Research Article

Immune Associated LncRNAs Identify Novel Prognostic Subtypes of Renal Clear Cell Carcinoma[1]

Seema Khadirnaikar[1], Pranjal Kumar[1], Sathiya N Pandi[2], Rohit Malik[3], Saravana M Dhanasekaran[2], Sudhanshu Kumar Shukla 0000-0002-5069-9708 0000-0002-5069-9708[1]

[1]Department of Biosciences and Bioengineering, Indian Institute of Technology Dharwad, Dharwad, Karnataka, 580011

[2]Michigan Center for Pathology, Department of Pathology, University of Michigan, Ann Arbor, Michigan, USA, 48005

[3]Bristol-Myers Squibb, Princeton, New Jersey, USA, 08540

Corresponding Author

Sudhanshu Kumar Shukla

[1]Department of Biosciences and Bioengineering, Indian Institute of Technology Dharwad, Dharwad, Karnataka, 580011

Phone number: +918362212853

Email: sudhanshu@iitdh.ac.in

Grant Support

None

[1] This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi:10.1002/mc.22949

Abbreviations

KIRC: Kidney Renal Clear Cell Carcinoma

LncRNA: Long non-coding RNA

TCGA-GDC: The Cancer Genome Atlas-Genomic Data Commons

GSEA: Gene Set Enrichment Analysis

FPKM: Fragments Per Kilobase of transcript per Million mapped reads

GO: Gene Ontology

HR: Hazard Ratio

CI: Confidence Interval

LIPS: LncRNA Immune Prognostic Signature

Keywords:

KIRC

Renal Clear Cell Carcinoma

Long non-coding RNA

Immune Genes

Prognostic signature

Abstract

Kidney Renal Clear Cell Carcinoma (KIRC) is a significant cause of cancer-related deaths. Here, we aim to identify the LncRNAs associated with the immune system and characterise their clinical utility in KIRC. A total of 504 patients' data was used from TCGA-GDC. *In–silico* correlation analysis identified 143 LncRNAs associated with immune-related genes ($r > 0.7$, p-value < 0.05). K-means consensus method clustered KIRC samples in three immune clusters, namely cluster C1, C2, and C3 based on the expression of 143 immune-related LncRNAs. Kaplan-Meier analysis showed that C3 patients survived significantly worse than the other two clusters (p-value < 0.0001). A comparison of TCGA miRNA, mRNA cluster with immune cluster showed the independence and robustness of immune clusters (HR = 2.02 and p-value = 2.12×10^{-8}). The GSEA and CIBERSORT analysis showed high enrichment of poorly activated T-cells in C3 patients. To define LncRNA immune prognostic signature, we randomly divided the TCGA sample into discovery and validation sets. By utilising multivariate Cox regression analysis, we identified and validated a seven LncRNA immune prognostic signature score (LIPS score) (HR = 1.43 and p-value = 2.73×10^{-6}) in KIRC. Comparison of LIPS score with all the clinical factors validated its independence and superiority in KIRC prognosis. In summary, we identified LncRNAs associated with the immune system and showed the presence of prognostic subtypes of KIRC patients based on immune-related LncRNA expression. We also identified a novel immune LncRNA based gene-signature for KIRC patients' prognostication.

Introduction:

Renal cell carcinoma (RCC) accounts for more than 2% of cancer-related deaths in the world [1,2]. Histologically, RCC tumors can be placed under several major and rare subtypes. KIRC subtype is the most common and aggressive subtype of kidney tumors [3, 4, 5]. Recent genomic studies of clear cell carcinoma have dramatically increased our understanding of these tumors [5–7]. These studies have also revealed high molecular heterogeneity in these tumors, which necessitates further sub-classification of these tumors to improve diagnosis and treatment strategies.

In the past few years, LncRNAs have attracted a lot of attention due to their striking tumor and tissue specificity [8, 9]. In the context of cancer, members of LncRNA group have been shown to function both as oncogene and tumor suppressor [10]. Similarly, many LncRNAs

have been implicated in the modulation of the immune system [11]. The immune system plays a vital role in tumor development and progression [12]. Thus, it is crucial to identify LncRNAs which are involved in the modulation of the immune system. The LncRNAs may prove to be useful not only in immunotherapy decisions but may also provide novel therapeutic options.

Here we have used guilt by association *in-silico* analysis to identify the immune-related LncRNAs. We also identified a novel prognostic subtype of KIRC patients. Further, we have defined the genetic and epigenetic status of these novel prognostic subtypes. Moreover, using various statistical methods, we developed and validated a LncRNA immune prognostic signature score (LIPS score) for the prognostication of KIRC patients.

Materials and Methods

Expression Analysis

RNA-sequencing data for tumors and normal kidney from KIRC patients was downloaded from TCGA-GDC website and analysed using Tuxedo pipeline [13]. In short, reads were mapped using TopHat, merged and assembled using cufflinks and cuffmerge [13, 14]. Read count and FPKM values were calculated.

Differential Expression, Correlation Analysis and K-Means Clustering

Differential expression (DE) analysis was carried out on raw counts using Limma package in R. Genes with average count more than zero in tumors and normal kidney were used for differential expression analysis. FDR correction was performed to identify the differentially expressed genes in KIRC compared to normal kidney. Genes with more than 5-fold expression and 0.05 corrected p-value were considered as differentially expressed.

To identify LncRNAs associated with immune modulating genes correlation analysis was performed in the statistical software R. List of genes involved in immune modulation was downloaded from innateDB database. LncRNAs with the correlation coefficient ≥ 0.7 and p-value ≤ 0.05 were used for further analysis. K-Means consensus clustering was done using 143 immune-related LncRNAs using R.

GSEA, CIBERSORT and Network Analysis

Gene set enrichment analysis was performed using GSEA software. A pre-ranked method with default parameters was used for the analysis. We used hallmark, immunogenic genes sets as input. CIBERSORT analysis was done on https://cibersort.stanford.edu/ using the beta version of absolute quantification. Average expression of the different clusters was used as mixture file and LM22 was used as a signature gene file. GO analysis was performed using http://www.geneontology.org/ website. GO biological function was used for analysis. For network analysis, https://genemania.org/ was used with default settings [15]. Genes differentially expressed in high vs low LIPS score group was used as input.

Cox Regression and Kaplan-Meier Analysis

Clinical data associated with KIRC was downloaded from TCGA-GDC website in January 2018. TCGA miRNA and mRNA cluster data and Chen *et al.* cluster data were obtained from the supplementary file supplied with TCGA–KIRC study and Chen *et al.* study respectively [5, 7]. *Survival* package of R was used for univariate and multivariate Cox regression analysis which provided hazard ratio, confidence interval (CI) and p-values. The assumptions of Cox model were validated using *Survival* package and the prognostic ability was cross-validated using *rms* package in R. Kaplan-Meier analysis was performed in GraphPad 7.0.

Development of LIPS score

Immune LncRNAs identified in correlation analysis were subjected to univariate Cox regression analysis in the R survival package. We used immune LncRNA expression data and clinical data from the discovery set and performed univariate Cox regression analysis. To define the best prognostic immune LncRNA, we used a hazard ratio cut-off of 1.5 and FDR < 0.05. A total of 62 immune LncRNAs showed significant correlation with survival (FDR < 0.05). To select the LncRNAs with highest prognostic ability, a hazard ratio cut-off of 1.5 was applied (either more than 1.5 or less than 0.67). Post application of these filtering criteria led to the prediction of seven LncRNAs. To develop LncRNA immune prognostic signature (LIPS) score, we combined the weighted expression of all the seven LncRNAs for each individual patient. The weights were obtained from Cox regression analysis. If the regression coefficient for LncRNAs are defined as β, the formula used for LIPS score development was:

LIPS score for Patient A = (β_1* expression of LncRNA1) + (β_2* expression of LncRNA2) + (β_3* expression of LncRNA3) + (β_4* expression of LncRNA4) + (β_5* expression of LncRNA5) + (β_6* expression of LncRNA6) + (β_7* expression of LncRNA7)

Results

Expression Analysis of Immune Genes in Normal Kidney and KIRC Samples

To delineate the expression pattern of immune genes, we analysed the TCGA RNA-Seq data of KIRC tumors and normal kidney samples. Interestingly, we found that 2378 genes (846 protein-coding genes and 1532 LncRNA genes) were expressed only in KIRC whereas 669 genes (235 protein-coding genes and 434 LncRNA genes) were exclusive to normal kidney (**Figure 1 A and Supplementary Table S1**). Nearly twice the number of LncRNAs showed KIRC specificity compared to protein-coding genes confirming the previous finding that the LncRNAs are more cancer-specific as compared to the protein-coding genes [8]. Limma with FDR correction (five-fold expression with the significance cut-off of 0.05) identified 2139 genes (1577 protein-coding genes and 562 LncRNA genes) as differentially expressed in KIRC (**Figure 1 B**). As expected, genes previously identified as a marker of KIRC, *CA9*, *ANGPTL4*, and *NDUFA4L2* were found to be overexpressed in KIRC compared to normal (**Supplementary Figure S1 A**).

Next, we downloaded the list of genes involved in modulation of the immune system from www.innatedb.com database. The differential expression analysis was carried out using Limma, which identified 130 immune-related genes differentially expressed in KIRC (112 overexpressed and 18 underexpressed) (**Figure 1 C**). As expected, in pathway analysis, all the enriched pathways indicated the immune modulation function of input genes (**Figure 1 D**).

In-silico Identification of LncRNAs Associated with Immune Pathways

Guilt by association is a standard method to identify the probable functions of unknown genes [16]. This principle states that genes that are associated are more likely to share a function. We utilised guilt by association method to identify the LncRNAs with a high probability of immune modulation role in KIRC samples. This analysis identified 143 LncRNAs associated with immune genes with more than 0.7 correlation coefficient and less than 0.05 p-value (**Supplementary Table S2**). Interestingly, LncRNA previously related to immune system

SNHG8 and *TP73-AS1* (correlation coefficient 0.72 and 0.84 respectively) were also identified in our analysis strengthening the analysis pipeline [17, 18].

Immune-Related LncRNAs are Associated with Prognostic Subtypes of KIRC

We performed K-means consensus clustering to identify immune LncRNAs associated subtypes in KIRC patients. The K-means consensus clustering placed samples into three groups that we named, clusters C1, C2 and C3 (**Figure 2 A and Supplementary Table S3**). The expression of immune-related LncRNAs in these novel clusters is shown in **Figure 2 B**. The variable expression of immune LncRNAs in patients belonging to different clusters prompted us to investigate the survival pattern of these patients. Strikingly, we discovered an association between the expression level of immune LncRNAs and survival where patients from C1 showed good prognosis and patients from C3 and C2 showed worst and intermediate prognosis respectively (**Supplementary Figure S1 B and Figure 2 C**). The Cancer Genome Atlas (TCGA) in its landmark paper on KIRC has identified miRNA and mRNA subtypes with prognostic association [5]. Hence, next, we performed a multivariate Cox regression analysis using TCGA miRNA, mRNA and newly identified immune clusters as a covariate. Interestingly, our immune cluster outperformed TCGA miRNA, mRNA clusters and predicted survival independently (**Figure 2 D**). Similarly, Chen *et al.* have identified five genomic clusters in KIRC with a significant difference in survival. Excitingly, in multivariate Cox regression analysis, immune cluster emerged as most significant and independent prognosticator (**Figure 2 D**). From the TCGA clinical data, we found that C3 patients have significant enrichment of a high-grade tumor (Grade-4), higher metastasis rate (M1), and lower average mutation (**Supplementary Table S3)**. As C3 patients showed inferior prognosis and different clinical and mutational features, we grouped C1, C2 patients and compared with C3 patients. In a Kaplan-Meier plot, C3 showed four times less median survival compared to combined C1 and C2 patients (**Figure 2 C)**.

We hypothesised that genetic and epigenetic difference between C3 patients and C1+C2 patients might explain the reason for differences in survival. To validate our hypothesis, we performed Gene Set Enrichment Analysis (GSEA) using differently regulated genes between these two groups (**Figure 3 A**). A pre-ranked GSEA analysis showed enrichment of genes associated with CD8 T up, effector CD8 cell down, T regulatory cell down, and Dendritic cell down in C3 patients (**Supplementary Figure S2 A**, **B, C, and D and Supplementary Table**

S4). KRAS signalling downregulation and upregulation of allograft rejection related genes in Cluster 3 (**Supplementary Figure S2 E and F and Supplementary Table S4**). We also found that patients belonging to C3 had high expression of PDCD1 (PD1) and CTLA4 expression compared to C1+C2 patients (**Supplementary Figure S3)**. Also, we did not find a significant difference of CD274 (PDL1) expression between C1+C2 and C3 patients (**Supplementary Figure S3)**. These results suggest the distinct immune cell infiltration in C3 and C1+C2 patients. We performed CIBERSORT analysis to identify the infiltration level of various immune cells. Interestingly, we found that cluster C3 patients had higher infiltration of CD8+ T-cells (p-value = 0.03) and C1+C2 patients had higher infiltration of Naïve B-cells, M0 and M2 type macrophages and neutrophils (p-values = 0.03, 0.02, < 0.0001, and 0.08) (**Figure 3 B).** We also showed that C3 patients showed the significantly lower number of somatic mutations compared to C1+C2 (p-value < 0.005) (**Figure 3 C and Supplementary Table S4**). Further, we explored common somatic mutations in C1+C2 and C3 patients using Chen *et al*. data. As expected, *VHL*, *PBRM1*, *SETD2* and *BAP1* were most mutated genes in both the clusters of patients (**Figure 3 D and Supplementary Table S5**). Interestingly, we found that C3 patients had much lower *MTOR*, *MLL3*, *PTEN*, *ARID1A*, *STAG2* and *NF2* mutations (**Figure 3 D**). Interestingly, inactivating mutation in the one of the frequently mutated genes in cancer, *TP53*, was absent in C3, (**Figure 3 D, Supplementary Table S5**). This data supports our previous finding that C3 patients, in general, have a lower mutation rate.

Comparison of DNA methylation changes showed that C3 patients group, in general, have more hyper-methylated genes compared to C1+C2 patients (**Supplementary Figure S4)**. This finding supports the previous observation of Chen *et al.* that KIRC patients with higher methylation level show poor survival.

Immune Associated LncRNA Signature is Robust Prognosticator of KIRC Patients:

In our earlier analysis, we found that immune-related LncRNAs are associated with prognosis of KIRC patients. Therefore, we developed a LncRNA based prognostic signature for KIRC patients. As TCGA RNA-Seq data is the only available LncRNA expression data with clinical data, we randomly divided the TCGA KIRC RNA-Seq into discovery and validation set (208 patients each). After applying the filtering criteria, we identified seven immune prognostic LncRNAs amongst which four were poor prognosis markers (risky LncRNA) and three were

good prognosis markers (protective LncRNA) **(Figure 4 A**). Further, we developed a LncRNA immune prognostic signature score (LIPS score) for each patient in the discovery set using the formula detailed in the method section. We validated the prognostic utility of LIPS score in a univariate Cox regression analysis (HR= 1.38, p-value = 5.14 X 10^{-8}) (**Table 1**). The proportionality assumptions of Cox model were tested and it was observed that none of the co-variates break the assumption and are time independent (**Supplementary Table S6**). Further, 10-fold cross-validation (c-index = 0.723) validated the prognostic ability of the proposed model. We also performed multivariate analysis of LIPS score with age, distance metastasis and tumor stage. As expected LIPS score was an independent predictor of survival in the discovery set (HR = 1.38, p-value = 6.00 X 10^{-4}) (**Table 1**). We next tested the prognostic power of LIPS score in the validation set. We performed univariate Cox regression analysis and showed that LIPS score was an independent predictor of survival of KIRC patients in the validation set as well (HR = 1.52, p-value = 2.47 X 10^{-9}) (**Table 1**). Furthermore, we also executed multivariate analysis of LIPS score with age, distance metastasis and tumor stage in the validation and showed again that LIPS score was a robust and independent prognosticator in the validation set (HR = 1.43, p-value = 2.73 X 10^{-6}).

The LIPS score distribution in the discovery and validation sets are presented in **Figure 4 B and 4 C**). Interestingly, we found that patients with lower LIPS score showed less mortality in both the set of patients (**Figures 4 D and E**). Next, we divided the patients into low and high LIPS score (negative and positive LIPS score) and performed Kaplan-Meier analysis to show that patients with low LIPS score survive significantly better than the patients with high LIPS score in the discovery set (HR = 3.02 and p-value < 0.0001) and validation set (HR = 3.04 and p-value < 0.0001) **(Figure 4 F and G)**.

We also compared the LIPS score with other previously published signature including five LncRNA signature [19, 20]. In a multivariate Cox regression analysis, LIPS score was both an independent and superior prognostic marker of KIRC patients (**Table 2**). The comparison of stage, metastasis and mutation between low and high LIPS score patients are shown in **Supplementary Table S7.**

Patients with High LIPS Score Have enrichment of inflammatory response Pathway

We noted above that LIPS score signature was consist of two different categories of LncRNAs that we termed Risky LncRNAs, for those LncRNA which were associated with poor survival, and protective LncRNA which were associated with good survival. As expected, high-risk LncRNAs were overexpressed in high LIPS score patients while low-risk LncRNAs were overexpressed in patients with lower LIPS score in both the discovery and validation set **(Figure 4 H and I**). We hypothesised that LncRNAs which form a part of the LIPS score modulate the global pattern of gene expression. As a result, altering various signalling pathways differently between low and high LIPS score patients, hence explaining the difference in survival. GO term analysis using overexpressed genes in high LIPS score versus low LIPS score (**Figure 5 A**), identified enrichment of pathways involved in acute-phase response, acute-inflammatory response, nucleobase-containing compound metabolic process, and cellular nitrogen compound metabolic and synthesis process **(Figure 5 B)**. Similarly, GO analysis using genes overexpressed in low LIPS score compared to high LIPS score (**Figure 5 C)** showed enrichment of pathways involved in transport of carboxylic acid, organic substance, ions and establishment of localization.

Discussion:

The immune system plays an important role in the development and progression of cancers including KIRC [21]. This phenomenon is being exploited to develop various immunotherapeutic strategies. However, reasons, why certain tumors are heavily immune infiltrated while others show minimal to no immune participation, are still not fully understood. Also, among the infiltrated tumors how the immune system is tamed or utilized by the tumors is only now being explored. Understanding immune composition and regulatory mechanisms are of primary importance to further innovate immunotherapeutic strategies. Recently, Şenbabaoğlu *et al* have shown the importance of immune infiltration on the KIRC patients' survival. Authors have also identified mRNA signature for the patients' prognosis [22]. Similarly, Giraldo *et al* have identified the role of Tumor-Infiltrating and Peripheral Blood T-cell in the prediction of early recurrence in localized RCC [23]. The high tissue and cancer specificity of LncRNAs which is useful as biomarker and therapeutic targets, also serves as a great tool to delineate the immune cell type/state composition of tumors [9]. Various LncRNAs have been shown to

modulate the immune system [24]. Here, we performed an *in-silico* analysis to identify the LncRNAs associated with an immune function in KIRC. We identified three prognostic clusters with different survival potential (C1, C2 and C3) using K-means clustering based on these immune-related LncRNAs. Interestingly, C1 with the best survival showed expression of 143 LncRNAs similar to normal while C3 with worst survival have higher expression of LncRNAs (**Figure 2 B**). Immune clusters showed significantly better and independent survival association compared to TCGA mRNA and miRNA clusters [5].

GSEA analysis revealed that C3 patients had overexpression of genes regulating T-cell population but the downregulation of genes involved in the activation of dendritic cells. Using CIBERSORT analysis, we also found that C3 patients have higher infiltration of CD8+ T-cells. These results suggest that although C3 patients, i.e. patients with worst survival, have higher enrichment of CD8+ T-cells, these cells were poorly activated probably due to lack of antigen presenting cells (dendritic cells). Interestingly, these patients also showed significantly lower mutational burden, hence lower number of neoantigens which corroborate with reduced activation of CD8+ T-cells [25]. These findings were further confirmed by the fact that patients belonging to C3 had high expression of PDCD1 (PD1) and CTLA4 expression compared to C1+C2 patients. Also, we did not find a significant difference of CD274 (PDL1) expression between C1+C2 and C3 patients **(Supplementary Figure S3)**. This difference in the level of PD1 and CTLA4 between C1+C2 and C3 indicating high infiltration but poor activation of T-cells. On the other hand, patients from C1+C2 had higher infiltration of Naïve B-cells, M0 and M2 type macrophages and neutrophils. The infiltration of neutrophils is associated with favourable prognosis of many cancer types [26, 27]. The high neutrophil level in C1+C2 suggests its involvement in a better prognosis.

To utilise the immune-associated LncRNAs in prognosis, we developed a seven LncRNA based immune score LIPS score. We showed that LIPS score was an independent marker of prognosis in the discovery and validation sets and high LIPS score was associated with poor survival. We also compared the prognostic ability of LIPS score with previously developed prognostic signature. In a multivariate analysis, LIPS score outperformed both the previous prognostic signatures. We hypothesise that gene expression differences between high and low LIPS score are the reason for the poor survival of high LIPS score patients. Interestingly, our

analysis showed that patients with high LIPS score had activation of immune-related pathways, suggesting the involvement of the immune system in the prognosis of these patients. In comparison, low LIPS score patients showed enrichment of pathways involved in transport and localization suggesting the differential role of these pathways in low LIPS score patients compared to high LIPS score patients.

In conclusion, RNA-Seq analysis identified novel prognostic immune subtype in KIRC patients characterised by differential expression of immune-related LncRNAs. The variation in the immune subtypes and the level of immune cell infiltration may explain the prognostic differences. More importantly, we have developed and validated a strong and robust immune LncRNA prognostic signature which was used to formulate LIPS score for KIRC patients.

Acknowledgements

The results shown here are in whole or part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/."

Funding

None

Disclosure

None to declare.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. CA: A Cancer Journal for Clinicians. 2017;67:7–30.

2. Gandaglia G, Ravi P, Abdollah F, Abd-El-Barr A-E-RM, Becker A, Popa I, et al. Contemporary incidence and mortality rates of kidney cancer in the United States. Can Urol Assoc J. 2014;8:247–52.

3. Muglia VF, Prando A. Renal cell carcinoma: histological classification and correlation with imaging findings. Radiologia Brasileira. 2015;48:166–174.

4. Moch H, Cubilla AL, Humphrey PA, Reuter VE, Ulbright TM. The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs—Part A: Renal, Penile, and Testicular Tumours. European Urology. 2016;70:93–105.

5. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature. 2013;499:43–9.

6. Gerlinger M, Horswell S, Larkin J, Rowan AJ, Salm MP, Varela I, et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. Nat Genet. 2014;46:225–33.

7. Chen F, Zhang Y, Şenbabaoğlu Y, Ciriello G, Yang L, Reznik E, et al. Multilevel Genomics-based Taxonomy of Renal Cell Carcinoma. Cell Rep. 2016;14:2476–89.

8. Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, et al. The landscape of long noncoding RNAs in the human transcriptome. Nature Genetics. 2015;47:199–208.

9. Sahu A, Singhal U, Chinnaiyan AM. Long Noncoding RNAs in Cancer: From Function to Translation. 2015.

10. Bhan A, Soleimani M, Mandal SS. Long Noncoding RNA and Cancer: A New Paradigm. Cancer Res. 2017;77:3965–81.

11. Chen YG, Satpathy AT, Chang HY. Gene regulation in the immune system by long noncoding RNAs. Nat Immunol. 2017;18:962–72.

12. Coventry BJ, Ashdown M, Henneberg M, Davies PCW. The Immune System and Responses to Cancer: Coordinated Evolution. F1000Research. 2015. doi:10.12688/f1000research.6718.1.

13. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols. 2013;7:562–578.

14. Trapnell C, Pachter L, Salzberg SL. TopHat: Discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25:1105–1111.

15. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010;38 Web Server issue:W214-220.

16. Petsko GA. Guilt by association. Genome Biology. 2009;10:104.

17. Wang W, Zhao Z, Yang F, Wang H, Wu F, Liang T, et al. An immune-related lncRNA signature for patients with anaplastic gliomas. J Neurooncol. 2018;136:263–71.

18. Hu H, Liu J-M, Hu Z, Jiang X, Yang X, Li J, et al. Recently Evolved Tumor Suppressor Transcript TP73-AS1 Functions as Sponge of Human-Specific miR-941. Molecular Biology and Evolution. 2018;35:1063–77.

19. Chen Y-L, Ge G-J, Qi C, Wang H, Wang H-L, Li L-Y, et al. A five-gene signature may predict sunitinib sensitivity and serve as prognostic biomarkers for renal cell carcinoma. J Cell Physiol. 2018.

20. Shi D, Qu Q, Chang Q, Wang Y, Gui Y, Dong D. A five-long non-coding RNA signature to improve prognosis prediction of clear cell renal cell carcinoma. Oncotarget. 2017;8:58699–708.

21. Vinay DS, Ryan EP, Pawelec G, Talib WH, Stagg J, Elkord E, et al. Immune evasion in cancer: Mechanistic basis and therapeutic strategies. Seminars in Cancer Biology. 2015;35:S185–98.

22. Şenbabaoğlu Y, Gejman RS, Winer AG, Liu M, Van Allen EM, de Velasco G, et al. Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. Genome Biology. 2016;17:231.

23. Giraldo NA, Becht E, Vano Y, Petitprez F, Lacroix L, Validire P, et al. Tumor-Infiltrating and Peripheral Blood T-cell Immunophenotypes Predict Early Relapse in Localized Clear Cell Renal Cell Carcinoma. Clin Cancer Res. 2017;23:4416–28.

24. Heward JA, Lindsay MA. Long non-coding RNAs in the regulation of the immune response. Trends Immunol. 2014;35:408–19.

25. Lanitis E, Dangaj D, Irving M, Coukos G. Mechanisms regulating T-cell infiltration and activity in solid tumors. Ann Oncol. 2017;28 suppl_12:xii18–32.

26. Treffers LW, Hiemstra IH, Kuijpers TW, van den Berg TK, Matlung HL. Neutrophils in cancer. Immunological Reviews. 2016;273:312–28.

27. Wikberg ML, Ling A, Li X, Öberg Å, Edin S, Palmqvist R. Neutrophil infiltration is a favorable prognostic factor in early stages of colon cancer. Hum Pathol. 2017;68:193–202.

Figure Legends

Figure 1. Expression of Immune Genes in KIRC: A) Venn diagram shows overlaps among the different categories of gene sets. **B**) Volcano plot represents the result of differentially expressed genes (DEGs) analysis of KIRC vs normal kidney (under-expressed: $\log_2(\text{Fold change}) < -2.32$ and FDR < 0.05; over-expressed: $\log_2(\text{Fold Change}) > 2.32$ and FDR < 0.05). **C)** The heat map shows the expression level of differentially expressed immune-related genes in KIRC samples compared to normal tissues (blue- lower expression, red- higher expression). **D)** Bar plot represents the enriched GO terms obtained using differentially expressed immune genes in KIRC.

Figure 2. K-means Clustering Identifies Novel KIRC Prognostic Subtype: A) Results obtained from K-means consensus clustering identified three different clusters namely C1(light blue), C2(dark blue) and C3(green) **B)** Heatmap of immune-related LncRNA expression in the three KIRC clusters and normal samples**. C)** Kaplan-Meier plot shows the difference in survival of patients belonging to C1+C2 and C3 clusters. **D**) The table summarises the results of the multivariate analysis performed to compare the prognostic value of TCGA miRNA, mRNA clusters with immune clusters identified in this study. The table also shows the results of the multivariate analysis conducted to compare the prognostic value of genomic clusters of KIRC as identified by Chen *et al.* with immune clusters identified in this study.

Figure 3. Characterization of immune C3 compared to C1+C2: A) Volcano plot represents the protein coding genes differentially expressed in C1+C2 compared to C3. **B**) Bar plot shows the level of different immune cell types in C1+C2 and C3. **C**) Box plot represents the degree of mutation in C1+C2 and C3 patients. **D)** Waterfall plot represents the frequency of the common mutations in C1+C2 and C3. The key is given to interpret the results.

Figure 4. Development of LIPS score: A) Volcano plot represents the prognostic immune LncRNAs that significantly predict KIRC patient's prognosis. Red spots indicate LncRNAs associated with poor survival (risky) and blue spots indicate LncRNAs that are associated with good survival marker (protective). A dot plot shows the distribution of LIPS score in **B)** discovery set and **C)** validation set. A dot plot shows the distribution of LIPS score with survival in **D)** discovery set and **E)** validation set (blue- patients alive, red- patients dead). A Kaplan-Meier plot shows the difference in survival of KIRC patients belonging to high and low LIPS score in **F)** discovery set and **G)** validation set. The patients were divided at 0 LIPS score. Heatmap showing the expression of seven immune prognostic LncRNA in **H)** discovery set and **I)** validation set.

Figure 5. Molecular Characterization of High and Low LIPS Patients: A) Heat map represents the expression of differentially expressed genes in high vs low LIPS patients (white-lower expression, red- higher expression (upregulated: 434, downregulated: 52)). **B**) An XY plot shows the GO terms enriched in patients with high LIPS score (blue- negative enriched GO terms, red- positive enriched GO terms, size represents the fold change). **C**) An XY plot shows the GO terms enriched in patients with low LIPS score (blue- negative enriched GO terms, red-positive enriched GO terms, size represents the fold change).

Table 1: Table showing the results of univariate and multivariate cox regression analysis performed to compare prognostic value of clinical variate with LIPS score developed in this study.

Discovery Cohort Cox regression analysis

Univariate analysis

Variable	HR	95% CI	P value

Distance Metastasis	M1	1.53	2.93	6.56	5.34×10^{-11}
	Stage II	1.54	0.64	3.744	0.33
	Stage III	3.1	1.7	5.65	2.00×10^{-4}
Stage	Stage IV	7.25	3.96	13.27	1.28×10^{-10}
Grade			Not significant		
Regional lymph nodes (N)			Not significant		
Age		1.03	1.009	1.05	4.20×10^{-3}
Score		1.38	1.23	1.55	5.14×10^{-8}
Multivariate analysis					
Distance Metastasis	M1	5.64	0.85	37.23	0.07
	Stage II	1.21	1.13	2.97	0.67
Stage	Stage III	3.1	0.14	4.08	0.01
	Stage IV	7.25	0.14	7.35	0.96
Age		1.01	0.99	1.03	0.13
Score		1.38	1.23	1.55	6.00×10^{-4}
Validation cohort Cox regression analysis					
Univariate analysis					
Variable		**HR**	**95% CI**		**P value**
Distance Metastasis	M1	3.39	2.03	5.67	3.18×10^{-6}
	Stage II	1.54	0.29	2.49	0.77
Stage	Stage III	3.1	0.84	3.2	0.14
	Stage IV	7.25	2.39	7.86	1.34×10^{-6}
Grade			Not significant		
Regional lymph nodes (N)			Not significant		
Age		1.05	1.03	1.07	9.76×10^{-6}
Score		1.52	1.32	1.75	2.47×10^{-9}
Multivariate analysis					
Distance Metastasis	M1	0.18	0.02	1.26	0.41
	Stage II	0.78	0.26	2.95	0.65
Stage	Stage III	1.44	0.73	2.84	0.28
	Stage IV	17.27	2.43	122.6	0.02
Age		1.06	1.03	1.08	3.03×10^{-6}
Score		1.43	1.24	1.65	2.73×10^{-6}

Table 2: Table showing the results of multivariate analysis performed to compare prognostic value of previously published gene signatures with LIPS score developed in this study.

Comparison of signatures			
Discovery set			
Variable	**HR**	**95% CI**	**P value**

		Lower	Upper	
LncRNA signature (Shi et. al.)	1.006	0.99	1.01	0.26
5 PcG signature (Chen et.al.)	1.03	1.01	1.04	0.01
LIPS	1.42	1.22	1.66	7.09×10^{-6}
Validation set				
LncRNA signature (Shi et. al.)	1.01	0.99	1.03	0.06
5 PcG signature (Chen et.al.)	1.01	0.99	1.03	0.12
LIPS	1.32	1.19	1.53	2.38×10^{-6}

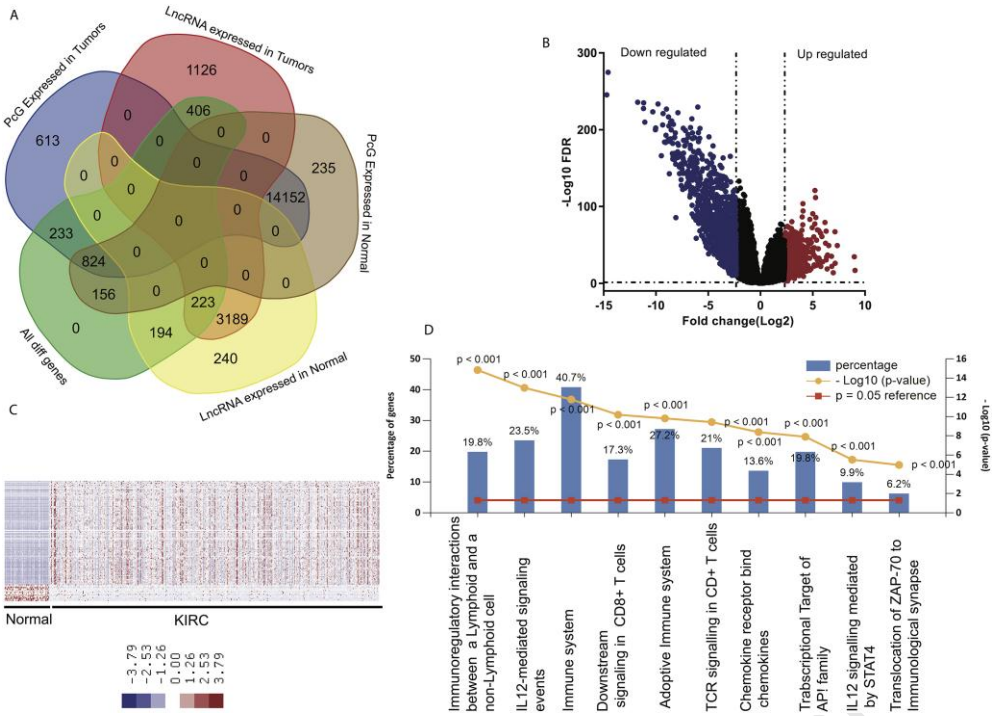
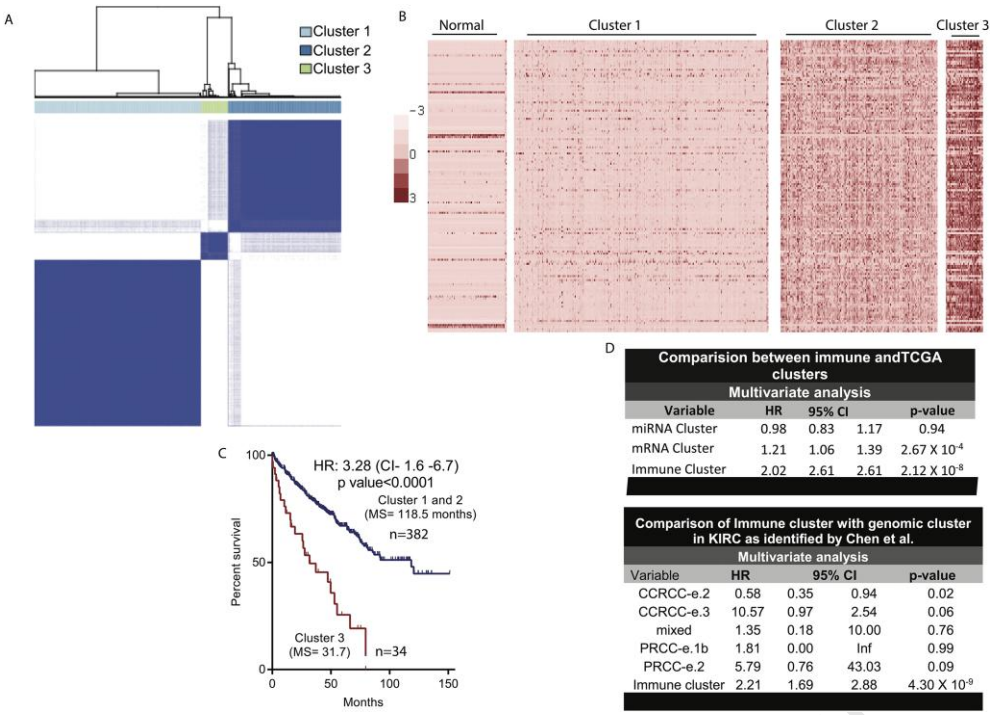


Figure 1 .



D

Comparision between immune andTCGA clusters

Multivariate analysis

Variable	HR	95% CI		p-value
miRNA Cluster	0.98	0.83	1.17	0.94
mRNA Cluster	1.21	1.06	1.39	2.67×10^{-4}
Immune Cluster	2.02	2.61	2.61	2.12×10^{-8}

Comparison of Immune cluster with genomic cluster in KIRC as identified by Chen et al.

Multivariate analysis

Variable	HR	95% CI		p-value
CCRCC-e.2	0.58	0.35	0.94	0.02
CCRCC-e.3	10.57	0.97	2.54	0.06
mixed	1.35	0.18	10.00	0.76
PRCC-e.1b	1.81	0.00	Inf	0.99
PRCC-e.2	5.79	0.76	43.03	0.09
Immune cluster	2.21	1.69	2.88	4.30×10^{-9}

figure 2 .

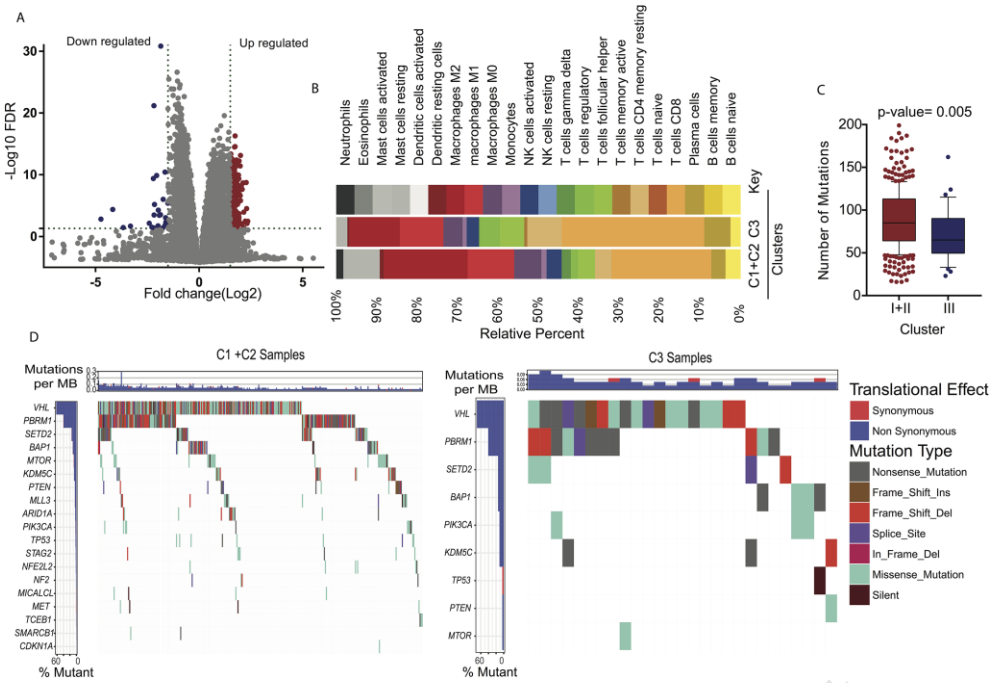


figure 3 .

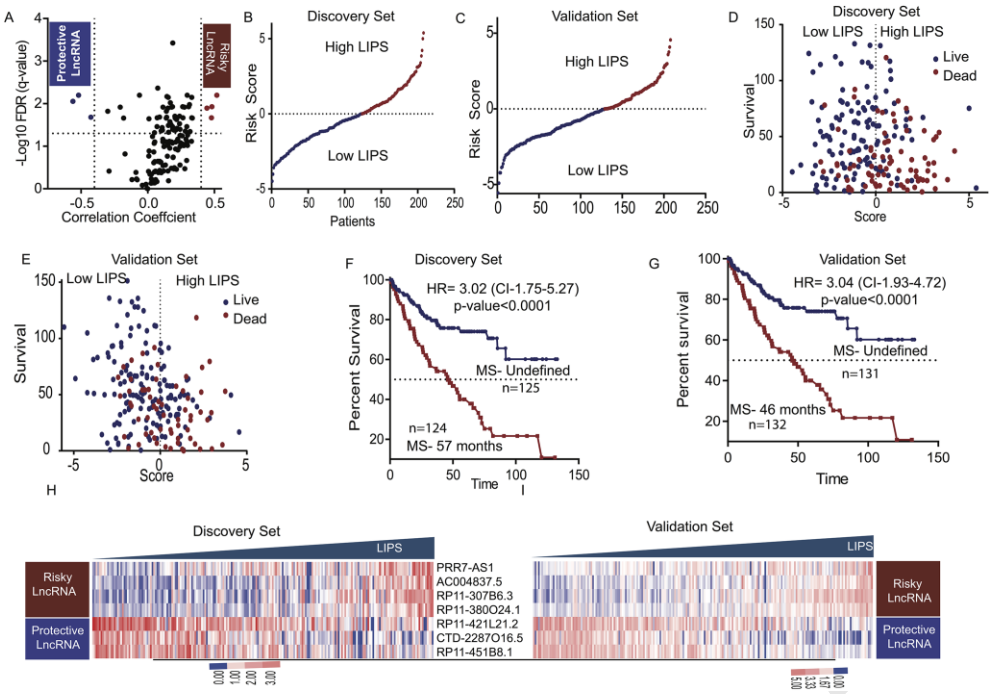


figure 4 .

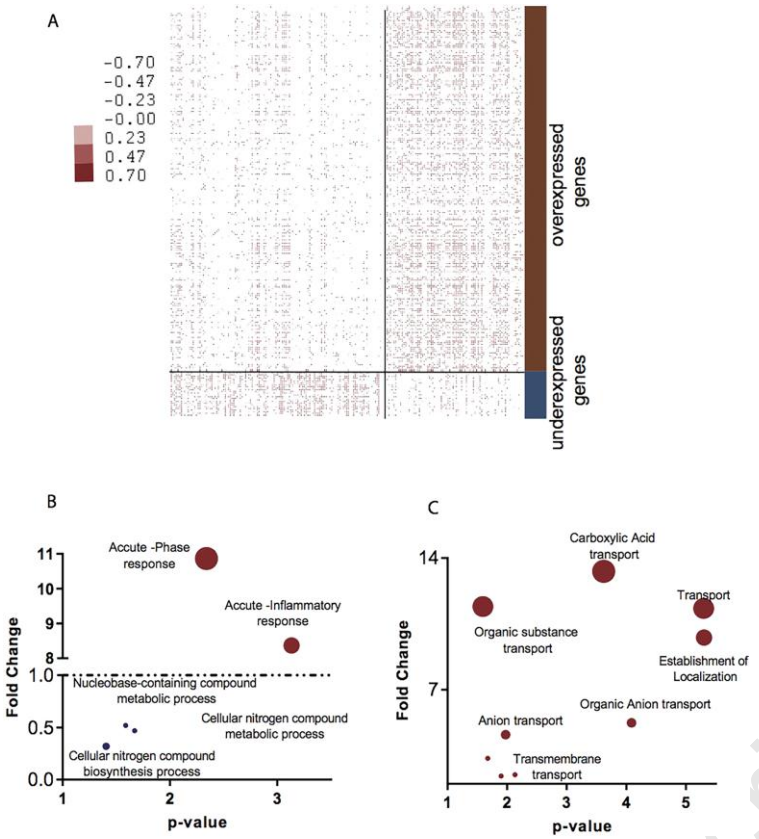


figure 5 .