Original Article

Construction and validation of a prognostic signature using CNV-driven genes for hepatocellular carcinoma

Jin Bian[1], Junyu Long[1], Xu Yang[1], Xiaobo Yang[1], Yiyao Xu[1], Xin Lu[1], Mei Guan[2], Xinting Sang[1], Haitao Zhao[1]^

[1]Department of Liver Surgery, State Key Laboratory of Complex Severe and Rare Disease, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College (CAMS & PUMC), Beijing, China; [2]Department of Medical Oncology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College (CAMS & PUMC), Beijing, China

Contributions: (I) Conception and design: J Bian, J Long, M Guan, X Sang, H Zhao; (II) Administrative support: X Lu, X Sang, H Zhao; (III) Provision of study materials or patients: J Bian, X Yang, X Yang, X Lu; (IV) Collection and assembly of data: J Bian, X Yang, Y Xu; (V) Data analysis and interpretation: J Bian, J Long, M Guan; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Mei Guan, MD. Department of Medical Oncology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College (CAMS & PUMC), 1 Shuaifuyuan, Wangfujing, Beijing 100730, China. Email: guanmei71@126.com; Xinting Sang, MD. Department of Liver Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College (CAMS & PUMC), 1 Shuaifuyuan, Wangfujing, Beijing 100730, China. Email: sangxt2010@163.com; Haitao Zhao, MD. Department of Liver Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College (CAMS & PUMC), 1 Shuaifuyuan, Wangfujing, Beijing 100730, China. Email: zhaoht@pumch.cn.

Background: Hepatocellular carcinoma (HCC) is one of the major causes of cancer-related deaths worldwide. Copy number variations (CNVs) affect the expression of genes and play critical roles in carcinogenesis. We aimed to identify specific CNV-driven genes and establish a prognostic model for HCC.

Methods: Integrative analysis of CNVs difference data and differentially expressed genes (DEGs) data from The Cancer Genome Atlas (TCGA) were conducted to identify critical CNV-driven genes for HCC. A risk model was constructed based on univariate Cox regression analysis, Least Absolute Shrinkage and Selection Operator (LASSO), and multivariate Cox regression analyses. The associations between CNV-driven genes signature and infiltrating immune cells were explored. The International Cancer Genome Consortium (ICGC) dataset was utilized to validate this model.

Results: After integrative analysis of CNVs and corresponding mRNA expression profiles, 568 CNV-driven genes were identified. Sixty-three CNV-driven genes were found to be markedly associated with overall survival (OS) after univariate Cox regression analysis. Finally, eight CNV-driven genes were screened to generate a prognostic risk model. Compared with low-risk group, the OS of patients in the high-risk group was significantly shorter in both the TCGA [hazard ratio (HR) =6.14, 95% confidence interval (CI): 2.72–13.86, P<0.001] and ICGC (HR =3.23, 95% CI: 1.17–8.92, P<0.001) datasets. Further analysis revealed the infiltrating neutrophils were positively correlated with risk score. Meanwhile, the high-risk group was associated with higher expression of immune checkpoint genes.

Conclusions: A novel signature based on CNV-driven genes was built to predict the survival of HCC patients and showed good performance. The results of our study may improve understanding of the mechanism that drives HCC, and provide an immunological perspective for individualized therapies.

Keywords: Copy number variation-driven genes (CNV-driven genes); hepatocellular carcinoma (HCC); prognosis; immune microenvironment

Submitted Oct 25, 2020. Accepted for publication Mar 10, 2021.
doi: 10.21037/atm-20-7101
View this article at: http://dx.doi.org/10.21037/atm-20-7101

^ ORCID: 0000-0002-3444-8044.

Introduction

Hepatocellular carcinoma (HCC) is a lethal malignancy and accounts for approximately 85% to 90% of primary liver cancers (1,2). Although targeted therapy and immunotherapy have emerged as potential therapies, curative therapies for HCC remain limited (3). Moreover, high post-operative recurrence rates and rare complete cures make it difficult for achieving long term survival. A study on natural history of HCC indicated that patients with advanced stage (Barcelona Clinic Liver Cancer Stage C) had a survival of only 3.4 months if untreated (4). HCC develops following a step-wise manner with abundant genetic and epigenetic molecular alterations (5). Therefore, it is crucial to achieve a better understanding of the underlying molecular mechanism that drives HCC occurrence and development. Exploring prediction model based on the factors that drive HCC can be useful for individualized therapy option and prognosis prediction for HCC patients.

As critical subclasses of somatic mutations, copy number variations (CNVs) refer to duplications or deletions of DNA segments, which are greater than 1 kb compared to a reference genome (6). CNVs account for the accumulation of genomic DNA aberrations, and play important role in cancer pathogenesis. Notably, CNVs can result in activation of oncogenes or inactivation of tumor suppressor genes, which drives cancer development (7,8). Multiple CNVs have been reported to be implicated in the pathogenesis and prognosis of cancers including HCC (9-12). Frequent CNVs of subpopulations of cancer cells were reported to contribute to HCC heterogeneity, indicating a critical role of CNVs in HCC development and progression (13). However, most previous studies focused on CNVs or transcriptome alterations separately, and a comprehensive study of how CNVs drives HCC is still lacking. Combining analysis of CNVs and corresponding gene expression will promote more accurate identification of the specific cancer signatures for HCC. In this study, we used transcriptomic and CNVs profiles to identify CNV-driven genes and aimed to construct a prognostic model for HCC. Our research may contribute to better understanding of the underlying mechanisms, and provide novel therapeutic targets for HCC treatment. We present the following article in accordance with the TRIPOD reporting checklist (available at http://dx.doi.org/10.21037/atm-20-7101).

Methods

Data collection

Gene expression profiles (374 tumor samples and 50 normal samples) and DNA CNVs data (379 HCC samples and 389 nontumor samples) of HCC patients were obtained from The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/, up to November 1, 2019). The corresponding clinical parameters were also obtained. HCC RNA-sequencing data were analyzed using the Illumina HiSeq 2000 RNA Sequencing platform, and CNVs data were analyzed with the Affymetrix SNP 6.0 platform. For validation cohort, RNA-sequencing profiles of 232 HCC patients with survival time and status were downloaded from the International Cancer Genome Consortium (ICGC) (https://dcc.icgc.org/, up to April 3, 2019). All analyses were performed according to relevant regulations and guidelines. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Identification of differentially expressed genes (DEGs) between tumor and normal tissues

To identify genes critical for HCC development, we used the "edgeR" R package to select DEGs between tumor and nontumor samples from TCGA (14). The |log2 (fold change [FC])| >2 and false discovery rate (FDR) <0.01were used as cutoff value for screening DEGs.

Integrative analysis of gene expression and DNA CNVs

Genes in CNV regions were annotated using Genome Research Consortium Human build 38 (GRCh38) as reference genome. The copy variation ratios of the genes both in normal and tumor samples were calculated and the gene-CNV matrix was constructed for Chi-square test. CNVs alteration rates between normal and tumor samples were then compared using Chi-square test, and CNVs data with adjusted P values less than 0.05 were chosen for next analysis. Then the CNVs data and DEGs data of the same sample were merged to construct a matrix. By using Kolmogorov-Smirnov test, those genes showing the same tendency both in CNVs and differential gene expression were selected as CNV-driven genes. Moreover, the differential expression of CNV-driven genes between tumor and normal samples was compared by utilizing the Wilcoxon rank-sum test method.

Development and validating the risk prognostic model

Prognostic CNV-driven genes were screened to construct a prognostic prediction model for the TCGA set. We employed univariate Cox proportional-hazards regression analyses to evaluate the associations between CNV-driven DEGs and prognosis. Genes with a P<0.0001 in univariate Cox regression analysis were selected for subsequent analysis. Least Absolute Shrinkage and Selection Operator (LASSO) regression analysis was employed to remove redundant variables and minimize overfitting (15). Then multivariate Cox proportional-hazards regression analysis was conducted to generate coefficients that were used as weights in the prognostic model. The prognostic prediction model including eight genes was built through a linear combination of mRNA expression level. The risk score = (0.06124 × CDCA8 mRNA level) + (0.05817 × AKR1B15 mRNA level) + (0.07457 × EZH2 mRNA level) + (0.02522 × EPS8L3 mRNA level) + (0.05672 × CBX2 mRNA level) + (0.02529 × TRIM16L mRNA level) + (0.11022 × FLVCR1 mRNA level) + (0.11982 × GPRIN1 mRNA level). Based on the risk score model, patients were divided into two groups with high- or low-risks. The optimal risk score cutoff value was obtained using X-tile software (16). Kaplan-Meier (KM) and log-rank methods were used to compare the overall survival (OS) between the two subgroups. The receiver operating characteristic (ROC) curves were plotted, and the external validation of the predictive model was conducted in the ICGC database.

Independence of risk score from other clinical features

In TCGA group, 235 patients with both clinical information and corresponding gene expression were included in the analysis. In ICGC group, 232 patients were included for independence analysis. Univariate and multivariate analyses of OS were employed to evaluate whether the risk score was independent of other clinical features.

Functional enrichment analysis and genome annotation

To explore the underlying biological functions, we performed gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analyses for CNV-driven genes. ClusterProfiler package (17) in R was used to plot the results.

Association between CNV-driven genes prognostic signature and tumor-infiltrating immune cells

Tumor Immune Estimation Resource (TIMER) is a comprehensive database for analyzing tumor-infiltrating immune cells (18). We utilized this database to estimate the abundance of major immune cell subpopulations in tumor immune microenvironment (CD4+ T cells, CD8+ T cells, B cells, macrophages, neutrophils, and dendritic cells) in HCC. The correlations between risk score and tumor-infiltrating immune cells were analyzed with Pearson test. Moreover, the expression levels of key immune checkpoint genes between high- and low-risk patients were compared by using Wilcoxon rank-sum test.

Statistical analysis

All the analyses were performed with R software (version 3.6.2). Unless otherwise specified, a P value less than 0.05 was considered statistically significant.

Results

DEGs in HCC

After data was collected as described in Methods, 3,598 DEGs between tumor and nontumor samples were identified. Among these DEGs, 3,298 genes were upregulated and 300 genes were downregulated. These DEGs were used for further analysis.

Identification of CNV-driven genes in HCC patients

By applying Chi-square test, 16,644 HCC-related CNV genes were identified (adjusted P<0.05). The distribution of HCC-related CNVs in the chromosomes is shown in *Figure 1*. Then CNV-driven genes were screened using Kolmogorov-Smirnov test. The Kolmogorov-Smirnov test identified 568 CNV-driven genes for HCC (Table S1). To illustrate the functional characteristics and biological effects of these CNV-driven genes, GO and KEGG analyses were conducted (*Figure 2A,B*). Results showed CNV-driven genes were significantly enriched in categories associated with cell division and proliferation, such as “nuclear division”, “chromosome, centromeric region”, and “ligand-gated ion channel activity”. These results indicate that the CNV-driven genes are involved in the dysregulation of tumor cell

proliferation, and are critical in the molecular mechanisms of HCC development. Results from KEGG analysis showed that the top six signaling pathways were cell cycle, melanoma, p53 signaling pathway, mineral absorption, oocyte meiosis and gastric cancer. Most of these signaling pathways are involved in tumor initiation and progression, indicating that the CNV-driven genes are critical in the molecular mechanisms of HCC development.

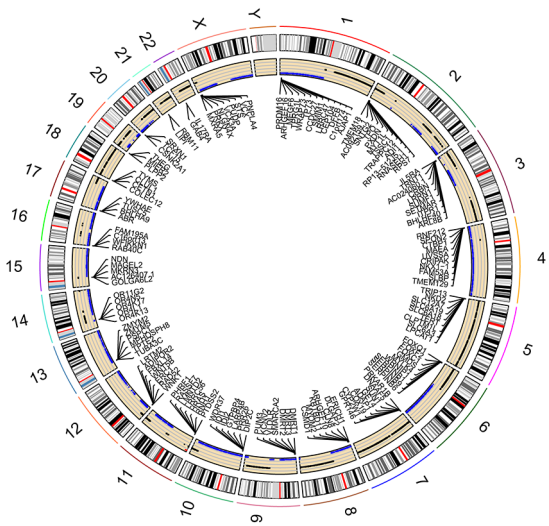

Figure 1 Distribution of HCC-related CNVs visualized by circos plot. The outside circle represents 24 chromosomes including sex chromosomes; the inside circle represents distribution of CNVs (the blue dots represent CNV deletions). HCC, hepatocellular carcinoma; CNV, copy number variation.

Screening of prognostic CNV-driven genes associated with survival

Wilcoxon rank-sum test was used to analyze the difference of the 568 CNV-driven genes between tumor and non-tumor tissues (FDR <0.05 and |log2 [FC]| >1), and 373 differentially expressed CNV-driven genes were finally selected (Table S2). Of the 373 CNV-driven genes, 63 CNV-driven genes were identified as potential prognostic biomarkers for OS after univariate analysis (P<0.0001, Table S3).

Generating and evaluating the HCC prognosis prediction model

The 63 selected CNV-driven genes were analyzed by LASSO regression method. Eight genes appeared 800 times of a total of 1000 repetitions and were detected as prognostic genes for building risk score (*Figure 3A*,*B*). Then, using the coefficients from multivariate Cox regression, we built a model based on the eight CNV-driven genes (Table S4). Hazard ratios of all the eight CNV-driven genes were greater than 1, indicating these genes were associated with shorter OS of HCC patients. Based on the optimal cutoff value of 2.43 for risk score, the

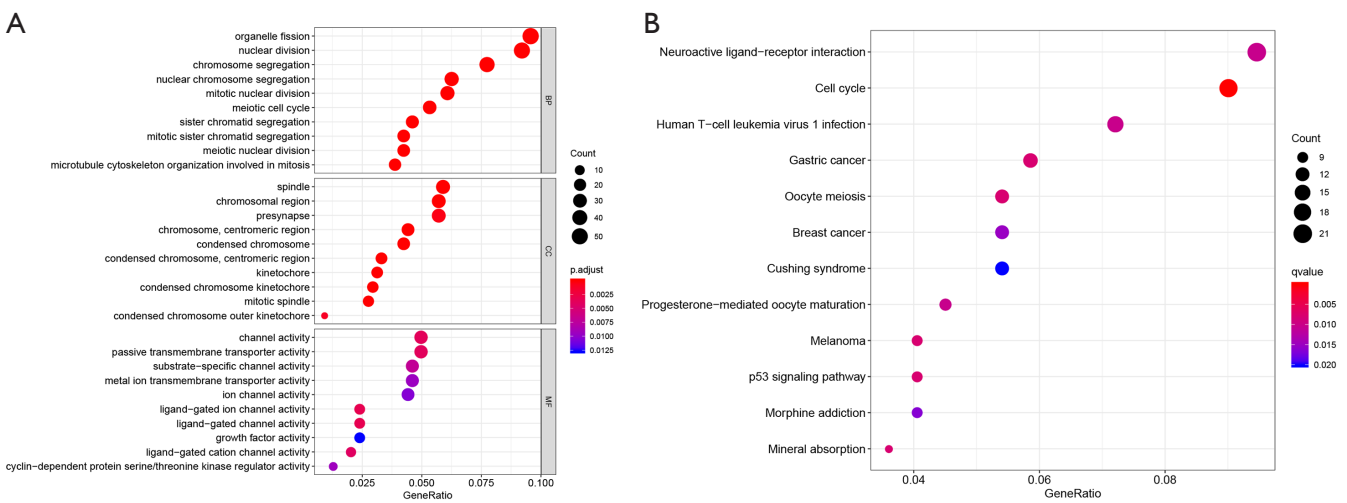

Figure 2 GO and KEGG enrichment for CNV-driven genes. (A) GO enrichment. (B) KEGG pathway enrichment. GO, gene ontology; KEGG, Kyoto encyclopedia of genes and genomes; CNV, copy number variation.

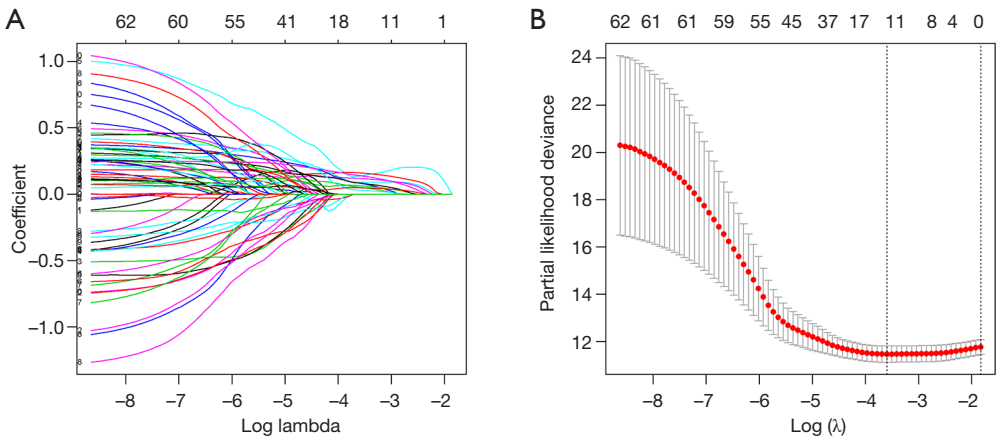

Figure 3 Selection of the prognostic CNV-driven genes for HCC patients by LASSO. (A) The LASSO coefficient changing profiles of 63 CNV-driven genes. (B) Determining the optimal lambda value in the LASSO by ten-fold cross-validation. Confidence intervals for each lambda was shown. CNV, copy number variation; HCC, hepatocellular carcinoma; LASSO, least absolute shrinkage and selector operation.

patients were grouped into two subgroups with high- and low-risks respectively. Patients with high-risk scores had significantly shorter OS (HR =6.14, 95% CI: 2.72–13.86, P<0.001) than patients in low-risk group (*Figure 4A*). The risk score distributions and expression of CNV-driven genes were plotted (*Figure 4B,C,D*). Expression levels for the eight CNV-driven genes increased as risk scores, indicating these CNV-driven genes were high risk factors for OS. The area under the ROC curve (AUC) curve for the 3-year OS was 0.704 (Figure S1A). The risk prognosis model was validated using external independent data from ICGC datasets. HCC patients in validating cohort were designated into high- and low-risk groups using the same risk score formula and cutoff obtained from the TCGA group. Compared to the low-risk group, high-risk group showed significantly poorer OS (HR =3.23, 95% CI: 1.17–8.92, P<0.001) (*Figure 4E*). The risk score distribution, vital statuses of patients, and expression levels of CNV-driven genes were shown in Figure 4F,G,H. The AUC of the 3-year OS was 0.768 for HCC patient in ICGC dataset (Figure S1B).

Independent of the prognostic model from other clinical features in TCGA and ICGC

Univariate and multivariate Cox proportional-hazards model were used to determine whether the risk score prognostic model was independent of clinical and pathological parameters (*Tables 1,2*). Among 235 patients in TCGA datasets, univariate analyses indicated that T category (primary tumor), TNM stage, and risk score were significantly correlated with OS (P<0.001). Multivariate analysis showed that the CNV-driven genes prognostic risk score was the only significant independent predictor for OS (P<0.001). Among 232 patients in ICGC, univariate analysis showed that risk score (P<0.001) and TNM stage (P<0.001) were related with OS, and further multivariate analysis showed risk score was still a predictor for OS independent of TNM stage (P<0.001).

Analysis of the tumor-infiltrating cells and immune genes with the CNV-driven risk signature

Tumor-infiltrating cells play critical roles in tumor immune balance and associate with cancer development. To explore whether the CNV-driven genes prognostic model was associated tumor-infiltrating cells, we analyzed the relationship between risk score and six immune cell subsets. Pearson correlation tests showed that the abundance of CD8+ T cells, dendritic cells, neutrophils and macrophages were positively correlated with risk scores (P<0.05, *Figure 5*). Of note, the Pearson correlation coefficient is largest in correlation between neutrophils and risk score, and the weakest correlation is observed between CD8+ T cells and risk score. There were no relations between risk score and B cells or CD4+ T cells. These results indicate that increased infiltrating neutrophils are associated with poorer survival and play a negative role in HCC immune balance. We further assessed the critical immune checkpoint genes expression between patients in high- and low-risk groups. As shown in *Figure 6*, high-risk cohort had higher expression

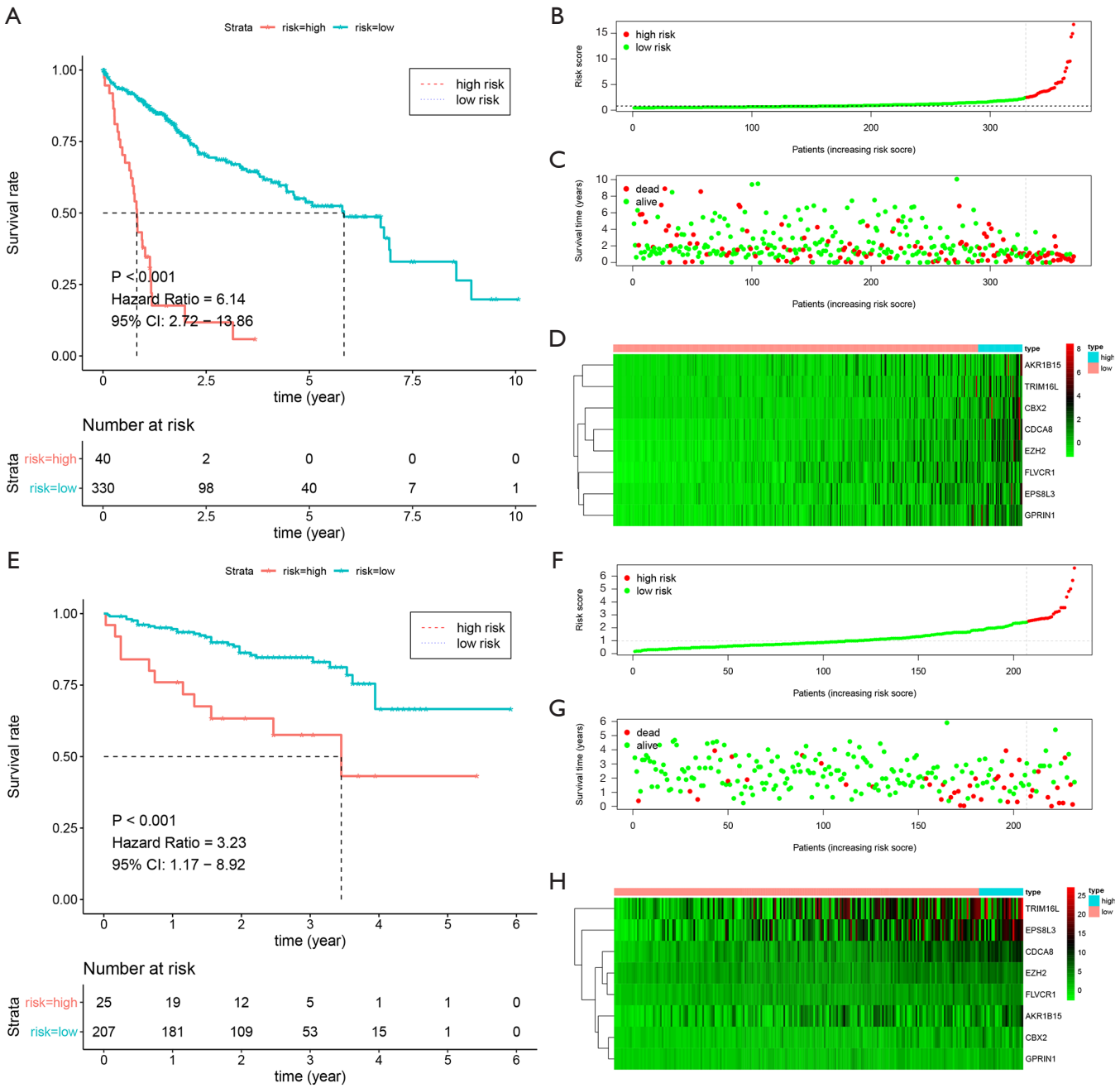

Figure 4 The Kaplan-Meier curves and the distribution of risk score, vital statuses, and CNV-driven genes expression. (A,E) K-M curves for the prognostic model in the TCGA set and ICGC set, respectively. (B,C,D) The distribution of risk score based on CNV-driven genes, the vital statuses of patients, and heatmap of the genes profiles in the TCGA set. (F,G,H) The risk score distribution, the vital statuses of patients, and heatmap of the gene profiles in the ICGC set. CNV, copy number variation; TCGA, The Cancer Genome Atlas; ICGC, International Cancer Genome Consortium.

Table 1 Univariate and multivariate regression analyses for TCGA group

Variables	Univariate analysis		Multivariate analysis	
	HR (95% CI)	P value	HR (95% CI)	P value
Age	1.005 (0.987–1.023)	0.591	1.008 (0.989–1.028)	0.421
Gender	0.780 (0.487–1.249)	0.301	0.849 (0.50–1.442)	0.544
Grade	1.017 (0.746–1.387)	0.914	1.077 (0.767–1.514)	0.668
Stage	1.864 (1.456–2.388)	8.07E–07	1.998 (0.670–5.957)	0.214
T	1.804 (1.434–2.270)	4.73E–07	0.843 (0.302–2.357)	0.746
M	3.850 (1.207–12.281)	0.023	1.437 (0.380–5.431)	0.593
N	2.022 (0.494–8.276)	0.328	0.838 (0.092–7.622)	0.876
Risk score	1.260 (1.180–1.347)	6.57E–12	1.242 (1.137–1.357)	1.404E–06

TCGA, The Cancer Genome Atlas; HR, hazard ratio; CI, confidence interval.

Table 2 Univariate and multivariate regression analyses for ICGC group

Variables	Univariate analysis		Multivariate analysis	
	HR (95% CI)	P value	HR (95% CI)	P value
Age	1.002 (0.972–1.033)	0.899	0.999 (0.965–1.033)	0.936
Gender	0.519 (0.278–0.966)	0.039	0.404 (0.213–0.764)	0.005
Stage	2.155 (1.493–3.110)	4.13E–05	1.865 (1.285–2.705)	0.001
Risk score	1.650 (1.363–1.998)	2.82E–07	1.484 (1.195–1.844)	0.000364

ICGC, International Cancer Genome Consortium; HR, hazard ratio; CI, confidence interval.

levels of CTLA4, TIM-3, LAG3 and CD39 compared to those in the low-risk cohort (P<0.05). These results suggest that high-risk patients had higher immunoinhibitory gene expression, and may benefit from immunotherapy based on immune checkpoint inhibitors.

Discussion

HCC remains a major cause of cancer-related deaths in the world, causing one of the highest public health burdens (19,20). Advancement in molecular analyses has facilitated deep understanding of the HCC mutation landscape and characteristics. Studies indicated that hepatocarcinogenesis was a multistep and multifactorial process caused by frequent aberrant gene alterations, including single nucleotide mutations and CNVs (21). Therefore, understanding the roles of CNVs in driving hepatocarcinogenesis is crucial for HCC prevention, treatment, and prognosis prediction.

Integrated genomic analysis can be an effective and essential method for identification of novel cancer driver genes. For instance, the widespread use of high-throughput sequencing has enabled more efficient and comprehensive analysis of CNVs, and provides opportunities for revealing new genes underlying the development of HCC (22). CNVs in oncogenes and tumor suppressor genes are involved in HCC malignant proliferation and transformation. Previous studies on HCC showed that oncogenic driver genes CCND1 and FGF19 had increased amplifications of copy numbers (23), while tumor suppressor genes CDKN2A and CDKN2B contained high frequency of deletions (12). Analysis of recurrent CNVs can also help to identify potential novel biomarkers such as IRF2, which is unique to hepatitis virus B-related HCC (24).

In this work, we conducted an integrative analysis of CNVs and gene expression profiles aiming to identify CNV-driven genes that associated HCC survival, and built a prognostic signature with CNV-driven genes. In multivariate Cox proportional-hazards analysis, the prognostic risk score proved to be an independent predictor

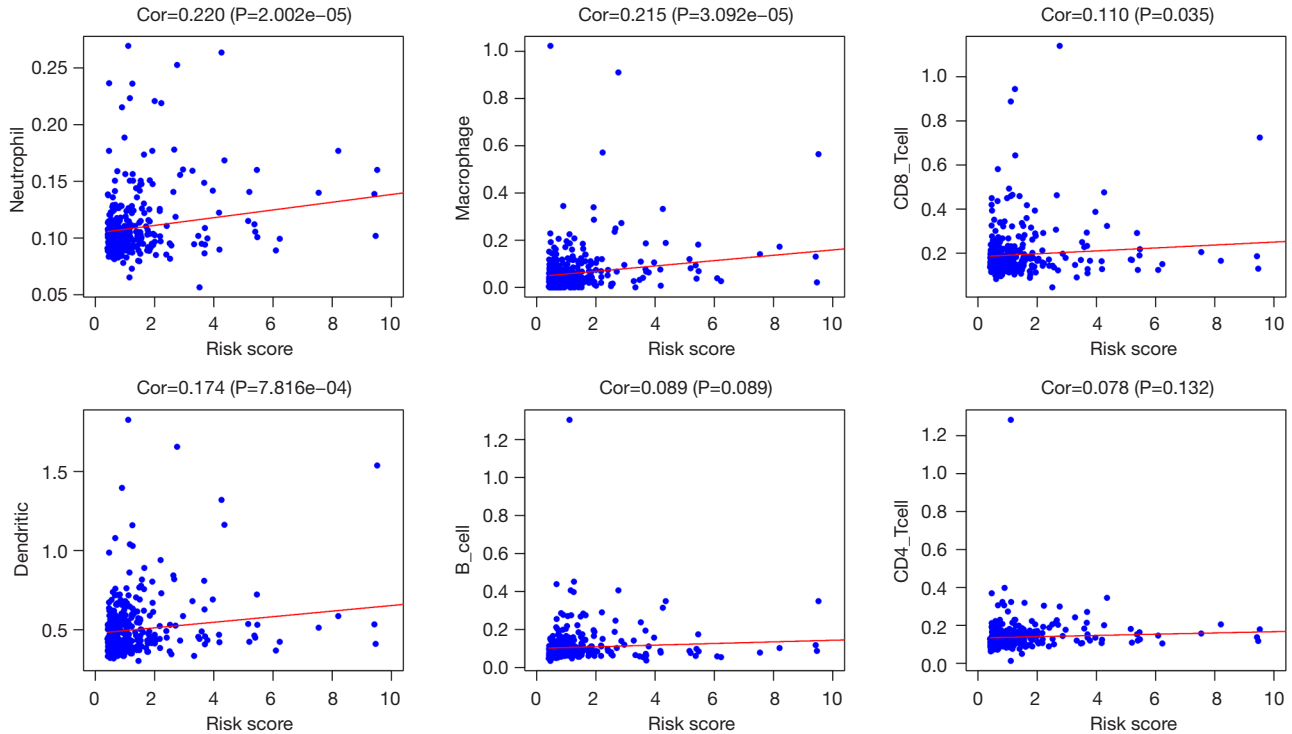

Figure 5 Relationships between the risk score and six immune cells for TCGA datasets. The Pearson correlation coefficients (Cor) are illustrated in each plot. TCGA, The Cancer Genome Atlas.

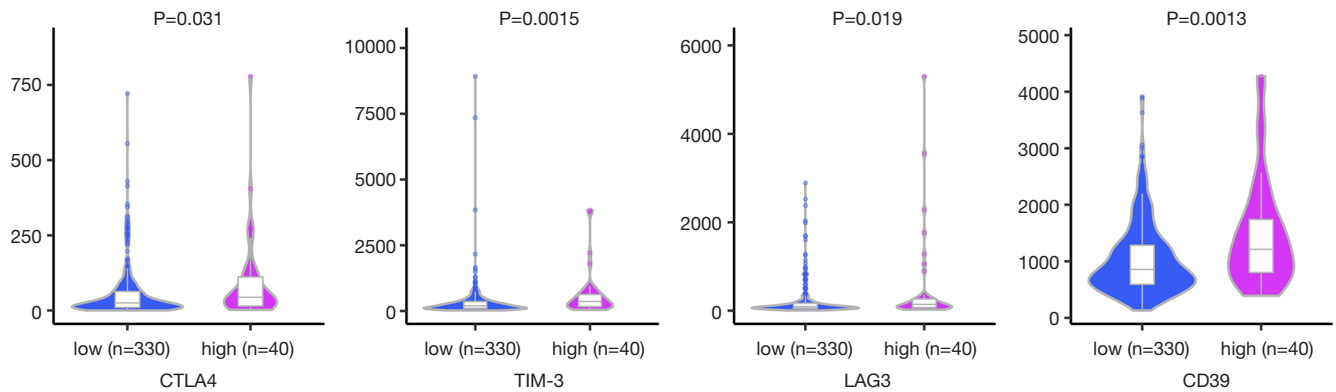

Figure 6 Comparison of immune checkpoints genes expression in the two risk subgroups.

for OS. Survival analysis showed the risk score prediction model had robust distinguishing ability, and might help to improve individualized prediction of OS in HCC patients.

Recent research results have highlighted the roles of tumor-infiltrating immune cells in HCC immune tolerance and survival prognosis (25,26). Since CNVs can result in alterations of key genes responsible for cancer immune surveillance, we therefore investigated the associations between tumor-infiltrating immune cells, immune checkpoint genes, and the risk score. The results showed there was the strongest positive correlation between neutrophils and risk score, indicating increased infiltrating neutrophils were risk factor for survival. Previous studies have demonstrated that neutrophil activation could drive tumor progression and metastasis (27), and infiltrating neutrophils contributed to HCC progression, and

associated with poor prognosis (28). These studies were consistent with our findings and implied that our risk prognostic signature was closely related with tumor immune microenvironment. Moreover, we explored the expression of immune checkpoint genes that had been proved critical in HCC. Based on the risk score, the gene expression of CTLA-4, TIM-3, LAG3, and CD39 were significantly higher in patients with high-risk. The high expression of immune checkpoint genes may be responsible for poorer survival of high-risk group. The prognostic signature can be utilized to identify high-risk populations who may benefit from cancer immunotherapy such as immune checkpoint inhibitors.

Our study has some limitations. Since the data is retrospective, the results need to be further confirmed in prospective studies. Moreover, the tumor-infiltrating immune cells and related genes expression remain to be further validated by experimental methods.

Conclusions

In summary, our study identified CNV-driven genes that related to HCC survival. A prognostic prediction model was established based on CNV-driven genes. Further analyses indicated that tumor-infiltrating immune cells and altered immune checkpoint genes might account for the model's prognostic capacity. These results contribute to the understanding of hepatocarcinogenesis from view of CNVs, and may improve outcome prediction for patients with HCC.

Acknowledgments

We thank Yu Lin (Statistician, Shenzhen Withsum Technology Limited) for assistance with the data interpretation.

Funding: This work was supported by the International Science and Technology Cooperation Projects, No. 2016YFE0107100; Capital Special Research Project for Health Development, No. 2014-2-4012; Beijing Natural Science Foundation, No. L172055 and No. 7192158; National Ten-thousand Talent Program, the Fundamental Research Funds for the Central Universities, No. 3332018032; and CAMS Innovation Fund for Medical Science (CIFMS), No. 2017-I2M-4-003 and No. 2018-I2M-3-001.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at http://dx.doi.org/10.21037/atm-20-7101

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/atm-20-7101). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The data used in the current study are obtained from The Cancer Genome Atlas database (TCGA) and the International Cancer Genome Consortium (ICGC), which are open to the public under some guidelines. Therefore, it is confirmed that all written informed consent was achieved and no ethical approval was needed.

References

1. Llovet JM, Zucman-Rossi J, Pikarsky E, et al. Hepatocellular carcinoma. Nat Rev Dis Primers 2016;2:16018.
2. Villanueva A. Hepatocellular Carcinoma. N Engl J Med 2019;380:1450-62.
3. Li ZL, Han J, Liu K, et al. Association of family history with long-term prognosis in patients undergoing liver resection of HBV-related hepatocellular carcinoma. Hepatobiliary Surg Nutr 2019;8:88-100.
4. Khalaf N, Ying J, Mittal S, et al. Natural History of Untreated Hepatocellular Carcinoma in a US Cohort and the Role of Cancer Surveillance. Clin Gastroenterol Hepatol 2017;15:273-81.e1.
5. Sia D, Villanueva A, Friedman SL, et al. Liver Cancer Cell of Origin, Molecular Class, and Effects on Patient Prognosis. Gastroenterology 2017;152:745-61.

6. Sebat J, Lakshmi B, Troge J, et al. Large-scale copy number polymorphism in the human genome. Science 2004;305:525-8.
7. Nik-Zainal S, Davies H, Staaf J, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 2016;534:47-54.
8. Hoang PH, Dobbins SE, Cornish AJ, et al. Whole-genome sequencing of multiple myeloma reveals oncogenic pathways are targeted somatically through multiple mechanisms. Leukemia 2018;32:2459-70.
9. Liu W, Sun J, Li G, et al. Association of a germ-line copy number variation at 2p24.3 and risk for aggressive prostate cancer. Cancer Res 2009;69:2176-9.
10. Diskin SJ, Hou C, Glessner JT, et al. Copy number variation at 1q21.1 associated with neuroblastoma. Nature 2009;459:987-91.
11. Park RW, Kim TM, Kasif S, et al. Identification of rare germline copy number variations over-represented in five human cancer types. Mol Cancer 2015;14:25.
12. Wang K, Lim HY, Shi S, et al. Genomic landscape of copy number aberrations enables the identification of oncogenic drivers in hepatocellular carcinoma. Hepatology 2013;58:706-17.
13. Hou Y, Guo H, Cao C, et al. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. Cell Res 2016;26:304-19.
14. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139-40.
15. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med 1997;16:385-95.
16. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. Clin Cancer Res 2004;10:7252-9.
17. Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics 2012;16:284-7.
18. Li T, Fan J, Wang B, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. Cancer Res 2017;77:e108-10.
19. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394-424.
20. Xie DY, Ren ZG, Zhou J, et al. 2019 Chinese clinical guidelines for the management of hepatocellular carcinoma: updates and insights. Hepatobiliary Surg Nutr 2020;9:452-63.
21. Niu ZS, Niu XJ, Wang WH. Genetic alterations in hepatocellular carcinoma: An update. World J Gastroenterol 2016;22:9069-95.
22. Chen CF, Hsu EC, Lin KT, et al. Overlapping high-resolution copy number alterations in cancer genomes identified putative cancer genes in hepatocellular carcinoma. Hepatology 2010;52:1690-701.
23. Sawey ET, Chanrion M, Cai C, et al. Identification of a therapeutic strategy targeting amplified FGF19 in liver cancer by Oncogenomic screening. Cancer Cell 2011;19:347-58.
24. Guichard C, Amaddeo G, Imbeaud S, et al. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. Nat Genet 2012;44:694-8.
25. Zheng C, Zheng L, Yoo JK, et al. Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. Cell 2017;169:1342-56.e16.
26. Zhang Z, Ma L, Goswami S, et al. Landscape of infiltrating B cells and their clinical significance in human hepatocellular carcinoma. Oncoimmunology 2019;8:e1571388.
27. Singel KL, Segal BH. Neutrophils in the tumor microenvironment: trying to heal the wound that cannot heal. Immunol Rev 2016;273:329-43.
28. Zhou SL, Yin D, Hu ZQ, et al. A Positive Feedback Loop Between Cancer Stem-Like Cells and Tumor-Associated Neutrophils Controls Hepatocellular Carcinoma Progression. Hepatology 2019;70:1214-30.

Cite this article as: Bian J, Long J, Yang X, Yang X, Xu Y, Lu X, Guan M, Sang X, Zhao H. Construction and validation of a prognostic signature using CNV-driven genes for hepatocellular carcinoma. Ann Transl Med 2021;9(9):765. doi: 10.21037/atm-20-7101

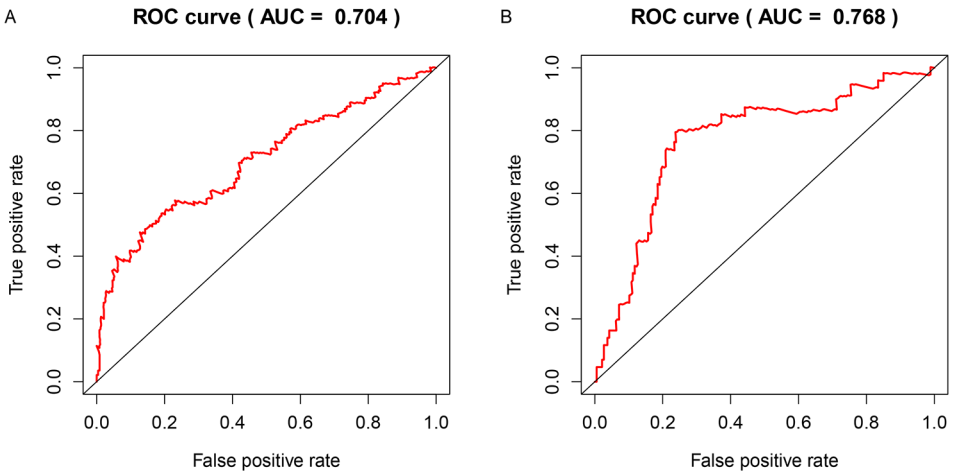

Figure S1 ROC curve of the risk prognostic model. (A) ROC of 3-year OS in TCGA set. (B) ROC of 3-year OS in ICGC set.

Table S1 Selection of 568 CNV-driven genes using K-S test

geneName	pvalue	geneName	pvalue
DSCC1	2.13E-23	CDC25A	0.00395
PITPNM3	5.69E-17	MDFI	0.003993
DCST2	8.93E-15	OLFM3	0.004041
CENPL	1.10E-13	KBTBD11	0.004056
FLVCR1	1.23E-13	WDR76	0.004066
CAP2	1.79E-13	E2F2	0.0041
TK1	7.19E-12	C19orf33	0.004205
ACTN2	8.03E-12	SSUH2	0.004253
TRIP13	1.10E-11	OIP5	0.004384
DCST1	1.39E-11	FABP4	0.004394
GSTZ1	4.64E-11	LY6K	0.004406
IQGAP3	9.71E-11	BEST4	0.004495
MSH5	9.86E-11	SHCBP1	0.004532
TRIM72	3.26E-10	DCX	0.004536
GMNN	8.36E-10	MNS1	0.00472
SBSPON	1.19E-09	SLC6A9	0.00477
AURKA	1.48E-09	NAT2	0.004791
NUF2	2.29E-09	GRIK4	0.004843
SAPCD1	2.30E-09	PRAMEF8	0.005041
CENPW	3.42E-09	HPDL	0.005155

Table S1 (*continued*)

Table S1 (*continued*)

geneName	pvalue	geneName	pvalue
LHX4	4.31E-09	DTL	0.005272
SLC26A6	5.12E-09	COL24A1	0.005405
E2F1	5.75E-09	DLX1	0.005471
SLC6A2	1.65E-08	PLK4	0.005512
TPX2	2.67E-08	PRDM9	0.005513
STAR	2.95E-08	PIF1	0.005853
MT2A	2.96E-08	GJA10	0.0059
EPHA2	3.35E-08	DMRT3	0.005936
KDM8	3.36E-08	CXCL17	0.006061
EME1	6.71E-08	PPFIA4	0.006076
MT1X	7.26E-08	TFAP2A	0.006084
RCAN1	7.69E-08	GRIN2B	0.006138
TRIM45	7.79E-08	SNCG	0.006261
MTFR2	8.44E-08	MCIDAS	0.006303
CCNE1	8.45E-08	NNMT	0.006503
TCF19	9.96E-08	HCRT	0.006559
MFSD2A	1.73E-07	CD109	0.006795
UBE2T	1.82E-07	AKAP14	0.006841
CNNM1	2.19E-07	RPRM	0.006974
ADRA1A	2.20E-07	MAD2L1	0.006995

Table S1 (*continued*)

http://dx.doi.org/10.21037/atm-20-7101

Table S1 (*continued*)

geneName	pvalue	geneName	pvalue
ACSM3	3.04E-07	OPRK1	0.00701
ARHGEF39	3.04E-07	ASB16	0.007061
NKD1	3.80E-07	SULT4A1	0.007136
CDKN2A	3.99E-07	PKMYT1	0.007136
TERT	4.29E-07	NEK2	0.007276
DNAJC6	7.76E-07	SPP1	0.007557
EGR3	8.85E-07	HIST2H4A	0.007651
ECT2	9.35E-07	TDGF1	0.007689
ZNF296	9.40E-07	TNFRSF19	0.007692
HP	1.30E-06	SERPINE1	0.007697
CAPN9	2.54E-06	CSRNP1	0.00781
ZNF648	2.69E-06	GAGE12J	0.007974
SLC2A5	2.86E-06	IQCD	0.007995
DPF1	3.70E-06	POPDC3	0.008109
GLUL	3.96E-06	NTF3	0.008207
GNAO1	4.16E-06	HHIP	0.008249
CLVS1	4.21E-06	PRAMEF7	0.008655
NDC80	4.64E-06	C1orf158	0.008699
TMEM145	4.70E-06	TLX1	0.008853
TNNT2	4.91E-06	TEDDM1	0.008955
GINS1	5.17E-06	CBX2	0.009009
ORC1	5.23E-06	JPH3	0.009072
BIRC5	5.35E-06	TAC3	0.009123
LENEP	5.64E-06	EXO1	0.009224
TTK	5.98E-06	EDIL3	0.009264
MESP2	6.08E-06	CSMD2	0.009306
FBXO43	6.63E-06	PZP	0.009467
S100A1	7.68E-06	BPIFB6	0.009526
ASPM	7.99E-06	RIPPLY2	0.009682
PBK	8.56E-06	TMEM190	0.009709
WDR62	8.64E-06	OR12D2	0.00982
CCDC78	8.89E-06	TNNI3	0.010009
PTP4A3	9.15E-06	MRAP2	0.010031
FBXL16	9.25E-06	GUCY2D	0.010194

Table S1 (*continued*)

Table S1 (*continued*)

geneName	pvalue	geneName	pvalue
GRIK2	1.05E-05	PAX2	0.010196
LRAT	1.05E-05	PDE1C	0.010331
AGBL4	1.07E-05	IL11	0.010425
DNASE1L2	1.11E-05	MYBPHL	0.010445
PRR11	1.14E-05	CETP	0.010487
CR1	1.14E-05	FAM183A	0.010659
SLC25A47	1.16E-05	TRAIP	0.010926
C5orf34	1.34E-05	LHFPL4	0.010995
MAEL	1.36E-05	FUT2	0.011094
DIRAS3	1.41E-05	MSLN	0.011203
C16orf89	1.59E-05	CD5L	0.01121
RDM1	1.60E-05	ADAMTS13	0.011247
PITX2	1.70E-05	TTC39A	0.011854
TPPP2	1.80E-05	DCC	0.01196
CLSPN	1.81E-05	MELK	0.012074
AC024361.1	1.94E-05	ECM1	0.012366
STRIP2	2.06E-05	CHGA	0.01239
SLC30A2	2.07E-05	PTHLH	0.012456
C20orf144	2.24E-05	OIT3	0.012494
TRAM1L1	2.37E-05	ANGPTL7	0.012827
MXD3	2.62E-05	C21orf62	0.012866
TOP2A	2.71E-05	CNTNAP4	0.012944
LPA	3.01E-05	UBD	0.01296
KIF14	3.14E-05	CCNB2	0.013309
EPS8L3	3.30E-05	SPINK4	0.01402
MCM10	3.86E-05	EGR2	0.014201
MRO	3.93E-05	RASGEF1A	0.014375
RND3	4.25E-05	VSIG10L	0.014389
ZIC2	4.65E-05	TRIM16L	0.014399
FGF4	4.67E-05	DMBT1	0.014437
HHIPL2	4.83E-05	EZH2	0.014555
KIFC1	4.85E-05	CCDC28B	0.014596
FANCI	4.94E-05	TGM4	0.014599
SKA1	4.98E-05	SIGLEC7	0.014646

Table S1 (*continued*)

http://dx.doi.org/10.21037/atm-20-7101

Table S1 (*continued*)

geneName	pvalue	geneName	pvalue
CCNE2	5.38E-05	HGF	0.014699
NGFR	5.98E-05	CFP	0.014755
FAM83D	5.98E-05	NTSR2	0.015076
C19orf67	6.33E-05	FBP1	0.015648
ANGPTL6	6.45E-05	CLEC1B	0.015697
PRC1	6.78E-05	DNTT	0.015697
C7	6.95E-05	DRGX	0.015788
SLC44A5	7.01E-05	SEZ6L2	0.015861
UBE2C	7.10E-05	PAQR4	0.015928
TRIM71	7.18E-05	RASD2	0.015981
LRRN3	7.28E-05	ZNF676	0.016049
POU3F2	7.71E-05	PADI3	0.016076
PRAMEF4	8.06E-05	FANCB	0.016291
RGS20	8.29E-05	SPATC1L	0.016313
KIF18B	8.36E-05	COCH	0.01653
ADCY8	8.48E-05	RGS9BP	0.016607
CDCA2	0.000102	DMP1	0.016619
DLL3	0.000104	SLC6A18	0.016933
AADAT	0.000104	ADIPOQ	0.017094
CDC6	0.000106	DKK1	0.017225
ADH4	0.000108	PLAC8	0.017327
CDCA8	0.000109	PAGE4	0.017676
DEPDC1	0.000111	WDHD1	0.017731
CDKN2C	0.000116	SAPCD2	0.017842
MND1	0.00012	NUSAP1	0.017873
CHRNB2	0.000128	PCDHA1	0.018016
DYDC2	0.000139	BCAN	0.018236
MUC12	0.000144	SPHK1	0.01825
GPRIN1	0.000144	CSN3	0.018537
IGSF3	0.000149	KCNH2	0.018574
MSX1	0.00015	RTKN2	0.018576
OR2B6	0.000153	SCN1A	0.018671
LCAT	0.000165	MCOLN3	0.018741
TBC1D26	0.000166	ASF1B	0.018794

Table S1 (*continued*)

Table S1 (*continued*)

geneName	pvalue	geneName	pvalue
SP5	0.00017	TEX19	0.019257
BLM	0.000206	TNFRSF9	0.019682
TNNC1	0.000207	MAFA	0.020213
FCN3	0.000211	AQP10	0.020458
MYOM2	0.000211	KCTD8	0.020701
ILDR2	0.000212	NPY1R	0.020904
FOS	0.000213	SH3GL3	0.021176
WNT3A	0.000222	CSTL1	0.021369
IL17D	0.000236	KCNH4	0.021465
MYBL2	0.00024	DLX5	0.021492
INSRR	0.000242	TSPAN5	0.021552
GDAP1L1	0.000248	UGT2B11	0.021569
TRIM16	0.000248	CELF3	0.021576
SFN	0.00025	CXCL12	0.021793
NEFL	0.000272	DYDC1	0.022052
RAD54L	0.000276	KEL	0.022118
PLK5	0.000279	ADRA2C	0.022513
FOXD2	0.00028	RPS6KL1	0.022855
AHNAK2	0.000283	ASPG	0.022965
WDR87	0.000308	LCN2	0.022966
CYP11B2	0.000309	PDZK1IP1	0.02298
CUZD1	0.000316	UTS2R	0.023074
SYT5	0.00032	C6orf223	0.023328
CRLF1	0.000355	PSMA8	0.023559
EFNA3	0.000365	TMEM266	0.023876
GNGT1	0.000383	RBPJL	0.023947
AMH	0.000393	KCNN1	0.024105
CDRT1	0.000409	PHYHIPL	0.024326
IL1RAP	0.000409	GJC1	0.024525
FGF20	0.000426	RNF224	0.024879
FLNC	0.000426	CLLU1	0.024925
CHAF1B	0.000432	REC114	0.025168
DIO2	0.000439	BUB1B	0.025173
HIST1H3H	0.000445	TMEM26	0.025593

Table S1 (*continued*)

http://dx.doi.org/10.21037/atm-20-7101

Table S1 (*continued*)

geneName	pvalue	geneName	pvalue
PRAMEF15	0.000447	CCNO	0.026383
CERS1	0.000453	SPTA1	0.026717
SPC25	0.000472	KIRREL2	0.026986
RXFP4	0.000495	ARHGAP11A	0.02749
CENPU	0.000499	FOXN4	0.028642
HIST1H2AM	0.0005	DDIT4L	0.028701
CDC20	0.000542	CLEC4M	0.029083
CPLX2	0.00058	MT1M	0.029124
MAPT	0.000582	GABRG2	0.029433
TMEM61	0.000583	SLC10A4	0.029756
PRR19	0.000585	ADAM18	0.0299
TICRR	0.000611	LHX8	0.029963
BHLHA9	0.000613	TMEM130	0.029969
PRAMEF11	0.000676	HTR3B	0.030061
MLANA	0.000711	TSLP	0.030157
RHBG	0.000722	FXYD3	0.030387
MT1G	0.000731	GFY	0.030621
MYH7B	0.000743	TWIST2	0.030742
AP1M2	0.000771	ARX	0.031148
NRCAM	0.000817	CDCA3	0.031234
MT1E	0.000818	TMEM178B	0.031289
FAM24B	0.000826	HIST1H2BL	0.031348
LEF1	0.000833	HELLS	0.031363
KIF2C	0.000862	DLX2	0.031769
SYN3	0.000866	SOHLH1	0.031942
STIL	0.000877	VASH2	0.03258
SPARCL1	0.000891	HIST1H4H	0.033016
NKPD1	0.000891	CDH22	0.03303
DMKN	0.000898	MCCD1	0.033099
CPEB3	0.000902	GABRA2	0.033139
GPM6A	0.000935	PCSK1N	0.03334
MYH13	0.000943	AKR1B15	0.033566
SKA3	0.000949	HJURP	0.033946
ZIC5	0.000957	C1orf61	0.034063

Table S1 (*continued*)

Table S1 (*continued*)

geneName	pvalue	geneName	pvalue
RRM2	0.000964	PXDNL	0.034163
CCDC185	0.000975	SCNN1G	0.034168
CCNA2	0.000989	OLFML2B	0.034288
PTGS2	0.001067	COMP	0.034717
LILRA5	0.001077	RBFOX1	0.034837
HIST1H4E	0.001079	ERC2	0.034892
CD24	0.001083	CYP17A1	0.035361
DNAH3	0.001094	COL7A1	0.035713
EXTL1	0.001175	MSH4	0.035842
DCAF12L2	0.001198	SCUBE1	0.036087
CACNA1B	0.0012	TUBAL3	0.036221
CELSR3	0.001218	THY1	0.036838
FGF19	0.001243	GPR19	0.036988
FPR2	0.00126	KIF11	0.037047
CDH24	0.001268	CLLU1OS	0.03705
PRRX1	0.001304	BAIAP2L2	0.0376
BCO2	0.00134	TRIM17	0.038129
RIMS2	0.00136	DRD4	0.038255
RGSL1	0.001462	PTGFR	0.038389
NKX1-2	0.001503	FTHL17	0.038397
RNF157	0.001511	FERMT1	0.038808
UNC5D	0.001567	HIST1H2AI	0.038906
TCF24	0.001626	IGFL2	0.039086
CRYBA4	0.001674	NCAPG	0.039138
CDK1	0.001683	NKX3-2	0.039567
ATP4A	0.001695	PHEX	0.039696
PRAMEF2	0.001698	TACC3	0.039811
PADI4	0.001727	HNRNPCL3	0.039879
MAMSTR	0.001798	CRH	0.039978
BDKRB1	0.0018	EDARADD	0.040143
C1QTNF3	0.001806	CTNND2	0.040363
SLC7A10	0.001818	KRT12	0.0405
AKR1B10	0.00182	GPRC6A	0.040554
CDT1	0.001827	SLC6A7	0.04062

Table S1 (*continued*)

http://dx.doi.org/10.21037/atm-20-7101

Table S1 (*continued*)

geneName	pvalue	geneName	pvalue
SORCS3	0.001905	CAGE1	0.04086
CENPF	0.001933	AMBN	0.041144
STMN1	0.001953	ZNF729	0.041209
CAPN8	0.00207	STOML3	0.041328
KCNA1	0.002138	PPP1R1B	0.041339
FANCD2	0.002144	FGF8	0.041766
KIF23	0.002221	COLCA2	0.041779
STK39	0.002222	SLCO1B3	0.041793
TRIM63	0.002265	EEF1A2	0.042188
UPK1A	0.002272	SFRP4	0.042195
GFAP	0.002313	MAGEB1	0.042222
ORC6	0.002334	CST2	0.042229
ANO2	0.002354	ATP6V0D2	0.042445
RNF151	0.002371	MAGEB16	0.042488
CDH12	0.0024	CCDC155	0.043075
RBM24	0.002474	SLCO1C1	0.043108
CDC7	0.002552	PLVAP	0.043201
GJA3	0.002599	PYDC1	0.04323
ZIC4	0.002637	PAGE1	0.043298

Table S1 (*continued*)

Table S1 (*continued*)

geneName	pvalue	geneName	pvalue
DMRTB1	0.002856	PPP1R14C	0.04356
ZWINT	0.002924	RHBDL3	0.043841
CRIP3	0.002929	ARID3A	0.04389
EGF	0.002939	KIF19	0.044518
PRR36	0.002949	MT1B	0.044538
DCAF4L1	0.00295	SDR16C5	0.044821
CPA6	0.003043	SLC30A8	0.044885
SLC5A11	0.003067	SYT2	0.045084
FHAD1	0.003071	NPHS1	0.045219
CCDC13	0.003075	PSAPL1	0.045438
CCL25	0.003123	HIST3H2BB	0.045642
MT1F	0.003136	KCNJ6	0.045655
RNFT2	0.003155	CCBE1	0.04567
MOS	0.003209	C1QL1	0.046086
FITM1	0.00321	FBXW10	0.046112
KCNJ5	0.003402	QRFPR	0.046277
TSPO2	0.003711	SLC12A1	0.047138
STMND1	0.003717	TMEM155	0.047734
GLP2R	0.00375	KPNA7	0.047942
TLL2	0.003847	IGFALS	0.048685
ARFGEF3	0.003936	ESR1	0.049811

Table S2 Wilcoxon rank-sum results

gene	conMean	treatMean	logFC	pValue	fdr
PITX2	0.014548	0.114562	2.977216	0.001589	0.001679
RBM24	0.271788	1.516507	2.4802	1.50E-16	2.69E-16
UBE2C	0.020923	8.417896	8.652209	1.02E-21	2.67E-21
OLFML2B	0.098781	2.417167	4.612942	2.35E-27	2.15E-26
ASB16	0.272465	0.641053	1.234375	4.44E-23	1.35E-22
CCNE1	0.010827	2.292889	7.726339	4.44E-25	1.79E-24
SYT5	0.042541	0.118585	1.478985	1.22E-08	1.56E-08
GINS1	0.178909	1.851394	3.371313	3.95E-25	1.61E-24
CDKN2C	0.52573	4.201133	2.998385	3.01E-28	5.90E-27
AMH	0.10028	0.326345	1.702369	6.50E-08	7.97E-08

Table S2 (*continued*)

http://dx.doi.org/10.21037/atm-20-7101

Table S2 (*continued*)

gene	conMean	treatMean	logFC	pValue	fdr
ASPG	9.642475	3.656328	-1.39901	5.59E-17	1.01E-16
GSTZ1	9.615745	4.162492	-1.20795	2.84E-21	6.93E-21
AKAP14	0.026238	0.09208	1.811205	8.14E-08	9.92E-08
TOP2A	0.034263	5.126495	7.225183	8.50E-25	3.21E-24
SLC2A5	0.183757	1.256253	2.773257	2.44E-12	3.58E-12
TRIM45	0.372207	1.097112	1.559535	1.92E-25	8.50E-25
EXO1	0.100456	1.082293	3.42945	2.45E-27	2.20E-26
CDCA8	0.065069	2.696066	5.372748	5.71E-28	7.59E-27
DCST2	0.389916	0.810804	1.05619	4.30E-22	1.18E-21
FBXW10	0.233479	0.559304	1.26034	2.24E-09	2.92E-09
TAC3	0.102345	0.239951	1.229298	3.07E-07	3.70E-07
TRIP13	0.18207	1.469062	3.012331	1.74E-28	5.90E-27
NCAPG	0.326387	1.929736	2.563746	7.79E-28	9.17E-27
MAEL	0.010256	0.772496	6.234976	4.34E-08	5.40E-08
PIF1	0.187355	0.781849	2.061115	1.78E-24	6.26E-24
DIO2	0.020267	0.368483	4.184362	2.15E-08	2.73E-08
MAPT	0.124496	0.56075	2.171254	3.34E-25	1.40E-24
TEX19	0.002376	0.113794	5.581928	4.87E-22	1.32E-21
ASPM	0.198747	1.684712	3.0835	1.98E-27	1.89E-26
GJA3	0.010813	0.046377	2.100646	0.003454	0.00364
GRIK4	0.056012	0.16358	1.546178	3.46E-06	3.96E-06
RASD2	0.037473	0.416474	3.47429	8.60E-27	6.33E-26
MYH7B	0.228411	0.768049	1.749567	1.39E-06	1.62E-06
LY6K	0.081836	0.363969	2.153011	0.000102	0.000112
LRRN3	0.446977	0.182201	-1.29467	5.18E-21	1.21E-20
EGF	0.006458	0.207539	5.006125	1.10E-05	1.24E-05
AKR1B15	0.221349	1.645843	2.894435	6.01E-09	7.72E-09
MT2A	890.5815	185.0133	-2.26712	5.34E-24	1.75E-23
HIST1H2AM	0.08472	0.534141	2.656443	2.63E-20	5.72E-20
ERC2	0.0172	0.06826	1.988584	4.37E-08	5.43E-08
HP	2316.924	656.6056	-1.81911	1.99E-21	5.00E-21
DLX1	0.008581	0.132206	3.94549	6.43E-12	9.10E-12
ZWINT	0.308956	5.3194	4.105791	6.61E-26	3.28E-25
CCNB2	0.070746	2.912539	5.363481	9.52E-28	1.06E-26

Table S2 (*continued*)

http://dx.doi.org/10.21037/atm-20-7101

Table S2 (*continued*)

gene	conMean	treatMean	logFC	pValue	fdr
IL1RAP	5.861583	2.910083	-1.01023	3.91E-16	6.80E-16
CAPN9	0.077565	0.240577	1.633023	2.09E-11	2.92E-11
BEST4	0.09446	0.326104	1.787566	7.28E-18	1.44E-17
CAPN8	0.21514	0.519395	1.271556	2.89E-12	4.20E-12
PTHLH	0.182931	0.947187	2.372349	2.49E-05	2.79E-05
KCNJ6	0.021697	0.062015	1.515103	7.02E-07	8.31E-07
ZIC2	0.001063	1.393709	10.35664	2.11E-20	4.66E-20
ECM1	10.22191	2.548523	-2.00393	4.24E-27	3.43E-26
CPEB3	3.781311	1.554166	-1.28275	9.00E-24	2.83E-23
NEK2	0.053849	2.17434	5.335514	2.89E-27	2.53E-26
LEF1	0.344472	1.357595	1.978592	3.47E-17	6.45E-17
VASH2	0.115679	0.438832	1.923535	1.02E-17	2.00E-17
ZNF648	0.143655	0.445529	1.632907	0.000105	0.000115
CD109	0.209051	1.246832	2.576337	4.03E-16	6.98E-16
FANCB	0.061653	0.216295	1.810756	1.30E-21	3.38E-21
C1QTNF3	0.002659	4.887746	10.84402	1.43E-12	2.13E-12
RTKN2	0.075386	0.27477	1.865848	5.34E-19	1.12E-18
TCF19	0.169691	3.861876	4.508322	1.53E-26	1.02E-25
CDK1	0.225714	3.206727	3.828531	7.09E-28	8.85E-27
MT1G	768.8577	195.446	-1.97595	2.78E-33	1.14E-30
PRR19	0.327069	0.79076	1.273644	9.25E-11	1.25E-10
GJC1	0.113938	0.576731	2.339646	4.53E-26	2.39E-25
HHIP	1.177773	0.375211	-1.65029	3.63E-26	1.96E-25
RNFT2	0.0626	0.366348	2.548969	1.10E-21	2.88E-21
TMEM61	0.060691	0.359659	2.567074	0.000139	0.000151
CSMD2	0.041075	0.150221	1.870734	2.21E-16	3.87E-16
LENEP	0.008534	0.065722	2.945029	5.23E-12	7.45E-12
ACSM3	7.327693	3.526277	-1.05521	3.51E-21	8.41E-21
KCNN1	0.03743	0.125001	1.739655	4.44E-12	6.38E-12
MT1X	245.77	51.49964	-2.25467	4.61E-24	1.52E-23
CRIP3	0.916265	3.012009	1.716889	7.54E-22	1.99E-21
ANO2	0.089137	0.240415	1.431432	3.11E-17	5.80E-17
EFNA3	0.29915	1.411479	2.238269	3.03E-22	8.44E-22
NNMT	298.152	77.54793	-1.94289	4.24E-19	8.96E-19

Table S2 (*continued*)

http://dx.doi.org/10.21037/atm-20-7101

Table S2 (*continued*)

gene	conMean	treatMean	logFC	pValue	fdr
TRIM17	0.081177	0.262592	1.693686	4.65E-11	6.39E-11
RHBDL3	0.02016	0.339687	4.074611	3.97E-09	5.12E-09
PRAMEF4	0.028001	0.452487	4.014341	0.000327	0.000349
RDM1	0.074991	0.491499	2.712394	1.70E-23	5.27E-23
ARID3A	0.240386	2.03473	3.08141	2.49E-18	5.09E-18
CUZD1	0.071638	0.430341	2.586677	7.37E-16	1.25E-15
NPHS1	0.021542	0.087244	2.01793	2.15E-07	2.61E-07
CLSPN	0.078489	0.406259	2.371843	4.73E-22	1.29E-21
PZP	9.670536	0.911966	-3.40654	2.27E-21	5.61E-21
GPM6A	1.842934	0.590992	-1.64079	6.42E-26	3.23E-25
TTK	0.172461	1.17707	2.770857	3.81E-27	3.27E-26
CHRNB2	0.017393	0.054507	1.647908	1.44E-11	2.03E-11
CDC20	0.00324	6.542496	10.97957	4.73E-21	1.12E-20
CD24	1.354992	28.50801	4.395012	1.35E-10	1.81E-10
CELSR3	0.099311	0.641211	2.690779	2.23E-27	2.09E-26
TNNC1	0.096207	1.799563	4.225359	1.41E-08	1.80E-08
CLEC4M	6.048182	0.865239	-2.80533	2.57E-30	3.53E-28
EZH2	0.755838	2.713735	1.84413	4.59E-28	7.20E-27
CENPF	0.082785	1.787432	4.432372	4.06E-28	6.97E-27
CDCA3	0.295806	1.512753	2.354454	1.84E-28	5.90E-27
BHLHA9	0.098621	0.021638	-2.18835	1.28E-14	2.03E-14
IGFL2	0.072625	0.150249	1.048824	3.83E-06	4.36E-06
FANCD2	0.360865	1.132547	1.650041	2.65E-25	1.14E-24
PLVAP	0.113872	25.64005	7.814837	2.74E-29	1.61E-27
CDKN2A	0.266405	3.129789	3.554375	1.87E-25	8.37E-25
CDC7	0.481091	1.407633	1.548889	3.18E-24	1.06E-23
C19orf33	0.220282	2.415676	3.455004	0.015246	0.015902
CLEC1B	5.102987	0.840832	-2.60145	1.77E-29	1.38E-27
RGS9BP	0.002441	0.033655	3.785376	2.92E-15	4.74E-15
GPRIN1	0.081052	0.807134	3.315885	6.69E-22	1.78E-21
RCAN1	16.90906	5.347118	-1.66096	5.46E-24	1.77E-23
KIF23	0.274751	1.206116	2.134175	8.04E-26	3.85E-25
TFAP2A	0.117091	0.363664	1.634983	5.48E-08	6.76E-08
UBD	2.782608	44.27614	3.99202	7.95E-16	1.34E-15

Table S2 (*continued*)

http://dx.doi.org/10.21037/atm-20-7101

Table S2 (*continued*)

gene	conMean	treatMean	logFC	pValue	fdr
SPHK1	0.539595	3.801175	2.816497	8.75E-05	9.59E-05
DIRAS3	2.731089	0.872223	-1.64671	2.04E-22	5.82E-22
NPY1R	2.265229	1.038644	-1.12496	1.64E-22	4.72E-22
DNAJC6	0.139482	0.729746	2.387314	1.97E-21	4.98E-21
LILRA5	1.921618	0.819385	-1.22971	1.99E-16	3.51E-16
RGS20	0.021651	0.095341	2.138638	1.57E-12	2.32E-12
FOXN4	0.166522	0.895339	2.426722	0.001048	0.001113
ORC1	0.122423	1.099514	3.166917	3.62E-26	1.96E-25
IQCD	0.081773	0.514416	2.653235	5.48E-21	1.28E-20
CSRNP1	18.98319	5.898589	-1.68628	1.32E-22	3.87E-22
ZIC5	0.017259	0.4382	4.666162	1.26E-20	2.85E-20
E2F1	0.042362	5.598552	7.046154	2.62E-23	8.06E-23
KIF14	0.121035	0.683004	2.496475	5.88E-26	3.03E-25
FHAD1	0.067709	0.184671	1.447543	1.15E-13	1.80E-13
FABP4	0.329511	6.321003	4.261756	3.49E-10	4.64E-10
EPS8L3	0.023725	3.821074	7.331422	1.93E-16	3.41E-16
DLX5	0.009925	0.220976	4.476685	3.86E-14	6.08E-14
CENPW	0.257762	5.643826	4.452562	8.05E-28	9.21E-27
CXCL12	21.69151	5.292733	-2.03505	1.08E-24	3.95E-24
TSLP	1.997337	0.84791	-1.2361	1.59E-20	3.55E-20
DLX2	0.003833	0.066805	4.123584	4.97E-07	5.91E-07
DSCC1	0.321833	1.383548	2.103988	2.52E-28	5.90E-27
ZNF676	0.061486	0.294662	2.260724	0.041389	0.042631
FPR2	0.61853	0.223964	-1.46558	6.95E-16	1.18E-15
SKA1	0.010313	1.430236	7.115643	2.87E-28	5.90E-27
TMEM145	0.063502	0.428228	2.75351	1.00E-19	2.15E-19
OIP5	0.321466	1.607628	2.322194	6.58E-24	2.10E-23
NRCAM	0.150986	1.097458	2.86168	0.000107	0.000116
NGFR	5.139406	2.249034	-1.1923	8.26E-18	1.63E-17
SLC6A9	0.262446	1.005621	1.937993	4.04E-15	6.50E-15
SERPINE1	82.85032	24.82337	-1.73881	1.78E-11	2.49E-11
S100A1	0.628054	2.948731	2.231134	7.33E-12	1.03E-11
MESP2	0.011331	0.579584	5.676711	2.87E-25	1.22E-24
HPDL	0.03376	0.272197	3.011271	1.75E-16	3.11E-16

Table S2 (*continued*)

http://dx.doi.org/10.21037/atm-20-7101

Table S2 (*continued*)

gene	conMean	treatMean	logFC	pValue	fdr
DEPDC1	0.09103	0.864537	3.247522	9.69E-27	6.88E-26
SLC44A5	0.205607	0.840544	2.031431	0.005659	0.005933
FAM183A	0.115749	0.390169	1.753102	4.93E-05	5.45E-05
DRD4	0.200916	0.849174	2.079467	1.06E-14	1.68E-14
CDCA2	0.088104	0.628759	2.835232	5.95E-25	2.38E-24
KIF11	0.267462	1.535667	2.52146	9.53E-26	4.51E-25
TDGF1	0.285436	1.506102	2.39958	3.96E-05	4.39E-05
E2F2	0.047971	0.486718	3.342856	3.43E-25	1.41E-24
MNS1	0.358531	1.890552	2.398639	5.34E-14	8.37E-14
UBE2T	0.071707	5.830857	6.345454	2.78E-28	5.90E-27
BAIAP2L2	0.14049	4.46118	4.988886	1.41E-18	2.88E-18
SFRP4	0.183978	1.272288	2.789819	3.73E-14	5.90E-14
COMP	0.116212	0.916884	2.97998	2.43E-05	2.73E-05
AHNAK2	0.09837	0.371052	1.915331	0.031477	0.032502
FLVCR1	0.472587	2.082588	2.139724	2.14E-28	5.90E-27
CAP2	0.431424	5.887226	3.77041	6.49E-25	2.52E-24
TRAIP	0.196064	1.132998	2.53075	4.53E-28	7.20E-27
PLK4	0.214192	0.728801	1.766623	3.38E-25	1.41E-24
C5orf34	0.165382	0.636554	1.944484	5.70E-28	7.59E-27
FAM83D	0.246566	3.68129	3.900168	8.12E-25	3.10E-24
GPR19	0.046778	0.241464	2.367916	2.04E-21	5.07E-21
KIFC1	0.017211	3.436522	7.641495	1.42E-26	9.62E-26
DDIT4L	0.056481	0.356477	2.657972	3.54E-05	3.94E-05
MELK	0.130939	1.979966	3.918504	4.98E-28	7.33E-27
TK1	0.255001	12.09733	5.56804	1.02E-24	3.75E-24
OIT3	21.30992	3.413868	-2.64205	7.53E-27	5.72E-26
ADAMTS13	4.936471	1.602208	-1.62342	5.33E-28	7.57E-27
HIST1H4H	0.102978	2.183295	4.406106	4.39E-18	8.86E-18
HIST2H4A	0.005339	0.11679	4.451322	1.74E-16	3.11E-16
RNF151	0.02758	0.079913	1.534797	2.18E-08	2.76E-08
MAMSTR	0.144142	0.578966	2.005994	2.02E-24	6.93E-24
ATP6V0D2	0.088933	0.407899	2.197414	3.37E-12	4.88E-12
FLNC	0.023016	1.835133	6.3171	4.03E-06	4.58E-06
SLC10A4	0.042695	0.095917	1.167719	9.75E-07	1.14E-06

Table S2 (*continued*)

http://dx.doi.org/10.21037/atm-20-7101

Table S2 (*continued*)

gene	conMean	treatMean	logFC	pValue	fdr
MT1M	64.23517	19.09479	-1.75018	7.24E-32	1.49E-29
KBTBD11	2.319411	0.872284	-1.41089	1.81E-20	4.01E-20
BIRC5	0.007384	4.530991	9.261281	8.83E-25	3.31E-24
ARHGAP11A	0.210571	1.226055	2.541644	1.02E-25	4.76E-25
C7	27.9382	9.474361	-1.56014	1.04E-17	2.01E-17
CDC6	0.180911	2.207426	3.609015	4.60E-27	3.64E-26
IGSF3	0.180897	1.730222	3.257714	8.68E-16	1.46E-15
NKX3-2	0.011058	0.087754	2.988365	5.92E-11	8.08E-11
RNF224	0.03374	0.098956	1.552336	6.20E-13	9.29E-13
EGR3	1.093695	0.526587	-1.05447	3.08E-13	4.66E-13
TNNI3	0.062335	0.26302	2.077055	0.000122	0.000133
SFN	0.079416	15.3129	7.591096	4.00E-11	5.54E-11
WDHD1	0.210959	0.753923	1.837454	1.33E-24	4.72E-24
TNFRSF9	0.107299	0.397206	1.888249	0.000211	0.000226
NDC80	0.276735	2.008922	2.859846	2.26E-28	5.90E-27
FBXO43	0.05482	0.457386	3.060631	2.11E-28	5.90E-27
ILDR2	0.084455	0.49759	2.558699	2.64E-06	3.04E-06
TLX1	0.211824	1.096112	2.371459	4.74E-16	8.17E-16
PAQR4	0.56139	2.260698	2.009693	7.33E-21	1.69E-20
PXDNL	0.055161	0.145364	1.397949	3.80E-22	1.05E-21
MDFI	0.22268	1.244204	2.482176	2.70E-12	3.95E-12
STMN1	1.142131	9.346068	3.032631	1.90E-27	1.86E-26
SLC5A11	0.169086	0.731083	2.11228	6.36E-06	7.18E-06
SYN3	0.051265	0.205736	2.004735	2.28E-16	3.97E-16
CCDC28B	0.581803	1.944585	1.74086	2.63E-24	8.89E-24
VSIG10L	0.526408	2.06782	1.973858	2.49E-11	3.46E-11
HIST3H2BB	0.022147	0.129457	2.547272	2.76E-06	3.17E-06
PBK	0.091128	2.213856	4.602525	1.71E-26	1.10E-25
CPA6	0.041481	0.323293	2.96232	9.17E-16	1.54E-15
NUSAP1	0.396866	5.824548	3.875421	6.10E-25	2.42E-24
CCDC78	0.167954	0.384122	1.193495	5.43E-18	1.09E-17
CCDC13	0.148696	0.562449	1.919357	0.004319	0.00454
CCBE1	1.440845	0.367424	-1.9714	2.54E-26	1.54E-25
PPFIA4	0.073195	0.176866	1.27283	1.52E-08	1.93E-08

Table S2 (*continued*)

http://dx.doi.org/10.21037/atm-20-7101

Table S2 (*continued*)

gene	conMean	treatMean	logFC	pValue	fdr
MCM10	0.090704	0.698713	2.945459	7.78E-25	3.00E-24
FBXL16	0.083861	0.657118	2.970079	0.00014	0.000151
ESR1	2.41054	0.964269	-1.32185	1.04E-22	3.14E-22
CDT1	0.040553	2.73458	6.07535	3.53E-26	1.96E-25
BUB1B	0.180706	1.390193	2.943566	2.75E-26	1.64E-25
TACC3	0.725177	3.666278	2.337912	7.44E-26	3.61E-25
TRIM72	0.007662	0.062565	3.02949	8.27E-05	9.09E-05
STIL	0.224356	0.793306	1.82209	2.79E-26	1.64E-25
HELLS	0.276158	0.910532	1.721215	3.84E-24	1.28E-23
SEZ6L2	0.093158	6.01878	6.013643	1.07E-09	1.41E-09
FCN3	35.31802	3.003514	-3.55568	1.09E-27	1.15E-26
HIST1H2AI	0.013629	0.381497	4.806957	4.56E-17	8.35E-17
NAT2	18.19904	3.44412	-2.40165	1.20E-25	5.56E-25
CBX2	0.154027	0.885284	2.522956	2.15E-24	7.32E-24
RIMS2	0.031793	0.104287	1.713792	0.010853	0.011349
DNASE1L2	0.274263	0.599029	1.127067	5.26E-18	1.06E-17
TSPO2	0.122683	0.656604	2.420088	1.83E-15	3.02E-15
HHIPL2	0.000103	1.102345	13.38249	8.41E-11	1.14E-10
CCNA2	0.193462	3.676458	4.248198	3.92E-27	3.29E-26
CENPL	0.253294	1.008503	1.993332	3.31E-28	6.20E-27
PRRX1	0.218561	0.551466	1.335236	5.61E-18	1.12E-17
TMEM26	0.698624	0.322499	-1.11522	2.63E-17	4.97E-17
CCNO	0.07178	0.779377	3.440668	2.40E-08	3.02E-08
IGFALS	40.76762	6.466022	-2.65647	1.19E-24	4.29E-24
CETP	9.589907	2.364191	-2.02017	2.59E-22	7.27E-22
WDR76	0.216672	1.370734	2.661362	1.89E-24	6.61E-24
ZNF296	0.189435	0.735477	1.956976	8.35E-18	1.64E-17
MRO	1.502704	0.60467	-1.31334	6.67E-20	1.44E-19
SHCBP1	0.141682	0.886207	2.644989	1.24E-25	5.66E-25
COCH	0.136112	1.37987	3.341665	2.91E-13	4.43E-13
SCUBE1	0.336968	1.41675	2.071901	0.023346	0.024228
TRIM16	0.328119	2.510745	2.935824	4.96E-23	1.50E-22
AURKA	0.227714	4.679457	4.361044	4.05E-28	6.97E-27
GDAP1L1	0.017947	0.0535	1.575766	3.14E-07	3.78E-07

Table S2 (*continued*)

http://dx.doi.org/10.21037/atm-20-7101

Table S2 (*continued*)

gene	conMean	treatMean	logFC	pValue	fdr
SKA3	0.056326	1.19191	4.403332	1.75E-27	1.80E-26
GUCY2D	0.093035	0.23911	1.361827	2.84E-05	3.17E-05
HIST1H2BL	0.01632	0.147016	3.171301	1.71E-12	2.51E-12
FBP1	247.6309	72.27755	-1.77657	3.04E-21	7.38E-21
RND3	22.85251	6.236732	-1.87349	7.42E-21	1.70E-20
KIF2C	0.05943	2.272652	5.257051	2.11E-28	5.90E-27
CSTL1	0.036414	0.116261	1.674811	0.000161	0.000173
STK39	0.525711	2.424077	2.205094	6.27E-14	9.78E-14
EGR2	2.816894	1.12102	-1.32929	1.71E-15	2.83E-15
POU3F2	0.002288	0.04978	4.443377	1.04E-10	1.40E-10
IL17D	0.160479	0.702411	2.129928	9.47E-13	1.41E-12
CNNM1	0.116347	0.816317	2.810697	0.022098	0.022991
QRFPR	0.03752	0.080415	1.099797	5.31E-06	6.01E-06
FOS	76.18826	11.97356	-2.66972	2.83E-21	6.93E-21
SLCO1B3	11.11193	4.362336	-1.34894	6.72E-16	1.15E-15
CDH24	0.279609	0.975768	1.803129	5.47E-27	4.25E-26
TLL2	0.04001	0.080026	1.000125	6.32E-10	8.34E-10
C20orf144	0.04771	0.142387	1.577447	8.94E-19	1.85E-18
TSPAN5	0.317436	1.182463	1.897259	2.15E-07	2.61E-07
COL7A1	0.413818	1.670616	2.013312	7.40E-19	1.54E-18
MAD2L1	0.524937	1.694201	1.690388	2.45E-25	1.06E-24
CFP	5.970888	1.515895	-1.97778	1.56E-26	1.02E-25
UGT2B11	0.100089	8.436219	6.397247	3.98E-11	5.51E-11
TPPP2	2.035941	0.942665	-1.11088	2.39E-20	5.23E-20
FOXD2	0.11756	0.536044	2.188959	2.36E-19	5.03E-19
DCC	0.014385	0.07741	2.427953	0.00034	0.000362
CRLF1	0.098063	0.715007	2.866175	1.02E-15	1.69E-15
PHEX	0.0582	0.183373	1.655698	5.04E-11	6.89E-11
ACTN2	0.005554	1.393166	7.970571	3.51E-08	4.40E-08
CD5L	19.14209	3.463174	-2.46658	5.66E-22	1.51E-21
TGM4	0.015961	0.038164	1.257625	2.15E-09	2.80E-09
FANCI	0.880719	2.13209	1.275514	2.32E-26	1.43E-25
UPK1A	0.074715	0.347748	2.218562	4.11E-08	5.13E-08
BCO2	5.329012	1.760738	-1.59769	6.26E-25	2.46E-24

Table S2 (*continued*)

http://dx.doi.org/10.21037/atm-20-7101

Table S2 (*continued*)

gene	conMean	treatMean	logFC	pValue	fdr
PRR11	0.109931	1.23872	3.494186	9.82E-24	3.06E-23
GMNN	0.548413	10.94823	4.319291	2.69E-28	5.90E-27
PRC1	0.329856	2.960605	3.165984	2.42E-28	5.90E-27
BCAN	0.173971	0.702088	2.012807	2.02E-21	5.04E-21
LCAT	45.88514	9.957143	-2.20422	1.77E-26	1.10E-25
SPARCL1	0.066112	26.77415	8.66171	2.74E-15	4.46E-15
WDR62	0.2259	0.849885	1.911585	1.21E-26	8.34E-26
KCNH4	0.053671	0.175715	1.711018	9.59E-15	1.53E-14
KIF18B	0.040063	1.188678	4.890942	7.58E-28	9.17E-27
KIF19	0.501837	0.24003	-1.064	6.92E-21	1.60E-20
AKR1B10	6.167714	208.7615	5.080976	1.15E-09	1.51E-09
TTC39A	0.248972	1.297075	2.381207	1.76E-17	3.37E-17
GLP2R	0.260314	0.114166	-1.18912	9.50E-19	1.96E-18
TERT	0	0.939635	Inf	9.15E-13	1.37E-12
LRAT	1.474371	0.481465	-1.6146	9.54E-25	3.54E-24
FERMT1	0.141558	0.863249	2.608387	1.75E-06	2.02E-06
CCNE2	0.45629	0.954661	1.065039	1.61E-21	4.13E-21
FITM1	3.128096	1.3153	-1.24989	1.39E-20	3.11E-20
SLC7A10	0.275615	0.984645	1.83695	7.61E-07	8.98E-07
HIST1H3H	0.170112	2.014558	3.565903	2.68E-15	4.39E-15
MT1F	60.20013	10.57133	-2.50961	9.52E-27	6.88E-26
DYDC2	0.059835	0.671964	3.489331	4.21E-11	5.79E-11
ADRA2C	0.00933	1.935463	7.696562	4.71E-12	6.74E-12
HJURP	0.133453	1.88652	3.82132	6.71E-28	8.64E-27
ANGPTL6	7.103794	2.625283	-1.43612	3.51E-29	1.81E-27
NKPD1	0.018546	0.044101	1.2497	1.74E-09	2.28E-09
PKMYT1	0.355972	1.437709	2.013935	4.72E-28	7.20E-27
ECT2	0.400966	2.222478	2.470618	1.23E-24	4.41E-24
TRIM16L	0.320882	4.18306	3.704443	2.21E-13	3.39E-13
BDKRB1	0.164212	0.39557	1.268376	5.68E-15	9.10E-15
SLC6A7	0.037054	0.094638	1.352795	4.36E-09	5.62E-09
CAGE1	0.046971	0.104224	1.149845	8.51E-07	9.96E-07
IQGAP3	0.230242	2.246248	3.286294	3.67E-26	1.96E-25
CCDC155	0.038716	0.093729	1.275566	0.001363	0.001443

Table S2 (*continued*)

http://dx.doi.org/10.21037/atm-20-7101

Table S2 (*continued*)

gene	conMean	treatMean	logFC	pValue	fdr
RAD54L	0.267144	0.97473	1.867387	1.82E-25	8.25E-25
CNTNAP4	0.076567	0.364232	2.250066	0.028635	0.029642
BLM	0.164885	0.661975	2.005317	2.11E-22	6.00E-22
MND1	0.241854	1.541771	2.672377	7.63E-27	5.72E-26
TMEM155	0.039806	0.098161	1.302166	8.40E-07	9.86E-07
NUF2	0.184491	1.831442	3.311358	1.74E-28	5.90E-27
ASF1B	0.093824	3.079952	5.036814	3.14E-26	1.82E-25
OR2B6	0.080812	0.330107	2.030299	2.38E-15	3.91E-15
HIST1H4E	0.218822	0.737377	1.752645	2.46E-17	4.67E-17
MFSD2A	24.37192	6.829485	-1.83537	3.44E-18	6.97E-18
MRAP2	0.015493	1.400257	6.497913	2.86E-17	5.37E-17
SLC26A6	0.739488	3.339663	2.175104	1.25E-29	1.29E-27
DTL	0.120552	1.880986	3.963764	3.35E-26	1.90E-25
TRAM1L1	0.044896	0.433655	3.271898	2.68E-08	3.36E-08
AADAT	8.699237	2.427923	-1.84117	7.44E-26	3.61E-25
CHAF1B	0.27502	1.271058	2.208421	6.03E-24	1.94E-23
CDC25A	0.175584	0.919295	2.388369	3.35E-21	8.07E-21
SPC25	0.176919	1.54862	3.129826	4.19E-27	3.43E-26
NTF3	1.871852	0.482896	-1.95468	1.89E-27	1.86E-26
EDARADD	0.204389	0.553282	1.436698	5.70E-05	6.28E-05
DPF1	0.0407	0.085679	1.073912	1.75E-16	3.11E-16
CDRT1	0.115322	0.248892	1.109848	3.33E-09	4.31E-09
MYBL2	0	5.756456	Inf	6.38E-19	1.34E-18
RNF157	0.292232	1.876634	2.682962	4.31E-17	7.92E-17
DYDC1	0.024611	0.082503	1.745123	5.23E-08	6.47E-08
C21orf62	0.158121	0.056999	-1.47201	1.36E-20	3.07E-20
ADH4	511.6001	124.0821	-2.04372	5.64E-22	1.51E-21
PADI4	0.268474	0.125414	-1.09809	3.70E-13	5.59E-13
GNAO1	2.292639	0.896881	-1.35402	4.13E-17	7.63E-17
RRM2	0.162038	4.357017	4.748936	4.70E-26	2.45E-25
COL24A1	0.089275	0.275864	1.627633	5.40E-12	7.68E-12
DCX	0.003992	0.029109	2.86612	7.84E-07	9.22E-07
PRDM9	0.006805	0.045411	2.738484	4.37E-10	5.79E-10
LPA	9.504567	3.082742	-1.62441	2.38E-22	6.72E-22

Table S2 (*continued*)

http://dx.doi.org/10.21037/atm-20-7101

Table S2 (*continued*)

gene	conMean	treatMean	logFC	pValue	fdr
KCNJ5	0.167049	0.755853	2.177838	1.42E-17	2.74E-17
CST2	0.058397	0.519969	3.154464	3.34E-07	3.99E-07
FUT2	0.066396	0.673373	3.342238	4.89E-16	8.39E-16
EPHA2	10.73155	4.072777	-1.39777	2.54E-13	3.88E-13
SLC25A47	162.0699	31.83659	-2.34786	2.37E-25	1.04E-24
RPS6KL1	0.242699	0.818577	1.753951	5.05E-21	1.19E-20
THY1	0.530496	6.199475	3.546732	1.05E-26	7.32E-26
RXFP4	0.001164	0.045579	5.291249	5.97E-07	7.09E-07
KPNA7	0.146004	0.689252	2.239022	1.91E-13	2.95E-13
PTP4A3	0.282298	6.63365	4.554514	1.18E-22	3.50E-22
EDIL3	0.090728	0.965943	3.412317	1.56E-22	4.53E-22
DLL3	0.011792	0.06191	2.392375	1.60E-06	1.86E-06
PHYHIPL	0.628596	2.246911	1.837737	5.85E-08	7.20E-08
ADRA1A	5.001141	1.346761	-1.89276	1.94E-24	6.72E-24
MYBPHL	0.002037	0.247092	6.92225	1.64E-10	2.20E-10
EXTL1	0.023205	0.087017	1.906826	2.00E-13	3.08E-13
ORC6	0.291223	0.99971	1.779385	1.73E-26	1.10E-25
MSX1	0.203942	0.939481	2.203703	8.45E-24	2.68E-23
SP5	0.139479	3.448012	4.627647	4.05E-12	5.83E-12
MT1E	302.6544	62.19184	-2.28287	3.38E-26	1.90E-25
SLCO1C1	0.053786	0.108955	1.018448	6.41E-26	3.23E-25
TPX2	0.310881	5.8912	4.244127	1.48E-21	3.82E-21
ZIC4	0.033878	0.293757	3.116201	3.82E-05	4.24E-05

http://dx.doi.org/10.21037/atm-20-7101

Table S3 Univariate Cox regression analysis for prognostic genes

geneName	HR	HR.95L	HR.95H	pvalue
CDCA8	1.161443	1.111036	1.214138	3.82E-11
KIF2C	1.149721	1.100378	1.201277	4.55E-10
NCAPG	1.241995	1.158143	1.331918	1.23E-09
EZH2	1.268453	1.172108	1.372716	3.63E-09
TPX2	1.063372	1.041855	1.085334	3.83E-09
TTK	1.42605	1.266796	1.605326	4.25E-09
HJURP	1.240704	1.154514	1.333328	4.33E-09
NDC80	1.235031	1.150845	1.325375	4.62E-09
MCM10	1.668145	1.404275	1.981599	5.73E-09
CDC20	1.035574	1.023397	1.047895	6.94E-09
PRR11	1.293862	1.184636	1.413159	1.03E-08
TRIP13	1.186421	1.113934	1.263624	1.07E-07
CDCA2	1.66532	1.37961	2.010199	1.09E-07
CBX2	1.272652	1.164052	1.391383	1.17E-07
GINS1	1.234745	1.141974	1.335052	1.21E-07
PBK	1.148177	1.089459	1.210059	2.48E-07
SPC25	1.276769	1.163513	1.40105	2.53E-07
ORC1	1.365331	1.211747	1.53838	3.15E-07
CDCA3	1.312349	1.181954	1.457129	3.56E-07
KIF23	1.24065	1.141357	1.348581	4.05E-07
SKA3	1.395679	1.225774	1.589135	4.81E-07
WDHD1	1.660105	1.362603	2.022561	4.89E-07
SHCBP1	1.286119	1.164907	1.419945	6.29E-07
UBE2C	1.02246	1.013382	1.031619	1.05E-06
RRM2	1.075061	1.04413	1.106909	1.18E-06
ZWINT	1.071171	1.041765	1.101406	1.29E-06
IQGAP3	1.197815	1.113011	1.28908	1.45E-06
MELK	1.195048	1.111444	1.284941	1.47E-06
EXO1	1.370201	1.2042	1.559085	1.75E-06
KIF18B	1.243475	1.135927	1.361204	2.34E-06
BUB1B	1.223424	1.124856	1.330629	2.54E-06
CENPL	1.518228	1.275717	1.806839	2.57E-06
GPRIN1	1.517847	1.275369	1.806427	2.62E-06
KIFC1	1.090491	1.051699	1.130714	2.77E-06
NUF2	1.16044	1.090022	1.235408	3.18E-06

Table S3 (*continued*)

http://dx.doi.org/10.21037/atm-20-7101

Table S3 (*continued*)

geneName	HR	HR.95L	HR.95H	pvalue
TACC3	1.092139	1.05226	1.13353	3.42E-06
AKR1B15	1.103243	1.057986	1.150437	4.28E-06
FAM83D	1.114742	1.063797	1.168127	5.33E-06
ECT2	1.165551	1.090537	1.245725	6.38E-06
MAD2L1	1.311657	1.165301	1.476395	6.98E-06
UBE2T	1.065509	1.036186	1.095661	8.33E-06
CLSPN	1.782421	1.381119	2.300326	8.95E-06
EPS8L3	1.057431	1.031663	1.083843	9.14E-06
ORC6	1.350585	1.181341	1.544077	1.08E-05
CDK1	1.082631	1.044804	1.121828	1.21E-05
FLVCR1	1.289836	1.150534	1.446004	1.27E-05
CDC7	1.334669	1.171751	1.520239	1.39E-05
DEPDC1	1.330772	1.167418	1.516985	1.90E-05
MYBL2	1.029306	1.015767	1.043025	1.91E-05
TRIM16L	1.04094	1.021875	1.06036	2.10E-05
STMN1	1.034763	1.018426	1.051362	2.57E-05
FANCB	3.487667	1.941422	6.265418	2.92E-05
TOP2A	1.047175	1.024689	1.070154	3.15E-05
SKA1	1.142587	1.072821	1.216889	3.37E-05
E2F2	1.831304	1.374415	2.440074	3.60E-05
CENPF	1.154581	1.077969	1.236639	4.08E-05
CDC25A	1.323112	1.15613	1.514212	4.75E-05
RAD54L	1.321091	1.153239	1.513373	5.91E-05
NEK2	1.129188	1.064147	1.198203	5.97E-05
TRAIP	1.336809	1.159584	1.541119	6.32E-05
ZIC2	1.178347	1.08611	1.278417	7.94E-05
BIRC5	1.037407	1.018654	1.056505	7.95E-05
DSCC1	1.211351	1.100579	1.333271	8.90E-05

http://dx.doi.org/10.21037/atm-20-7101

Table S4 Risk score model for OS

geneName	coef	HR	HR.95L	HR.95H	pvalue
CDCA8	0.192995	1.212877	0.941793	1.561987	0.13483
AKR1B15	0.175446	1.191777	1.031897	1.37643	0.016978
EZH2	0.141074	1.15151	0.882519	1.50249	0.298674
EPS8L3	0.147772	1.159249	0.979418	1.372099	0.085769
CBX2	0.071536	1.074157	0.898003	1.284866	0.433762
TRIM16L	0.183376	1.201267	1.015561	1.42093	0.03234
FLVCR1	0.153859	1.166326	0.958221	1.419626	0.124939
GPRIN1	0.102045	1.107433	0.902035	1.359601	0.329594

http://dx.doi.org/10.21037/atm-20-7101