



AARHUS UNIVERSITY  
NATURAL SCIENCES  
BIOINFORMATICS RESEARCH CENTRE



---

# Pan-Cancer Exploration of rDNA Copy Number and Its Clinical and Immunological Significance

---

AUTHOR:  
Chongming Chen  
202303402

SUPERVISOR :  
Nicolai Juul BIRKBAK

June 15, 2025

## Abstract

Cancer remains one of the leading causes of death worldwide, and uncovering its underlying mechanisms continues to be a central focus of biomedical research. A representative trait of cancer cells is their unlimited proliferative capacity, which is tightly linked to ribosome biogenesis. This study aimed to investigate the role and mechanism of ribosomal DNA copy number (rDNA CN), which serves as the foundation for ribosome metabolism. I developed a computational pipeline to estimate rDNA CN from whole-genome sequencing (WGS) data and applied it to seven TCGA cancer cohorts to perform a pan-cancer analysis. The analytical framework involved statistical assessments and survival modeling, including Kaplan–Meier and Cox proportional hazards models, to examine the association between rDNA CN and clinical outcomes. Additionally, RNA-seq data from the TCGA cohorts were utilized for immunological profiling and transcriptomic analysis. Notably, a higher 45S rDNA copy number was associated with improved survival outcome in bladder urothelial carcinoma (BLCA) and pancreatic adenocarcinoma (PAAD), although such associations were not found in all other cancer types. Results from immunological and transcriptomic analyses further suggest that the functional impact of rDNA CN is tumor-specific, yet broadly associated with immunological characteristics.

**Keywords:** Ribosomal DNA (rDNA), Copy number, Pan-cancer, Survival analysis

# Contents

<b>Contents</b>	<b>I</b>
<b>List of Figures</b>	<b>III</b>
<b>List of Tables</b>	<b>V</b>
<b>Abbreviations</b>	<b>VI</b>
<b>1 Introduction</b>	<b>1</b>
1.1 The Structure of Ribosome DNA and Its Function	1
1.1.1 Ribosome	1
1.1.2 The Structure of rDNA	3
1.1.3 Transcription and Processing of rDNA	4
1.2 Roles of Ribosomes in Cancer	5
1.2.1 Ribosome Biogenesis and Dysregulation in Cancer	6
1.2.2 Nucleolar Stress Response	7
1.2.3 Ribosomal Proteins	7
1.2.4 Ribosomal Heterogeneity in Cancer	8
1.3 Variation of rDNA Sequence and Its Implications in Cancer	8
1.4 Research Objectives and Content	9
<b>2 Materials and Methods</b>	<b>11</b>
2.1 Datasets	11
2.2 Calculation of rDNA Copy Number	12
2.2.1 Obtaining rDNA array reads	12
2.2.2 Selecting reference rDNA sequences	12
2.2.3 Optimization of rDNA Read Alignment: De Novo Mapping vs. Slicing	13
2.2.4 Estimation of Background Reads Depth (BRD) for Single Copy Exons and Introns	13
2.2.5 Validation and Calculation of rDNA Copy Number	14
2.3 Immune Profiling Based on RNA-Seq and WGS	14
2.4 RNA Sequencing Profile	15
2.5 Statistical Analysis	16
2.5.1 Group Comparison and Correlation Analysis	16

---

2.5.2	Kaplan-Meier (KM) Survival Analysis . . . . .	16
2.5.3	Cox Proportional Hazards Model (CoxPH) . . . . .	17
2.6	Code Availability . . . . .	18
<b>3</b>	<b>Results . . . . .</b>	<b>19</b>
3.1	Pan-Cancer Landscape of rDNA Copy Number Variation and Its Clinical Correlation . . . . .	19
3.1.1	Distribution of rDNA Copy Number Across Cancer Types . . . . .	19
3.1.2	Clinical Characteristic and Correlation . . . . .	20
3.2	Survival Outcome Analysis . . . . .	23
3.3	Immune Analysis . . . . .	28
3.3.1	Correlation between rDNA Copy Number and TIL Scores . . . . .	28
3.3.2	Correlation between rDNA Copy Number and T/B Cell Fractions . . . . .	32
3.4	Transcriptomic Analysis . . . . .	35
3.4.1	Differentially Expressed Genes (DEGs) . . . . .	35
3.4.2	Functional Enrichment Analysis . . . . .	38
<b>4</b>	<b>Discussion . . . . .</b>	<b>43</b>
<b>5</b>	<b>Conclusion . . . . .</b>	<b>46</b>
	<b>Bibliography . . . . .</b>	<b>47</b>
<b>A</b>	<b>Appendix . . . . .</b>	<b>51</b>

## List of Figures

1.1	<b>The structure of Ribosomes.</b> (Adapted from [Gibbons <i>et al.</i> , 2015]) The large 60S subunit is shown in red, and the small 40S subunit is shown in blue. . . . .	2
1.2	<b>The Structure of Ribosome DNA.</b> . . . .	3
1.3	<b>Schematic diagram illustrating the processing of ribosomal pre-rRNA.</b> . . . .	5
1.4	<b>The Workflow of this Study.</b> Red indicates using whole-genome sequencing (WGS) data, while green indicates using RNA sequencing (RNA-seq) data. . . . .	9
2.1	<b>Validation of rDNA Copy Number Calculation Methods.</b> The x-axis shows the 5S rDNA copy number estimated using the slicing method , while the y-axis shows the estimates from de novo mapping (WGS). A strong linear correlation was observed ( $R^2 = 1.00$ , $p = 3 \times 10^{-10}$ ). . . . .	15
3.1	<b>Pan-cancer Distribution of rDNA Copy Number in Blood Samples from TCGA Cohorts.</b> (A) Distribution of 45S rDNA copy number across PAAD, BLCA, COAD, LUAD, LUSC, GBM, and PRAD; (B) Distribution of 5S rDNA copy number in the same TCGA cancer cohorts. One-way ANOVA tests reveal significant differences in both 28S and 5S rDNA copy numbers across tumor types. (45S: $p < 2 \times 10^{-16}$ ; 5S: $p < 2 \times 10^{-16}$ ) . . . . .	20
3.2	<b>Age distribution and Its Correlation with rDNA Copy Number across TCGA Cohorts.</b> (A) Distribution of patient age at diagnosis across the entire dataset and within individual TCGA cancer cohorts; (B–H) Spearman correlation analyses between age and rDNA copy number for 45S across different cohorts. Weak but statistically significant negative correlations were observed in several cancer types. . . . .	21
3.3	<b>Analysis of Sex Differences in Copy Number of 45S rDNA.</b> (A) All datasets ( $p = 0.0045$ , Wilcoxon test). (B) Across different TCGA cancer cohorts, including BLCA ( $p = 0.60$ ), COAD ( $p = 0.25$ ), LUAD ( $p = 0.044$ ), LUSC ( $p = 0.28$ ), GBM ( $p = 0.49$ ), and PAAD ( $p = 0.86$ ). . . . .	22
3.4	<b>Analysis of Tumor Stage Differences in Copy Number of 45S rDNA.</b> (A) All datasets (Wilcoxon rank-sum test, $p = 0.15$ ). (B–F) Across different TCGA cancer cohorts using Wilcoxon rank-sum test, including BLCA ( $p = 0.06$ ), COAD ( $p = 0.40$ ), LUAD ( $p = 0.45$ ), PAAD ( $p = 0.61$ ), and LUSC ( $p = 0.34$ ). No statistically significant differences were observed. . . . .	24

3.5	<b>KM Survival Curves for Overall Survival (OS) and Cox regression analysis.</b> (A) OS; (B–H) Forest plot displaying HRs and corresponding p-values across all TCGA datasets. . . . .	25
3.6	<b>KM Survival Curves for Progression-Free Interval (PFI) and Cox regression analysis.</b> (A) PFI; (B–H) Forest plot displaying HRs and corresponding p-values across all TCGA datasets. . . . .	26
3.7	<b>Cox Regression Analysis after Adjusting for Sex and Age.</b> Forest plot displaying HRs and corresponding p-values across all TCGA datasets after correcting for sex, age and tumor stage. . . . .	27
3.8	<b>Bubble Plots of Correlation between rDNA Copy Number and Tumor Infiltrating Lymphocytes across TCGA Cohorts.</b> Spearman correlation coefficients (R) are shown for seven TCGA cancer cohorts. Circle color indicates the direction and strength of the correlation (blue for positive, red for negative), and circle size reflects the absolute correlation coefficient ( $ R $ ). Squares represent statistically significant correlations ( $p < 0.05$ ), while Circles represent non-significant results ( $p \geq 0.05$ ). . . . .	29
3.9	<b>Correlation of Tumor Infiltrating Lymphocytes and 5S rDNA CN in PAAD and PRAD.</b> Fifteen tumor infiltrating lymphocytes (TIL) and a total TIL score are evaluated for correlation with Copy Number; DC: dendritic cells. . . . .	31
3.10	<b>Correlation of T/B Cell Fraction and 45S rDNA CN across TCGA Cohorts.</b> Spearman r was used to evaluate the correlation. . . . .	33
3.11	<b>Correlation of T/B Cell Fraction and 5S rDNA CN across TCGA Cohorts.</b> Spearman r was used to evaluate the correlation. . . . .	34
3.12	<b>Volcano Plots of DEGs between High and Low 45S rDNA Copy Number Groups across TCGA Cohorts.</b> Red and blue points denote significantly down-regulated and up-regulated genes, respectively. . . . .	36
3.13	<b>Differentially Expressed Genes Frequency across TCGA Cohorts Associated with High versus Low 45S rDNA Copy Number.</b> This plot shows genes that were significantly differentially expressed ( $p < 0.05$ and $ \log_2 FC  \geq 1$ ) in at least two cancer cohorts. . . . .	37
3.14	<b>GO and KEGG Pathway Enrichment of 45S rDNA in the BLCA cohort.</b> Dot-plots show significantly enriched Gene Ontology terms including biological process (BP), cellular component (CC), and molecular function (MF), along with KEGG, for DEGs between high and low 45S rDNA copy number groups from tumor RNA-seq of TCGA-BLCA cohort. . . . .	39
3.15	<b>GO and KEGG Pathway Enrichment Frequency Associated with 45S rDNA Copy Number across TCGA Cohorts.</b> Positive bars indicate up-regulated pathways while negative bars indicate down-regulated pathways. Each colored segment corresponds to a TCGA cancer cohort. . . . .	41

## List of Tables

2.1	TCGA cohorts Overview. . . . .	11
-----	--------------------------------	----

## Abbreviations

TCGA	The Cancer Genome Atlas
BLCA	Bladder Urothelial Carcinoma
COAD	Colon Adenocarcinoma
GBM	Glioblastoma Multiforme
LUAD	Lung Adenocarcinoma
LUSC	Lung Squamous Cell Carcinoma
PAAD	Pancreatic Adenocarcinoma
PRAD	Prostate Adenocarcinoma
rDNA	Ribosomal DNA
rRNA	Ribosomal RNA
RP	Ribosomal Protein
CNV	Copy Number Variation
OS	Overall Survival
PFI	Progression-Free Interval
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
WGS	Whole Genome Sequencing
RNA-seq	RNA Sequencing
DEG	Differentially Expressed Gene
TME	Tumor Microenvironment
TCR	T Cell Receptor
IGH	Immunoglobulin Heavy Chain
TIL	Tumor-Infiltrating Lymphocyte
CoxPH	Cox Proportional Hazards Model
HR	Hazard Ratio
CI	Confidence Interval