision medicine. Despite the enormous potential of molecular biomarkers, the research of prognostic prediction still has a long way to go. The combination of reliable clinical data, multidimensional genomic measurements and mature bioinformatics algorithms may confer more robust prognostic value that will inform clinical decision making.

Abbreviations

adrenocortical carcinoma, ACC; bladder urothelial carcinoma, BLCA; breast invasive carcinoma, BRCA; cervical and endocervical cancers, CESC; cholangiocarcinoma, CHOL; colon adenocarcinoma, COAD; lymphoid neoplasm diffuse large B-cell lymphoma, DLBC; esophageal carcinoma, ESCA; glioma, GBMLGG; head and neck squamous cell carcinoma, HNSC; kidney chromophobe, KICH; kidney renal clear cell carcinoma, KIRC; kidney renal papillary cell carcinoma, KIRP; acute myeloid leukemia, LAML; liver hepatocellular carcinoma, LIHC; lung adenocarcinoma, LUAD; lung squamous cell carcinoma, LUSC; mesothelioma, MESO; ovarian serous cystadenocarcinoma, OV; pancreatic adenocarcinoma, PAAD; pheochromocytoma and paraganglioma, PCPG; prostate adenocarcinoma, PRAD; rectum adenocarcinoma, READ; sarcoma, SARC; skin cutaneous melanoma, SKCM; stomach adenocarcinoma, STAD; testicular germ cell tumors, TGCT; thyroid carcinoma, THCA; thymoma, THYM; uterine corpus endometrial carcinoma, UCEC; uterine carcinosarcoma, UCS; uveal melanoma, UVM; the Cancer Genome Atlas, TCGA; concordance index, c-index; Karnofsky Performance Status Scale, KPS; hazard ratio, HR; confidence interval, CI.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethical Committee of Renji Hospital, Shanghai Jiao Tong University School of Medicine.

Consent for publication

Not applicable.

Availability of data and materials

The TCGA data were accessed through the Broad Institute's Firehose System (<http://gdac.broadinstitute.org/>) and the the Pan-Cancer Atlas repository (<https://gdc.cancer.gov/node/905/>). The authors declare that all data supporting the findings of this study are available within the Main text and its Supplementary Data, or from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by the Natural Science Foundation of China (No. 81402022), and the Fund of China Scholarship Council (No. 201406280106).

Authors' contributions

X.X.S: conceptualization, funding acquisition, formal analysis, and drafted the work. W.W: methodology, review and editing. H.M, Y.L.H, C.W: data curation. W.J: conceptualization, review and editing.

Acknowledgements

We acknowledge contributions from the TCGA Research Network. We are also grateful to Dr. Yang Liu and Dr. Nilay S. Sethi from Department of Medical Oncology, Dana-Farber Cancer Institute, Boston for critical reading of the manuscript and helpful revision advice.

References

1. Ludwig JA, Weinstein JN: Biomarkers in cancer staging, prognosis and treatment selection. *Nat Rev Cancer* 2005, 5(11):845-856.
2. Roth AD, Delorenzi M, Tejpar S, Yan P, Klingbiel D, Fiocca R, d'Ario G, Cisar L, Labianca R, Cunningham D *et al*: Integrated analysis of molecular and clinical prognostic factors in stage II/III colon cancer. *J Natl Cancer Inst* 2012, 104(21):1635-1646.
3. Yuan Y, Van Allen EM, Omberg L, Wagle N, Amin-Mansour A, Sokolov A, Byers LA, Xu Y, Hess KR, Diao L *et al*: Assessing the clinical utility of cancer genomic and proteomic data across tumor types. *Nat Biotechnol* 2014, 32(7):644-652.
4. Zhao Q, Shi X, Xie Y, Huang J, Shia B, Ma S: Combining multidimensional genomic measurements for predicting cancer prognosis: observations from TCGA. *Brief Bioinform* 2015, 16(2):291-303.
5. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kaminska B, Huelsken J, Omberg L, Gevaert O *et al*: Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell* 2018, 173(2):338-354 e315.
6. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, Shen R, Taylor AM, Cherniack AD, Thorsson V *et al*: Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* 2018, 173(2):291-304 e296.
7. Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM: The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 2013, 45(10):1113-1120.

8. Zhu B, Song N, Shen R, Arora A, Machiela MJ, Song L, Landi MT, Ghosh D, Chatterjee N, Baladandayuthapani V *et al*: Integrating Clinical and Multiple Omics Data for Prognostic Assessment across Human Cancers. *Sci Rep* 2017, 7(1):16954.
9. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, Colaprico A, Wendl MC, Kim J, Reardon B *et al*: Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* 2018, 173(2):371-385 e318.
10. Eckel-Passow JE, Lachance DH, Molinaro AM, Walsh KM, Decker PA, Sicotte H, Pekmezci M, Rice T, Kosel ML, Smirnov IV *et al*: Glioma Groups Based on 1p/19q, IDH, and TERT Promoter Mutations in Tumors. *N Engl J Med* 2015, 372(26):2499-2508.
11. Ceccarelli M, Barthel FP, Malta TM, Sabedot TS, Salama SR, Murray BA, Morozova O, Newton Y, Radenbaugh A, Pagnotta SM *et al*: Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell* 2016, 164(3):550-563.
12. Iranzo J, Martincorena I, Koonin EV: Cancer-mutation network and the number and specificity of driver mutations. *Proc Natl Acad Sci U S A* 2018, 115(26):E6010-E6019.
13. Xu J, Li Z, Wang J, Chen H, Fang JY: Combined PTEN Mutation and Protein Expression Associate with Overall and Disease-Free Survival of Glioblastoma Patients. *Transl Oncol* 2014, 7(2):196-205 e191.
14. Han F, Hu R, Yang H, Liu J, Sui J, Xiang X, Wang F, Chu L, Song S: PTEN gene mutations correlate to poor prognosis in glioma patients: a meta-analysis. *Onco Targets Ther* 2016, 9:3485-3492.
15. Akiyama-Abe A, Minaguchi T, Nakamura Y, Michikami H, Shikama A, Nakao S, Sakurai M, Ochi H, Onuki M, Matsumoto K *et al*: Loss of PTEN expression is an independent predictor of favourable survival in endometrial carcinomas. *Br J Cancer* 2013, 109(6):1703-1710.

Figures

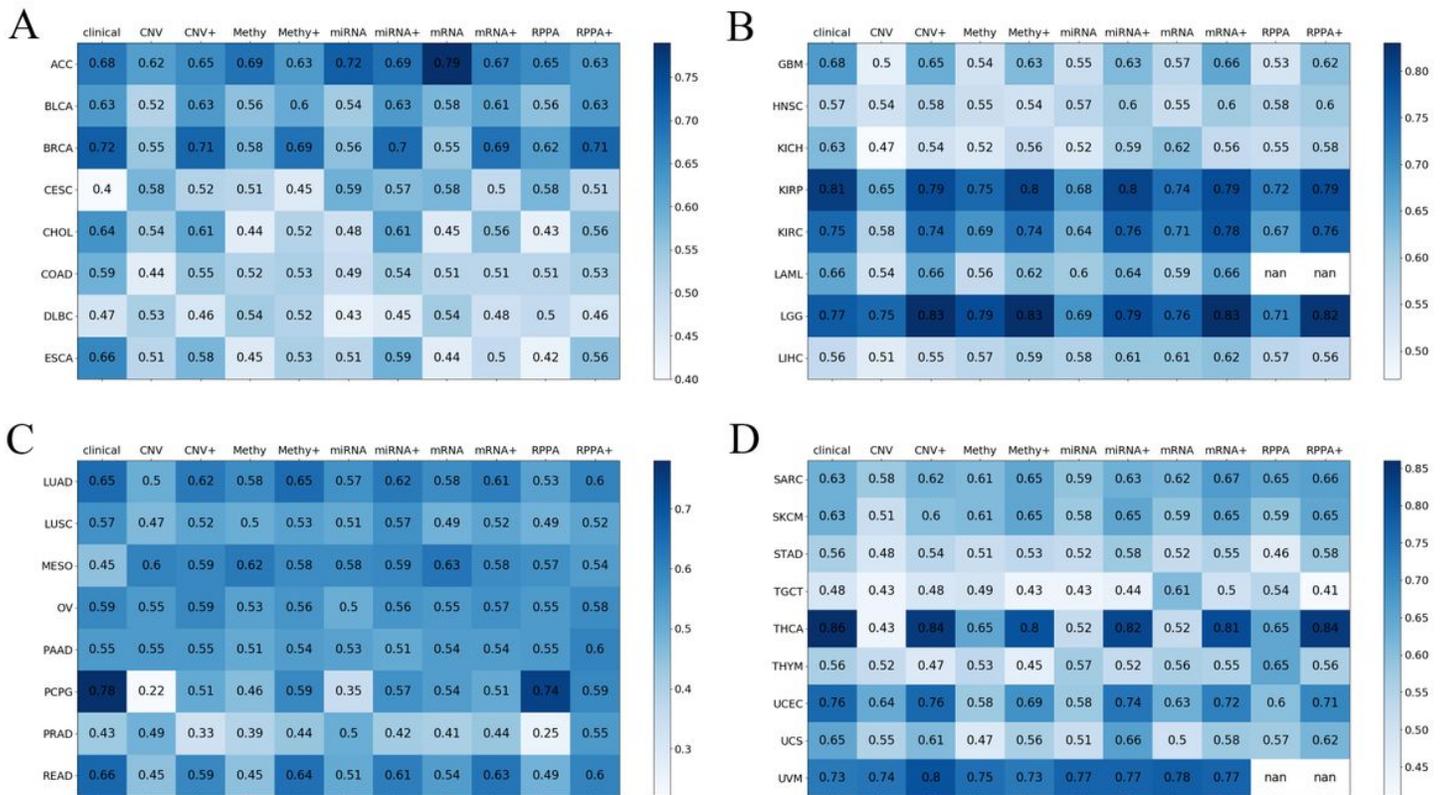


Figure 1

Predictive performance of the clinical data, molecular data and their combinations across cancer types. The c-index value was represented by color depth, that dark blue corresponded to high c-index value, while light blue meant low c-index value. Clinical: c-index calculated from the clinical data alone; CNV: c-index calculated from the copy number data alone; CNV+: c-index calculated from the combination data of copy number data and clinical data; Methy: c-index calculated from the methylation data alone; Methy+: c-index calculated from the combination data of methylation data and clinical data; miRNA: c-index calculated from the miRNA data alone; miRNA+: c-index calculated from the combination data of miRNA data and clinical data; mRNA: c-index calculated from the mRNA data alone; mRNA+: c-index calculated from the combination data of mRNA data and clinical data; RPPA: c-index calculated from the protein data alone; RPPA+: c-index calculated from the combination data of protein data and clinical data.

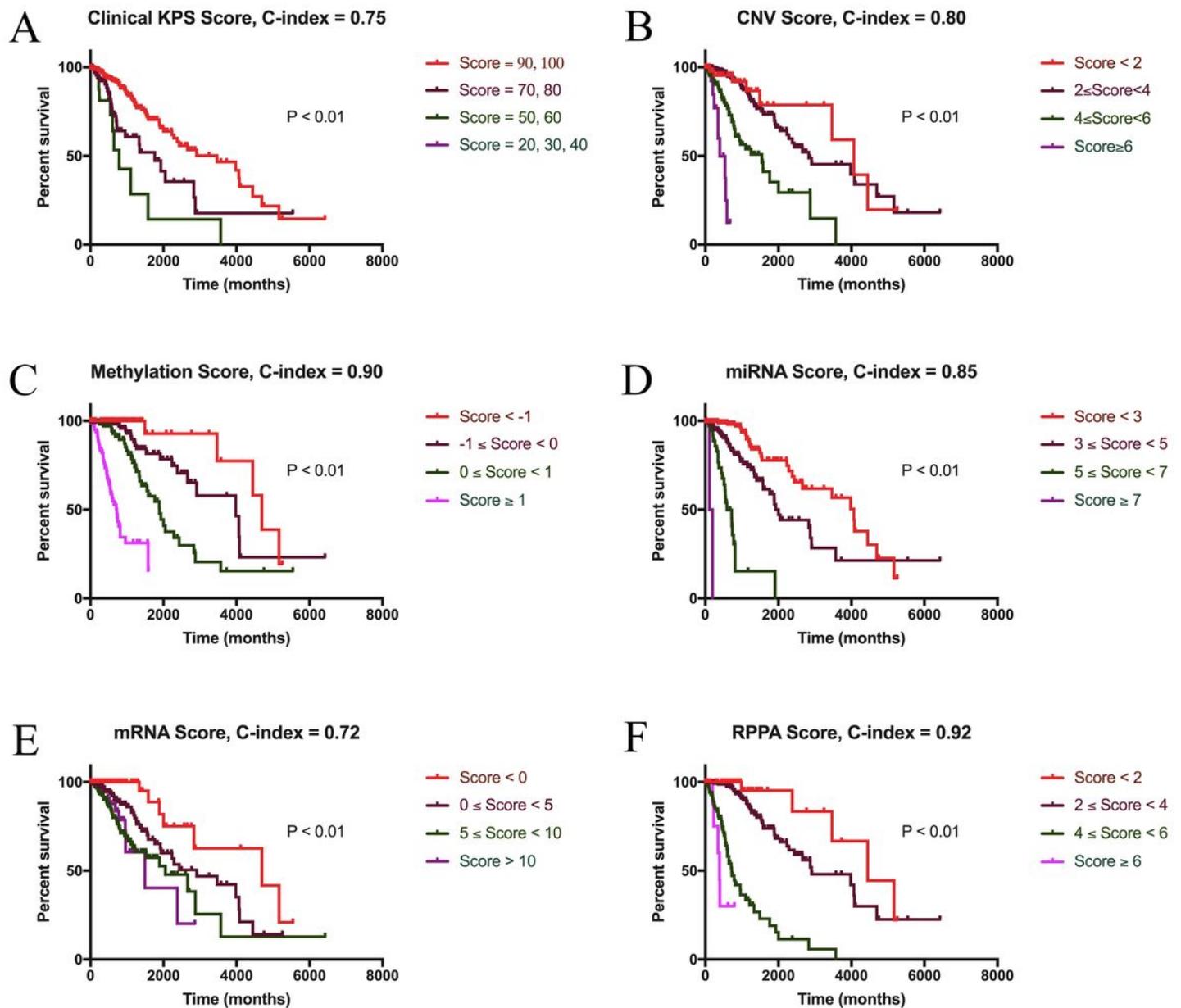


Figure 2

Kaplan-Meier survival analysis in the patients with brain lower grade glioma. (A) Patients were split into “high risk”, “high-intermediate risk”, “low-intermediate risk” and “low risk” groups according the clinical Karnofsky Performance Status Scale. (B) Patients were split into the four risk groups according the integrative model combining clinical data and CNV data. (C) Patients were split into the four risk groups according the integrative model combining clinical data and methylation data. (D) Patients were split into the four risk groups according the integrative model combining clinical data and miRNA data. (E) Patients were split into the four risk groups according the integrative model combining clinical data and mRNA data. (F) Patients were split into the four risk groups according the integrative model combining clinical data and RPPA data.

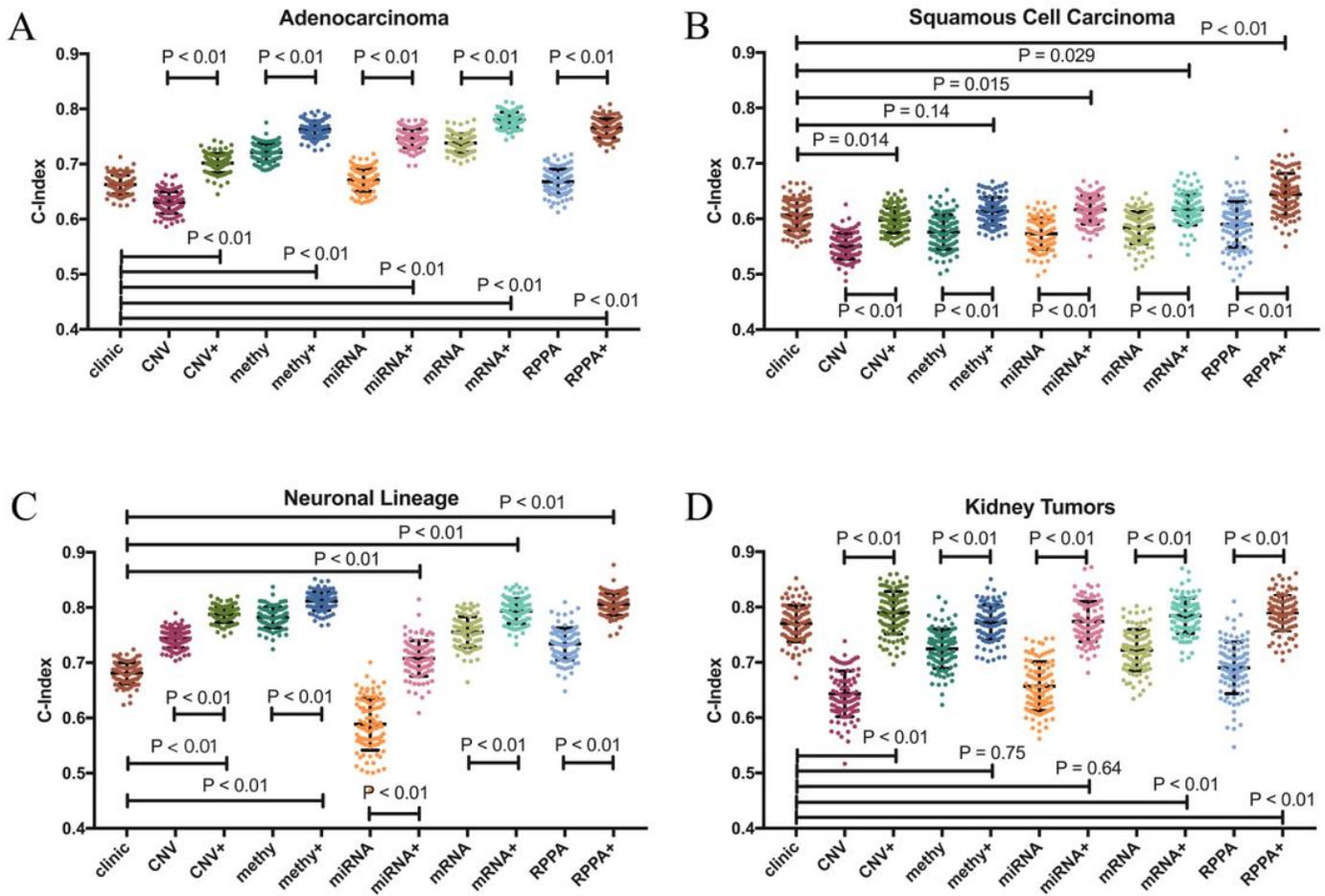


Figure 3

Predictive performance of the clinical data, molecular data and their combinations in adenocarcinoma (A), squamous cell carcinoma (B), neuronal tumors (C) and kidney tumors (D). Clinical: c-index calculated from the clinical data alone; CNV: c-index calculated from the copy number data alone; CNV+: c-index calculated from the combination data of copy number data and clinical data; Methy: c-index calculated from the methylation data alone; Methy+: c-index calculated from the combination data of methylation data and clinical data; miRNA: c-index calculated from the miRNA data alone; miRNA+: c-index calculated from the combination data of miRNA data and clinical data; mRNA: c-index calculated from the mRNA data alone; mRNA+: c-index calculated from the combination data of mRNA data and clinical data; RPPA: c-index calculated from the protein data alone; RPPA+: c-index calculated from the combination data of protein data and clinical data.

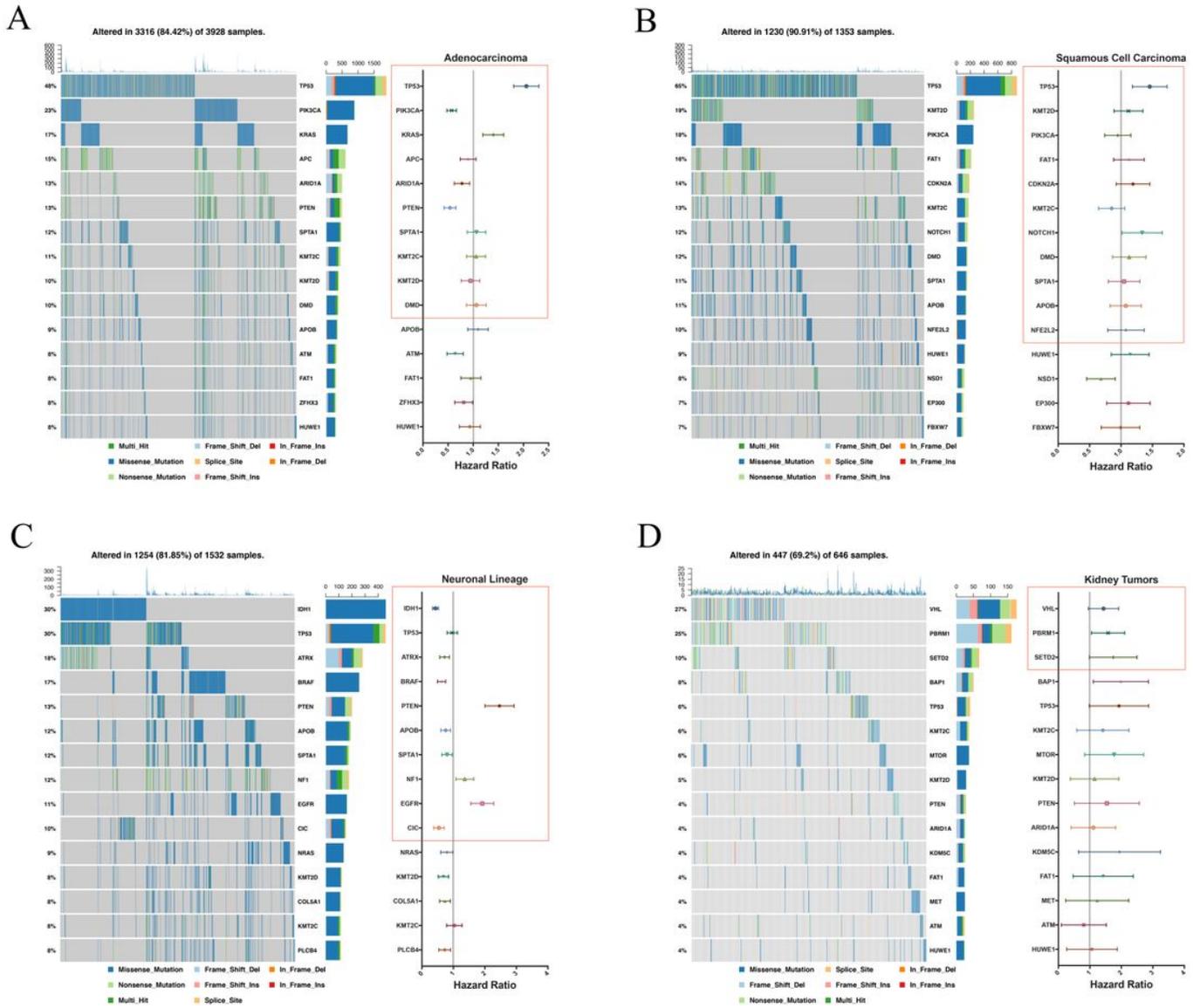


Figure 4

Oncoplots and forest plots showing top significantly mutated genes and their univariate prognostic value in adenocarcinoma (A), squamous cell carcinoma (B), neuronal tumors (C) and kidney tumors (D). Genes in oncoplots were arranged according to the mutation frequency, with the mutation count and mutation type of each gene shown on the right. The total number of mutations detected in each patient was shown at the top, and color keys for the mutation type were provided in the bottom panel. Prognostic value of each gene was depicted as the hazard ratio (HR ± 95% confidence interval) for ‘mutation’ vs ‘wild type’ in the forest plot. The red boxes highlighted the top mutated driver genes with high mutation frequencies (>10%).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigure1.jpg](#)
- [SubmissionSupplementaryTables.docx](#)