

# Differential Genomic Instability-Associated LncRNAs Predict Differences of Clinical Outcome and Immunity in Left- And Right- Sided Colon Adenocarcinoma

**Junnan Guo**

Harbin Medical University Third Clinical College: Tumor Hospital of Harbin Medical University

**Tianyi Xia**

Harbin Medical University Third Clinical College: Tumor Hospital of Harbin Medical University

**Shenhui Deng**

Fourth Affiliated Hospital of Harbin Medical University

**Binbin Cui**

Harbin Medical University Third Hospital: Tumor Hospital of Harbin Medical University

**Yanlong Liu** (✉ [2355@hrbmu.edu.cn](mailto:2355@hrbmu.edu.cn))

Harbin Medical University Third Hospital: Tumor Hospital of Harbin Medical University

<https://orcid.org/0000-0002-5254-3897>

---

## Primary research

**Keywords:** Colon adenocarcinoma, left-sided, right-sided, genomic instability, immunity, prognosis

**Posted Date:** February 4th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-175488/v1>

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

[Read Full License](#)

---

# Abstract

**Background:** The left-sided and right-sided colon adenocarcinoma (LCCs and RCCs, respectively) have unique characteristics in various aspects, particularly molecular features and clinical heterogeneity. The purpose of our study was to develop a prognostic risk model based on differential genomic instability-associated (DGIA) Long non-coding RNAs (lncRNAs) of LCCs and RCCs, therefore the prognostic key lncRNAs could be identified.

**Methods:** We adopted two independent gene data-sets, corresponding somatic mutation and clinical information from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases. Identification of differential DGIA lncRNAs from LCCs and RCCs were conducted with appliance of "Limma" analysis. Then, we screened out key lncRNAs based on univariate and multivariate Cox proportional hazard regression analysis. Meanwhile, DGIA lncRNAs related prognostic model (DRPM) was established. We employed the DRPM in the model group and internal verification group from TCGA for the purpose of risk grouping and accuracy verification of PRSM. We also verified the accuracy of key lncRNAs with GEO data. Finally, the differences of immune infiltration and functional pathways were analyzed within different risk groups.

**Results:** A total of 123 DGIA lncRNAs were screened out by differential expression analysis. We obtained 6 DGIA lncRNAs by the construction of DRPM, including AC004009.1, AP003555.2, BOLA3-AS1, NKILA, LINC00543 and UCA1. After the risk grouping by these DGIA lncRNAs, we found the prognosis of high-risk group (HRG) was significantly worse than that in low-risk group (LRG) (all  $p \leq 0.05$ ). In all TCGA samples and model group, the expression of CD8<sup>+</sup> T cells in HRG was lower than that in LRG (all  $p \leq 0.05$ ). The functional analysis indicated that there was significant up-regulation with regard of pathways related to both genetic instability and immunity in LRG, including cytosolic DNA sensing pathway, response to dsRNA, RIG-I like receptor signaling pathway and Toll-like receptor signaling pathway.

**Conclusion:** Through the analysis of the DGIA lncRNAs between LCCs and RCCs, we established a DRPM which could predicate prognosis of LCCs and RCCs, and 6 key DGIA lncRNAs were identified as well. They can not only predict the prognostic risk of patients, but also serve as biomarkers for evaluating the differences of genetic instability and immune infiltration.

## Background

Colon cancer (CC) are one of the leading types of cancer occurred in humans and globally more than 1.8 million people acquired this disease each year [1, 2]. Lately, immunotherapy has achieved breakthroughs and fetching much consideration as a leading therapy against tumors. However, some CC patients shows a low response and drug resistance [3]. The traditional treatments such as surgery, radiotherapy, and chemotherapy are used to rectify the cancers, but their durable effects are also difficult to predict. The differences of above phenomenon are more obvious in left-sided and right-sided CCs (LCCs and RCCs, respectively) [4]. It is well known that the molecular and clinical heterogeneity

differences between LCCs and RCCs are very complex, as are their occurrence, development, and response to treatment and prognosis [5]. It is important to understand the potential molecular biological mechanisms, as these influences treatment decisions.

Recently, researchers have revealed that LCCs and RCCs are different in clinical and genomic characteristics [6]. In addition to the microsatellite instability status, the identified differences include APC, TP53, RAS and BRAF mutations, etc [7–9]. The dissimilarities in gene expression patterns could be used to analyze LCCs and RCCs. Mainly, doctors can be benefited by selecting the most effective individualized treatment via the degree and nature of these molecular mutations. Therefore, while we are looking for novel and precise prognostic biomarkers, their use is more vital for guiding targeted therapy.

Prior to cell division, the fidelity of humans genome replication, exhibits a high degree of consistency over time [10]. However, various genome mutations causes to alter this replica, which leads to the occurrence and development of tumors [11]. In CC, mutations in mismatch repair genes led to functional defects, which can cause microsatellite instability (MSI). The distinction in MSI status is also one of the factors helps to differentiate between LCCs and RCCs [12]. Also, a variety of biological processes are related to genome instability, including abnormal transcription and post-transcriptional regulation, DNA damage regulation, etc [13]. Latest findings disclosed that variations in the instability of the genome produces new antigens, which affects the immune phenotype and immunotherapy response [14]. Long non-coding RNAs (lncRNAs) are considered to be incapable to encode proteins, and they play an indispensable regulatory role in tumors. Currently, lncRNAs have been shown to be related to genome stability [15],but in LCCs and RCCs, the influence of differential genomic instability-associated (DGIA) lncRNAs on tumor-associated immune microenvironment has not been explored yet.

Therefore, in the current investigation we proposed to create a prognostic model and risk clustering, containing key lncRNAs based on the differentially expressed genes and genomic instability in the LCCs and RCCs in The Cancer Genome Atlas (TCGA) database; the leading goal of this study was to analyze the differences in immune infiltration between high- and low-risk groups (HRG and LRG, respectively) along with the verification in the Gene Expression Omnibus (GEO) database. Moreover, to screen out new prognostic biomarkers related to genetic instability in LCCs and RCCs and to provide a molecular basis for identifying immunotherapy.

## **Materials And Method**

### **Data collection**

In this research, we ratify two independent gene data-sets from different high-throughput platforms, including 473 colon adenocarcinoma (COAD) samples from TCGA (<https://portal.gdc.cancer.gov/>) and 156 COAD samples from GEO (<http://www.ncbi.nlm.nih.gov/geo/>) (GSE103479). The downloaded data included paired lncRNA and mRNA expression profiles, somatic mutation information, and clinical information. The CRCs in the cecum, ascending colon and hepatic flexure were defined as LCCs and CRCs

in splenic flexure, descending colon, sigmoid colon, and rectosigmoid junction was defined as RCCs. After screening based on CRCs location, there were a total of 411 samples with complete information available for analysis, of which 322 from TCGA and 89 from GEO. In this series of analysis, the TCGA samples were divided into two groups randomly, as the model group and the internal verification group. To ensure the undifferentiated clustering, we performed an analysis to determine the differences in stratification of various clinical factors. The GEO sample was used as the external validation group to verify the accuracy of prognostic lncRNAs. The analysis excluded RNA that was undetectable in more than 10% of the samples. Concerning each data-set, the gene ID was converted to the corresponding gene symbol according to the corresponding annotation package.

## **Identification of differential genomic instability-associated lncRNAs from LCCs and RCCs**

Initially, we examined differentially expressed genes (DEGs) in LCCs and RCCs from TCGA by the R package “Limma” ( $|\log_2\text{foldchange}| > 0.5$ , false discovery rate (FDR)  $< 0.05$ ) [16]. These DEGs were distributed by the human genome annotation package into mRNAs and lncRNAs. In order to assess the genomic instability, we proposed a mutator hypothesis-derived calculation method: we determined the cumulative number of somatic mutations (CNSMs) on the basis of number of changed sites for each gene on each sample and categorized the patients in descending order. The top 25% of patients were titled with genomic unstable like (GU) group and the last 25% as genomic stable like (GS) group. The differentially expressed lncRNAs between the two groups was evaluated and called as DGIA lncRNAs from LCCs and RCCs ( $|\log_2\text{foldchange}| > 0.5$ , FDR  $< 0.05$ ).

## **Cluster and analyze the TCGA samples according to DGIA**

Hierarchical cluster test (HCA) was performed to verify the grouping effect of DGIA lncRNAs and to batch all TCGA samples according to DGIA by the R package “sparcl” [17]. HCA is the approach commonly used to classify similar samples or variables using Euclidean distances and Ward’s linkage method. The samples were bundled into GU group and GS group by clustering. Subsequently, we explored the two groups on the CNSMs by univariate analysis.

## **Functional enrichment analysis**

To estimate the potential functions of DGIA lncRNAs, two methods were applied to identify mRNAs that were more likely to be co-expressed with them. The first one was used to analyze the co-expression relationship between lncRNAs and mRNAs by Pearson’s correlation tests. Here we designated that the top 10 mRNAs with the highest coefficients have a strong co-expression relationship with each DGIA lncRNAs.

The second one was used to analyzed the DEGs between LCCs and RCCs through weighted gene co-expression network assessment by the R package “WGCNA” [18]. At first, construction of an adjacency matrix (AM) of genes was done using the power function and an appropriate power index was selected. Later AM was converted into a topological overlap matrix. Finally, the gene consensus modules were

collected and performed the correlation analysis with CNSMs. The mRNAs in the modules with the highest absolute correlation coefficient with CNSMs were selected for further examination.

Overall, intersection of the mRNAs was screened by the two methods and accomplished Gene Ontology (GO) functional annotations and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis by R package "clusterProfiler" [19].

## **Construction of DGIA lncRNAs related prognostic model (DRPM)**

To check the effect of DGIA lncRNAs on the prognosis, DGIA lncRNAs was selected by univariate Cox proportional hazard regression analysis (COX). lncRNAs with  $P < 0.05$  in univariate COX were retained and multivariate COX was performed in the model group by the R package "glmnet" [20]. The risk scores (RS) of the model group and the internal validation group were estimated according to the coefficient of each lncRNAs within the model. The patients in TCGA were separated into HRG and LRG with poor prognosis.

## **Validation of the DGIA lncRNAs in DRPM**

Log-rank test was used to expose the difference in survival of HRG and LRG in the model group and internal validation group by R packages "survcomp" [21]. Simultaneously, the predictive effect of DRPM were figured out through the receiver operating characteristics curve (ROC) and the area under the ROC curve (AUC) by the R package "survivalROC" [22]. Additionally, univariate and multivariate COX were utilized to verify the independent predictive effect of the RS obtained by the model.

In the external verification group, DGIA lncRNAs was employed in DRPM to construct a prognostic model again by multivariate COX. The Log-rank test was also used for survival analysis, and time-dependent ROC (timeROC) of 1, 3, and 5 years were plotted. The purpose was to verify the accuracy of prognostic DGIA lncRNAs.

The survival curves of DGIA lncRNAs in DRPM were plotted and the differences were analyzed by log-rank test. The R package "maxstat" [23] was performed to get the best cut-off value.

## **Immune infiltration and gene set enrichment analysis (GSEA) in HRG and LRG**

R package "CIBERSORT" [24] was employed in the TCGA samples to estimate the relative infiltration abundance of 22 immune cells and to assess the variations in the various immune cells' infiltration of HRG and LRG. The results with  $p < 0.05$  were retained. "CIBERSORT" calculated the  $p$ -value of the deconvolution for each sample by Monte-Carlo simulation to provide the assessed confidence. The differences in the abundance of 22 immune cells in HRG and LRG were examined by the Wilcoxon rank-sum test.

Besides, to study the differences in biological functions of genes between HRG and LRG, we downloaded the biological process (BP), molecular function (MF) data-sets related to GO, and KEGG data-set. GSEA was performed using the Bioconductor package “fgsea” [25] with 10,000 permutations between LRG and HRG. The threshold values were  $p < 0.05$ .

## Result

### DEGs and DGIA lncRNAs in LCCs and RCCs

We firstly selected, separated and bundled TCGA samples in furtherance to segregate DGIA lncRNAs from DEGs of LCCs and RCCs. Soon after, the corresponding gene expression data were standardized and analyzed. We obtained 1724 DEGs (Fig. 1a), including 1325 mRNAs and 399 lncRNAs. According to the CNSMs, the top 25% ( $n = 75$ ) and last 25% ( $n = 62$ ) patients were labeled as GU group and GS group, respectively. By analyzing in contrast in lncRNAs between these two groups, 123 DGIA lncRNAs were attained. Among them, 63 lncRNAs showed upregulation whereas 60 exhibited downregulation in the GU group (Fig. 1b, 1c). Based on these DGIA lncRNAs, we carried out unsupervised HCA on TCGA specimens and distributed them into GU group and GS group (Fig. 1d). The CNSMs in both groups were significantly different with its median higher in the GU group comparing with GS group (Fig. 1e). These findings conclusively depicted that the selected DGIA lncRNAs had a marvelous classification effect.

### Functional enrichment analysis for DGIA lncRNAs

To explore the functionalities and pathways concerned with 123 DGIA lncRNAs, we operated functional enrichment analysis on protein-coding genes (PCGs) co-expressed with DGIA lncRNAs. The first method's procedure included a correlation analysis between the selected DGIA lncRNAs and 1,325 differential mRNAs from LCCs and RCCs. The PCGs of DGIA lncRNAs, that is, the top 10 mRNAs with the strongest correlation with each lncRNAs, were achieved.

The second method disclosed, after constructing the co-expression network (Additional file 1: Figure S1), the cognizance of blue module with highest positive correlation, and the turquoise module with highest negative correlation (Fig. 2a). We intersected the selective mRNAs from the blue and turquoise modules with PCGs chosen by the first method. Thereupon, these genes were used to construct a lncRNAs-mRNA co-expression network (Fig. 1f).

The results of functional enrichment analysis with the intersection PCGs comprised DNA-binding transcription activator activity, RNA polymerase II-specific and various phospholipase-related enzyme activities. These molecular functions are closely associated with the formation and development of genomic instability. More importantly, the enrichments of biological processes are mainly related to immune processes, such as T cell activation, lymphocyte differentiation, regulation of T cell activation, etc. (Fig. 2b). KEGG enrichment analysis displayed that both regulating pluripotency of stem cells signaling pathways, as well as immune-related pathways, including Th17 cell differentiation, Th1, and

Th2 cell differentiation, PD – L1 expression, PD – 1 checkpoint pathway in cancer, etc., were significantly enriched (Fig. 2c). These results indicated that 123 DGIA lncRNAs not only cause genomic instability but also influence the regulation of immune system. The variation in the expression of these 123 DGIA lncRNAs potentially disturbs the balance of co-expressed PCGs regulatory network, consequently causing instability in the cell genome, and also affecting the killing of tumors by immune cells, mostly by the proliferation, differentiation, activation, and receptor recognition of T cells. Thus, these DGIA lncRNAs possess astounding potential in immune regulation while affecting gene instability.

## Construction of DRPM using DGIA lncRNAs

The samples from TCGA were randomly and uniformly arranged into model group (n = 162) and validation group (n = 160). The clinical factors were not statistically significantly different in both groups (all  $p > 0.05$ ) (Additional file 2: Table S1). In the model group, we accomplished univariate and multivariate COX to assort and construct DRPM with 123 DGIA lncRNAs, and 6 prognostic-related DGIA lncRNAs with corresponding risk coefficients were determined (Table 1). All patients in TCGA were divided into HRG and LRG on the basis of median of RS (0.851) measured by DRPM in the model group (Additional file 2: Table S2).

Table 1  
DRPM information including 6 DGIA lncRNAs

lncRNAs	Coefficients	HR	95% CI lower	95% CI upper	P value
NKILA	0.198	1.219	1.027	1.447	0.024
AC004009.1	0.316	1.371	1.128	1.668	0.002
AP003555.2	0.377	1.457	1.227	1.731	0.001
BOLA3-AS1	0.329	1.390	1.038	1.860	0.027
LINC00543	0.074	1.077	1.007	1.152	0.030
UCA1	0.013	1.014	1.004	1.023	0.004

DGIA: Differential genomic instability-associated; lncRNAs: Long non-coding RNAs; DRPM: DGIA lncRNAs related prognostic model; HR: Hazard ratio; CI: Confidence interval.

## Validation of the DRPM

To confirm the anticipated effects of DRPM, we conducted Kaplan-Meier test to plot a survival curve. The results demonstrated that the survival outcomes of HRG were worse than those in LRG (all  $p < 0.05$ ) (Fig. 3a). The ROC curves plotted for patients in different groups confirmed the consistency and satisfying categorization effects of DRPM, the AUC were shown in the figures respectively (Fig. 3b). Using RS, we organized the patients in different groups and detected changes in expression level of the prognostic DGIA lncRNAs. The heat map presented the increment in the expression levels of 6 lncRNAs in HRG (Fig. 3c).

To verify the independent predictive effects of RS, we combined RS with clinical factors for univariate and multivariate COX analysis. These clinical factors were age, gender, and TNM stage. The results indicated that RS was an independent prognostic factor (Table 2). Besides, to assess the risk clustering ability of DRPM in different strata, we separately stratified age (< 65 years and  $\geq$  65 years), gender (male and female), and clinical stage (Stage I-II and Stage III-IV). The survival curves of HRG and LRG were plotted through stratification of different clinical factors. HRG and LRG exhibited significant difference in overall survival up to the stratification of age, gender, and Stage I-II (all  $p < 0.05$ ) (Additional file 3: Figure S2). In stratification of Stage III-IV, the difference was very close to reaching statistical significance ( $p = 0.077$ ) (Additional file 3: Figure S2). In summary, the DRPM revealed a consistent and promising prognostic evaluation ability in different stratification.

Table 2  
Univariate and multivariate COX of prognostic factors in different groups

Factors	Univariate COX				Multivariate COX			
	HR	95% CI lower	95% CI upper	P value	HR	95% CI lower	95% CI upper	P value
All patients in TCGA (n = 322)								
Age	1.031	1.010	1.053	0.004	1.043	1.021	1.066	0.001
Gender	1.330	0.828	2.134	0.238	1.149	0.698	1.891	0.584
T	3.555	2.155	5.864	0.001	2.897	1.601	5.240	0.001
N	1.898	1.438	2.506	0.001	1.132	0.804	1.595	0.477
M	4.484	2.740	7.337	0.001	2.985	1.603	5.559	0.001
Risk score	1.228	1.161	1.300	0.001	1.196	1.127	1.269	0.001
Model group (n = 162)								
Age	1.024	0.996	1.053	0.094	1.033	1.002	1.065	0.038
Gender	0.734	0.393	1.369	0.331	0.758	0.388	1.480	0.417
T	2.359	1.235	4.507	0.009	1.465	0.702	3.060	0.309
N	1.904	1.315	2.756	0.001	1.216	0.763	1.940	0.411
M	4.654	2.451	8.838	0.001	3.405	1.500	7.726	0.003
Risk score	1.235	1.158	1.317	0.001	1.160	1.078	1.249	0.001
Internal validation group (n = 160)								
Age	1.040	1.006	1.076	0.021	1.063	1.026	1.102	0.001
Gender	3.060	1.361	6.878	0.007	2.179	0.904	5.248	0.083
T	6.649	3.078	14.362	0.001	9.612	3.662	25.230	0.001
N	2.074	1.313	3.276	0.002	1.378	0.744	2.555	0.308
M	4.349	1.992	9.493	0.001	2.299	0.767	6.891	0.137

COX: Cox proportional hazard regression analysis.

Factors	Univariate COX				Multivariate COX			
	HR	95% CI lower	95% CI upper	<i>P</i> value	HR	95% CI lower	95% CI upper	<i>P</i> value
Risk score	1.178	1.033	1.343	0.015	1.250	1.088	1.436	0.002

COX: Cox proportional hazard regression analysis.

## Validation of the prognostic DGIA lncRNAs

To verify the accuracy of prognostic DGIA lncRNAs, we plotted survival curves for these lncRNAs in TCGA samples. In AC004009.1, AP003555.2, BOLA3-AS1, NKILA, LINC00543 and UCA1, the prognosis of the high expression group was worse as compared to the low expression group (all  $p \leq 0.05$ ) (Fig. 4a). Also, we investigated the correlation between these lncRNAs and stages, and the results indicated that the expression levels of AP003555.2, BOLA3-AS1, NKILA, LINC00543, and UCA1 were significantly different between at least two stages (Fig. 4b).

Meanwhile, in the external validation group from GEO, we constructed a model and grouped patients with the four prognostic DGIA lncRNAs, comprising BOLA3-AS1, NKILA, LINC00543, and UCA1. The prognosis of HRG was also worse than that of LRG ( $p \leq 0.001$ ) (Fig. 5a). The timeROC of 1, 3, and 5 years proved that the model had a promising classification effect, and that of 3 years displayed optimum effects (AUC = 0.83) (Fig. 5b).

## Immune infiltration and GSEA within different risk groups

The above-mentioned enrichment investigation demonstrated that DGIA lncRNAs also influences immune regulation. Hence, we evaluated the differences in the infiltration of 22 immune cells in HRG and LRG according to the results of CIBERSORT. The expression of CD8<sup>+</sup> T cells, out of all TCGA samples and model group, was lower in HRG than in LRG (all  $p \leq 0.05$ ) (Fig. 6). CD8<sup>+</sup> T cells are the cytotoxic immune cells that are capable of directly killing tumor cells, and their abundance differences indicated the immune-related reasons for the prognostic difference in HRG and LRG.

To explore the significantly altered MF, BP, and pathways in HRG and LRG, we performed GO- and KEGG-related GSEA. Mainly immune and genomic instability related pathways in LRG were significantly enriched. In GO enrichment terms, immune-related pathways encompassed response to type I interferon (IFN- $\alpha$ ), natural killer cell activation, T cell activation involved in immune response, etc. Simultaneously, some genomic instability-related pathways were also significantly enriched, including structural constituent of ribosome, transcription elongation from RNA polymerase II promoter, response to dsRNA, and some energy-related pathways in glucose metabolism (Fig. 7a, 7b). In KEGG enrichment terms, besides the regulation of autophagy and cytosolic DNA sensing pathway, which are related to genomic instability, there were also immune-related pathways, including antigen processing and presentation, and

cytokine receptor interaction enriched (Fig. 7c). Finally, we also noticed that the CNSMs of LRG was significantly higher comparing with HRG (all  $p < 0.05$ ) (Additional file 4: Figure S3).

## Discussion

Lately, it has been proclaimed that genomic instability is one of the key prognostic factors for most cancers [26]. Various assays are used to assess the genomic instability by the expression of certain characteristic proteins and mutations of genes [27, 28]. Moreover, with the development of gene sequencing technology, the detection of genomic instability has achieved increased resolution [29]. In recent times, researchers have put a great effort to identify PCGs and microRNAs and to find biomarkers related to genomic instability and prognosis [30, 31]. Simultaneously, the intensive study on lncRNAs also makes researchers aware of their role in genomic stability. Although many works have been done by scientists, identification of DGIA lncRNAs in LCCs and RCCs and their relationship with immunity are still rare. Therefore, we explored the influence of genomic instability in LCCs and RCCs, as well as the key prognostic DGIA lncRNAs.

We identified 123 DGIA lncRNAs from the DEGs of LCCs and RCCs, and the functional analysis on their co-expressed PCGs surprisingly affirmed that these DGIA lncRNAs potentially influenced the genomic instability and immune functions through PCGs. In molecular functions, the accumulation of errors during the transcription of DNA under the action of RNA polymerase is the source of genomic instability in all organisms [32]. Additionally, phospholipase C participates in numerous physiological processes within the cell, especially signal transduction pathways that regulates cell functions and proliferation [33, 34]. These processes also involve genomic mutations and even leads to cancers [35]. Other enriched pathways are mainly related to the positive activation and differentiation of T cells. Thus, we suspected that genomic instability could potentially cause differences in prognosis and immunity of LCCs and RCCs.

Moreover, we investigated whether DGIA lncRNAs can identify differences in immunity while predicting clinical outcomes. Upon identification of DRPM containing 6 key lncRNAs, we successfully divided the patients into HRG and LRG with poor prognosis. From the study of differences in immune infiltration and the GSEA between HRG and LRG, it has been concluded that some pathways related to genetic instability in the LRG are significantly enriched, including regulation of autophagy and glucose metabolism-related pathways (Fig. 7c). Genomic instability could cause the production of a large number of misfolded proteins, and autophagy participate in the degradation of ubiquitinated and misfolded proteins [36]. Autophagy has also been reported to be involved in the regulation of number of centrosomes during cell division to maintain genomic instability [37]. Besides, some pathways are associated with both genetic instability and immunity, such as the cytosolic DNA sensing pathway, response to dsRNA/RIG-I like receptor signaling pathway and Toll-like receptor signaling pathway (Fig. 7a, 7c). In somatic cells, cytoplasmic DNA sensors, after identifying the double-strand DNA (dsDNA), activate the cytosolic DNA sensing pathway and innate immune responses [38]. These dsDNA may be endogenous and are the yields of inaccurate replication of mitochondrial DNA (mtDNA) or micronuclear DNA [39, 40]. DsDNA can

be transformed into double-strand RNA (dsRNA) by the action of RNA polymerase III for recognition by the RNA sensor RIG-I [41, 42]. DsRNA can also be recognized by Toll-like receptors to induce inflammatory cytokines and IFN- $\gamma$  [39, 43]. Some pieces of research disclosed that dsRNA accumulates in the mitochondria as a result of gene deletion during transcription of mtDNA, which promote the production of IFN- $\gamma$  (eliciting innate immune response) after being recognized [40]. These shreds of evidence and our enrichment results explain the potential mechanism of immune activation by genomic instability.

The immune-related pathways are also significantly enriched and CD8<sup>+</sup> T cell infiltration is higher in LRG. Antigen processing and presentation, response to IFN- $\gamma$ , and cytokine related pathways also provide the basis for the immune system activation processes (Fig. 7b). CNSM in LRG is significantly higher than that in HRG (Additional file 4: Figure S3), which indicates that the degree of genetic instability is higher. Mardis suggested that genomic instability could predict immunotherapy response more accurately. Various forms of genomic instability in cancers produce new antigens and immune-responsive phenotypes eventually [14]. In the analysis of LCCs and RCCs, we have confirmed that genomic instability does affect the immune response and the prognosis through a series of potential mechanisms.

Main limitations are as follows: firstly, while using GEO data for the verification of key lncRNAs predictive effects, two key lncRNAs were missed. Although the classification effect was propitious, the verification was not sufficient enough. In this regard, we examined the prognosis and clinical characteristics of all key lncRNAs to support the evidence. Secondly, we defined a mutator hypothesis-derived calculation method to evaluate genomic instability. In depth study is required to corroborate the functionality and significance of this method.

## Conclusion

We constructed DRPM based on the DGIA lncRNAs of LCCs and RCCs. While using DRPM to predict the prognosis, we also deeply investigated the influence and mechanism of genetic instability on immunity. Meanwhile, 6 key DGIA lncRNAs were identified. They can not only predict the prognostic risk of patients but also serve as biomarkers for evaluating the differences of genetic instability and immune infiltration. The findings our research can provide some basis for identifying the benefits of immunotherapy through genetic instability.

## Abbreviations

CC: Colon cancer; LCCs and RCCs: Left-sided and right-sided CCs; MSI: microsatellite instability ; lncRNAs: Long non-coding RNAs ; DGIA: Differential genomic instability-associated ; TCGA: The Cancer Genome Atlas ; HRG and LRG: High- and low-risk groups; GEO: Gene Expression Omnibus ; COAD: Colon adenocarcinoma ; DEGs: Differentially expressed genes ; FDR: False discovery rate ; CNSMs: Cumulative number of somatic mutations ; GU: Genomic unstable like ; GS: Genomic stable like ; HCA: Hierarchical cluster test ; AM: Adjacency matrix ; GO: Gene Ontology ; KEGG: Kyoto Encyclopedia of Genes and Genomes ; DRPM: DGIA lncRNAs related prognostic model ; COX: Cox proportional hazard regression

analysis ; RS: Risk scores ; ROC: Receiver operating characteristics curve ; AUC: Area under the ROC curve ; TimeROC: Time-dependent ROC ; GSEA: Gene set enrichment analysis ; BP: Biological process ; MF: Molecular function ; PCGs: Protein-coding genes ; IFN- $\gamma$ : Type I interferon ; DsDNA Double-strand DNA ; MtDNA: Mitochondrial DNA ; DsRNA: Double-strand RNA.

## Declarations

### Acknowledgements

Not applicable.

### Author contributions

BC and YL designed the study. JG and TX drafted the manuscript. JG and SD collected, analyzed, and interpreted the data. JG, TX and SD drew the figures and tables. BC and YL helped with the final revision of the article. All authors have read and approved the final manuscript.

### Funding

This work was supported by Natural Science Foundation of Heilongjiang Province of China (ZD2017019), Nn10 Program of Harbin Medical University Cancer Hospital (Nn102017-02), the Post-doctoral Scientific Research Developmental Fund of Heilongjiang (LBH-Q18085) and Harbin Medical University Cancer Hospital Preeminence Youth Fund (JCQN2019-04).

### Availability of data and materials

Publicly available datasets were analyzed in this study. This data can be found here: <https://portal.gdc.cancer.gov>, <https://www.ncbi.nlm.nih.gov/gds/>.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

All authors read the final manuscript and agreed to publish it.

### Competing interests

The authors declare no conflict of interest.

### Author details

<sup>1</sup> Department of Colorectal Surgery, Harbin Medical University Cancer Hospital, Harbin 150086, P. R. China. <sup>2</sup> Department of Anesthesiology, The Fourth Affiliated Hospital of Harbin Medical University,

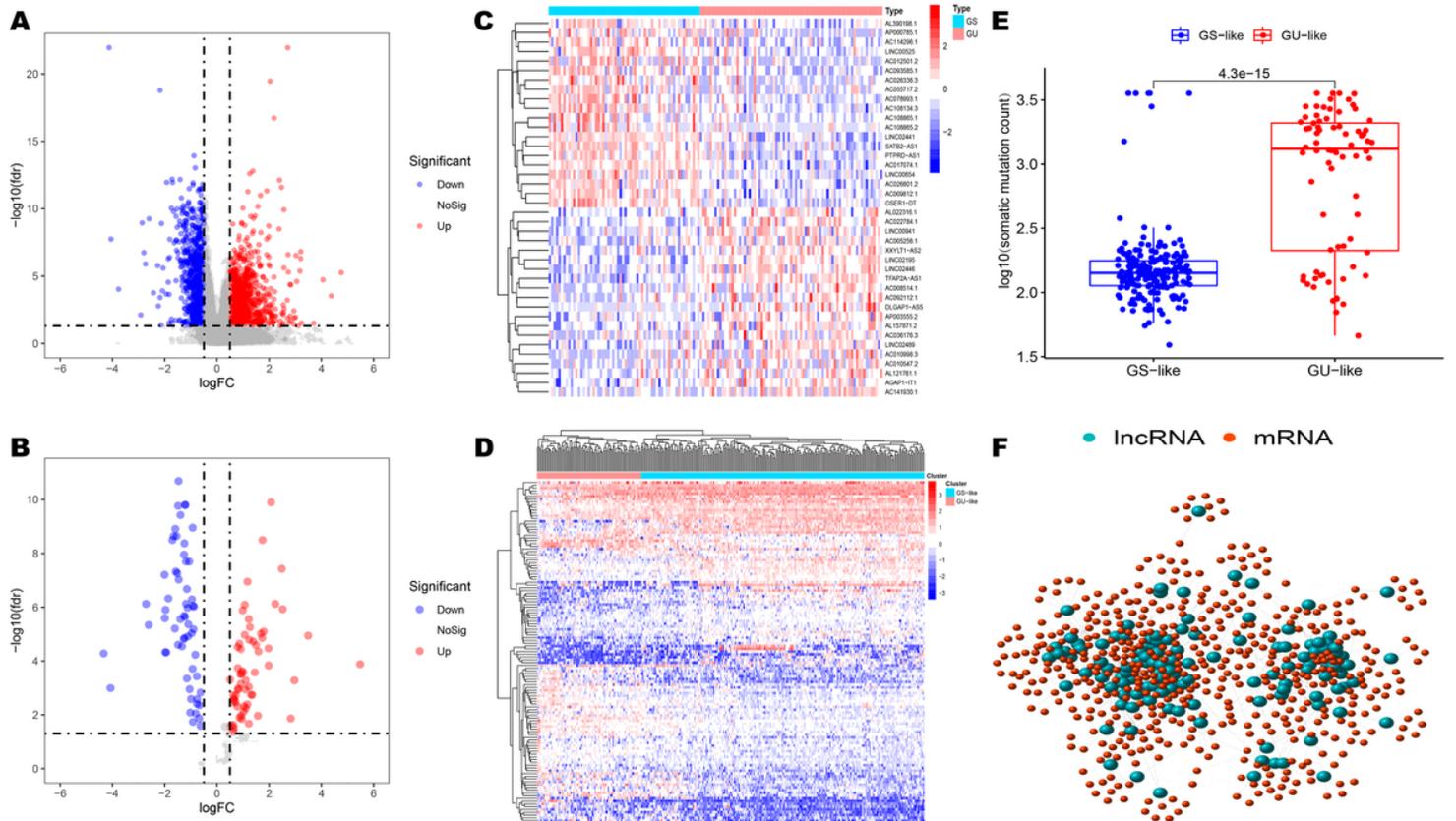
## References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* 2020;70(1):7–30.
2. Verkuijl SJ, Jonker JE, Trzpis M, Burgerhof JGM, Broens PMA, Furnee EJB. Functional outcomes of surgery for colon cancer: A systematic review and meta-analysis. *Eur J Surg Oncol* 2020.
3. Wu C. Systemic Therapy for Colon Cancer. *Surg Oncol Clin N Am.* 2018;27(2):235–42.
4. Grass F, Lovely JK, Crippa J, Ansell J, Hubner M, Mathis KL, et al. Comparison of recovery and outcome after left and right colectomy. *Colorectal Dis.* 2019;21(4):481–6.
5. Blakely AM, Lafaro KJ, Eng OS, Ituarte PHG, Fakih M, Lee B, et al. The Association of Tumor Laterality and Survival After Cytoreduction for Colorectal Carcinomatosis. *J Surg Res.* 2020;248:20–7.
6. Mjelle R, Sjursen W, Thommesen L, Saetrom P, Hofslie E. Small RNA expression from viruses, bacteria and human miRNAs in colon cancer tissue and its association with microsatellite instability and tumor location. *BMC Cancer.* 2019;19(1):161.
7. Guinney J, Dienstmann R, Wang X, de Reynies A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med.* 2015;21(11):1350–6.
8. Molina-Cerrillo J, San Roman M, Pozas J, Alonso-Gordoa T, Pozas M, Conde E, et al. BRAF Mutated Colorectal Cancer: New Treatment Approaches. *Cancers (Basel)* 2020, 12(6).
9. Requena DO, Garcia-Buitrago M. Molecular Insights Into Colorectal Carcinoma. *Arch Med Res.* 2020;51(8):839–44.
10. Bebenek A. [DNA replication fidelity]. *Postepy Biochem.* 2008;54(1):43–56.
11. Kane DP, Shcherbakova PV. A common cancer-associated DNA polymerase epsilon mutation causes an exceptionally strong mutator phenotype, indicating fidelity defects distinct from loss of proofreading. *Cancer Res.* 2014;74(7):1895–901.
12. Lichtenstern CR, Ngu RK, Shalapour S, Karin M. Immunotherapy. Inflammation and Colorectal Cancer. *Cells* 2020, 9(3).
13. Boulianne B, Feldhahn N. Transcribing malignancy: transcription-associated genomic instability in cancer. *Oncogene.* 2018;37(8):971–81.
14. Mardis ER. Neoantigens and genome instability: impact on immunogenomic phenotypes and immunotherapy response. *Genome Med.* 2019;11(1):71.
15. Thapar R. Regulation of DNA Double-Strand Break Repair by Non-Coding RNAs. *Molecules* 2018, 23(11).
16. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.

17. Witten DM, Tibshirani R. A framework for feature selection in clustering. *J Am Stat Assoc.* 2010;105(490):713–26.
18. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.
19. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284–7.
20. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw.* 2010;33(1):1–22.
21. Schroder MS, Culhane AC, Quackenbush J, Haibe-Kains B. survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics.* 2011;27(22):3206–8.
22. Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics.* 2005;61(1):92–105.
23. Laska E, Meisner M, Wanderling J. A maximally selected test of symmetry about zero. *Stat Med.* 2012;31(26):3178–91.
24. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* 2015;12(5):453–7.
25. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545–50.
26. Andor N, Maley CC, Ji HP. Genomic Instability in Cancer: Teetering on the Limit of Tolerance. *Cancer Res.* 2017;77(9):2179–85.
27. Cortes-Ciriano I, Lee S, Park WY, Kim TM, Park PJ. A molecular portrait of microsatellite instability across multiple cancers. *Nat Commun.* 2017;8:15180.
28. Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, Zou X, et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat Med.* 2017;23(4):517–25.
29. Stadler ZK, Battaglin F, Middha S, Hechtman JF, Tran C, Cercek A, et al. Reliable Detection of Mismatch Repair Deficiency in Colorectal Cancers Using Mutational Load in Next-Generation Sequencing Panels. *J Clin Oncol.* 2016;34(18):2141–7.
30. Habermann JK, Brucker CA, Freitag-Wolf S, Heselmeyer-Haddad K, Kruger S, Barenboim L, et al. Genomic instability and oncogene amplifications in colorectal adenomas predict recurrence and synchronous carcinoma. *Mod Pathol.* 2011;24(4):542–55.
31. Mettu RK, Wan YW, Habermann JK, Ried T, Guo NL. A 12-gene genomic instability signature predicts clinical outcomes in multiple cancer types. *Int J Biol Markers.* 2010;25(4):219–28.
32. Khristich AN, Mirkin SM. On the wrong DNA track: Molecular mechanisms of repeat-mediated genome instability. *J Biol Chem.* 2020;295(13):4134–70.
33. Owusu Obeng E, Rusciano I, Marvi MV, Fazio A, Ratti S, Follo MY, et al. Phosphoinositide-Dependent Signaling in Cancer: A Focus on Phospholipase C Isozymes. *Int J Mol Sci* 2020, 21(7).

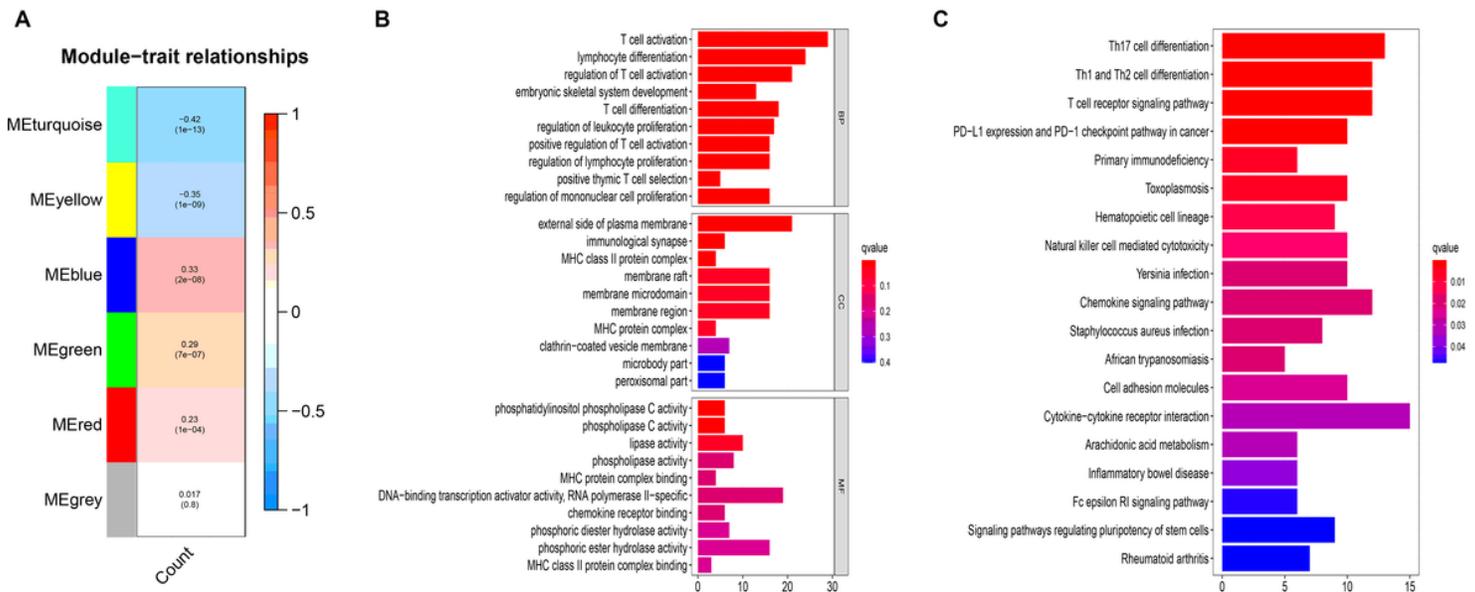
34. Jang HJ, Suh PG, Lee YJ, Shin KJ, Cocco L, Chae YC. PLCgamma1: Potential arbitrator of cancer progression. *Adv Biol Regul.* 2018;67:179–89.
35. Koss H, Bunney TD, Behjati S, Katan M. Dysfunction of phospholipase Cgamma in immune disorders and cancer. *Trends Biochem Sci.* 2014;39(12):603–11.
36. Matsumoto G, Wada K, Okuno M, Kurosawa M, Nukina N. Serine 403 phosphorylation of p62/SQSTM1 regulates selective autophagic clearance of ubiquitinated proteins. *Mol Cell.* 2011;44(2):279–89.
37. Watanabe Y, Honda S, Konishi A, Arakawa S, Murohashi M, Yamaguchi H, et al. Autophagy controls centrosome number by degrading Cep63. *Nat Commun.* 2016;7:13508.
38. Kwon J, Bakhoun SF. The Cytosolic DNA-Sensing cGAS-STING Pathway in Cancer. *Cancer Discov.* 2020;10(1):26–39.
39. Kawai T, Akira S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat Immunol.* 2010;11(5):373–84.
40. Dhir A, Dhir S, Borowski LS, Jimenez L, Teitell M, Rotig A, et al. Mitochondrial double-stranded RNA triggers antiviral signalling in humans. *Nature.* 2018;560(7717):238–42.
41. Samuel CE. Adenosine deaminase acting on RNA (ADAR1), a suppressor of double-stranded RNA-triggered innate immune responses. *J Biol Chem.* 2019;294(5):1710–20.
42. Hur S. Double-Stranded RNA. Sensors and Modulators in Innate Immunity. *Annu Rev Immunol.* 2019;37:349–75.
43. Miyake K, Shibata T, Ohto U, Shimizu T, Saitoh SI, Fukui R, et al. Mechanisms controlling nucleic acid-sensing Toll-like receptors. *Int Immunol.* 2018;30(2):43–51.

## Figures



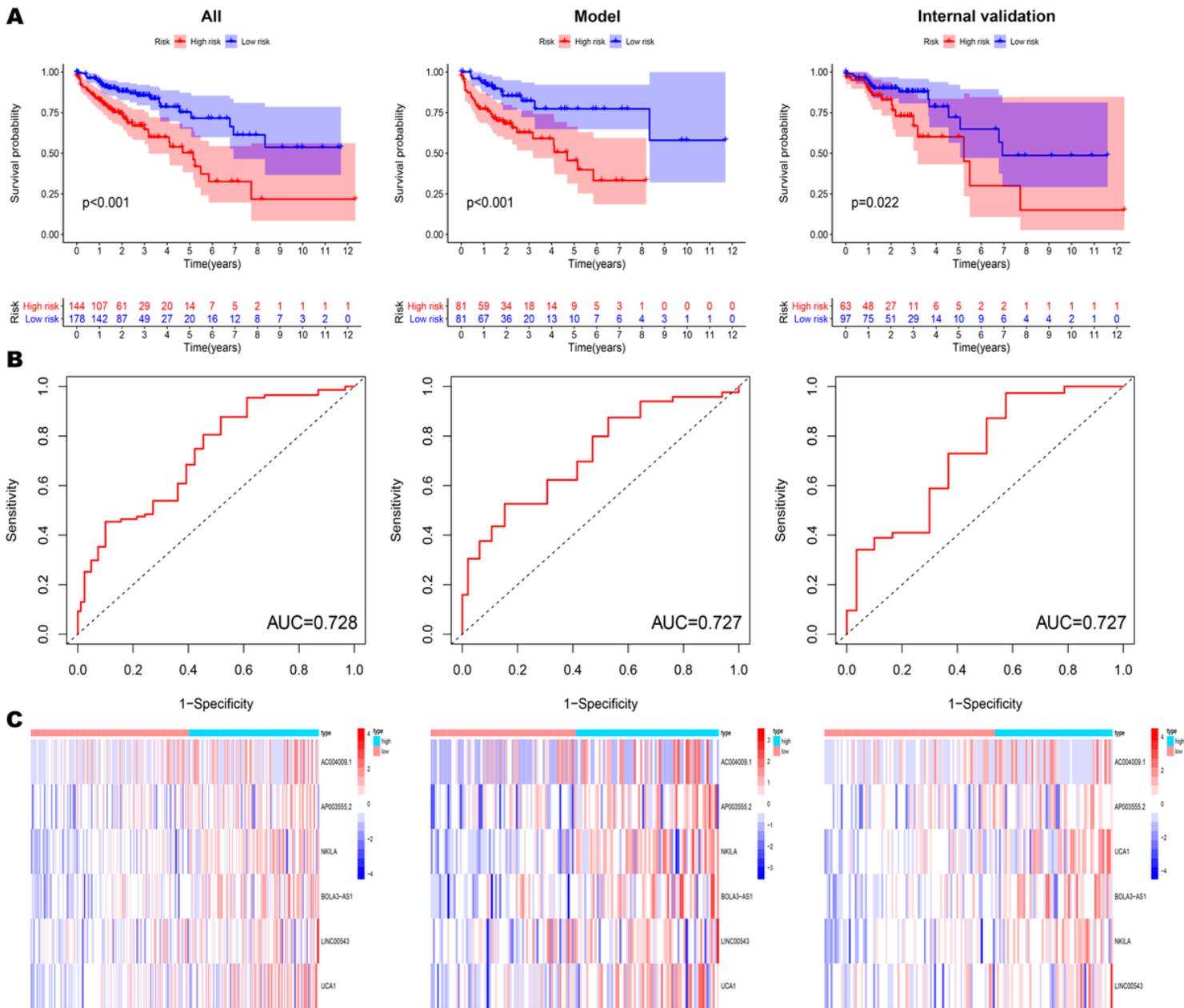
**Figure 1**

a Differentially expressed genes between LCCs and RCCs. b Differentially expressed DGIA IncRNAs between GU group and GS group. Red and blue circles indicate high and low genes expression, respectively. c Heat map depicts the differentially expressed DGIA IncRNAs in TCGA patients. d Unsupervised clustering of TCGA patients based on the expression pattern of 128 candidate DGIA IncRNAs. e In boxplot, cumulative number of somatic mutations in the GU-like group are significantly higher than those in the GS-like group. f Co-expression network of DGIA IncRNAs and intersection of mRNAs based on two methods. The red circles represent IncRNAs, and the blue circles represent mRNAs. LCCs and RCCs: Left and right-sided colon adenocarcinoma; DGIA IncRNAs: Genomic instability-associated long non-coding RNAs; GU: Genomic unstable; GS: Genomic stable.



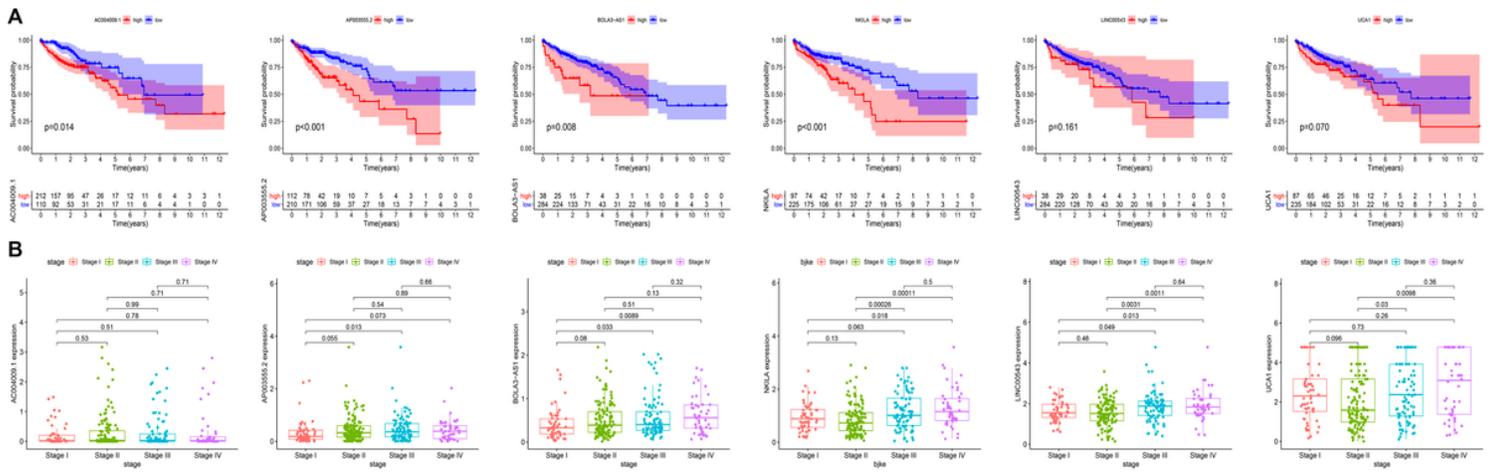
**Figure 2**

a Correlation between the gene modules and CNSMs. Each cell contains corresponding correlation coefficient and p-value. The correlation coefficient decreased in size from red to blue. b GO functional annotations for mRNAs co-expressed with lncRNAs. c KEGG enrichment analysis for mRNAs co-expressed with lncRNAs. CNSMs: Cumulative number of somatic mutations; GO: Gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes.



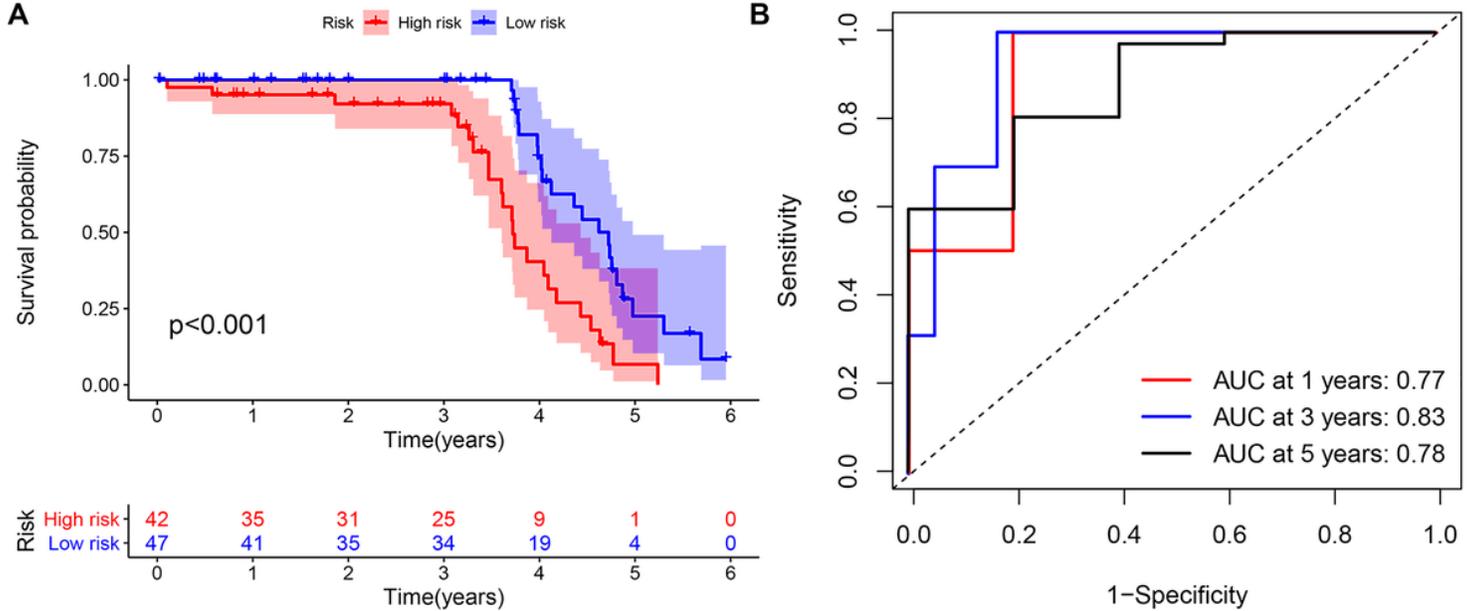
**Figure 3**

a Kaplan-Meier curves of overall survival in different groups. b ROC curves in different groups. c The heat map of 6 key lncRNA expression patterns with increasing risk score. ROC: Receiver operating characteristics.



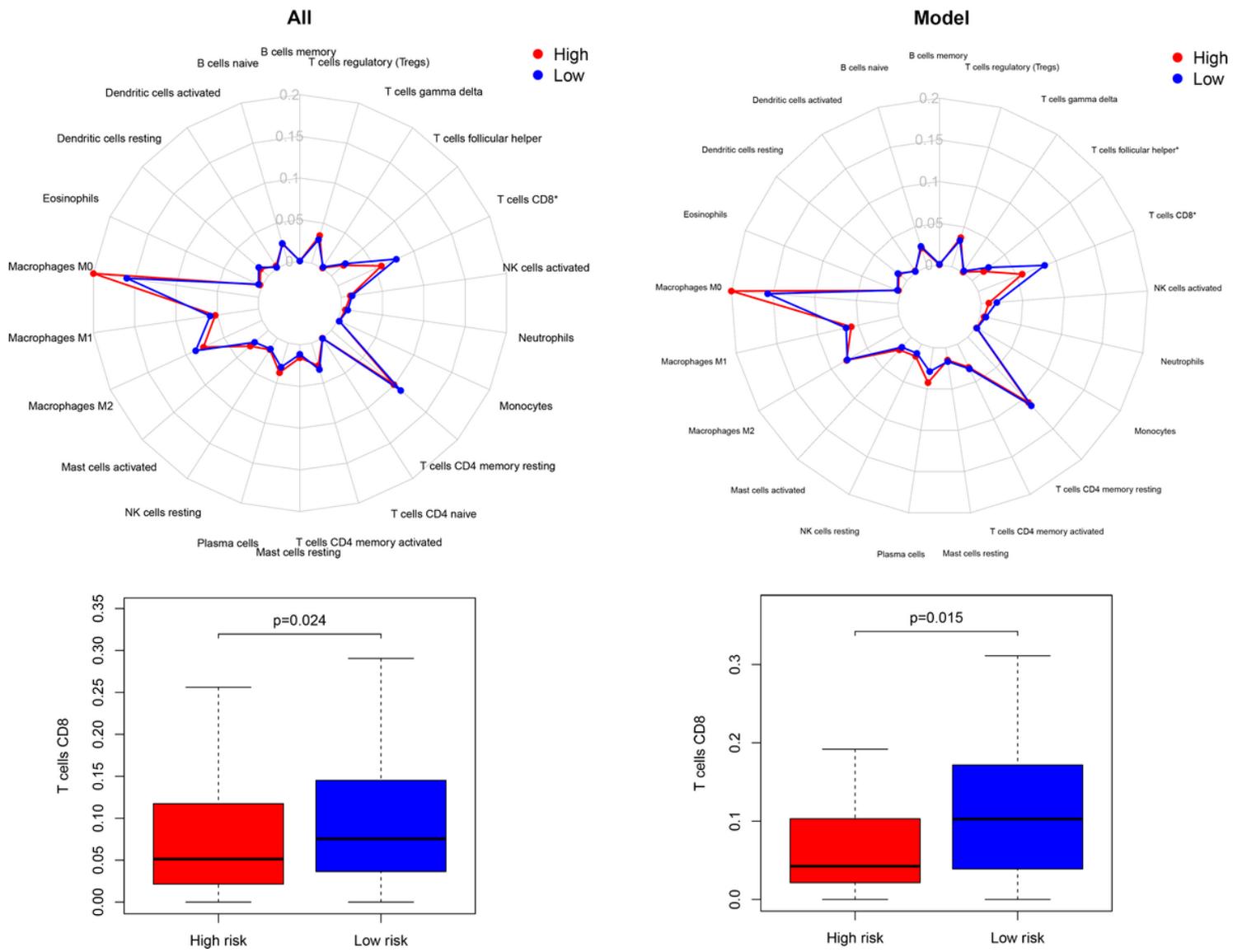
**Figure 4**

a Kaplan-Meier curves of overall survival in 6 key DGIA lncRNA. b The correlation between 6 key DGIA lncRNA and pathologic stages.



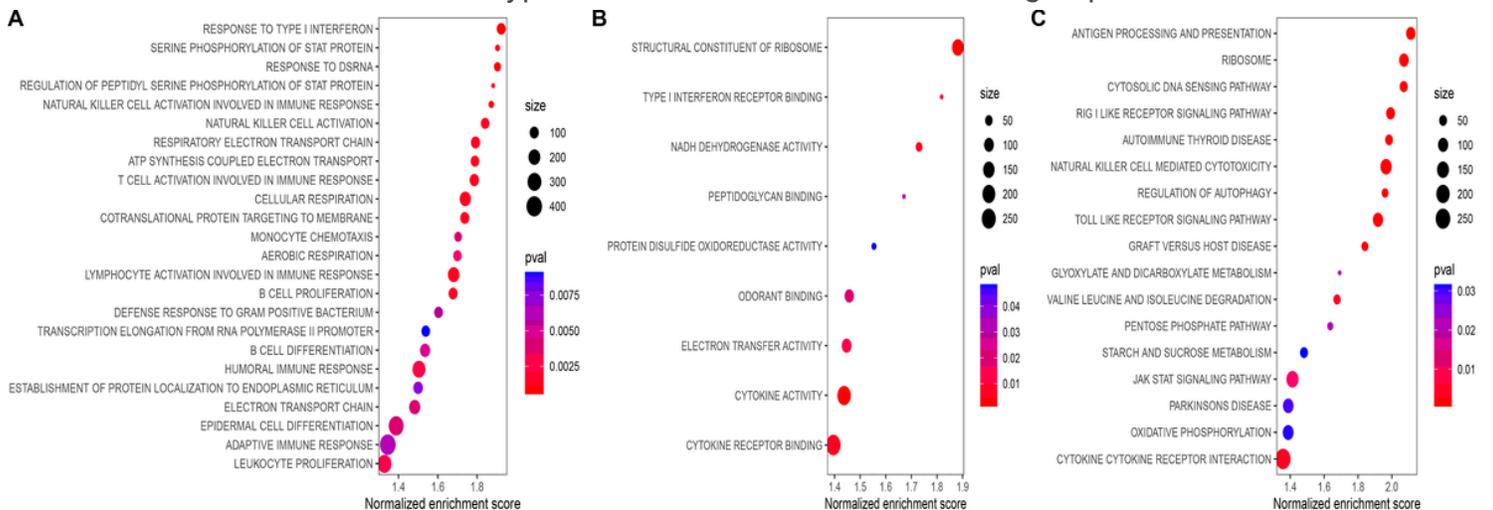
**Figure 5**

a Kaplan-Meier curve of overall survival in external validation group. b TimeROC curves for 1, 3, 5 years in external validation group. TimeROC: Time-dependent ROC.



**Figure 6**

The differences of 22 immune cell types abundance within different risk groups.



**Figure 7**

a BP of GO-related GSEA between different risk groups. b MF of GO-related GSEA between different risk groups. c KEGG-related GSEA between different risk groups. BP: Biological process; MF: Molecular function.

## Supplementary Files

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

- [FigureS1.tif](#)
- [TableS1andS2.docx](#)
- [FigureS2.tif](#)
- [FigureS3.tif](#)