

Replication Protein A in Cancer: A Pan-Cancer Analysis Focusing on Lower-Grade Glioma and Liver Hepatocellular Carcinoma

Ravi Bhushan

Centre of Biomedical Research

Arunima Dhurve

Centre of Biomedical Research

Prachi Gupta

Centre of Biomedical Research

Shishir K Gupta

Centre of Biomedical Research

Prabhaker Yadav

`prabhaker@cbmr.res.in`

Centre of Biomedical Research

Research Article

Keywords: Pan-cancer, RPA, Therapeutics, DNA damage, DNA Repair, Mutation

Posted Date: July 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10187322/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Egyptian Liver Journal on September 21st, 2026. See the published version at <https://doi.org/10.1186/s43066-026-00555-y>.

Replication Protein A in Cancer: A Pan-Cancer Analysis Focusing on Lower-Grade Glioma and Liver Hepatocellular Carcinoma

Ravi Bhushan^{†1}, Arunima Dhurve^{†1}, Prachi Gupta¹, Shishir K Gupta^{2,3}, Prabhaker Yadav^{1,3*}

[†]Contributed equally to this work

¹Department of Systems Biology, Centre of Biomedical Research, Sanjay Gandhi Post-Graduate Institute of Medical Sciences Campus, Raebareli Road, Lucknow, 226014, Uttar Pradesh, India

²Department of Data Science, Centre of Biomedical Research, Sanjay Gandhi Post-Graduate Institute of Medical Sciences Campus, Raebareli Road, Lucknow, 226014, Uttar Pradesh, India

³Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, Uttar Pradesh, India

Running Title: A Comparative Pan-cancer Analysis of RPA Gene Family

***Correspondence:**

Dr. Prabhaker Yadav
Assistant Professor
Department of Systems Biology
Centre of Biomedical Research
Email: prabhaker@cbmr.res.in
Mob.: - 8176933350

Abstract

Background

Replication Protein A (RPA) is an important single-stranded DNA-binding complex that is crucial for DNA repair and thereby maintains genome stability. However, a comprehensive pan-cancer analysis revealing the role of the RPA gene family in cancer progression remains lacking.

Methods In this study, we performed a comprehensive pan-cancer analysis of the RPA gene family. Briefly, gene expression analysis, genetic mutation analysis, survival prediction analysis, functional annotation, tumor microenvironment analysis, and drug sensitivity analysis of RPA genes were performed using the TCGAplot R package and various web tools, including GEPIA3, TIMER3, HPA, cBioPortal, UALCAN, and GSCALite. TCGA, CPTAC, GTEx, GDSC, and CTRP datasets were utilized for various analysis. All statistical analysis were performed using R.

Results The results revealed that RPA genes were overexpressed across multiple cancer types, including liver hepatocellular carcinoma (LIHC), and brain lower-grade glioma (LGG). High expression of RPA genes was associated with worse prognosis in LGG and LIHC. Correlation analysis with immunomodulators revealed positive correlations. Poor prognosis, along with a positive correlation, indicates an immune-inflamed yet immunosuppressive tumor microenvironment. TMB/MSI analysis revealed a very weak positive correlation, indicating low immunogenicity; however, drug sensitivity analysis showed a strong positive correlation, indicating susceptibility to targeted therapeutic strategies.

Conclusions Overexpression of RPA genes was significantly associated with a worse prognosis, an immunosuppressive environment, and a weak immunotherapeutic response in liver cancers and gliomas. Furthermore, RPA expression was sensitive to various drugs, such as doxorubicin; hence, RPA genes may serve as potential prognostic and therapeutic biomarkers in LGG and LIHC.

Keywords ·Pan-cancer ·RPA ·Therapeutics ·DNA damage ·DNA Repair ·Mutation

32 **Introduction**

33 Maintaining genomic integrity is fundamental to cellular survival. Human cells employ DNA
34 damage surveillance and repair systems that monitor the genome and signal if any error occurs.
35 Various endogenous and exogenous factors, such as reactive oxygen species (ROS), can disrupt
36 this balance and induce DNA damage, leading to genomic instability. Cells rely on multiple
37 DNA damage-repair pathways that work in an integrated manner to restore normal DNA
38 structure and function. However, if the damage exceeds the repair capacity, errors can
39 accumulate as mutations. Over time, unrepaired mutations may lead to cellular transformation
40 and increase the risk of cancer [1]. Among all DNA damages, double-strand breaks (DSBs) are
41 the most severe, as unrepaired DSBs can cause chromosome loss, rearrangements, apoptosis,
42 or malignant transformation. Cells primarily resolve DSBs through three main pathways:
43 nonhomologous end joining (NHEJ), homologous recombination (HR), and single-strand
44 annealing (SSA). NHEJ is a low-fidelity process that directly rejoins broken ends. The other
45 two mechanisms are homology-based; they start with generating 3' single-stranded DNA tails.
46 While HR uses these tails to invade a homologous sequence, SSA relies on annealing between
47 complementary ss-tails formed on opposite sides of the break [2]. Single-stranded DNA
48 (ssDNA) plays a key role as an intermediate during DNA metabolic processes. This ssDNA,
49 being unstable, is vulnerable to nucleases digestion and forms secondary structures that can
50 halt normal DNA processing. Consequently, the unprotected ssDNA jeopardizes the stability
51 of genetic material. Replication Protein A (RPA) is the primary ssDNA-binding protein in
52 human cells that binds ssDNA with exceptionally high affinity ($K_d \sim 10^{-9}$ – 10^{-11} M). The
53 formation of ssDNA signals the recruitment of RPA, which in turn recruits additional factors
54 to the site of damage to support efficient replication, transcription, repair, and recombination.
55 RPA protects ssDNA digestion from nucleases, destabilizes secondary structures, and facilitates

checkpoint activation. Due to these vital functions, disruption of RPA activity can lead to the accumulation of mutations, chromosomal instability, and oncogenic progression [3].

RPA is a heterotrimer composed of three subunits: RPA1, RPA2, and RPA3, with molecular weights of 70, 34, and 14 kDa, respectively. The RPA complex contains six oligonucleotide-binding (OB) folds that bind to DNA via linkers called DNA-binding domains (DBD A-F). RPA1 can have up to four DBDs (A, B, C, and F), while RPA2 and RPA3 have one each, DBD-D and E, respectively. The trimerization core is formed by the interaction of one domain from each subunit (DBD-C, D, and E) with the other domains (DBD-A, B, and F), and the latter are involved in RPA's association with ssDNA to form a stable structure. Studies suggest that RPA3 directly interacts with ssDNA. DBD-C contains a zinc-finger motif that helps stabilize the structure and enhance RPA's DNA-binding activity. Protein-protein interactions of RPA are mediated through the N-terminal domains of RPA1 and RPA2, whereas the primary phosphorylation site for proteins resides on the N-terminal domain of RPA2 [4]. RPA4, a 30 kDa subunit of the RPA complex, is a homolog of the RPA2 subunit. It does not drive DNA replication but can bind to DNA repair sites and reduce the levels of DNA damage markers. Thus, the principal role of RPA4 is to maintain genomic stability under stress or damage, rather than to initiate replication.

Materials and methods

RPA Data Retrieval

Transcriptomic, clinical, and proteomic data for RPA genes across 33 cancer types were accessed in October 2025 from The Cancer Genome Atlas [5], Human Protein Atlas (HPA), and Clinical Proteomic Tumor Analysis Consortium (CPTAC). In accordance with the respective data access policies, data was retrieved.

RPA Gene Expression Analysis

Differential expression analyses of the RPA gene across various cancer types were performed using the TCGAplot R package with TCGA datasets. Protein expression levels of RPA genes were assessed through the UALCAN web tool [6] utilizing CPTAC proteomic datasets [7].

Genetic Mutation Mapping

cBioportal [8] was used for conducting mutational profiling. Somatic mutation and variant data were obtained from the TCGA PanCancer Atlas Studies and the UniProt database [9]. Three-dimensional structure of RPA1, RPA2, RPA3 and RPA4 were downloaded from the RCSB Protein Data Bank [10].

Prognostic Analysis of Survival

GEPIA2 [11] The TIMER3 website was used for survival analysis. GEPIA was accessed, and the “Survival Map” under the “Survival Analysis” module was navigated to, and gene symbols were entered into the search interface to obtain overall survival (OS) data across all TCGA cancer types. Similarly, in TIMER, “Gene” module under the immune association tab was accessed, and PFS (Progression Free Survival), PFI (Progression Free Interval), and DSS (Disease Specific Survival) were analysed for all RPA genes.

Functional Analysis of RPA-associated genes

The GEPIA web server was utilized to identify the top 100 genes similar to RPA1, RPA2, RPA3, and RPA4. Platforms like Gprofiler [12] and Metascape [13] were used to perform all the enrichment analyses.

Sub-cellular localization of RPA genes

Sub-cellular localization data for RPA genes were obtained from the Human Protein Atlas (HPA) database [14]. Immunofluorescent images from multiple cell lines, including A431

103 (squamous carcinoma), A549 (Lung adenocarcinoma), U-2 OS (osteosarcoma), U251MG
104 (Glioblastoma), HPA1 (near-haploid cell line derived from the KBM-7 cell line), HeLa
105 (cervical adenocarcinoma), and U2CS (osteosarcoma), demonstrating RPA gene localization
106 were retrieved, analyzed, and interpreted.

107 **Genomic instability and mutation burden analysis**

108 In cancer immunogenomics, tumor mutational burden (TMB) and microsatellite instability are
109 widely used to predict immunotherapy response. Immunotherapy response associated with
110 RPA genes was evaluated using TMB and MSI analyses. All analyses and data visualisation
111 were performed using TCGAplot (R package) in R version 4.0.3 and ggplot2 (v3.3.2).

112 **Tumor immunoregulators correlation mapping**

113 To investigate associations with immune checkpoint molecules, including immunoinhibitors,
114 immunostimulators, chemokines, and chemokine receptors, the relationship between RPA gene
115 expression and tumor immunoregulatory genes was examined. TCGA plot R package and
116 ggplot2 in R version 4.5.1 were used to perform analysis and visualizations.

117 **Immune Infiltration Profiling**

118 To characterize the immune landscape within the tumor microenvironment (TME), the
119 ESTIMATE algorithm [15] was employed to evaluate levels of immune cell infiltration in
120 TCGA samples. Data analysis and visualization were performed using the TCGAplot R
121 package; ggplot2 was used to generate graphical output.

122 **Drug sensitivity analysis**

123 Gene Set Cancer Analysis (GSCA) [16] was used for analyzing the correlation between gene
124 expression and drug sensitivity. In which data from GDSC and CTRP is available. The Half
125 Maximal Inhibitory Concentration (IC50) of 265 small molecules in 860 cell lines and their

corresponding mRNA gene expression were analysed. The mRNA expression data and drug sensitivity data were merged.

Pearson correlation analysis was performed to get the correlation between gene mRNA expression and drug IC50. FDR adjusted P-value. GSCA is an integrated platform for genomic, pharmacogenomic, and immunogenomic cancer analysis using gene sets.

Statistical Analysis and Correlation Assessment

Differential gene expression comparison was performed using the Wilcoxon rank-sum or Kruskal-Wallis test. Kaplan-Meier survival and multivariate Cox regression were used to assess prognostic relationships. To evaluate relationships among continuous variables, Pearson correlation was used. Immune infiltration, TMB, MSI, and immunoregulatory gene correlations were carried out using the TCGA plot R package to ensure uniformity across all computational evaluations. P-values were corrected using the False Discovery Rate (FDR). Figure 1 depicts the pipeline used for this study.

Results

Gene Expression Profiling of RPA genes

The transcriptomic and proteomic expression levels of RPA1, RPA2, RPA3, and RPA4, members of the RPA gene family, were systematically analyzed and compared across 33 cancer types and corresponding normal tissues using data from the TCGA, GTEx, and CPTAC datasets.

Data revealed significant expression of RPA1, RPA2, RPA3, and RPA4 across all cancer types (**Fig. 2A-D, Supplementary Fig. 1**). RPA1, RPA2, and RPA3 were significantly overexpressed across the largest number of cancers, including those related to the kidney, liver, and brain (**Fig. 2A-D**). Overexpression of RPA1, RPA2, and RPA3 was significant in 10, 12, and 13 cancers,

respectively (Table 1). Overall, a significant increase in the expression of RPA1, RPA2, and RPA3 was observed across the largest number of cancers, particularly kidney cancer (KICH, KIRC, KIRP), liver cancer (LIHC), and gliomas (LGG). Similarly, the anatomical expression analysis also revealed that RPA genes were broadly overexpressed among multiple organs (**Fig. 2E-1H**).

Table 1 Summary of RPA gene expression in all cancer types

Gene	Number of Cancers with Significant higher expression in Paired Box-dot-Plot
RPA1	10
RPA2	12
RPA3	13
RPA4	12

Furthermore, RPA1, RPA2, and RPA3 were overexpressed at the protein level, consistent with their corresponding mRNA expression in the majority of cancers, including LGG and LIHC (**Supplementary Fig. 2**).

Mutational Landscape of RPA genes

Alteration frequency data were analysed to characterise the mutational landscape of RPA1, RPA2, RPA3, and RPA4. Across the majority of cancer types (in 13 cancers), RPA1 exhibited alteration frequencies exceeding 2% (**Fig. 3A**). In contrast, alteration frequencies for RPA2, RPA3, and RPA4 were below 2% in most cancer types (**Fig. 3C, 3E, 3G**). Additionally, these mutations are associated with worse prognosis in case of RPA1 (HR = 1.04) (**Fig. 3B**), and RPA2 (HR = 1.1) (**Fig. 3D**), while better prognosis in case of RPA3 (HR = 0.49) (**Fig. 3F**) and RPA4 (HR = 0.94) (**Fig. 3H**).

Further analysis of TCGA mutation data revealed distinct mutational patterns across the RPA gene family. RPA1 exhibited 62 missense, 4 truncating, 12 splice 3 fusion, and no in-frame

mutation. RPA2 showed 30 missense, 3 truncating, 1 splice, 1 fusion, and no in-frame mutation. RPA3 harboured 13 missense, 4 truncating, 1 splice, 2 fusion, and no in-frame mutation. In contrast, RPA4 presented with 76 missense and 6 truncating mutations, with no in-frame, splice, or fusion mutations detected (**Supplementary Fig. 3, Table 2**). Further, alpha-missense variant prediction also revealed the highest number of pathogenic variants in RPA1 (3278) and RPA2 (1553) compared to RPA3 and RPA4 (**Table 2**).

Table 2 Summary of different types of mutations present in RPA genes

Gene	Missense	Truncating	In-frame	Splice	Fusion	Total	Alpha Missense Variants (Pathogenic)
RPA1	62	4	0	12	3	81	3278
RPA2	30	3	0	1	1	35	1553
RPA3	13	4	0	1	2	16	314
RPA4	76	6	0	0	0	82	87

Clinical Prognostic Value of RPA

Using expression data for each of the four RPA genes, univariate and multivariate Cox proportional hazards regression analysis were conducted to evaluate the prognostic significance of RPA genes in various cancers. For this OS, PFS, PFI, and DSS were performed across 33 cancer types. Univariate analysis revealed that elevated expression of RPA1, RPA2, and RPA3 is associated with worse OS in LGG, while their high expression is associated with better OS in KIRC (**Fig. 4A**). Multivariate analysis also revealed similar findings in the case of PFI (**Fig. 4B**). Further, RPA2, RPA3, and RPA4 were associated with worse survival in LIHC (**Fig. 4A**). A summary of overall survival analysis is given in **Table 3**.

Table 3 Summary of overall survival (OS) analysis

	Cancer with worse overall survival (OS)	Cancer with better OS
RPA1	ACC, LGG, KICH, MESO	KIRC

RPA2	LAML, LGG, LIHC	KIRC, CESC
RPA3	GBM, KICH, LGG, LIHC	KIRC
RPA4	KIRP, LIHC, KIRC, SARC	OV

The prognostic relevance of RPA1, RPA2, RPA3, and RPA4 was further illustrated in these cancers using a Kaplan-Meier (KM) survival plot (**Supplementary Fig. 4**). Collectively, these results demonstrate that RPA genes are strong negative prognostic biomarkers in LGG and LIHC.

Localisation and Functional Annotation of RPA genes

We next investigated the subcellular localization and functional role of the RPA gene across multiple cancer cell lines, with particular emphasis on their distribution within the nucleus, microtubules, and endoplasmic reticulum (ER). For RPA1, subcellular localization was examined in A431 (Human epidermoid carcinoma cell line), A549 (human lung adenocarcinoma cell line), and U2OS (human osteosarcoma) cell lines (**Fig. 5A**). RPA2 localization was evaluated in A431, U251MG, and U2OS cells (**Fig. 5B**). In contrast, RPA3 localization was assessed in HPA1, HeLa (cervical adenocarcinoma), and U2OS cell lines (**Fig. 5C**).

The distinct subcellular localization patterns of RPA genes reflect their diverse functional roles. RPA1, RPA2, and RPA3 exhibited strong nuclear staining, with prominent punctate signals indicative of nucleolar accumulation. This localization pattern suggests their involvement in essential nuclear processes, including DNA replication, DNA repair, and nucleic acid metabolism, consistent with the established role of the RPA complex.

To further elucidate the biological functions and pathways of RPA genes in cancer, functional annotation was performed. Functional enrichment revealed that all three RPA paralogs are closely aligned with their known biological roles. RPA1, RPA2, and RPA3 showed shared roles in DNA replication, spliceosome, cell cycle, nucleotide excision repair, and base excision repair

(**Fig. 6A, 6C, 6E**). KEGG pathway enrichment analysis revealed DNA replication, mismatch repair, base-excision repair, and nucleotide excision repair as major significant pathways (**Fig 6B, 6D, and 6F**). Enrichment using g-profiler also revealed similar results (**supplementary Fig. 5**).

Chemokine Expression Analysis and Immunomodulatory Pathways

Understanding tumor microenvironment (TME), immunological landscape, and immune surveillance is essential in pan-cancer analyses. Accordingly, RPA gene expression was evaluated in relation to immunomodulatory factors, including immune stimulators, immune inhibitors, and chemokines and their corresponding receptors. Correlation analyses revealed significant associations between RPA genes and multiple immunomodulators across diverse cancer types.

In RPA1, several malignancies, including LIHC, ACC, PAAD, KICH, KIRC, STAD, COAD, LGG, BRCA, BLCA, READ, HNSC, UVM, PCPG, PRAD, and KIRP, showed strong positive correlations with immune-stimulatory molecules. In contrast, negative correlations were observed in LUSC, LUAD, ESCA, UCEC, TGCT, UCS, SKCM, SAARC, UVM, DLBC, READ, COAD, LAML, and GBM (**Fig. 7A**). In the case of RPA2, cancers such as PAAD, LIHC, KIRC, KIRP, ACC, UVM, KICH, CHOL, COAD, BRCA, LUAD, LUSC, LAML, PRAD, LGG, STAD, SKCM, READ, ESCA, BLCA, OV, UCEC, TGCT, and HNSC exhibited positive correlation with the majority of immunostimulators. Conversely, GBM, DLBC, UCS, MESO, SARC, THYM, PCPG, and THCA demonstrated negative correlations (Fig. 7B). For RPA3, positive correlations with immunostimulatory factors were observed in UVM, LGG, ACC, and KICH. In contrast, negative correlations were identified in THYM, DLBC, OV, LUSC, GBM, LAML, READ, COAD, and BRCA (Fig. 7C). Regarding RPA4, malignancies, including LIHC, KIRP, KIRC, THCA, ESCA, and KICH, showed positive correlations with

immunostimulators. At the same time, COAD, LUAD, LAML, SARC, STAD, OV, CESC, and LUSC exhibited negative correlations (**Fig. 7D**).

With respect to immunoinhibitory molecules, RPA1 exhibited positive correlations across multiple cancer types, including COAD, READ, STAD, PAAD, KICH, KIRC, LIHC, ACC, and LGG, whereas negative correlations were observed in ESCA, GBM, LUSC, SARC, THYM (**Fig. 8A**). For RPA2, positive associations with immunoinhibitors were identified in KICH, UVM, CHOL, PAAD, HNSC, PCPG, LAML, KIRP, KIRC, LGG, LIHC, and READ, while fewer cancer types demonstrated negative correlations (**Fig. 8B**). In the case of RPA3, positive correlations were observed in UVM, ACC, LGG, KICH, PCPG showed positive correlation while THYM, DLBC, LAML exhibited negative correlations with immunoinhibitory molecules (**Fig. 8C**). Regarding RPA4, the majority of cancers demonstrated positive correlations with immunoinhibitors, including THYM, LUAD, COAD, BLCA, UCEC, ESCA, and KIRC. In contrast, negative correlations were noted in LAML, SARC, STAD, and OV (**Fig. 8D**). Analysis of the association between RPA genes and chemokines (**Fig. 9**) as well as chemokine receptors (**Fig. 10**) demonstrated that RPA1 and RPA2 showed positive correlations with the majority of chemokines and their receptors, whereas RPA3 and RPA4 exhibited minimal or negligible positive correlations.

Overall, results indicated that overexpression of RPA genes was positively correlated with various immunomodulators associated with T-cell co-stimulation, TNF/NF- κ B pathways, and cytokine signalling in cancers such as LGG and LIHC, reflecting an immune-inflamed microenvironment; however, worse prognosis in these cancers reflects an immunosuppressive tumour microenvironment.

Immune Infiltration Analysis

Immune infiltration analysis revealed a significant positive correlation of RPA genes (RPA1, RPA2, and RPA3) with stromal, immune, and ESTIMATE scores in LGG. Positive correlation indicates that higher RPA gene expression is associated with increased immune and stromal infiltration and low tumor purity (**Fig 11A**). Further analysis revealed that overexpression of these genes is associated with increased infiltration of pro-tumor immune cells, such as M0 and M2 macrophages and neutrophils (**Fig 11B**). No significant association with RPA genes was observed in LIHC. Only RPA3 showed negative correlations with the stromal and ESTIMATE scores, but no association was found with the immune score. Negative correlations of RPA genes with these scores were observed in several cancers, including GBM, TGCT, ACC, and others.

Immune Response Associated with RPA Genes

Mirosatellite instability (MSI) and tumor mutational burden (TMB) are key indicators of an effective anti-tumor immune response. Elevated TMB/MSI is generally associated with a high mutation burden, leading to increased neoantigen production and thereby enhancing immunogenicity and improving immunotherapy response. Elevated TMB was observed in UCEC, ACC, COAD, KICH, LUAD, and READ for RPA1 (**Fig. 12A**). In contrast, increased TMB associated with RPA2 expression was noted in ACC, COAD, DLBC, GBM, KICH, LAML, MESO, PAAD, READ, SARC, SKCM, STAD, UCS (**Fig. 12C**). For RPA3, higher TMB was detected in UCEC, UCS, ACC, BLCA, KICH, LAML, LGG, LIHC, LUAD, MESO, PRAD, and SARC (**Fig. 12E**). Additionally, ACC and DLBC exhibited increased TMB in association with RPA4 expression (**Fig. 12G**).

Elevated MSI was observed in UVM, COAD, KIRC, READ, SARC, SKCM, STAD, TGCT, and UCEC in association with RPA1 expression (**Fig. 12B**). In the case of RPA2, increased MSI was detected in UCEC, UCS, UVM, ACC, COAD, DLBC, KIRC, MESO, READ, SARC,

and STAD (**Fig. 12D**). For RPA3 higher MSI was observed in UVM, CHOL, DLBC, and SARC (**Fig. 12F**) whereas ACC and exhibited elevated MSI in association with RPA4 expression (**Fig. 12H**). Table 4 shows the number of cancers with high TMB and MSI for all four RPA genes.

However, in LGG and LIHC, a very weak positive correlation (in some cases, even negative) was observed between the overexpression of RPA family genes and TMB/MSI. Low TMB/MSI in LGG and LIHC, associated with RPA gene overexpression, indicates decreased immune response, as evidenced by a very weak positive correlation. The immune infiltration analysis, as well as immunomodulatory association analysis, also suggested an immunosuppressive environment, suggesting the therapeutic potential of RPA genes in these cancers.

Table 4 Summary of immune response analysis

	RPA1	RPA2	RPA3	RPA4
TMB	6	13	10	2
MSI	9	11	4	2

Drug Sensitivity Analysis

Drug sensitivity analysis was conducted using the GSCA platform to explore the relationship between gene expression and therapeutic response. The results show that the expression levels of the RPA genes are significantly associated with the sensitivity to multiple anticancer drugs. Specifically, a negative correlation was observed between RPA gene expression and drug exposure, indicating increased sensitivity at higher RPA expression levels (**Fig. 13A, 13B**). Several kinase inhibitors exhibit enhanced efficacy in high-RPA gene expression. Given the prognostic and immunological outcomes, including poor survival and an immunosuppressive environment, these results are relevant, as high RPA expression in LGG and LIHC makes them susceptible to targeted therapeutic strategies, making RPA a potential therapeutic biomarker.

304

305 **Discussion**

306 RPA is a heterotrimeric complex that acts as a single-strand DNA-binding protein in eukaryotes
307 and plays an important role in DNA replication, recombination, repair, and the DNA damage
308 response (DDR) [17, 18]. Dysfunctional DDR causes genomic instability, leading to cancer.
309 RPA is a crucial player in the DDR pathway. Very few studies have reported the role of RPA in
310 cancer. High RPA expression has been linked with poor survival of glioma patients, and RPA
311 knockdown killed GBM cancer stem-like cells [19]. It has been shown that RPA recruits
312 RAD52 to repair sites, and their association is critical for RAD52 activity. Furthermore,
313 RAD52 inactivation causes synthetic lethality in homologous recombination (HR)-deficient
314 cancer cell lines. Inhibition of the RPA: RAD52 protein complex by mitoxantrone and
315 doxorubicin leads to selective killing of BRCA1-mutated breast cancer cells, and BRCA1/2-
316 mutated ovarian cancer cells [20]. RPA has also been shown to interact with the Enhancer of
317 mRNA Decapping protein 4 (EDC4) and causes cisplatin resistance in cervical cancer by
318 alleviating DNA damage [21]. Despite these studies, there is a need for pan-cancer analysis of
319 RPA genes to elucidate their role in cancer progression.

320 In this pan-cancer analysis, we examined the expression profiles, mutational status, prognostic
321 outcomes, and functional roles of RPA paralogs across 34 cancer types. Furthermore, the
322 associations of RPA genes with immune regulators, immune score, stromal score, ESTIMATE
323 score, and drugs were studied. The expression profile revealed that RPA genes were
324 significantly overexpressed across the greatest number of cancer types at both the RNA and
325 protein levels. High expression of RPA1, RPA2, and RPA3 has been reported in early stages of
326 gastric cancers, which are consistent with our findings. However, in stage III and stage IV
327 gastric cancer, expression of RPA genes was found to be low, indicating the use of RPA as an

early diagnostic biomarker for gastric cancer [22]. However, a recent study has identified low expression of RPA genes in 776 pure ductal carcinomas in situ (DCIS), 239 DCIS that co-exist with invasive breast cancer (IBC). Further, low RPA was associated with aggressive DCIS, aggressive IBC, and poor survival outcomes [23]. This apparent discrepancy may be attributed to several factors, such as tumor heterogeneity, distinct molecular landscapes, and multiple tumor subtypes; nevertheless, this divergent expression of RPA underscores the complexity of RPA gene regulation and demand cancer-specific RPA gene investigation. Tumor heterogeneity results from the accumulation of mutations over time, which support cancer cell survival, proliferation, and metastasis. In this context, the somatic mutation status of RPA genes was analysed. Results revealed that RPA1 harbours the maximum number of missense mutations. Alteration frequency exceeded 2% for RPA1 across 13 cancer types. Further, these mutations of RPA1 and RPA2 were associated with worse survival outcomes. Together, high expression and an increased mutation burden suggest a link to genomic instability, tumor evolution, and heterogeneity, thereby contributing to aggressive disease behaviour. Mutation of a conserved leucine residue to proline in the high-affinity DNA-binding site of RPA (residue L221 in human RPA) has been shown to cause high rates of cancer in mice and DNA-repair defects in yeast [24]. These findings suggest that RPA has a context-dependent function. RPA may function as a tumor initiator when altered and may promote tumor progression when overexpressed. Furthermore, RPA overexpression associated with increased mutation burden may result from replication stress or the DNA damage response. Prognostic analysis revealed that RPA overexpression is significantly associated with worse prognosis in several cancers. Overexpression of RPA1, RPA2, and RPA3 was associated with worse OS and PFI in LGG, whereas worse OS, PFI, PFS, and DSS were observed in LIHC. The prognostic significance of RPA has been reported in several cancers, including Hepatocellular Carcinoma, esophageal carcinoma [25], colon cancer [26], ovarian adenocarcinomas [27] gastric cancer, [28] lung

cancer [29], Clear Cell Renal Cell Carcinoma [30], and bladder urothelial carcinoma [31]. A study reported that transient knockdown of RPA3 inhibited cell proliferation, colony formation, and invasion in HepG2 cells. The same group performed a stable knockdown of RPA3 and found that tumor growth was significantly inhibited in the xenograft mouse model. Another study reports significant upregulation of RPA1 and RPA2 in esophageal tumors compared to adjacent non-tumor tissue, and RPA1 expression was associated with worse patient outcomes [32]. All these findings are consistent with our prognostic findings, implicating a strong prognostic role of RPA genes in cancers such as LGG and LIHC.

Tumor microenvironment (TME) plays an important role in tumor progression, survival, angiogenesis, proliferation, and metastasis [33]. It consists of stromal cells (cancer-associated fibroblasts, endothelial cells, mesenchymal stem cells), immune cells (CD8⁺ T cells, Tregs, M0 macrophages, etc.), and tumor cells [34]. Given its importance, we explored the role of RPA in TME modulation. Results revealed a positive association of immune score, stromal score, and ESTIMATE score with increased RPA1, RPA2, and RPA3 expression in LGG. Additionally, the expression of RPA1, RPA2, and RPA3 was positively correlated with infiltration of immune cells, such as M0/M2 macrophages, while negatively correlated with T cells and B cells in LGG. M0/M2 macrophages are pro-tumor immune cells that are immune suppressive, while T cells and B cells are anti-tumor immune cells that are immune activators [35]. Positive association of RPA genes with pro-tumor cells and negative association with anti-tumor cells indicate an immune-suppressive microenvironment that supports tumor growth in LGG. The worst prognosis associated with RPA overexpression in LGG, despite a positive correlation with immune regulators, is supported by findings from immune infiltration analysis. Thus, in LGG, TME is immune-inflamed but immune-suppressed due to infiltration of pro-tumor immune cells, such as M0/M2 macrophages. Overall, these findings highlight the role of RPA in tumor growth by modulating the tumor microenvironment. Furthermore,

TMB/MSI analysis also revealed a weak positive correlation of RPA genes in LGG and LIHC, indicating a decreased immune response. However, RPA genes were sensitive to several drugs in our drug-sensitivity analysis. RPA gene overexpression was correlated with sensitivity to several drugs like docetaxel, doxorubicin, and teniposide. Interestingly, a study reported that inhibition of RPA with drugs like quinacrine, mitoxantrone, and doxorubicin caused selective killing of breast and ovarian cancer cell lines [36]. This reinforces our findings and establishes RPA as a novel drug target for cancer therapy, particularly in cancers such as LGG and LIHC.

Although this pan-cancer study involves several analyses, it suffers from a number of limitations. The data used for the analysis are publicly available datasets that may be subject to inherent biases. Multiple datasets have been analysed using multiomics approaches, but our study lacks in vitro, in vivo, and other functional studies to validate our findings. To elucidate the mechanistic role of RPA in the tumor microenvironment, further functional studies involving a xenograft mouse model are required.

In conclusion, our first pan-cancer study of RPA genes revealed high expression across various cancers, including LGG and LIHC. High RPA expression was associated with a worse prognosis in LGG and LIHC, suggesting that it may serve as a prognostic and therapeutic biomarker. Overexpression of RPA genes was positively correlated with various immune regulators, the immune score, the stromal score, and the ESTIMATE score, reflecting the role of RPA in modulating the tumor microenvironment. Furthermore, RPA overexpression was sensitive to various drugs like docetaxel, doxorubicin, and teniposide, making RPA a novel drug target. However, further in-vitro and functional studies are required to validate these findings.

402 **Acknowledgements**

403 The authors would like to acknowledge the Director of the Centre of Biomedical Research for
404 administrative support, the Academy of Scientific and Innovative Research (AcSIR) and
405 Council of Science & Technology, Uttar Pradesh.

406 **Patient and public involvement**

407 Not applicable

408 **Authors' contributions**

409 **PY** and **RB** - Conceptualisation, data analysis and visualisation, writing, and reviewing; **AD** -
410 Writing, referencing, editing, and data visualisation. **PG** - Reviewing and Image formatting,
411 and **SKG** - Writing and reviewing. All authors reviewed the manuscript.

412 **Funding**

413 Council of Science & Technology, Uttar Pradesh. Grant number: CST/D-726 and CST/D-2507

414 **Data availability**

415 All data are contained within the manuscript. The datasets used and/or analysed during the
416 current study are available from the corresponding author on reasonable request.

417 **Declarations**

418 NA

419 **Ethics approval and consent to participate**

420 Since public datasets are used for analysis in this paper, no ethical considerations are required.

421 **Consent for publication**

422 Not applicable.

423 **Competing interests**

424 The authors have no competing interests.

425 **Author details**

426 NA

427

428

429

References

1. Jensen RB, Rothenberg E (2020) Preserving genome integrity in human cells via DNA double-strand break repair. *MBoC* 31:859–865. <https://doi.org/10.1091/mbc.E18-10-0668>
2. Yan H, Toczylowski T, McCane J, et al (2011) Replication protein A promotes 5'→3' end processing during homology-dependent DNA double-strand break repair. *Journal of Cell Biology* 192:251–261. <https://doi.org/10.1083/jcb.201005110>
3. Gan X, Zhang Y, Jiang D, et al (2023) Proper RPA acetylation promotes accurate DNA replication and repair. *Nucleic Acids Research* 51:5565–5583. <https://doi.org/10.1093/nar/gkad291>
4. Dueva R, Iliakis G (2020) Replication protein A: a multifunctional protein with roles in DNA replication, repair and beyond. *NAR Cancer* 2:zcaa022. <https://doi.org/10.1093/narcan/zcaa022>
5. Wu W-T, Li Y-J, Feng A-Z, et al (2021) Data mining in clinical big data: the frequently used databases, steps, and methodological models. *Mil Med Res* 8:44. <https://doi.org/10.1186/s40779-021-00338-z>
6. Chandrashekar DS, Bashel B, Balasubramanya SAH, et al (2017) UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia* 19:649–658. <https://doi.org/10.1016/j.neo.2017.05.002>
7. Ellis MJ, Gillette M, Carr SA, et al (2013) Connecting genomic alterations to cancer biology with proteomics: the NCI Clinical Proteomic Tumor Analysis Consortium. *Cancer Discov* 3:1108–1112. <https://doi.org/10.1158/2159-8290.CD-13-0219>
8. Gao J, Aksoy BA, Dogrusoz U, et al (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 6:p11. <https://doi.org/10.1126/scisignal.2004088>
9. Wu CH (2006) The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Research* 34:D187–D191. <https://doi.org/10.1093/nar/gkj161>
10. Feng Z, Verdigué N, Di Costanzo L, et al (2020) Impact of the Protein Data Bank Across Scientific Disciplines. *CODATA* 19:25. <https://doi.org/10.5334/dsj-2020-025>
11. Tang Z, Kang B, Li C, et al (2019) GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Research* 47:W556–W560. <https://doi.org/10.1093/nar/gkz430>
12. Kolberg L, Raudvere U, Kuzmin I, et al (2023) g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res* 51:W207–W212. <https://doi.org/10.1093/nar/gkad347>

13. Zhou Y, Zhou B, Pache L, et al (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 10:1523. <https://doi.org/10.1038/s41467-019-09234-6>
14. Uhlén M, Björling E, Agaton C, et al (2005) A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol Cell Proteomics* 4:1920–1932. <https://doi.org/10.1074/mcp.M500279-MCP200>
15. Yoshihara K, Shahmoradgoli M, Martínez E, et al (2013) Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 4:2612. <https://doi.org/10.1038/ncomms3612>
16. Liu C-J, Hu F-F, Xie G-Y, et al (2023) GSCA: an integrated platform for gene set cancer analysis at genomic, pharmacogenomic and immunogenomic levels. *Briefings in Bioinformatics* 24:bbac558. <https://doi.org/10.1093/bib/bbac558>
17. Madru C, Martínez-Carranza M, Laurent S, et al (2023) DNA-binding mechanism and evolution of replication protein A. *Nat Commun* 14:2326. <https://doi.org/10.1038/s41467-023-38048-w>
18. VanderVere-Carozza PS, Gavande NS, Jalal SI, et al (2022) In Vivo Targeting Replication Protein A for Cancer Therapy. *Front Oncol* 12:826655. <https://doi.org/10.3389/fonc.2022.826655>
19. Pedersen H, Anne Adanma Obara E, Elbæk KJ, et al (2020) Replication Protein A (RPA) Mediates Radio-Resistance of Glioblastoma Cancer Stem-Like Cells. *Int J Mol Sci* 21:1588. <https://doi.org/10.3390/ijms21051588>
20. Al-Mugotir M, Lovelace JJ, George J, et al (2021) Selective killing of homologous recombination-deficient cancer cell lines by inhibitors of the RPA:RAD52 protein-protein interaction. *PLoS One* 16:e0248941. <https://doi.org/10.1371/journal.pone.0248941>
21. Wu X, Zhong Y, Chen Q, et al (2020) Enhancer of mRNA Decapping protein 4 (EDC4) interacts with replication protein a (RPA) and contributes to Cisplatin resistance in cervical Cancer by alleviating DNA damage. *Hereditas* 157:41. <https://doi.org/10.1186/s41065-020-00154-w>
22. Fourtziala E, Givalos N, Alexakis N, et al (2020) Replication Protein A (RPA1, RPA2 and RPA3) expression in gastric cancer: correlation with clinicopathologic parameters and patients' survival. *J BUON* 25:1482–1489
23. Algethami M, Toss MS, Woodcock CL, et al (2023) Unravelling the clinicopathological and functional significance of replication protein A (RPA) heterotrimeric complex in breast cancers. *NPJ Breast Cancer* 9:18. <https://doi.org/10.1038/s41523-023-00524-3>
24. Hass CS, Gakhar L, Wold MS (2010) Functional characterization of a cancer causing mutation in human replication protein A. *Mol Cancer Res* 8:1017–1026. <https://doi.org/10.1158/1541-7786.MCR-10-0161>

25. Dahai Y, Sanyuan S, Hong L, et al (2013) A relationship between replication protein A and occurrence and prognosis of esophageal carcinoma. *Cell Biochem Biophys* 67:175–180. <https://doi.org/10.1007/s12013-013-9530-y>
26. Givalos N, Gakiopoulou H, Skliri M, et al (2007) Replication protein A is an independent prognostic indicator with potential therapeutic implications in colon cancer. *Mod Pathol* 20:159–166. <https://doi.org/10.1038/modpathol.3800719>
27. Levidou G, Ventouri K, Nonni A, et al (2012) Replication protein A in nonearly ovarian adenocarcinomas: correlation with MCM-2, MCM-5, Ki-67 index and prognostic significance. *Int J Gynecol Pathol* 31:319–327. <https://doi.org/10.1097/PGP.0b013e31823ef92e>
28. Zhang Y, Yu C (2020) Distinct expression and prognostic values of the replication protein A family in gastric cancer. *Oncol Lett* 19:1831–1841. <https://doi.org/10.3892/ol.2020.11253>
29. He J, Wang Z, Wang Y, et al (2022) The Roles of EXO1 and RPA1 Polymorphisms in Prognosis of Lung Cancer Patients Treated with Platinum-Based Chemotherapy. *Dis Markers* 2022:3306189. <https://doi.org/10.1155/2022/3306189>
30. Gola M, Kieżun J, Kraziński BE, et al (2025) RPA1, RFC1, and POLE Expression in Clear Cell Renal Cell Carcinoma: Immune and Clinical Relevance. *Anticancer Res* 45:5267–5286. <https://doi.org/10.21873/anticancer.17867>
31. Levidou G, Gakiopoulou H, Kavantzias N, et al (2011) Prognostic significance of replication protein A (RPA) expression levels in bladder urothelial carcinoma. *BJU Int* 108:E59-65. <https://doi.org/10.1111/j.1464-410X.2010.09828.x>
32. Dahai Y, Sanyuan S, Hong L, et al (2013) A relationship between replication protein A and occurrence and prognosis of esophageal carcinoma. *Cell Biochem Biophys* 67:175–180. <https://doi.org/10.1007/s12013-013-9530-y>
33. Wang Q, Shao X, Zhang Y, et al (2023) Role of tumor microenvironment in cancer progression and therapeutic strategy. *Cancer Med* 12:11149–11165. <https://doi.org/10.1002/cam4.5698>
34. Anderson NM, Simon MC (2020) The tumor microenvironment. *Curr Biol* 30:R921–R925. <https://doi.org/10.1016/j.cub.2020.06.081>
35. Zhang W, Wang M, Ji C, et al (2024) Macrophage polarization in the tumor microenvironment: Emerging roles and therapeutic potentials. *Biomed Pharmacother* 177:116930. <https://doi.org/10.1016/j.biopha.2024.116930>
36. Al-Mugotir M, Lovelace JJ, George J, et al (2021) Selective killing of homologous recombination-deficient cancer cell lines by inhibitors of the RPA:RAD52 protein-protein interaction. *PLoS One* 16:e0248941. <https://doi.org/10.1371/journal.pone.0248941>

Figure Legends

Fig. 1: Flowchart illustrating the different steps involved in the analysis: Analysis performed during pan-cancer study involves gene expression profiling, mutational screening, prognostic survival analysis, functional annotation, immune-infiltration analysis, as well as immunotherapy response analysis and drug sensitivity analysis.

Fig. 2: Paired dot-box plot showing upregulation and downregulation of RPA across various tumor types. (A, B, C, D) The mRNA expression profiles of RPA genes were analyzed across all normal samples and tumor tissues using the TCGA datasets. **(E, F, G, H)** Anatomical Body plot showing the spatial distribution levels of RPA genes across various organs for different cancer types using TCGA and GTEx datasets. Expression values are represented as the logarithm of transcript per million (TPM).

Fig. 3: Genetic mutational analysis of RPA genes: (A, C, E, G) Frequencies of genetic alterations of RPA genes across various cancer types, and **(B, D, F, H)** Kaplan–Meier survival graphs and log-rank tests comparing cancer patients' survival outcomes with and without genetic mutations in RPA genes.

Fig. 4: Prognostic relevance of RPA genes in cancers: A. Univariable Cox regression forest plot assessing the prognostic effect of RPA gene expression across multiple cancer types. **B.** Multivariate Cox regression analysis involving PFI, PFS, and DSS analysis.

Fig. 5: Immunostaining of RPA genes. Fluorescent antibody staining showing the subcellular localization of RPA proteins in the endoplasmic reticulum (ER), nucleus, and microtubules of A431 (squamous carcinoma), A549 (Lung adenocarcinoma), U-2 OS (osteosarcoma), U251MG (Glioblastoma), HPA1 (near-haploid cell line derived from the KBM-7 cell line), HeLa (cervical adenocarcinoma), and U2CS (osteosarcoma) cells supporting their roles in DNA repair and cell cycle.

Fig. 6: Gene Ontology and KEGG pathway analysis of RPA genes. (A, C, E) Gene ontology enrichment analysis revealed significant biological processes related to RPA1, RPA2, and RPA3. **(B, D, F)** KEGG pathway enrichment analysis of RPA1, RPA2, and RPA3 showing significant pathway enriched.

Fig. 7: Correlation of RPA gene expression with tumor immune-activating markers across TCGA cancers. (A, B, C, D) The heatmap illustrates the relationship between RPA gene expression and major immunostimulatory genes across multiple cancer types. Positive and negative correlations are shown in red and blue, respectively. Color intensity indicates the strength of the association; statistical significance is indicated by * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$.

Fig. 8: Relationship between RPA gene expression and tumor immunoinhibitory activity across TCGA cancers. (A, B, C, D) The heatmap summarizes the relationship between RPA gene expression and key immunoinhibitory genes across the TCGA cancer cohort. Correlation coefficients are

represented by color intensity (red: positive; blue: negative). Asterisks denote statistical significance (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Fig. 9: Association of RPA gene expression with chemokines in pan-cancer TCGA analysis. (A, B, C, D) The heatmap depicts Spearman correlations between RPA and chemokine-related genes across multiple cancers. Red and blue indicate positive and negative correlations, respectively; asterisks denote significance (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

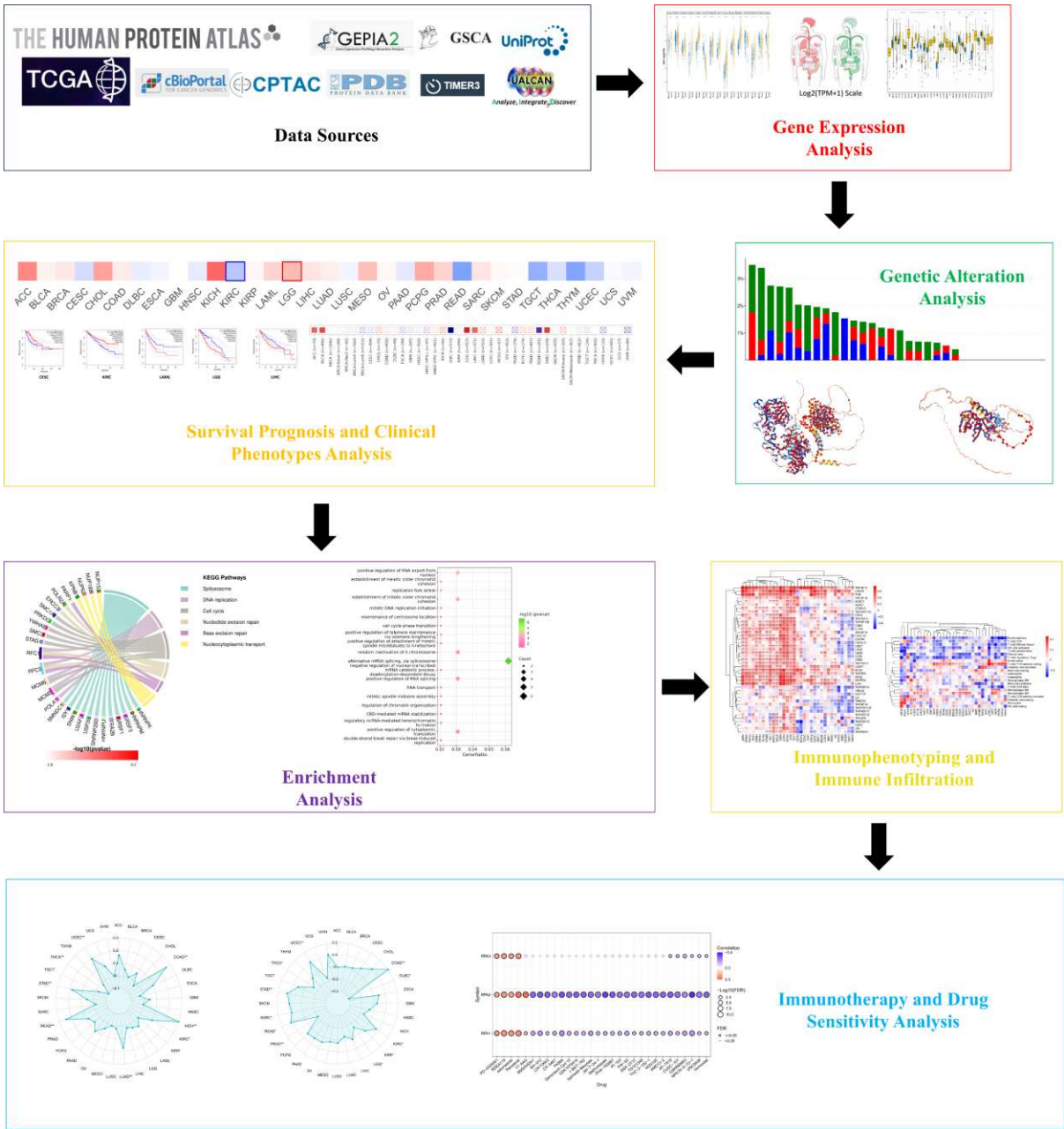
Fig. 10: Link between RPA gene expression and chemokine receptors across TCGA cancers. (A, B, C, D) This heatmap reveals associations between RPA gene expression and chemokine receptor genes across different cancer types. Spearman correlation coefficients are shown by color, with red indicating positive and blue indicating negative correlations. The intensity of the color reflects the strength of the association, and statistical significance is indicated by asterisks ($p < 0.05$, * $p < 0.01$, ** $p < 0.001$).

Fig. 11: Immune infiltration analysis of RPA genes. A Correlation of immune score, stromal score, and ESTIMATE score with RPA genes. **B.** Analysis of immune infiltration patterns in relation to RPA gene expression.

Fig. 12: Evaluation of the relationship between tumor mutation burden (TMB) and microsatellite instability (MSI) with RPA gene expression in multiple cancer types. (A, C, E, G) Spearman correlation analysis of TMB with RPA1, RPA2, RPA3, and RPA4, respectively. **(B, D, F, H)** Spearman correlation analysis of MSI with RPA1, RPA2, RPA3, and RPA4, respectively. Significant correlations are marked with (*).

Fig. 13: Drug Sensitivity Analysis: A. Correlation between GDSC drug sensitivity and mRNA expression of RPA genes. **B.** Correlation between CTR drug sensitivity and mRNA expression of RPA genes

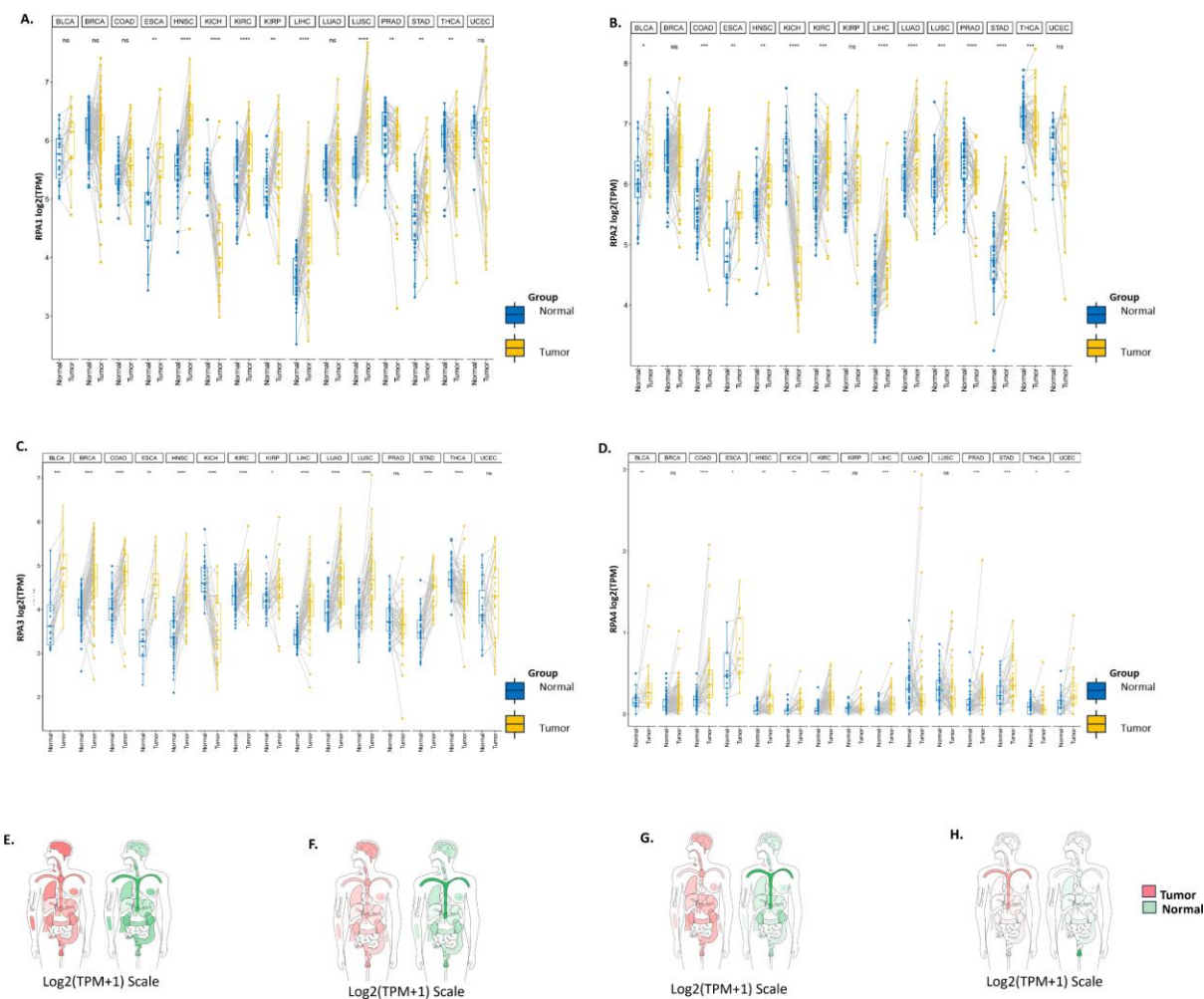
618 **Fig. 1**



619

620

Fig. 2



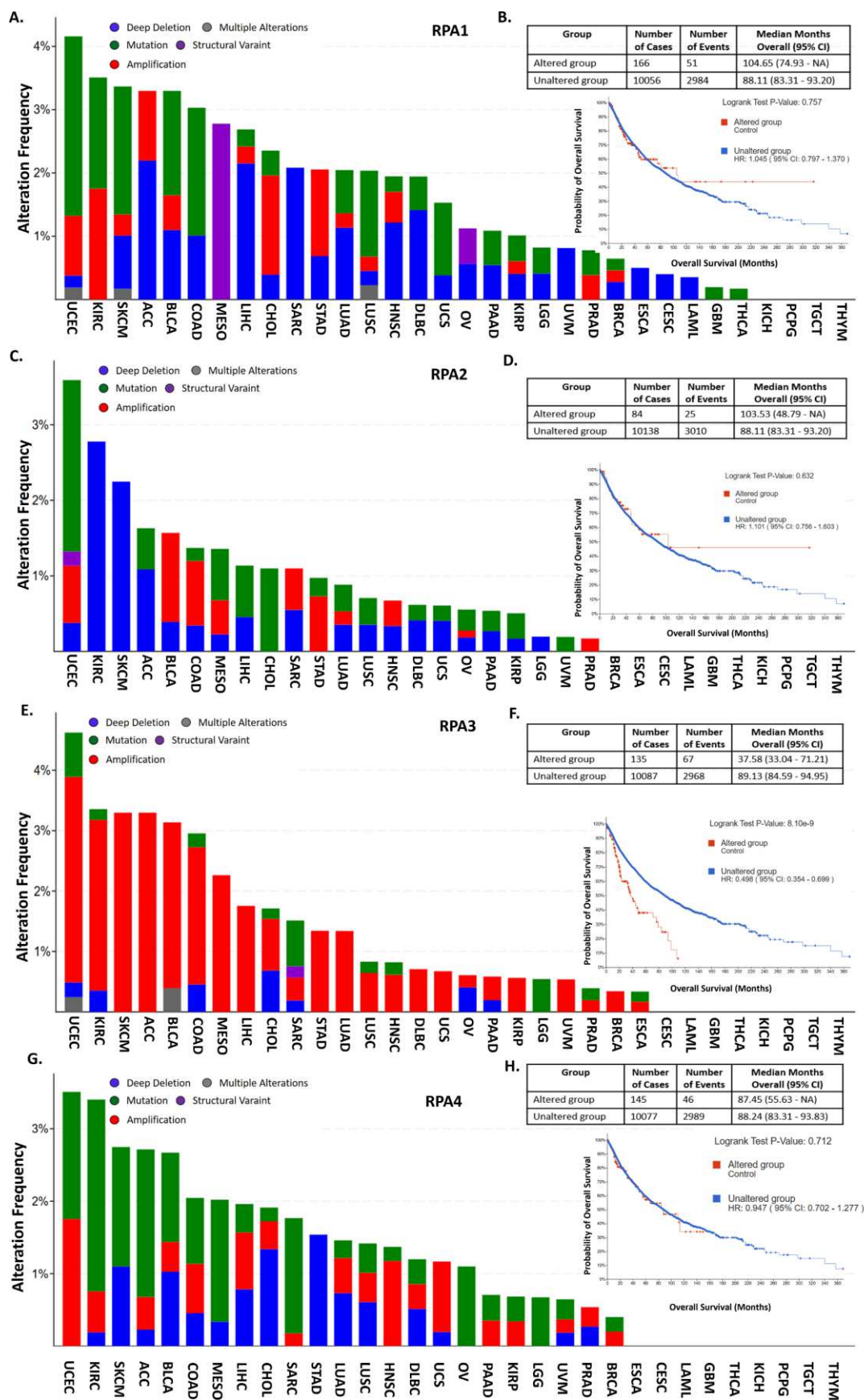


Fig. 4

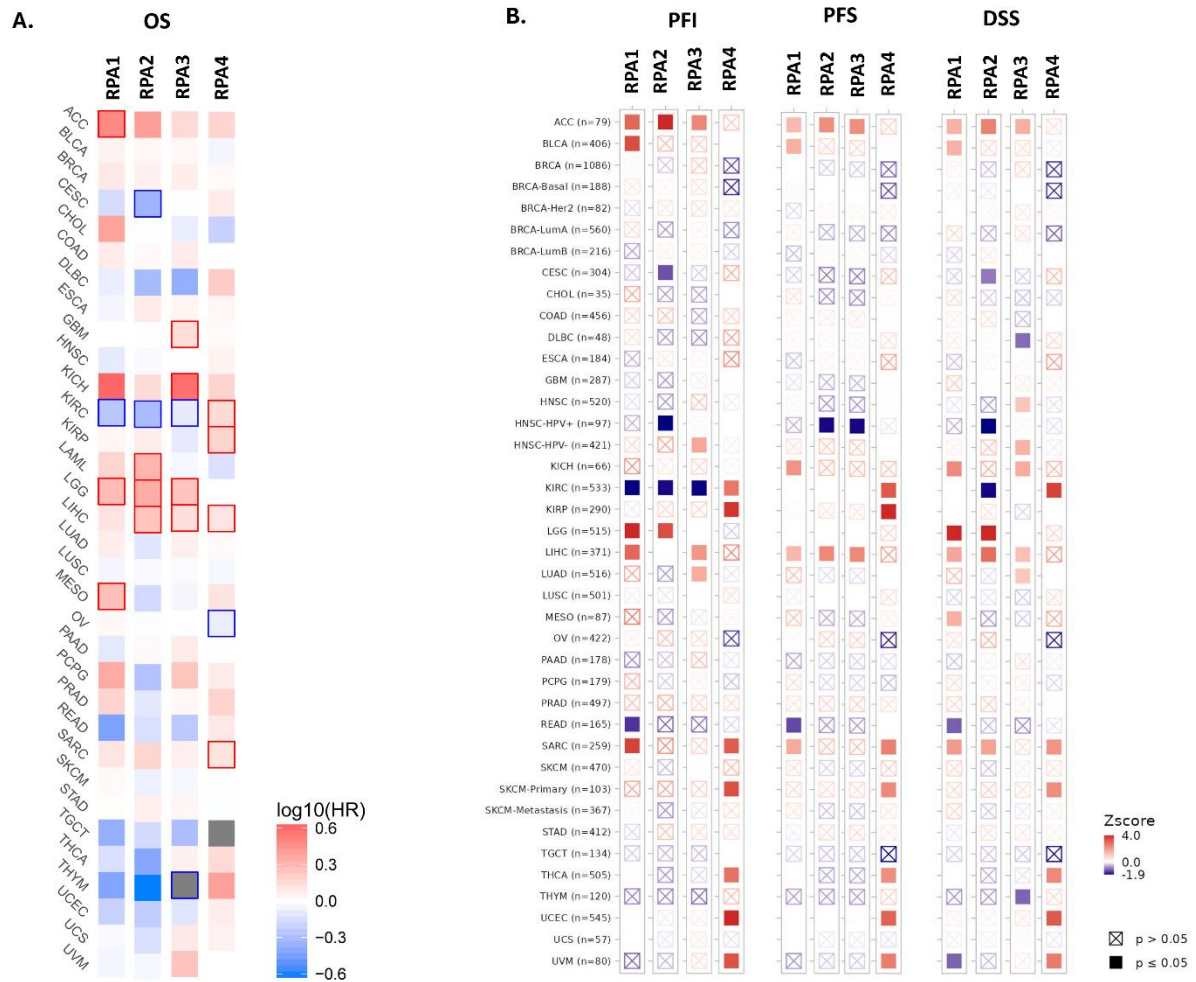
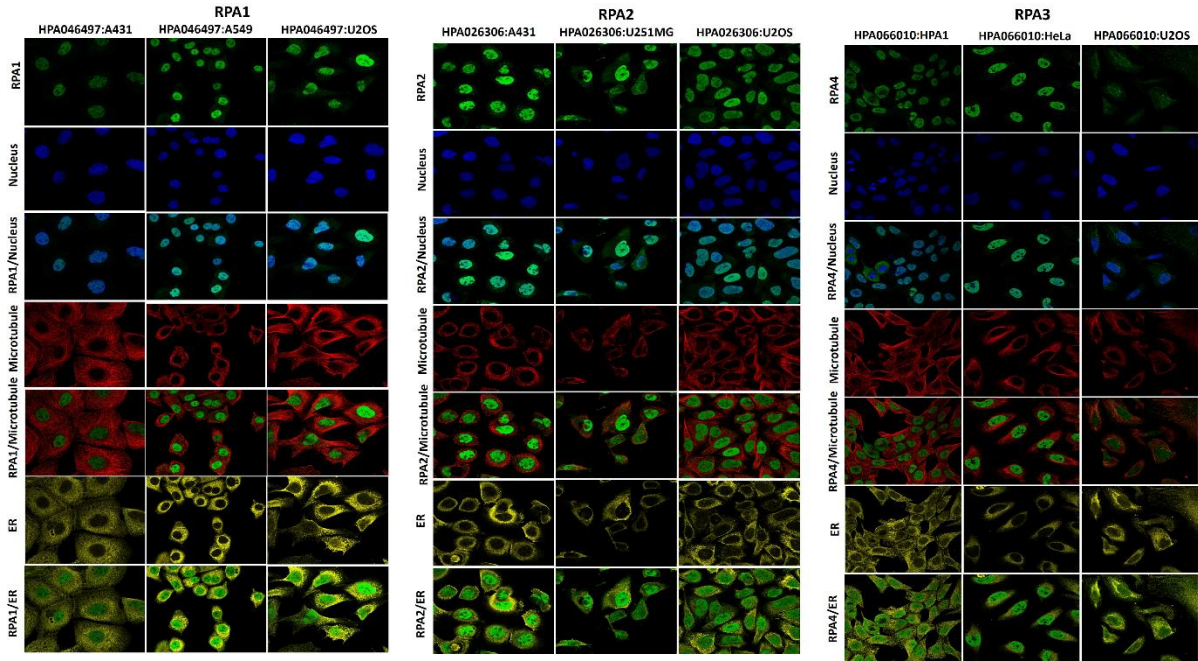
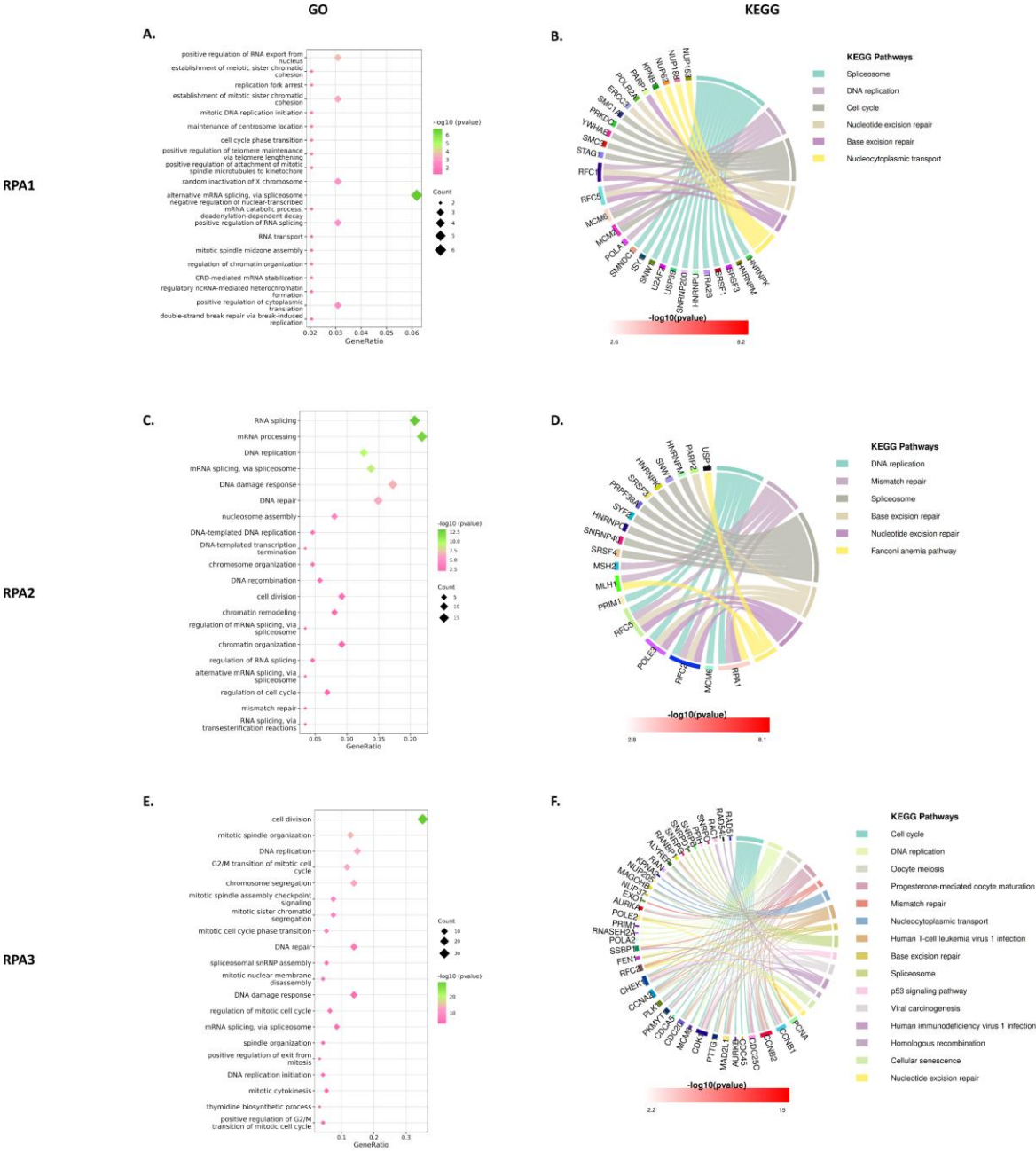
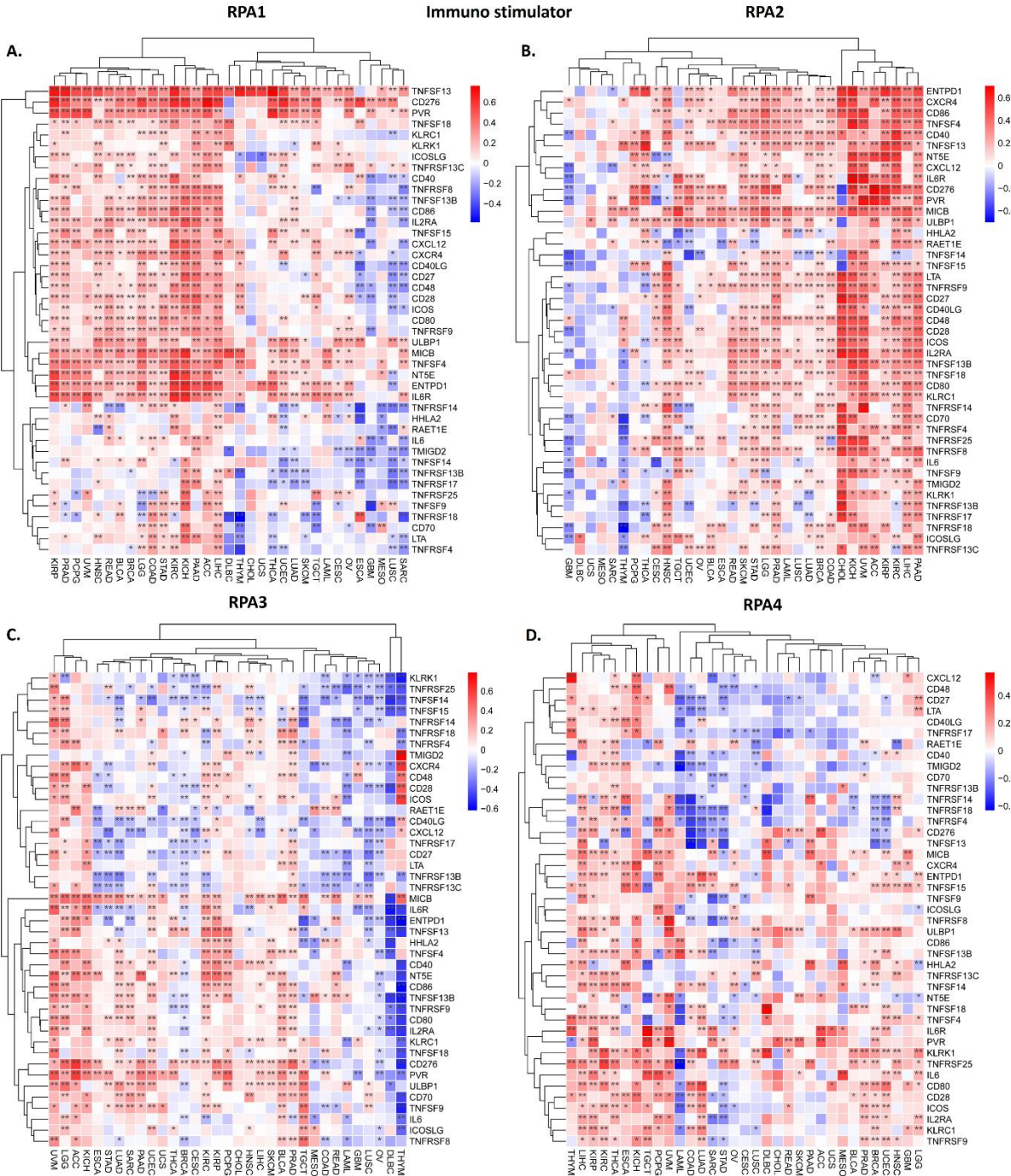


Fig. 5







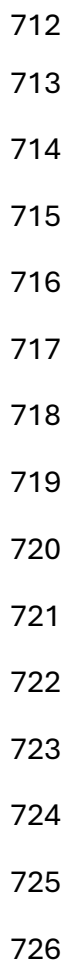


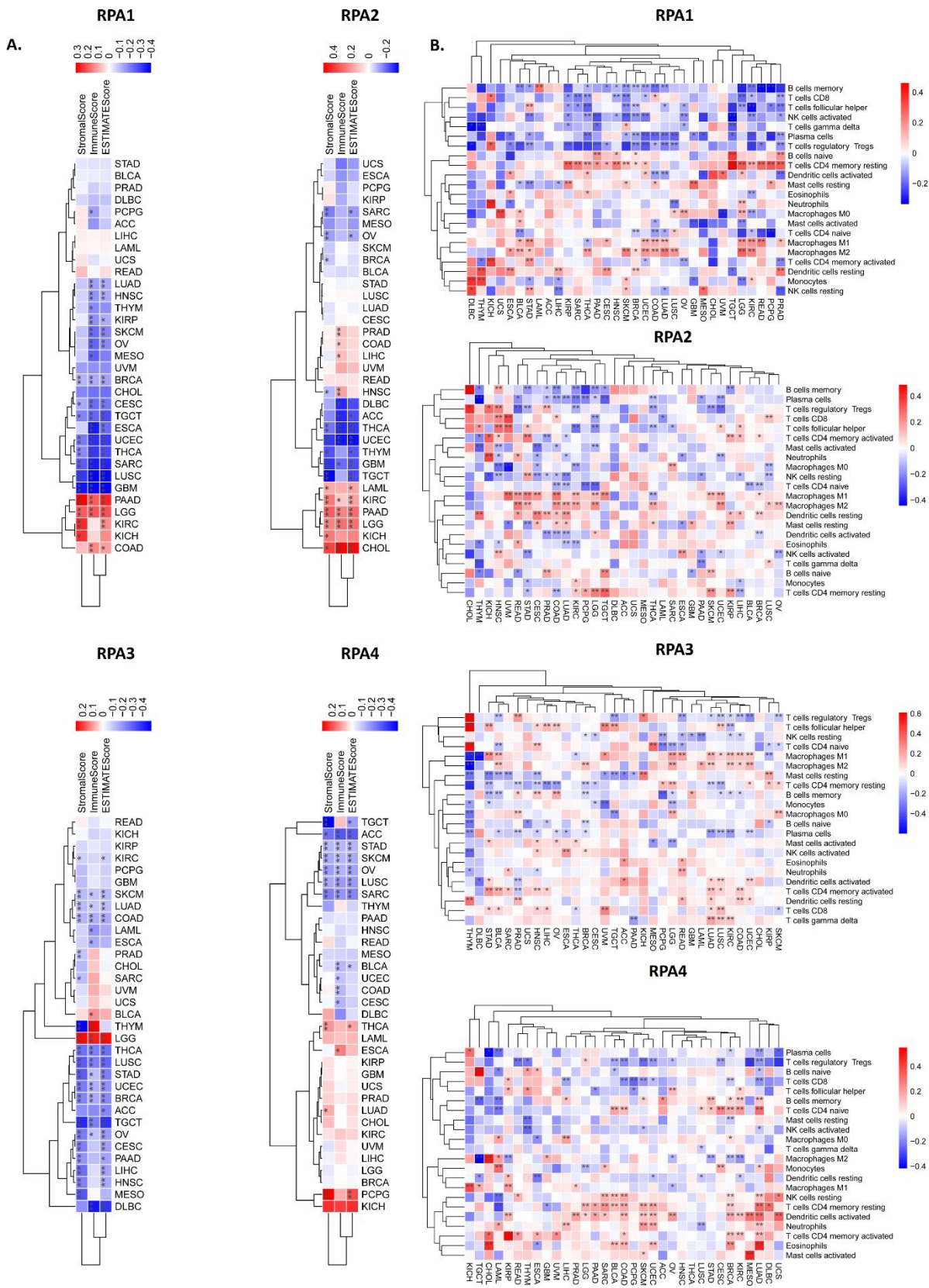
702

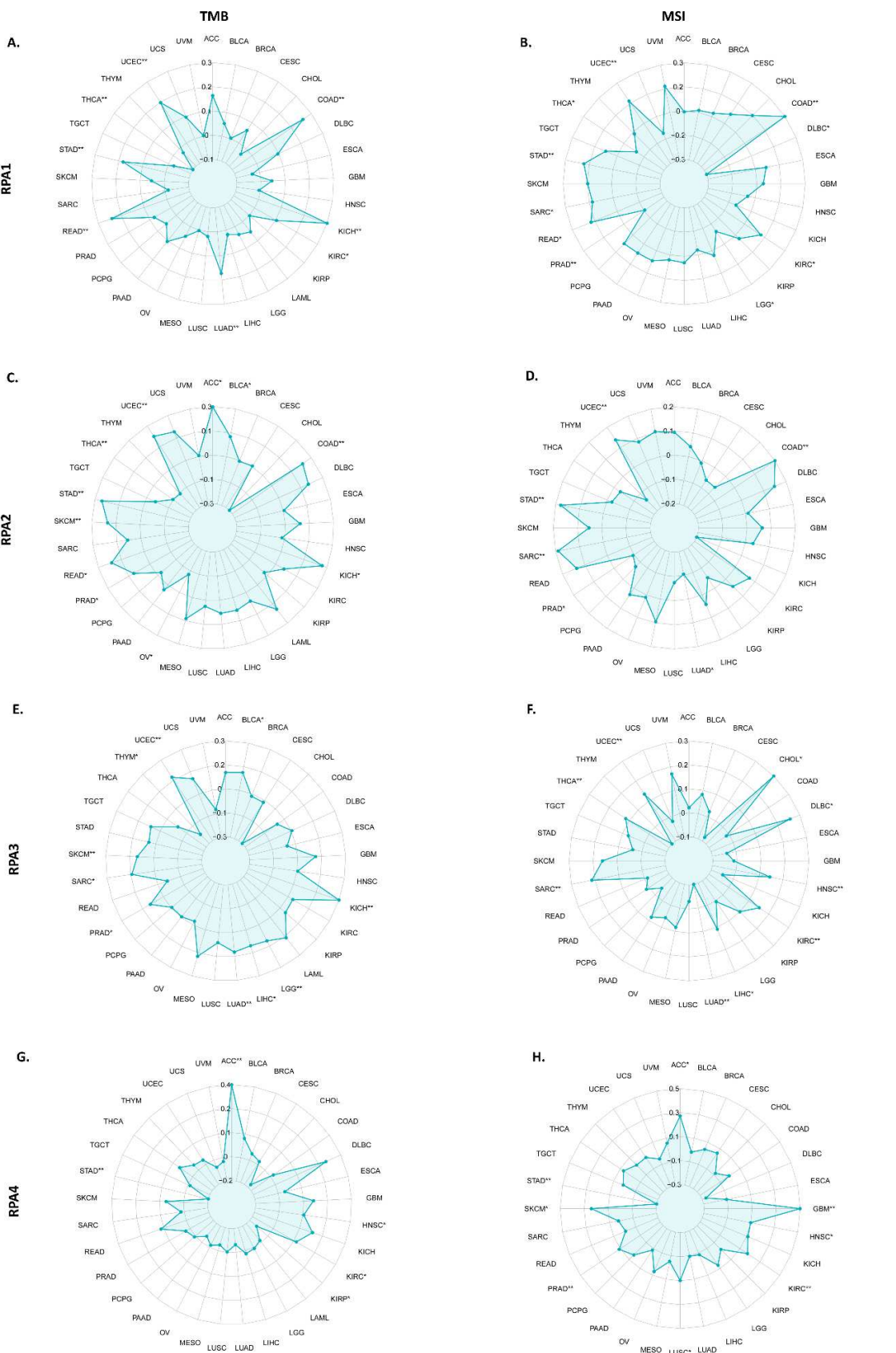


710

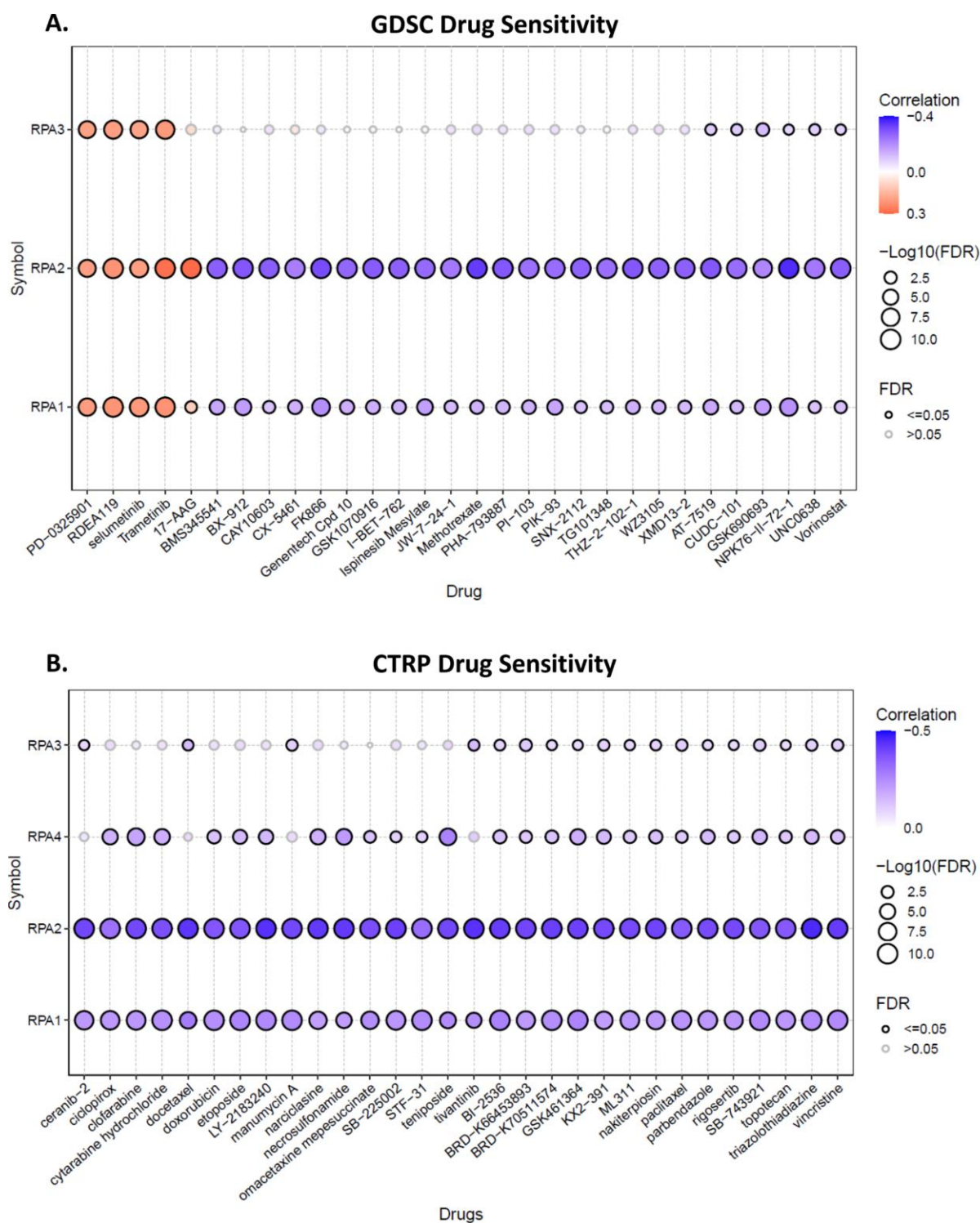
711







732 **Fig. 13**



733
734
735
736
737

738

739 **Table 1:** Summary of RPA gene expression in all cancer types

Gene	Number of Cancers with Significant higher expression in Paired Box-dot-Plot
RPA1	10
RPA2	12
RPA3	13
RPA4	12

740

741 **Table 2:** Summary of different types of mutations present in RPA genes

Gene	Missense	Truncating	In-frame	Splice	Fusion	Total	Alpha Missense Variants (Pathogenic)
RPA1	62	4	0	12	3	81	3278
RPA2	30	3	0	1	1	35	1553
RPA3	13	4	0	1	2	16	314
RPA4	76	6	0	0	0	82	87

742

743 **Table 3:** Summary of overall survival (OS) analysis

	Cancer with worse overall survival (OS)	Cancer with better OS
RPA1	ACC, LGG, KICH, MESO	KIRC
RPA2	LAML, LGG, LIHC	KIRC, CESC
RPA3	GBM, KICH, LGG, LIHC	KIRC
RPA4	KIRP, LIHC, KIRC, SARC	OV

744

745 **Table 4:** Summary of immune response analysis

	RPA1	RPA2	RPA3	RPA4
TMB	6	13	10	2
MSI	9	11	4	2

746

Supplementary Files

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

- [SupplementaryfileRPA.docx](#)