A Sexual Bias in mitochondrial protein-coding gene expression across different tissues and the prognostic value in multiple cancers

Alan Tardin da Silva[1], **Cristina dos Santos Ferreira**[2] **and Enrique Medina-Acosta**[1*]

[1] Núcleo de Diagnóstico e Investigação Molecular, Laboratório de Biotecnologia, Universidade Estadual do Norte Fluminense, Campos dos Goytacazes, RJ, Brazil.
[2] Laboratório Nacional de Computação Científica, Petrópolis, RJ, Brazil.

*** Correspondence**:
Enrique Medina-Acosta
quique@uenf.br

Keywords: Mitochondrial protein-coding genes; Sexual Dimorphism; mtDNA Genes; Aging; Survival outcome;

Abstract: words
Total number of words (with citations):
Number of figures in the main typescript:
Number of supplementary figures:
Supplementary data:

ABSTRACT

Mitochondria in mammalian cells provide ATP through oxidative phosphorylation. The overproduction of reactive oxygen species (ROS) in mitochondrial cells promotes cancer by modifying gene expression or function. Mating introduces competing mitochondrial (mtDNA) and nuclear DNA (nDNA) gene products, leading to biological differences between males and females for diseases and disorders such as cancer. There is a significant sex bias in aging-related conditions. We aimed to investigate whether sex and age affect mitochondrial protein-coding gene expression in cancer and, if so, to determine the prognosis value in survival outcomes, stemness, and immune cell infiltrates. We compared normal versus primary tumor transcriptomes (bulk RNA-Seq) from The Cancer Genome Atlas (TCGA), and the Genotype-Tissue Expression (GTEx) projects to test these hypotheses. Correlations between gene expression, survival, protective or risk factor, stemness, and immune cell infiltrate were performed in RStudio using UCSC Xena Shiny. Eleven mitochondrial protein-coding genes were altered in brain cancer (*MT-ND2*, *MT-ND1*, *MT-ATP8*, *MT-ATP6*, *MT-CO2*, *MT-CYB*, *MT-CO3*, *MT-ND4L*, *MT-ND4*, *MT-ND3*, *MT-CO1*). MT-ND5 and MT-ND6 are disproportionately expressed in female brain tissues. Mitochondrial global polymorphic expression sites of variation were more significant in the 50-59 and 60-79-year-old age groups than in the 20-49-year-old age groups. Pan-cancer survival analysis revealed a 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *and MT-ND6*) downregulated in low-grade glioma (LGG). This gene signature increased LGG overall survival, disease-specific survival, and progression-free interval without sex-specific association. However, the correlation with disease-free interval survival was female-specific. The 4-component gene signature was protective in LGG but risky in thymoma cancer and uterine corpus endometrial carcinoma. In LGG, the 4-component gene signature positively correlated with immune monocyte, NK, and B cell infiltrates and negatively correlated with T cell CD4+ Th2, macrophage M1 and M2, myeloid dendritic cell, and neutrophil. We identified a 13-component mitochondrial protein-coding gene signature associated with stemness in kidney chromophobe. A sex-biased effect was observed in mitochondrial protein-coding for brain tissues, with a female bias. However, an aging effect with higher polymorphic site expression was observed in male tissues. We conclude that the differentially expressed mitochondrial protein-coding genes provide new insights into carcinogenesis, helping to identify new prognostic markers. The overexpression of the 4-component gene signature is associated with a better prognosis in LGG, with positive and negative correlations with immune cell infiltrates.
.

Keywords: Mitochondrial protein-coding genes; Sexual Dimorphism; mtDNA Genes; Aging; Survival outcome;

INTRODUCTION

The mammalian mitochondrial DNA (mtDNA) encodes 13 key proteins essential for the functioning of the oxidative phosphorylation system (OXPHOS) (Larsson NG, 1998;Foster et al., 2006;Gustafsson et al., 2016). The mitochondria are therefore considered a vital component of all cells with a nucleus, so mtDNA-associated disorders or diseases can affect many tissues, and clinical features are variable. Moreover, defects in mitochondrial metabolism cause a wide range of human diseases, such as cancer, which include examples from all medical fields. Thus, mtDNA disorders are associated with a range of multi-systemic diseases in humans, where the understanding of the spectrum of these diseases has been expanded by the recognition of mutations in protein-coding genes in mitochondria caused by reactive oxygen species (ROS), which cause not only several neurodegenerative diseases but also other disorders as cancer (Harman, 1992b;Wallace, 1992;Geromel et al., 2001;Taylor and Turnbull, 2005;Schapira, 2006). Studies in the last ten years have shown that the human mtDNA of 13 protein-coding acts in several more functions than was described (Faure et al., 2011;Lee et al., 2013;Breton et al., 2014;Capt et al., 2016). Thus, investigating these genes' expression using RNA-Seq experiments is very useful in understanding the dynamics of expression in a population. It is also known that an RNA-Seq analysis allows the identification of gene expression signatures, which can be altered by external factors such as the environment, health status, or disease (da Silva Francisco Junior et al., 2019).

Using RNA-Seq analysis to investigate the extent of allelic specific expression (ASE) becomes a challenging task for mitochondrial protein-coding genes, because mitochondria have a haploid genome, and it is not possible to perform a genome-wide ASE, only for the detection of microheteroplasmy. Since microheteroplasmy is a form of heteroplasmy, this would also damage the mtDNA, but the levels of microheteroplasmy are about 2−5% in mitochondrial genomes. Moreover, the microheteroplasmy and heteroplasmy are present in mitochondrial diseases but also in normal individuals, where heteroplasmic variants without apparent functional consequences are observed in samples from individuals without any overt mitochondrial disease (Calloway et al., 2000;Kirches et al., 2001;Santos et al., 2008;He et al., 2010). Furthermore, there is a variation in pattern expression between mtDNA and the nuclear DNA variants. Consequently, calculating the ASE for mitochondrial protein-coding genes involves using a different methodology to calculate this specific expression in a population (Gemmell et al., 2004;Kassam et al., 2016;da Silva Francisco Junior et al., 2019;Stenton and Prokisch, 2020;Stewart and Chinnery, 2021).

The mtDNA is in the mitochondrial matrix and is closed to the respiratory chains, which are the primary source of ROS from the OXPHOS system. ROS induce somatic mutations in mtDNA due to the lack of protection by histones (Richter, 1995;Shokolenko et al., 2009). Thus, cumulative mutations in mtDNA can result in dysfunction of the OXPHOS system, leading to diseases associated with mitochondria, which is a hallmark of many diseases, such as cancer (van Oven and Kayser, 2009;Srinivasan et al., 2017). Additionally, the mitochondria are both producers and targets of oxidative stress induced by exposure to ROS, resulting in mitochondrial dysfunction and interfering in electron transfer activity in oxidative phosphorylation. Defects in electron transfer activity increase ROS production, thus establishing a “vicious cycle” with aging that interferes with production and target properties (Miquel et al., 1980;Linnane AW, 1989;Papa, 1996;Ozawa, 1997).

Aging is a universal process whose manifestations are familiar and unmistakable, and old age in humans and even animals can be readily recognized after a minimal assessment. Despite this, an accepted definition of aging and a detailed understanding of the biological mechanisms underpinning aging are elusive. So, aging is described as the progressive loss of function accompanied by decreased

fertility and increased mortality and disability (Kirkwood, 2000). Thus, aging is associated with evidence of deleterious changes in the molecular structure of DNA, proteins, lipids, and prostaglandins, all markers of oxidative stress (Harman, 1992a;1993). It has been recognized that ROS also play a role in normal signaling processes and that their generation is essential for maintaining homeostasis and cellular responsiveness (Droge, 2002). Since mitochondria are involved in oxidative stress, a mitochondrial theory of aging has been proposed, where an accumulation of somatic mutations in mtDNA induced by exposure to ROS leads to errors in the mtDNA-encoded polypeptides, affecting the electron transfer and oxidative phosphorylation system (Miquel et al., 1980;Linnane AW, 1989). Declines in the activity of the mitochondrial respiratory system and its constituent enzymes have been reported with advancing age, notably in cytochrome c oxidase, in several tissues, including skeletal muscle, heart, and liver (Muller-Hocker, 1989). Thus, the integrity of mitochondrial DNA in these tissues is gradually reduced with age, evidenced by the accumulation of deletions, duplications, and some point mutations in the mtDNA (Nagley and Wei, 1998;Kopsidas et al., 2002).

Having this perspective on the mammalian mtDNA, which encodes 13 proteins essential for the functioning of other systems, like the OXPHOS system, we considered building a computational meta-analysis of the mitochondrial protein-coding genes to investigate the existence of differentially expressed genes (DEG) in healthy tissues versus cancer tissues. We also consider exploring differences in the expression of these 13 genes in different tissues, at different age groups, and between sexes (females and males). In the same way, a possible punctual polymorphic point of variable expression of these genes was also investigated. This study approach explores the possibility of sex bias in the expression of the mitochondrial protein-coding gene expression and the impact on survival outcomes in several cancers. Thus, we hypothesize that a sex-bias expression of the 13 mitochondrial protein-coding genes may exist across different tissues and in the survival outcomes. Additionally, we hypothesize that the polymorphic site expression in mitochondria protein-coding genes may differ between the sexes at different age groups impacting pathologies such as cancer.

MATERIAL AND METHODS

Analysis of Differential Expression of Mitochondrial Protein-Coding Genes in Different Tissues

We used 19,131 RNA-Seq experiments samples from 18.972 donors (8516 Female and 10456 Male) from The Genotype-Tissue Expression (GTEx) for 52 non-diseased tissue sites across nearly 1000 individuals and The Cancer Genome Atlas (TCGA) for over 20,000 primary cancer and matched samples from healthy subjects spanning 33 cancer types, to search for a differential expressed genes (DEG). We used the UCSC Xena browser (https://xena.ucsc.edu/) to construct the base of this analysis and selected the TCGA TARGET GTEx cohort study and the phenotypic variables primary_site, study, sample type, main category, and gender as our first variable to organize our data based on these features. Posteriorly, we selected our second variable, genomic, using the mitochondrial protein-coding genes as input. Here, we used a gene expression RNA-Seq - RSEM norm_count, to access the expression of the 13 mitochondrial protein-coding genes. The RSEM was used because it relates to gene abundances from single-end or paired-end RNA-Seq data; therefore, RSEM outputs abundance estimates, 95% credibility intervals, and visualization files and can also simulate RNA-Seq data. After, we analyzed the DEG sets (above 1.5 and below -1.5 log2Fold Change, FDR < 0.05) of each primary health tissue compared to the same tissue in cancer (the primary tumor) to search if there is a tissue with DEG in health samples against tumor samples. We used the pipeline from the Ma'ayan lab's Appyter bulk RNA-Seq analysis and included the L1000FWD analysis, Limma-Voom Differential Gene Expression. All this analyzed data was downloaded from the UCSC Xena browser and taken to RStudio to organize the data.

Selection of Mitochondrial Protein-Coding Genes and Polymorphic sites

The mitochondrial protein-coding genes used in this analysis were those already described in the literature and validated by MITOMAP, a human mitochondrial genome database used as a reference for mitochondrial analysis (Table S1), and by NCBI genes. Data collection was performed through the RStudio interface, R programming language (https://cran.r-project.org/), using the data packages from the "BiomaRt", "Bioconductor," and "BSgenome" libraries to filter the desired information. The analysis involved extracting the physical positions of the 13 mitochondrial protein-coding genes and global polymorphic sites in the population contained within the genes. For this analysis, we used the UCSC Genome Browser of Santa Cruz, a public repository of databases, to construct a custom track with information on the physical positions of genes and polymorphic sites in the mitochondrial genome. The version genome used was the GRCh38/hg38 because it is the only one with data about genes and global polymorphic sites mapped in a population for the mitochondrial genome. We performed a computational meta-analysis to determine the physical position of each mitochondrial protein-coding gene and the global polymorphic sites (in bp). After, we search for polymorphic sites in UCSC genome browser Data base SNPs (dbSNPs151) to construct a site-specific expression based on polymorphic sites of variation already known by the dbSNPs151. This data was downloaded, and this information was integrated into the RStudio environment.

Expression Analysis of Mitochondrial Protein-Coding Genes and Polymorphic sites

For the analysis of polymorphic site expression at the mitochondrial protein-coding genes, a data collection was built based on the physical positions of the mitochondrial genome, using samples from public repositories of databases, such as UCSC (https://genome.ucsc.edu/). UCSC is an online genome browser with an extensive collection of assemblies and annotations of data about various model organisms, along with a large set of tools for visualizing, analyzing, and downloading data. In this Browser, we work with the GRCh38/hg38 genome version. We accessed the clade option, for Mammal and the genome option, for Human. We selected the group option to access GTEx RNA-Seq Signal Hub, where we extract normalized transcript expression levels (RPKM - Reads Per Kilobase of transcript, per Million mapped reads). Using the RPKM levels, we perform an intersection of these expression data with a custom track, previously obtained, containing the physical position of the polymorphic sites and the 13 mitochondrial protein-coding genes. Lastly, using the table option, we cross all this information with each patient's sample available in all 52 tissues (Table S2), going to the option output format where data points were selected. The Signal hub data were also separated according to other criteria, such as age groups, in 20-49 years, 50-59 years, and 60-79 years, by sex, male and female.

After obtaining this data collection, some packages were used in RStudio, such as stringr, plyr, rtracklayer, BSgenome, and XML, to perform another computational meta-analysis filtering the mitochondrial protein-coding genes polymorphic sites. These points were correlated by their physical positions within the protein-coding mitochondrial genes to use these polymorphic points as a filter to obtain the RPKM of each patient sample that crossed the physical position of these points. We averaged the expression of each signal hub in all tissues based on the physical position of the polymorphic sites so that only the signal hubs contained within the physical position were selected. The final data set was separated into two categories based on sex (Female or Male); and based on age, with three great groups, 20-49 years, 50-59 years, and 60-79 years, also considering the sex (Female or Male). We estimate the RPKM mean of the genes and the polymorphic sites in each tissue sample. We also performed the statistical Tukey test for multiple comparisons, using the mean expression of each gene or polymorphic sites in each tissue for the three ages to calculate the RPKM mean of each age group. The False Discovery Rate was also calculated to conceptualize multiple comparisons in the analyzed data. Therefore, the FDR tests using the Benjamini-Hochberg method had a cut-off of ≤ 0.05. This analysis was performed for females and males separately.

Pan-Cancer Survival Analysis of Mitochondrial Protein-Coding Genes

We used data from the GEPIA2 (http://gepia2.cancer-pku.cn/#index), a platform containing public data for interactive analysis of gene expression profiles in RNA-Seq samples from primary cancers from the TCGA (The Cancer Genome Atlas Program https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga), a reference cancer genomics program, covering 33 cancer types (Table S3), with a total of 9,736 samples. We also used data from samples from RNA-Seq experiments available in nontumor primary tissues from the Genotype-Tissue Expression (GTEx) project (8.587 samples) from 52 tissues (Table S2). A computational meta-analysis was performed to determine a pan-cancer overall survival map of the mitochondrial genes in 33 cancer types (Table S3). We set the following parameters: Survival Time Units: months, significance level: 0.05, P-Value Adjustment: False Discovery Rate (FDR), group cutoff: median, cutoff-high: 50% used.

We define gene signatures as the additive expression profiles of at least three genes exhibiting the prognostic value per tissue.

Prognostic Factors of Pan-Cancer Survival Outcomes

The prognostic factor values of the 13 mitochondrial protein-coding genes were estimated by the Kaplan-Meier method using a computational meta-analysis in RStudio, P (version 2022.02.3 build 492 for Windows) (Allaire, 2012) using the UCSC Xena Shiny standalone application (WANG et al., 2021) (http://xena.ucsc.edu/). We queried the expression profiles of each gene using four methods of survival outcomes: overall survival (OS), disease-specific survival (DSS), disease-free interval survival (DFI), and progression-free interval (PFI). OS refers to the percentage of patients alive after cancer diagnosis throughout the study period. DSS describes the percentage of patients who have not died from a cancer type in a defined period, such as patients who died from causes other than the specific cancer type is not counted in this measurement. DFI is the length of time after primary treatment for cancer that the patient survives without any signs or symptoms of that cancer. PFI is the length of time, during and after the treatment of a cancer type, that a patient lives with the disease but does not worsen. We set the parameters s described in the prior session and did not categorize the patients by age. We also investigated whether there is a sex-specific effect on survival outcomes.

Molecular Profile Analysis for Risk or Protective Effects in Different Cancers

For this analysis, we used mRNA data expression for 33 different types of cancer available through the TCGA for all samples (Table S3). In this analysis, we constitute a gene signature that presents at least three genes per tissue. Essentially, a statistical regression model commonly used in medical research, the Cox model, was used to investigate the association between patient survival time and one or more predictor variables, in this case, the risk or protective effect. Therefore, the effect of covariates estimated by any proportional hazard model can be reported as a hazard ratio. In this way, we aim to identify whether the mitochondrial protein-coding genes have risk or protective effects based on OS (Overall Survivor) having an adjusted threshold of 0.5. We define genes as risky (log (Hazard Ratio) > 0) or protective (log (Hazard Ratio) < 0), or NS (No statistical significance, P-value > 0.05)).

Analysis of the association of the molecular profile to immune response signatures and Stemness

Subsequently, we investigated the association between the molecular profile of mitochondrial protein-coding genes and stemness (the impact of a stem cell-like tumor phenotype). Here, we aim to obtain a possible relationship, positive (+1), negative (-1), or neutral (0), between gene signatures and Stemness in all available cancer types (Table S3). We also investigate the existence of a correlation between the mitochondrial protein-coding genes and the immune infiltrate signatures. In both analyses, genetic signatures presented at least three genes per tissue or at least three genes per tissue in each immune cell. For both analyses, we used mRNA expression data from patients of Pan-Cancer Atlas for 33 different types of cancer. Spearman's test was performed to assess the correlation of the signatures, positive (≥0.50) or negative (≤ -0.50). We also filter this data considering its p-value, admitting only data with a p-value lower than 0.05.

RESULTS

Differential Expression of Mitochondrial Protein-Coding Genes

We compared the RNA expression profiles of the 13 mitochondrial protein-coding genes in tumor tissues from the TCGA project and matched (n=17) normal tissues from the GTEX project to identify disease conditions in which the genes are differentially expressed. Eleven mitochondrial protein-coding genes are DEG in brain tissue (*MT-ND2*, *MT-ND1*, *MT-ATP8*, *MT-ATP6*, *MT-CO2*, *MT-CYB*, *MT-CO3*, *MT-ND4L*, *MT-ND4*, *MT-ND3*, *MT-CO1*) with LogFC ≥1.5 (p ≤0.05). We also noticed other differentially expressed genes with LogFC ≥1.5 in tissues like the breast (*MT-ATP8*, *MT-ND1*, *MT-ND2*, *MT-CYB*, p ≤0.05), liver (*MT-ND4*, *MT-ATP6*, *MT-ATP8*, *MT-ND4L*, *MT-CYB*, *MT-ND2*, p ≤0.05), pancreas (*MT-ND2*, p ≤0.05) and testis (*MT-ND1*, p ≤0.05). We also observed a LogFC ≤ -1.5 in tissues such as peripheral blood (*MT-ND4*, *MT-ND5*, *MT-CO1*, *MT-ND4L*, p ≤0.05), ovary (*MT-ND6*, *MT-ND1*, *MT-CO1*, p ≤0.05) and thyroid (*MT-ND6*, p ≤0.05) (Figure 1) (Table S4).

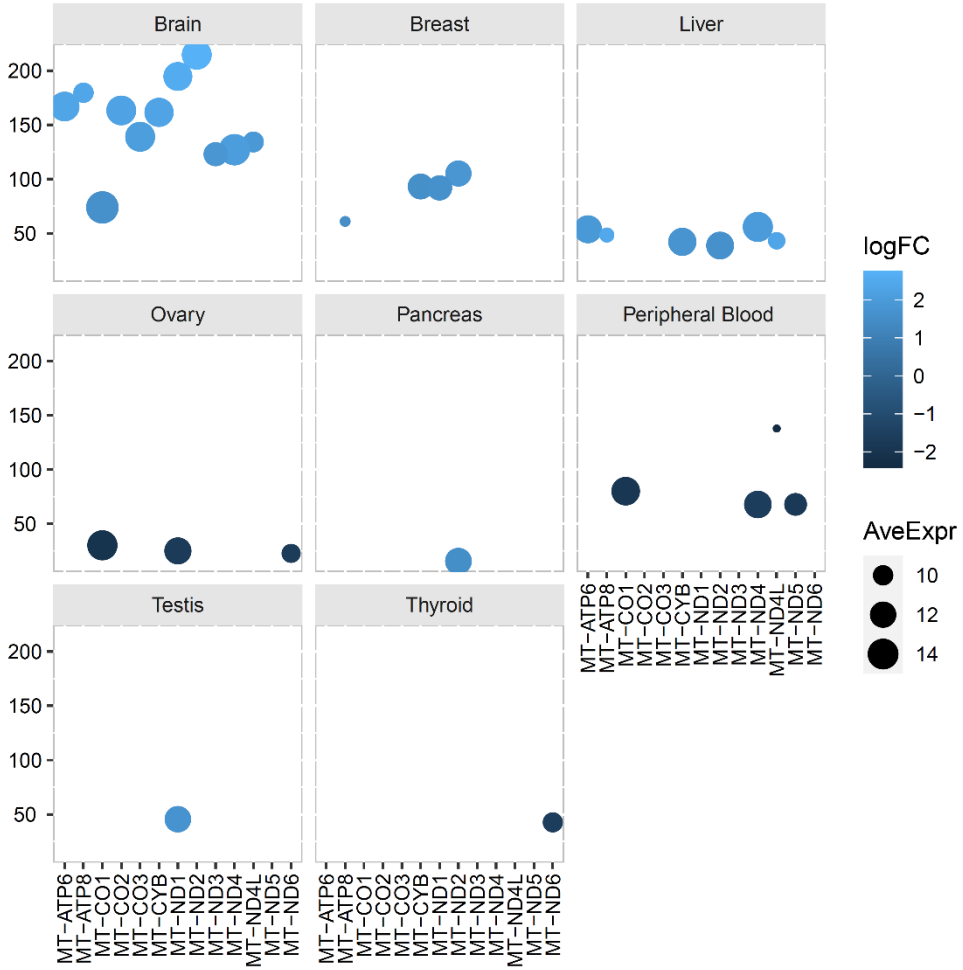


Figure 1. Balloon plot of the differential expressed mitochondrial protein-coding genes in normal versus cancer tissues. Balloon size refers to the average expression (norm_counts); blue color intensity refers to the Log2FC. Overexpression (+); underexpression (-). X-axis: the 13 mitochondrial protein-coding genes; Y-axis, FDR (-log10). The tissues are indicated in each box.

In Silico Consolidation of the Mitochondrial Protein-Coding Genes and Polymorphic sites

From the names of the mtDNA genes, their physical positions, and to which chain within the mitochondrial genome they belonged (H-heavy; L-light) were obtained, as well as the product of each gene. Subsequently, still using the physical positions of the mtDNA genes, 2297 polymorphic sites were obtained in the mitochondrial genome, of which only 1540 were within the 13 mitochondrial genes (MT-ND1 = 156, MT-ND2 = 126, MT-ND3 = 46, MT-ND4 = 157, MT-ND4L = 39, MT-ND5 = 230, MT-ND6 = 81, MT-CO1 = 162, MT-CO2 = 82, MT-CO3 = 108, MT-CYB = 191, MT- ATP6 = 112, MT-ATP8 = 37) (Figure S1).

Expression Profile of Sex-specific Aging-related in Mitochondrial Protein-Coding Genes and Polymorphic sites

We compared the expression profiles of the 13 mitochondrial protein-coding genes and polymorphic sites in males and females according to three age groups (20-49-year-old, 50-59-year-old, and 60-79-year-old) in 52 GTEx tissues. We used two query strategies: first, we used the physical coordinates of each gene to extract the normalized transcript expression levels (RPKM) as a metric of the abundance of each donor in the mitochondria genome, and second, we used the physical position of polymorphic sites in population from the dbSNP151 database to capture site-specific abundance. We observed sex-biased effects on expression profiles across different tissues.

Eleven genes (*MT-ND6*, *MT-ND5*, *MT−ND4*, *MT-ND4L*, *MT−ND3*, *MT−ND2*, *MT−ND1*, *MT−CYB*, *MT−CO3*, *MT−CO1*, *MT−ATP6*; Tukey test, $p \leq 0.05$) showed differences in expression profile of males in thirty-four tissues across different ages (Figure 2) (Table S5-S6). The most significant male-biased expression across different ages was noticed in *MT-ND5* for Artery – Aorta, Brain – Substantia Nigra, and Esophagus – Mucosa (Wilcox test, $p \leq 0.001$) and in the *MT-ND6* for Brain – Hypothalamus and Kidney – Cortex (Wilcox test, $p \leq 0.001$) (Table S5-S6). For females, ten genes (*MT−ND6*, *MT−ND5*, *MT−ND4*, *MT−ND2*, *MT−ND1*, *MT−CYB*, *MT−CO3*, *MT−CO1*, *MT−ATP8*, *MT−ATP6*; Tukey test, $p \leq 0.05$) exhibited differences in expression profile for nineteen tissues across different ages (Figure 3) (Table S5-S6). The most significant female-biased expression across different ages was observed in *MT-ATP6*, *MT-ATP8*, and *MT-CO1* for Artery – Coronary, and in *MT-CO1*, *MT-CO3*, and *MT-CYB* for Adipose – Visceral (Omentum) (Wilcox test, $p \leq 0.05$). We also noticed that *MT-ND2* exhibit a highly significant female-biased expression in Brain – Frontal Cortex (BA9) and Brain – Substantia Nigra across different ages (Wilcox test, $p \leq 0.001$) and in *MT-ND1* only for Brain – Substantia Nigra (Wilcox test, $p \leq 0.001$). Interestingly, the *MT-ND5* also exhibited a female-biased expression in Artery – Aorta, Brain – Frontal Cortex (BA9), and Colon – Transverse (Wilcox test, $p \leq 0.001$) at different ages, and the same was noticed in *MT-ND6* for Brain – Frontal Cortex (BA9) and Brain – Substantia Nigra (Wilcox test, $p \leq 0.001$) (Table S5-S6).

Considering possible differences in expression between the 52 different tissues analyzed for the 13 mitochondrial protein-coding genes in the three age groups, we observed sex-biased and age-group-dependent effects on expression profiles (Figure 4-6). In the 20-49-year-old group, thirty-four different tissues showed significantly different expression profiles for *MT-ATP6*, *MT-ATP8*, *MT-ND1*, *MT-ND2*, *MT-ND3*, *MT-ND4*, *MT-ND4L*, *MT-ND5*, *MT-ND6*, *MT-CYB*, *MT-CO1*, *MT-CO2*, *MT-CO3* (Wilcox test, $p \leq 0.05$). In this age group, eight genes exhibited the most significant expression profiles in Artery - Coronary (*MT-ATP6*, *MT-ATP8*, *MT-CO1*, *MT-CO2*, *MT-ND1*, *MT-ND2*, *MT-ND4*, and *MT-ND5*; Wilcox test, $p \leq 0.001$). For the 50-59-year-old group, we noticed nineteen different tissues

where 13 genes exhibited the same effect (Wilcox test, p≤0.05). For this age group, the most significant expression profile between the tissues occurred in *MT-CO3*, *MT-CYB*, and *MT-ND6* in Adipose – Subcutaneous (Wilcox test, p≤0.001); *MT-ATP6*, *MT-ATP8*, *MT-CO1*, *MT-CO3*, and *MT-ND5* in Adipose – Visceral (Omentum) (Wilcox test, p≤0.001); *MT-ATP6*, *MT-ATP8*, *MT-CO1*, *MT-CO2*, *MT-ND1*, *MT-ND2*, *MT-ND5*, and *MT-ND6* in Brain – Cerebellar Hemisphere (Wilcox test, p≤0.001); *MT-CO1*, *MT-ND4*, *MT-ND4L*, and *MT-ND6* in Brain – Hippocampus (Wilcox test, p≤0.001); *MT-ATP8*, *MT-CO1*, *MT-CO3*, *MT-CYB*, *MT-ND4* and *MT-ND5* in Brain – Hypothalamus (Wilcox test, p≤0.001); *MT-ND1*, *MT-ND4*, and *MT-ND6* in Brain – Nucleus accumbens (basal ganglia) (Wilcox test, p≤0.001); *MT-ND1* in Brain – Substantia Nigra, (Wilcox test, p≤0.001). The last age group, 60-79-year-old, exhibited thirty-four tissues with differences in expression for the same 13 mitochondrial protein-coding genes (Wilcox test, p ≤0.05). In this last age group, the most significant expression profile between the tissues and the mitochondrial protein-coding genes happens in *MT-CO1*, *MT-CO3*, *MT-CYB*, and *MT-ND5* in Adipose – Visceral (Omentum) (Wilcox test, p≤0.001); *MT-ND4L* and *MT-ND5* in Artery – Coronary (Wilcox test, p≤0.001); *MT-ATP6*, *MT-ATP8*, *MT-CO1*, *MT-CO2*, *MT-ND1*, *MT-ND2*, *MT-ND4*, *MT-ND4L*, *MT-ND5* and *MT-ND6* in Brain – Amygdala (Wilcox test, p≤0.01); *MT-ND4* and *MT-ND5* in Brain – Putamen (basal ganglia) (Wilcox test, p≤0.05); *MT-CO1*, *MT-CO3*, *MT-ND4* and *MT-ND5* in Colon – Sigmoid (Wilcox test, p≤0.01); *MT-ATP6*, *MT-ATP8*, *MT-CO1*, *MT-ND1*, *MT-ND2*, *MT-ND4* and *MT-ND6* in Esophagus – Muscularis (Wilcox test, p≤0.01); MT-ND1, MT-ND2, MT-ND5 and MT-ND6 in Kidney – Cortex (Wilcox test, p≤0.05); *MT-ATP8*, *MT-CO1*, *MT-CO2*, *MT-CYB*, *MT- ND1*, *MT-ND3*, *MT-ND4*, *MT-ND5* and *MT-ND6* in Minor Salivary Gland (Wilcox test, p≤0.001); *MT-CO1*, *MT-ND1*, *MT-ND5* and *MT-ND6* in Small Intestine – Terminal Ileum (Wilcox test, p≤0.01); *MT-CYB*, *MT-ND4* and *MT-ND5* in Thyroid (Wilcox test, p≤0.01).

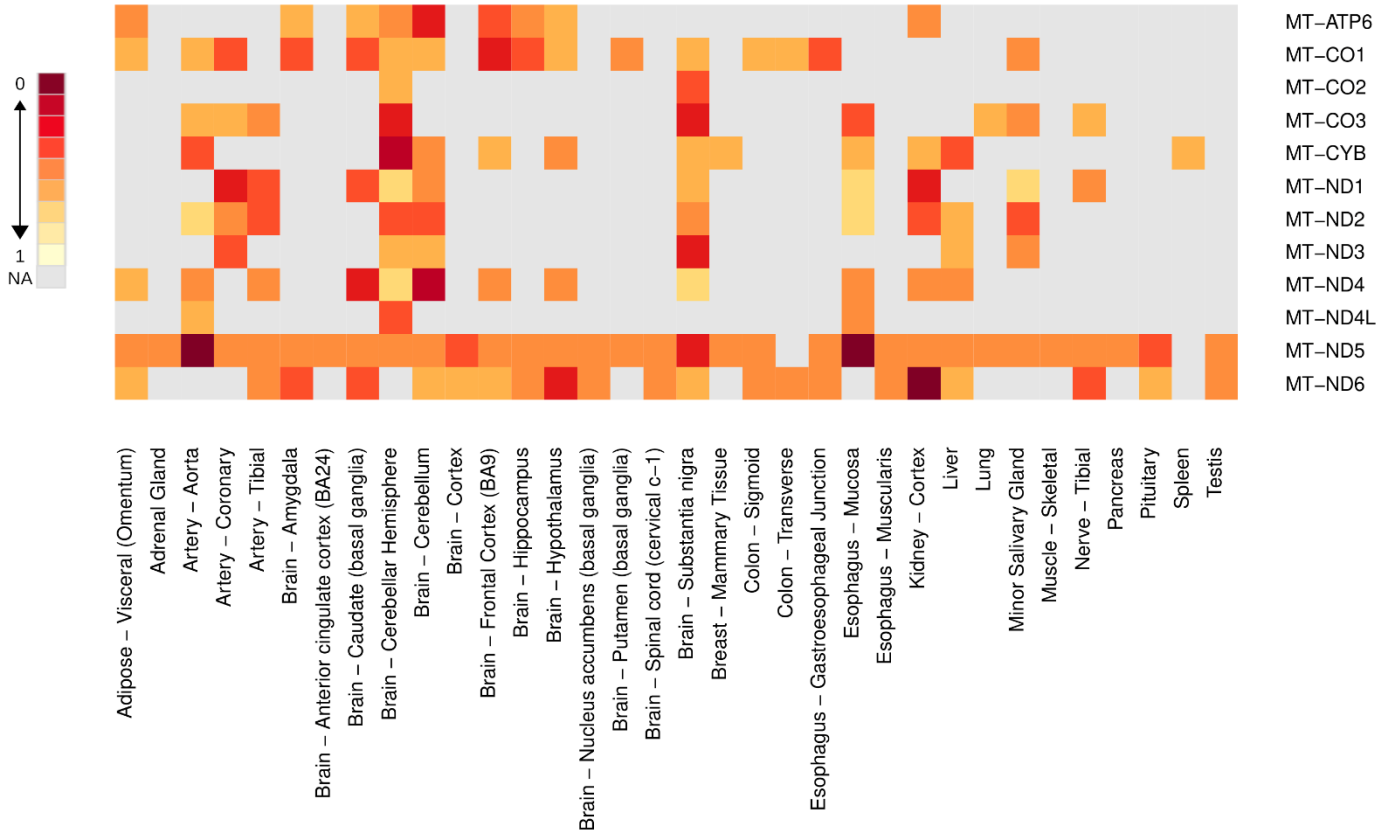


Figure 2. Heatmap of significant mitochondrial protein-coding gene expression across different ages for each tissue in males. The expression of all mitochondrial protein-coding genes was measured using RPKM (reads per kilobase million), and the color scale denotes the level of significance of gene expression in each tissue. The different ages consisted of three groups (20-49 years; 50-59 years, and 60-79 years). The significance level was measured based on the Tukey test and the FDR adjustment, Benjamin-Hochberg.

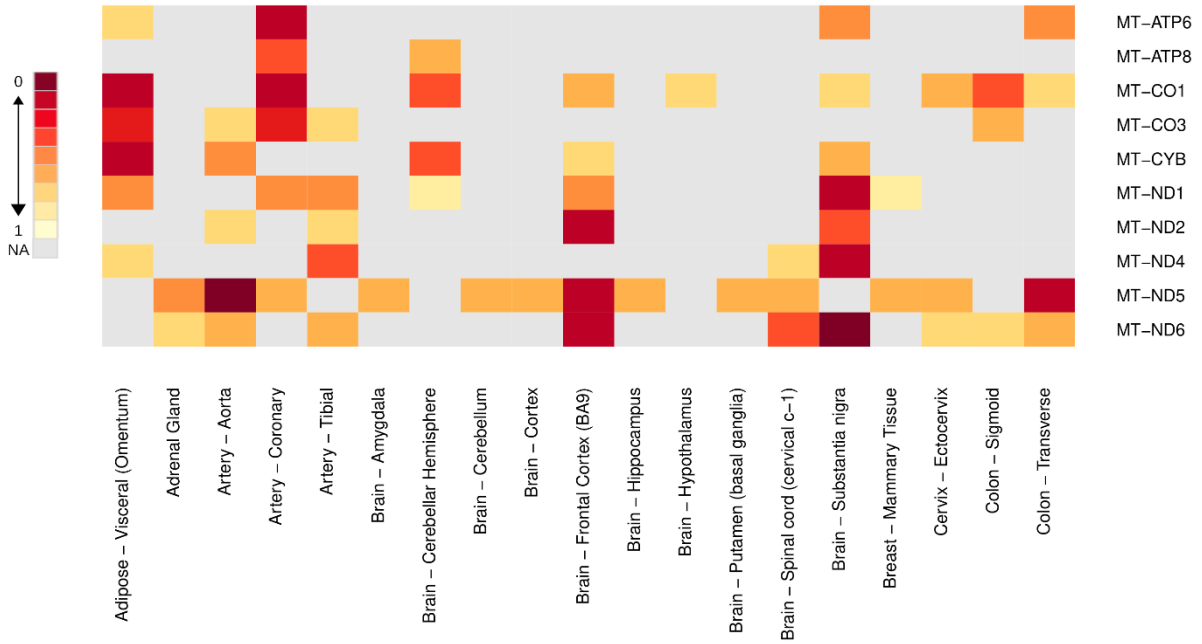


Figure 3. Heatmap of significant mitochondrial protein-coding gene expression across different ages for each tissue in females. The expression of all mitochondrial protein-coding genes was measured using RPKM (reads per kilobase million), and the color scale denotes the level of significance of gene expression in each tissue. The different ages consisted of three groups (20-49 years; 50-59 years, and 60-79 years). The significance level was measured based on the Tukey test and the FDR adjustment, Benjamin-Hochberg.

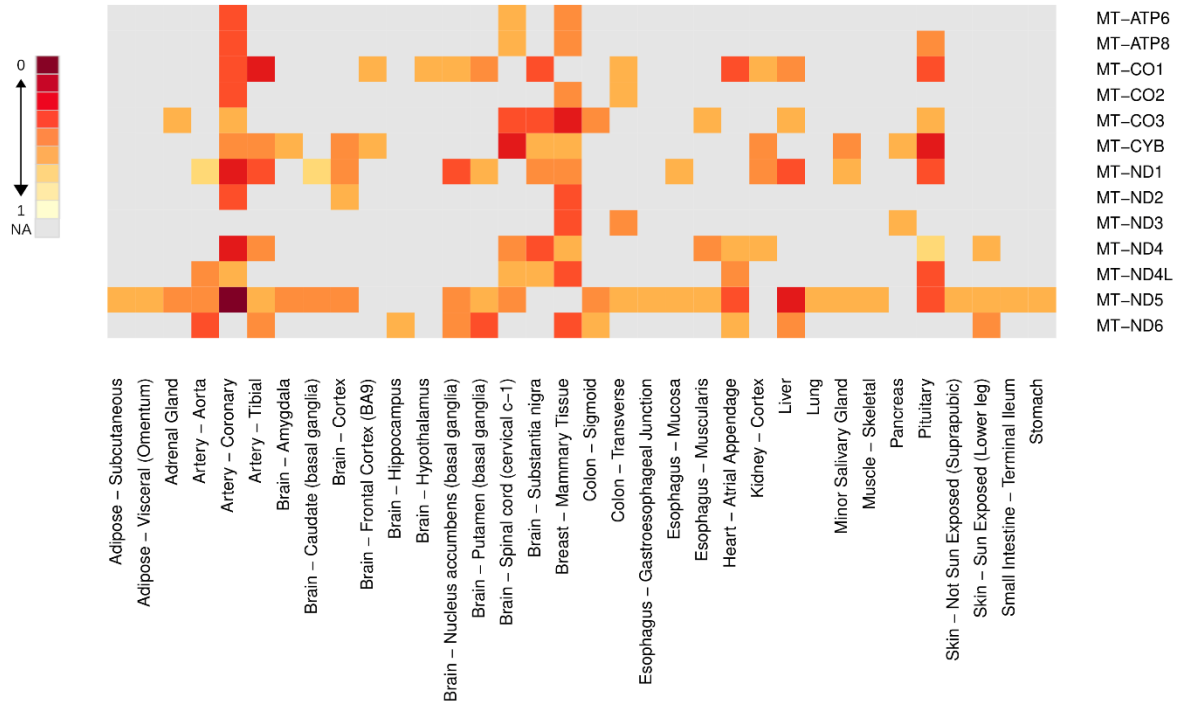


Figure 4. Heatmap of significant mitochondrial protein-coding gene expression between sexes in 20-49-year-old group. All mitochondrial protein-coding genes were measured between the sexes (female and male) using RPKM (reads per kilobase million) across different tissues, and the color scale denotes the level of significance of gene expression in each tissue. The sex-specific tissue samples were excluded, and the significance level was measured based on the Wilcox test and the FDR adjustment, Benjamin-Hochberg.

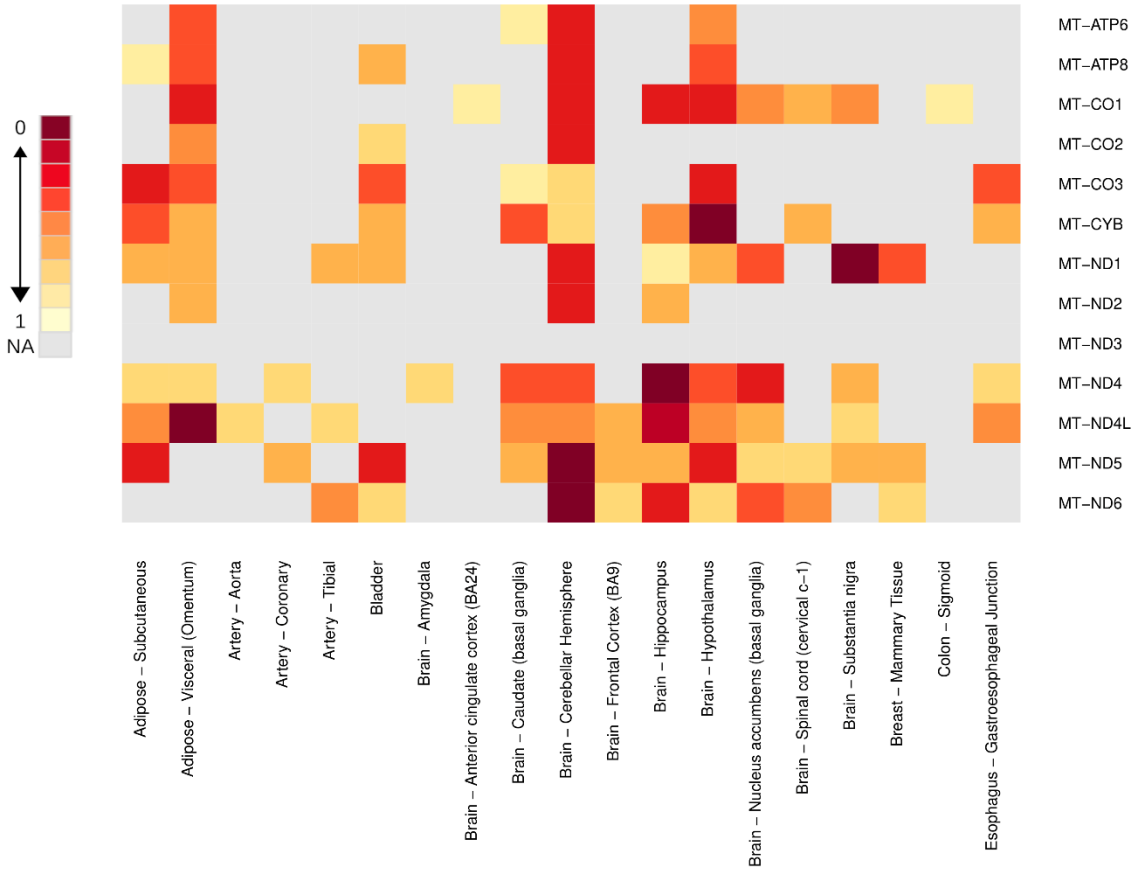


Figure 5. Heatmap of significant mitochondrial protein-coding genes expression between sexes in 50-59-year-old group. All mitochondrial protein-coding genes were measured between the sexes (female and male) using RPKM (reads per kilobase million) across different tissues, and the color scale denotes the level of significance of gene expression in each tissue. The sex-specific tissue samples were excluded, and the significance level was measured based on the Wilcox test and the FDR adjustment, Benjamin-Hochberg.

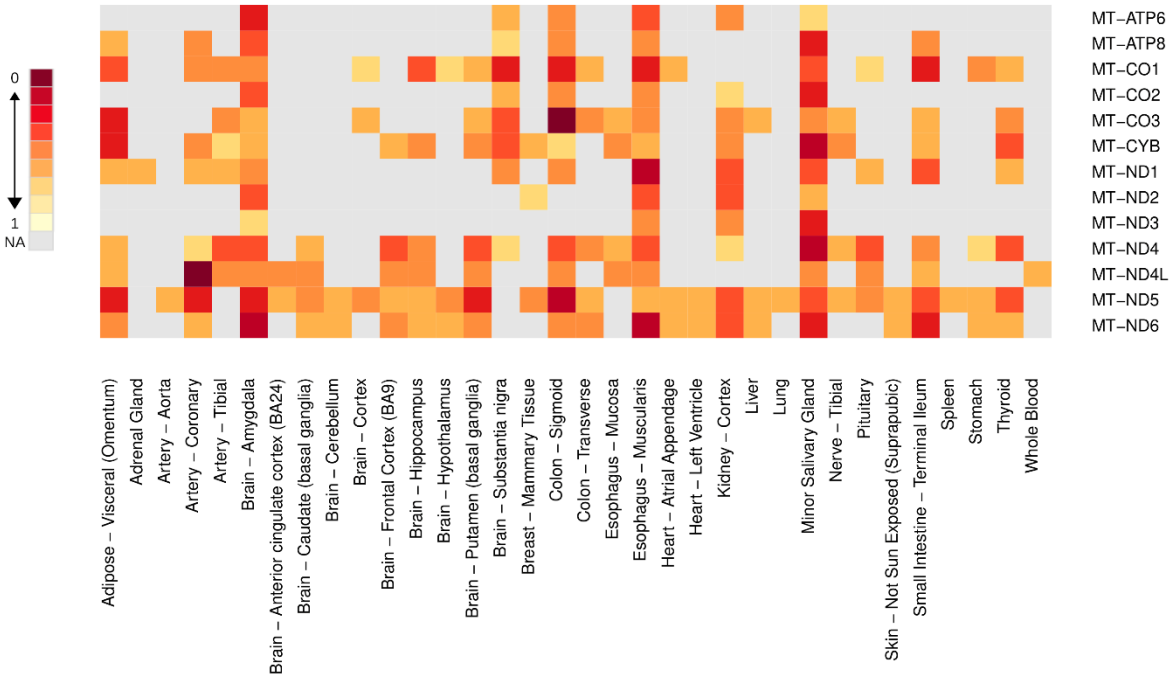


Figure 6. Heatmap of significant mitochondrial protein-coding genes expression between sexes in 60-79-year-old group. All mitochondrial protein-coding genes were measured between the sexes (female and male) using RPKM (reads per kilobase million) across different tissues, and the color scale denotes the level of significance of gene expression in each tissue. The sex-specific tissue samples were excluded, and the significance level was measured based on the Wilcox test and the FDR adjustment, Benjamin-Hochberg.

The polymorphic profile of the variation points in population for the 13 mitochondrial protein-coding genes was clustering in 52 tissues from the GTEx project comprising three different ages (20-49-year-old, 50-59-year-old, 60-79-year-old), which allowed building a landscape of these polymorphic sites related to sexes (males and females) (Figure 7 and 8). This landscape will provide information about the variation of polymorphic points along the transcripts and whether this variation is global or sustained at all points in the population by using the RPKM transcript. We identify an age-group-dependent sex-biased effect on the polymorphic sites in females in ten different tissues among the three age groups (Adipose – visceral (Omentum), Adrenal Gland, Artery Coronary, Brain – Amygdala, Brain – Frontal cortex (BA9), Brain – Hippocampus, Brain – Putamen (basal ganglia), Brain – Spinal cord (cervical c-1), Brain – Substantia nigra, Breast – Mammary tissue, Cervix – Ectocervix; Tukey test, $p \leq 0.05$). In the same way, an age-male-biased of polymorphic sites was found in eight tissues (colon–sigmoid, colon – transverse, esophagus – gastroesophageal junction, liver, lung, muscle-skeletal, nerve tibial, prostate; Tukey test, $p \leq 0.05$).

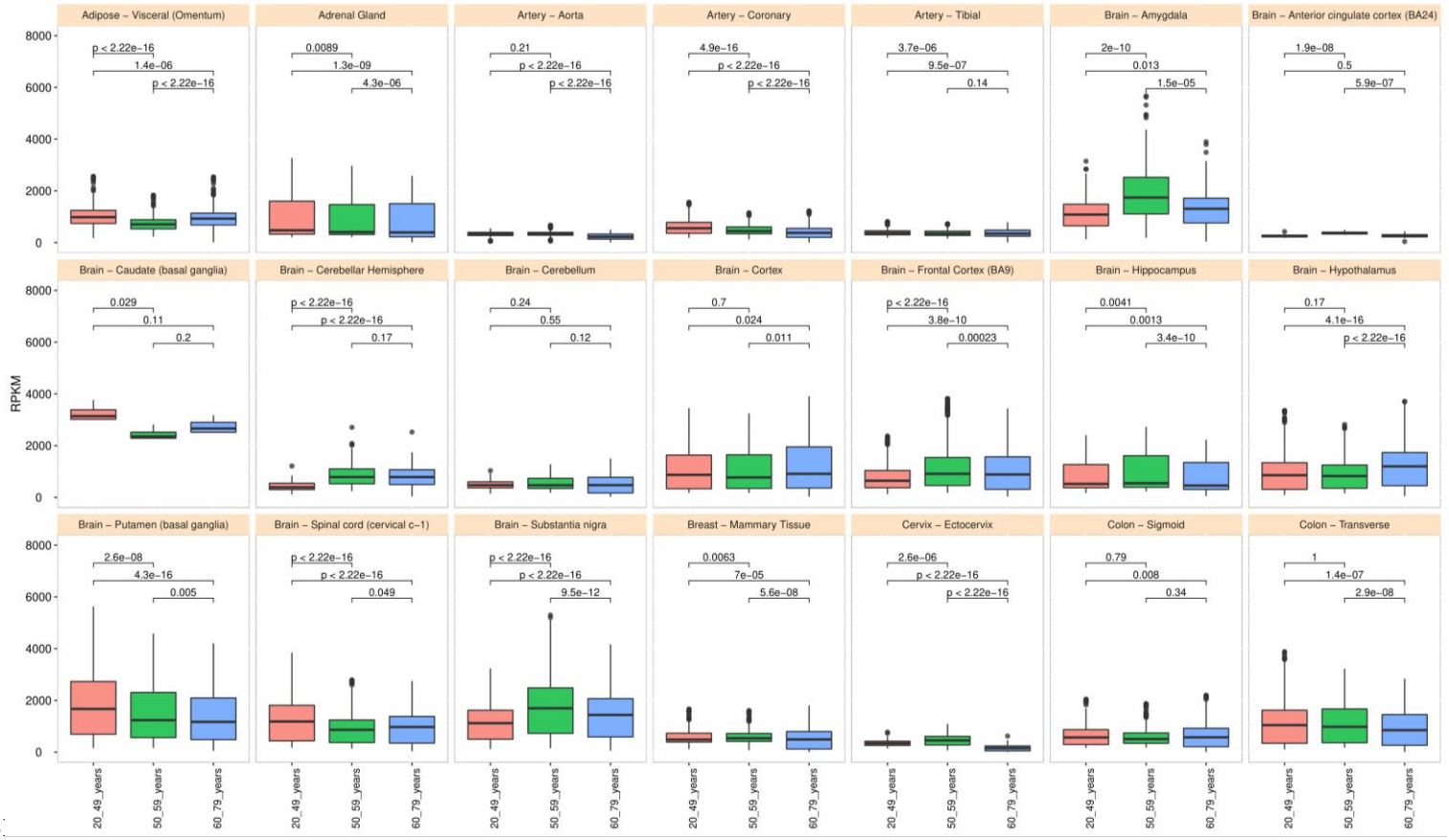


Figure 7. Boxplot of mitochondrial polymorphic site expression in females across tissues at different age groups. The polymorphic site expression at the mitochondrial protein-coding genes was measured using RPKM (reads per kilobase million) in three age groups, 20-49 years (red), 50-59 years (green), and 60-79 years (blue). The Tukey test and the FDR adjustment, Benjamini-Hochberg assessed significant differences between age groups.

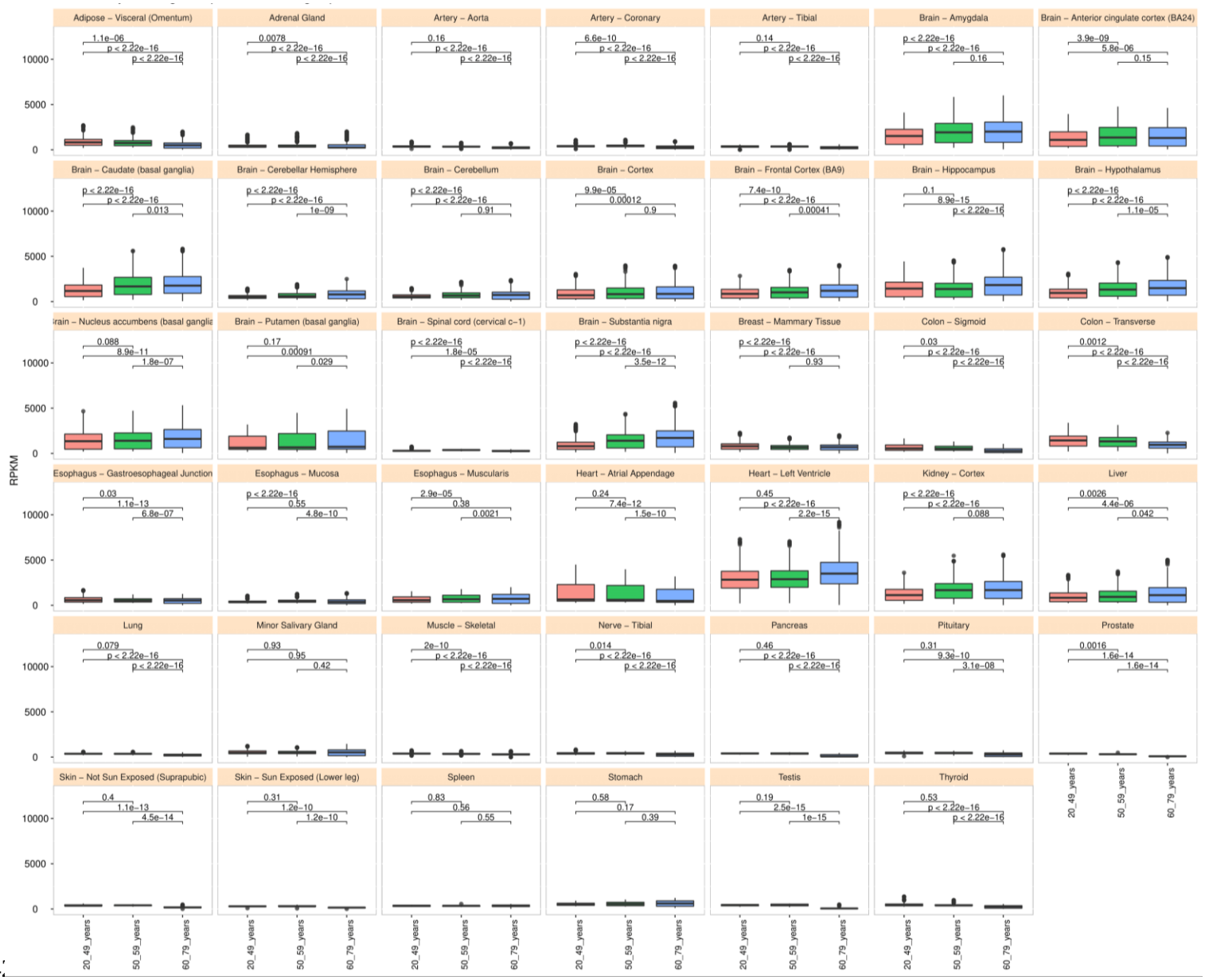


Figure 8. Boxplot of mitochondrial polymorphic site expression in males across tissues at different age groups. The polymorphic site expression at the mitochondrial protein-coding genes was measured using RPKM (reads per kilobase million) in three age groups, 20-49 years (red), 50-59 years (green), and 60-79 years (blue). The Tukey test and the FDR adjustment, Benjamini-Hochberg assessed significant differences between age groups.

Pan-Cancer Survival Map Analysis Based on the Expression Status of the 13 Mitochondrial Protein-coding Genes

We estimated the overall survival outcome for each of the molecular expression profiles of the 13 mitochondrial protein-coding genes in 33 cancer types to map the hazard risk based on the Cox proportional-hazards model for overexpression and under-expression patient groups. The under-expression of the *MT-CO1*, *MT-CO2*, *MT-ND5*, and *MT-ND6* genes is a hazard risk in lower-grade glioma (hazard ratio range 0.41-0.51, Logrank p ≤0.00026, p(HR) ≤0.00033), while the under-expression of the MT-CO3 gene is a hazard risk in pancreatic adenocarcinoma (PAAD) (hazard ratio 0.45, p(HR)=0.00022, Log-rank p=0.00017), Figure 9 (Table S10).

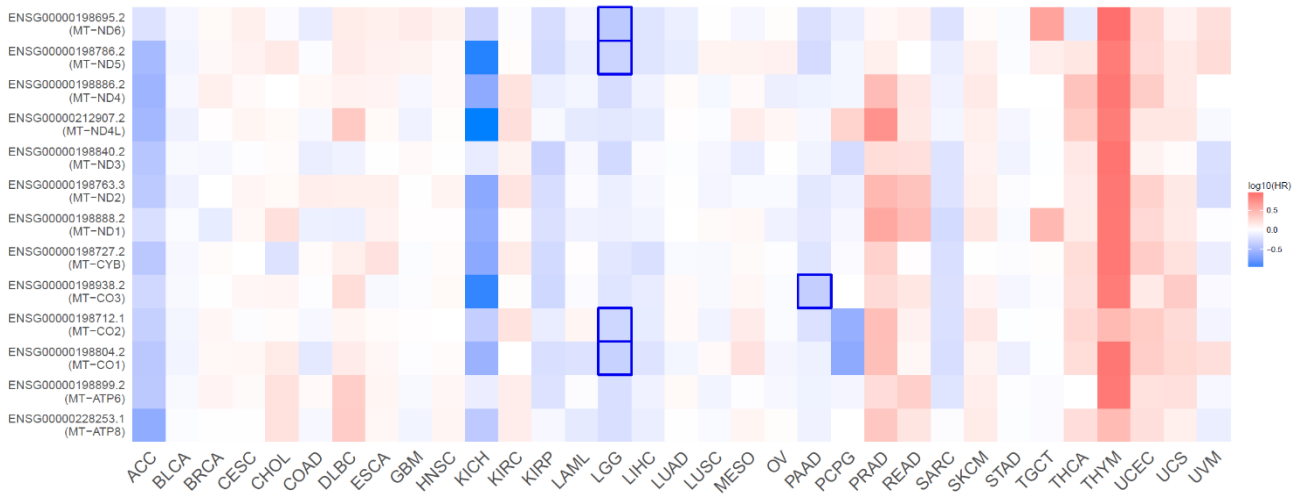


Figure 9. Overall survival heatmap for the 13 mitochondrial protein-coding genes. Squares denote overall survival associated with overexpression (red) or under-expression (blue). Highlighted frames indicate significant associations with overall survival outcomes based on hazard ratios (logarithmic scale log10). Y-axis: mitochondrial genes (Ensemble transcripts and UCSCXena names), X-axis: pan-cancer types. The under-expression profiles of the *MT-CO1*, *MT-CO2*, *MT-ND5*, and *MT-ND6* genes have a significant overall survival risk in LGG (hazard ratio range 0.41-0.51, Logrank p ≤0.00026, p(HR) ≤0.00033), while the under-expression of the MT-CO3 gene is a hazard risk in PAAD (hazard ratio 0.45, p(HR)=0.00022, Logrank p=0.00017).

The 4-component gene signature is a Prognostic Factor of Survival Outcomes in LGG

We estimated the prognostic factor value of the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, and *MT-ND6*) by correlating the molecular expression profile with rates of survival outcomes in LGG patients (n=521). We used four methods of survival outcomes: overall survival (OS) and disease-specific survival (DSS), disease-free interval (DFI), and progression-free-interval (PFI). Overexpression of the 4-component gene signature was significantly correlated with OS (Log-rank p < 0.0001) (Figure 10, Table S11), DSS (Log-rank p < 0.0001) (Figure 11, Table S12), and PFI (Log-rank p = 2e−04) (Figure 12, Table S13). The favorable prognostic factor of the overexpression was not sex-specific since, in either males or females, the overexpression was significantly correlated with OS, DSS, and PFI (Log-rank p ≤ 0.0076) (Figure S2-S7; Tables S14-S19). However, for the DFI outcome, we observed a sex-specific association in that the downregulation of the 4-component gene signature in female patients exhibited a greater prognostic factor value (Log-rank p= 0.019) compared to males (Log-rank p= 0,14), Figure S8 and Figure S9 (Tables S20 and S21). There are no sufficient data (n ≤10 samples) about the clinical-pathologic stages (Stages I-IV) in LGG subjects. Therefore, we could not estimate the prognostic values with the clinical staging.

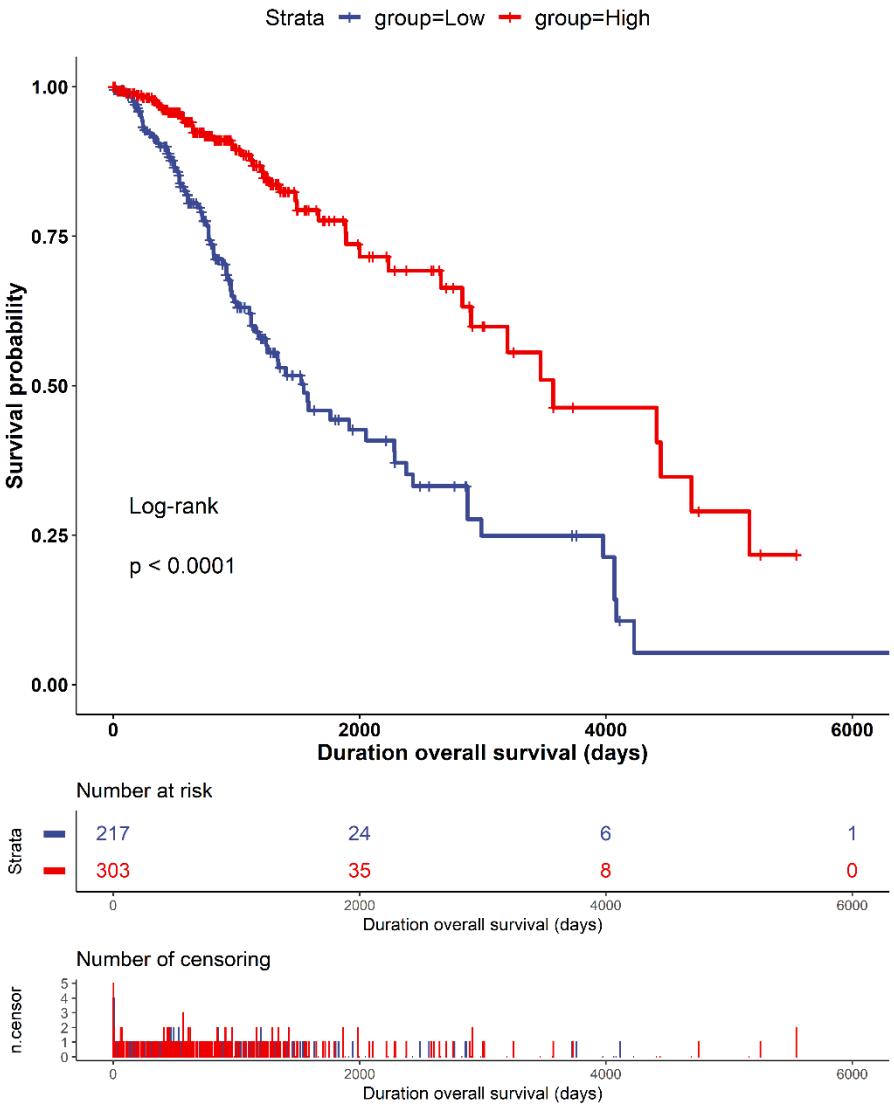


Figure 10. Kaplan-Meier plot of overall survival outcome in low-grade glioma. The curves denote the contribution of the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*) to the overall survival outcome in patients from the low (blue – downregulated) and high (red – upregulated) expression groups. The censoring number refers to patients who did not suffer the outcome of interest during the specified study period. The overexpression of the signature has a more significant prognostic factor value.

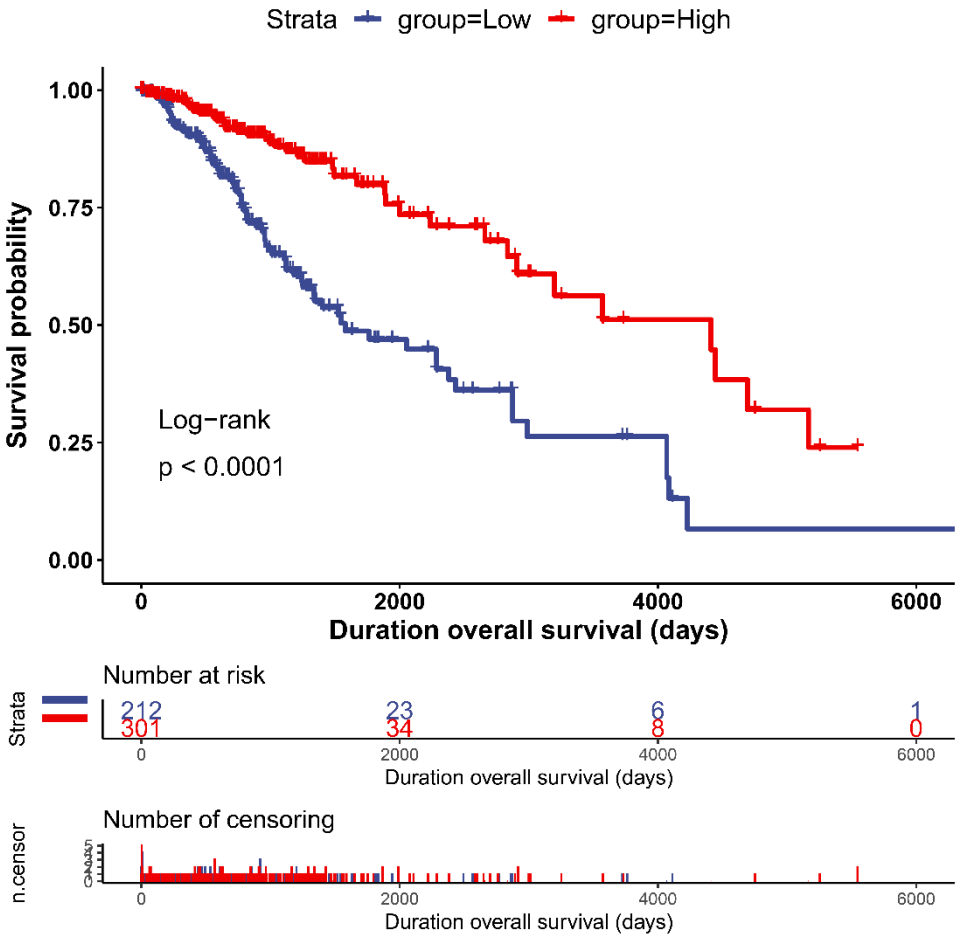


Figure 11. Kaplan-Meier plot of disease-specific outcome in low-grade glioma. The curves denote the contribution of the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*) to the disease-specific outcome in patients from the low (blue – downregulated) and high (red – upregulated) expression groups. The censoring number refers to patients who did not suffer the outcome of interest during the specified study period. The overexpression of the signature has a more significant prognostic factor value.

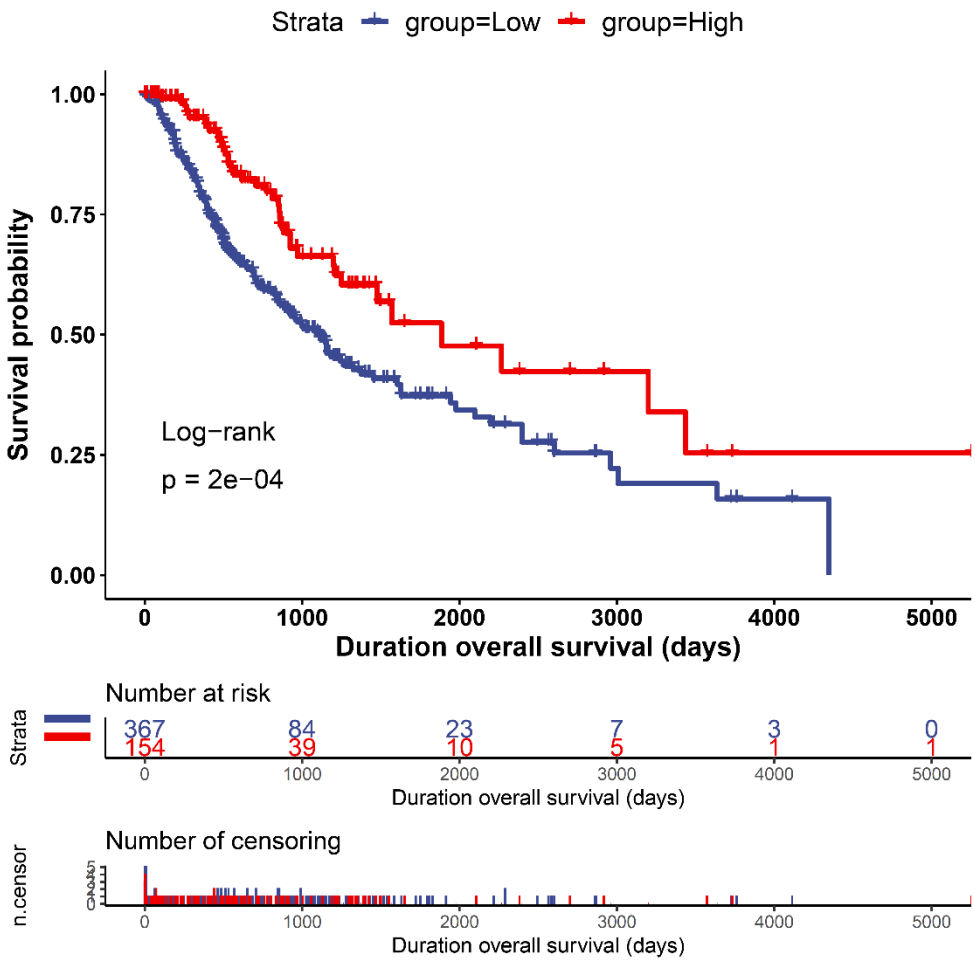


Figure 12. Kaplan-Meier plot of progression-free-interval outcome in low-grade glioma. The curves denote the contribution of the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*) to the progression-free-interval outcome in patients from the low (blue – downregulated) and high (red – upregulated) expression groups. The censoring number refers to patients who did not suffer the outcome of interest during the specified study period. The overexpression of the signature has a more significant prognostic factor value.

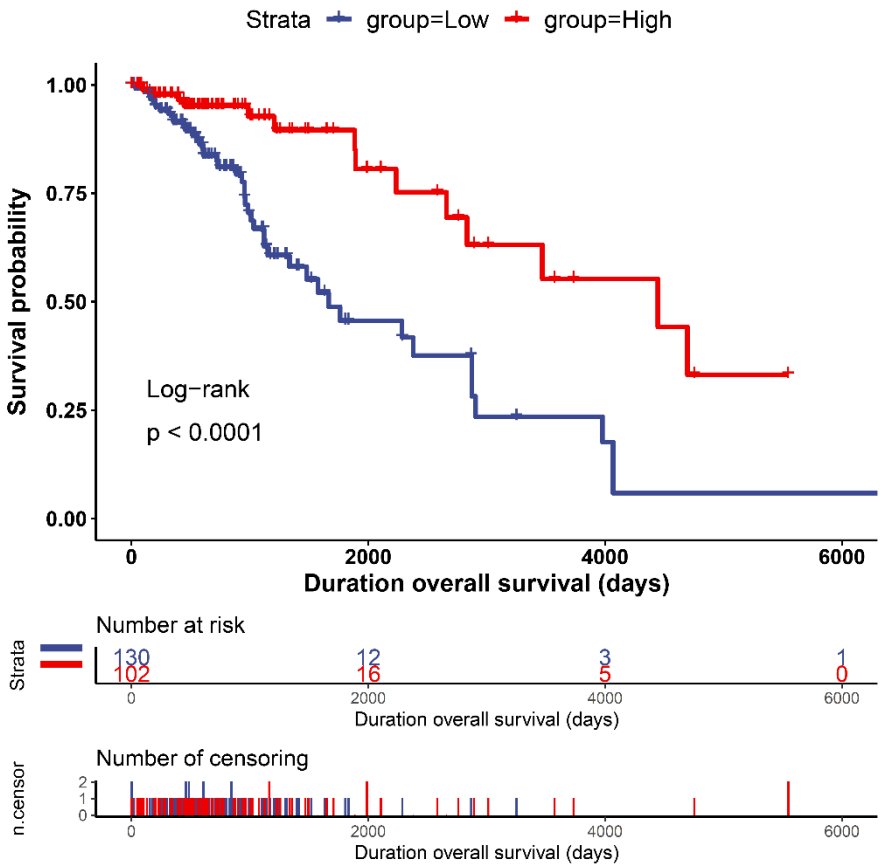


Figure S2. Kaplan-Meier plot of overall survival outcome in low-grade glioma. The curves denote the contribution of the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*), in females, to the overall survival outcome in patients from the low (blue – downregulated) and high (red – upregulated) expression groups. The censoring number refers to patients who did not suffer the outcome of interest during the specified study period. The overexpression of the signature has a more significant prognostic factor value.

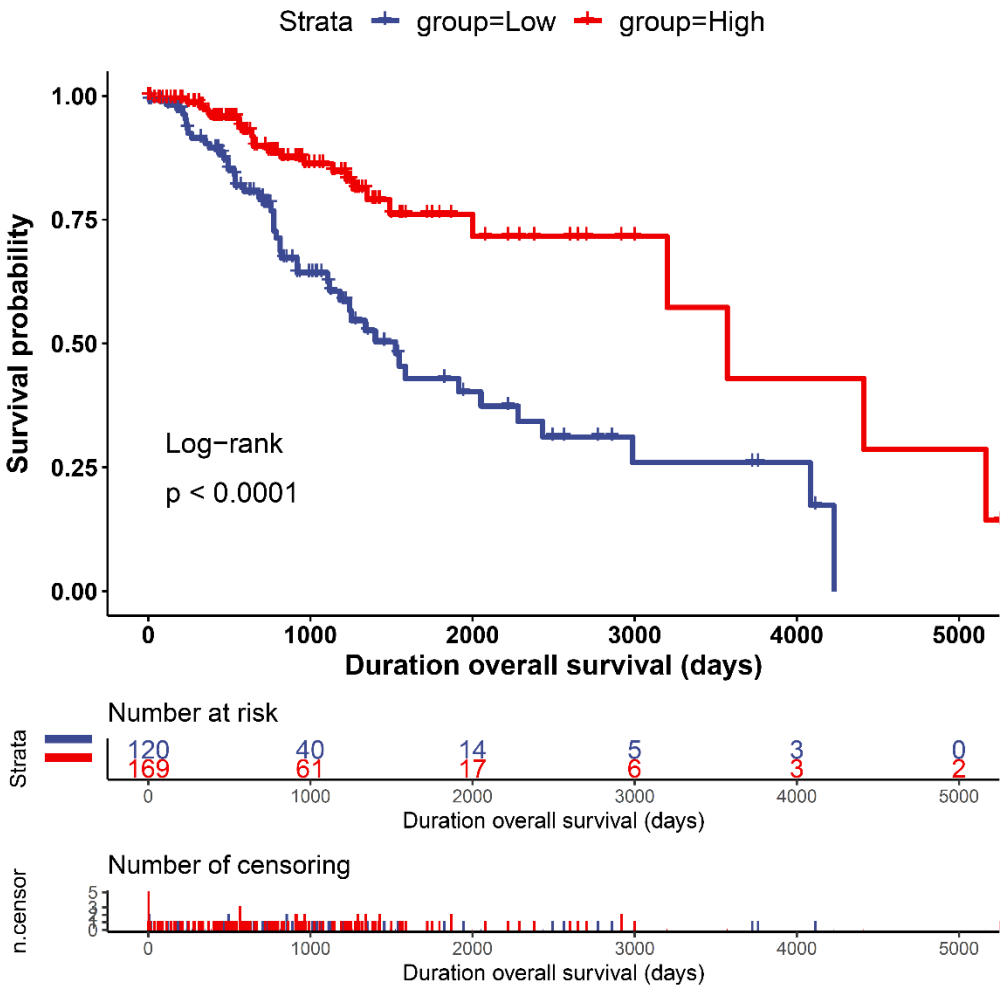


Figure S3. Kaplan-Meier plot of overall survival outcome in low-grade glioma. The curves denote the contribution of the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*), in males, to the overall survival outcome in patients from the low (blue – downregulated) and high (red – upregulated) expression groups. The censoring number refers to patients who did not suffer the outcome of interest during the specified study period. The overexpression of the signature has a more significant prognostic factor value.

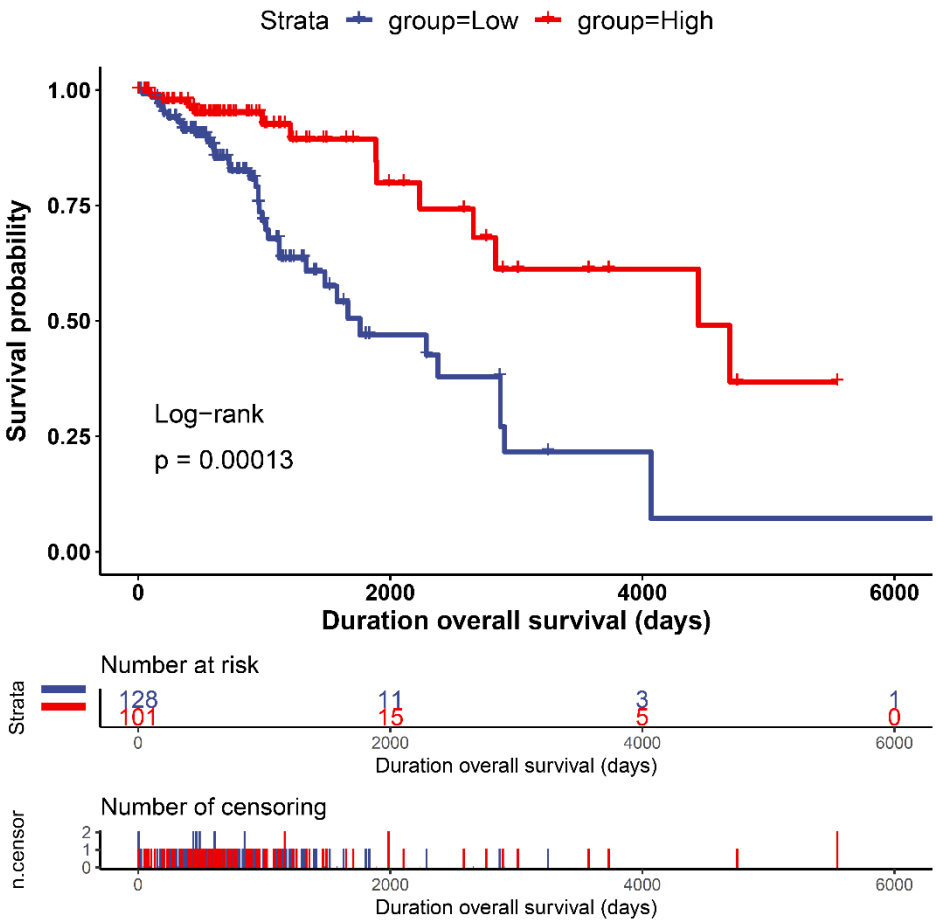


Figure S4. Kaplan-Meier plot of disease-specific survival outcome in low-grade glioma. The curves denote the contribution of the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*), in females, to the disease-specific survival outcome in patients from the low (blue – downregulated) and high (red – upregulated) expression groups. The censoring number refers to patients who did not suffer the outcome of interest during the specified study period. The overexpression of the signature has a more significant prognostic factor value.

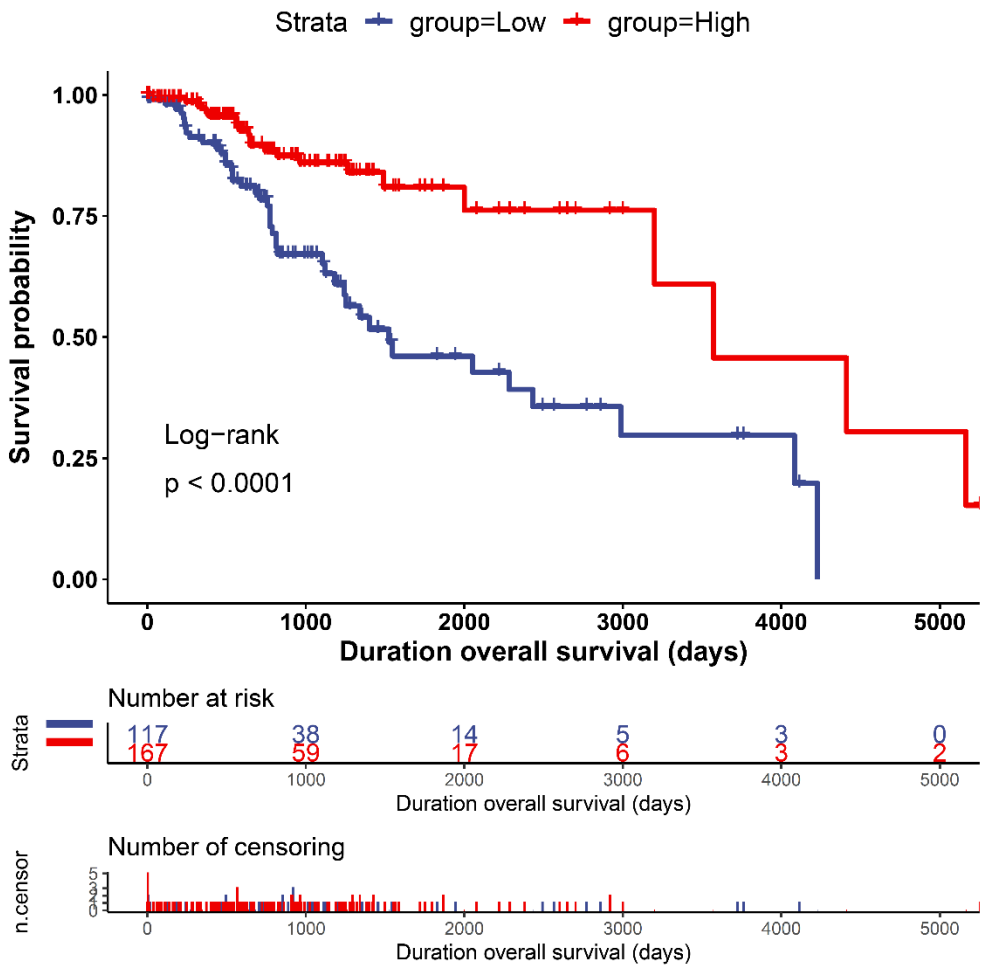


Figure S5. Kaplan-Meier plot of disease-specific survival outcome in low-grade glioma. The curves denote the contribution of the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*), in males, to the disease-specific survival outcome in patients from the low (blue – downregulated) and high (red – upregulated) expression groups. The censoring number refers to patients who did not suffer the outcome of interest during the specified study period. The overexpression of the signature has a more significant prognostic factor value.

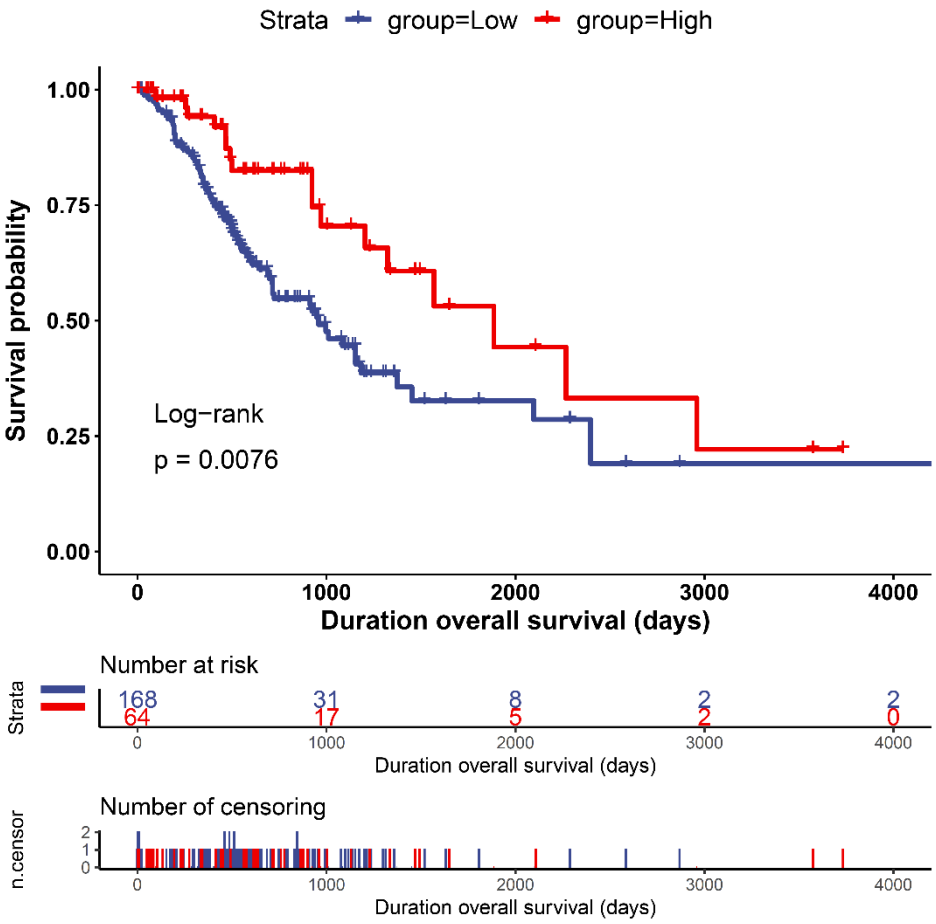


Figure S6. Kaplan-Meier plot of progression-free-interval outcome in low-grade glioma. The curves denote the contribution of the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*), in females, to the progression-free-interval outcome in patients from the low (blue – downregulated) and high (red – upregulated) expression groups. The censoring number refers to patients who did not suffer the outcome of interest during the specified study period. The overexpression of the signature has a more significant prognostic factor value.

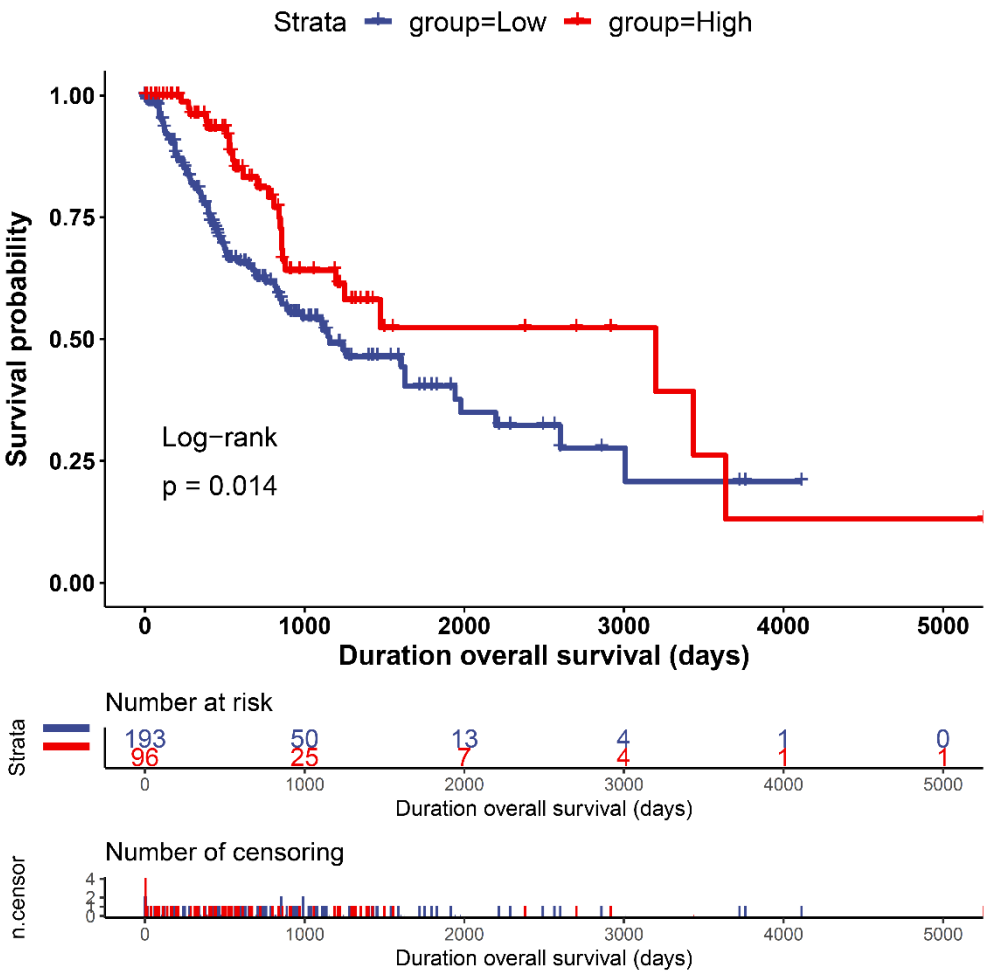


Figure S7. Kaplan-Meier plot of progression-free-interval outcome in low-grade glioma. The curves denote the contribution of the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*), in males, to the progression-free-interval outcome in patients from the low (blue – downregulated) and high (red – upregulated) expression groups. The censoring number refers to patients who did not suffer the outcome of interest during the specified study period. The overexpression of the signature has a more significant prognostic factor value.

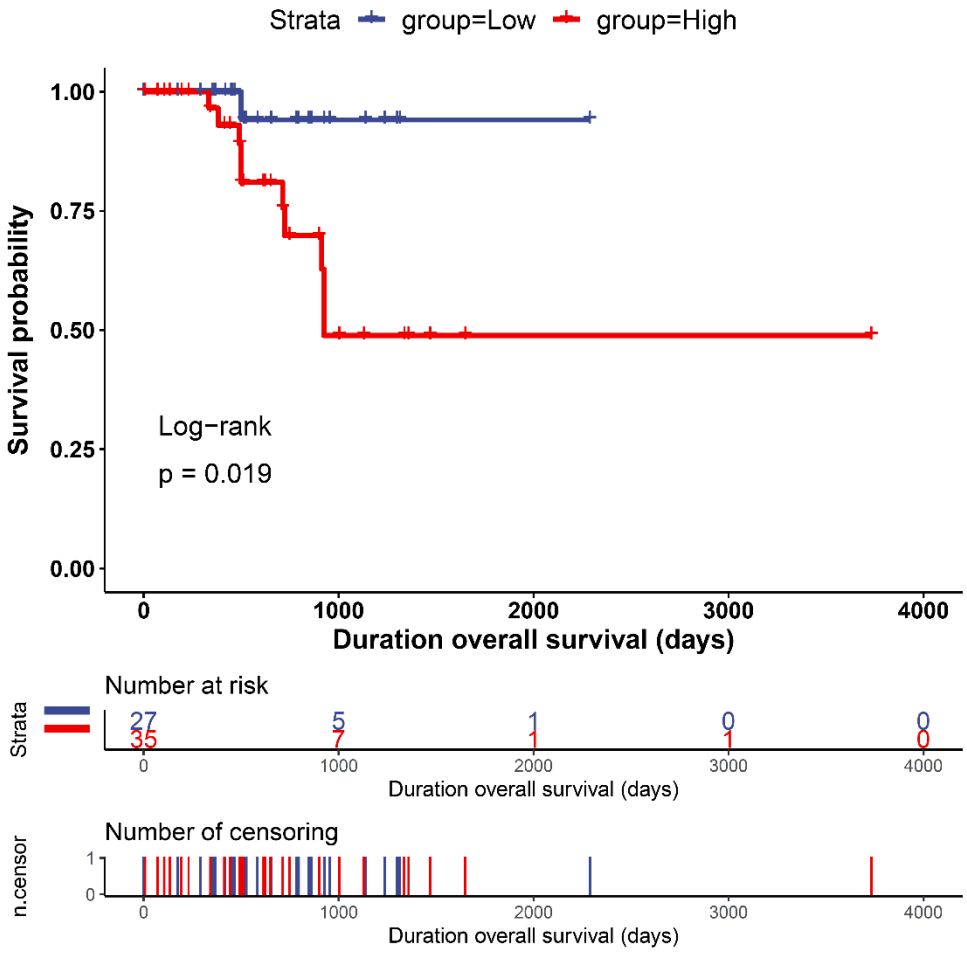


Figure S8. Kaplan-Meier plot of disease-free interval outcome in low-grade glioma in females. The curves denote the contribution of the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*) to the disease-free interval outcome in female patients from the low (blue – downregulated) and high (red – upregulated) expression groups. The censoring number refers to patients who did not suffer the outcome of interest during the specified study period. The downregulation of the signature has a more significant prognostic factor value.

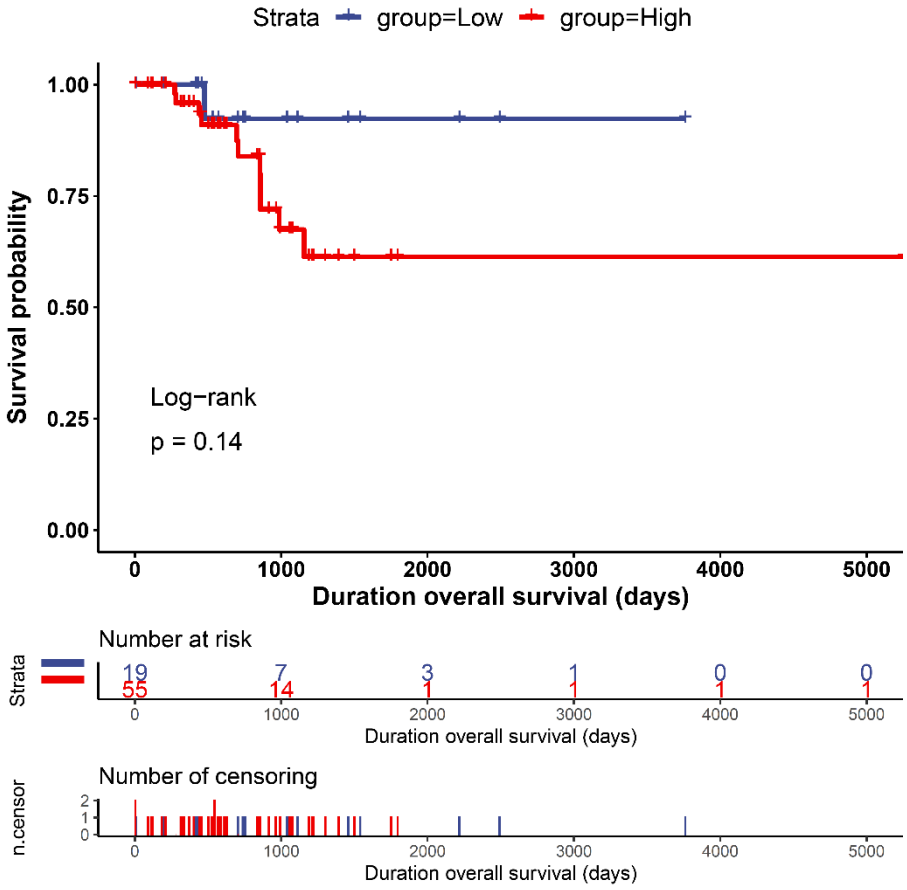


Figure S9. Kaplan-Meier plot of disease-free interval outcome in low-grade glioma in males. The curves denote the contribution of the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*), in males, to the disease-free interval outcome in patients from the low (blue – downregulated) and high (red – upregulated) expression groups. The censoring number refers to patients who did not suffer the outcome of interest during the specified study period. The overexpression of the signature has a more significant prognostic factor value.

Protective or Risk Effects of the 4-component Gene Signature in Different Cancers

We observed a protective effect of the 4-component gene signature in LGG (hazard ratio -0.17), kidney chromophobe (KICH – hazard ratio -0.32), kidney renal papillary cell carcinoma (KIRP - hazard ratio -0.10), adrenocortical carcinoma (ACC - hazard ratio -0.19), pancreatic adenocarcinoma (hazard ratio -0.08) and hepatocellular liver carcinoma (LIHC - hazard ratio -0.05) (Figure 12) (Table S12). In contracts, the signature has a risk effect on thymoma (THYM – hazard ratio 0.25) and uterine corpus endometrial carcinoma (UCEC – hazard ratio 0.06) (Figure 13; Table S22).

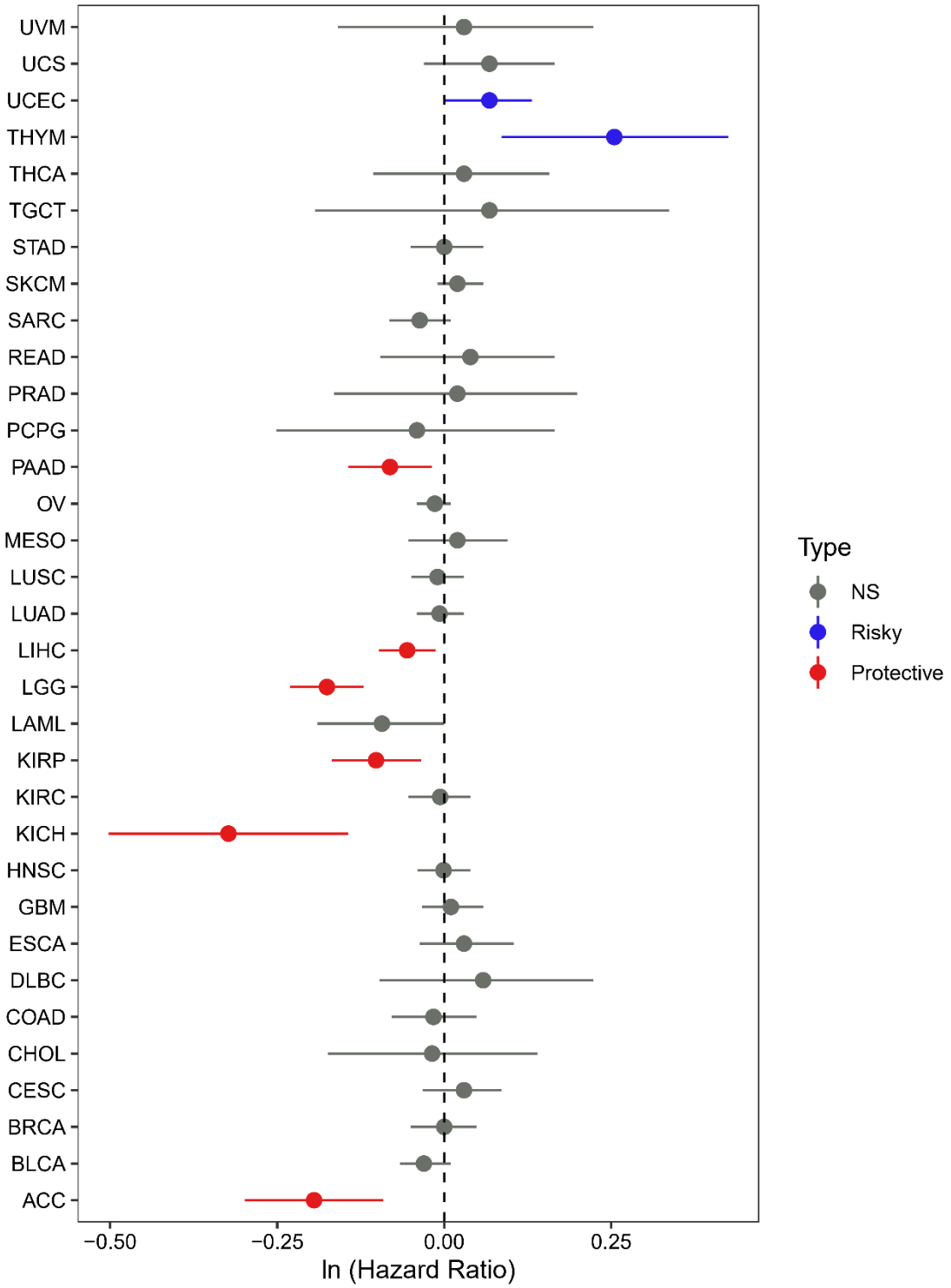


Figure 13. Boxplot of the protective or risk effects of the 4-component gene signature based on hazard ratio. Cox model analysis for the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*) in 33 cancer types. The signature is protective (red) in LGG, ACC, KICH, KIRP, LIHC, and PAAD cancers (p-value < 0.0113), but it has a risk effect (blue) in THYM and UCEC cancers (p-value < 0.0465).

Correlation Analysis Between Molecular Profiles and Immune Infiltrates

We estimated the degrees of the correlation between the mRNA expression levels of the 13 mitochondrial protein-coding genes and the molecular expression profiles indicative of immune cell infiltrates in 33 cancer types (Table S23). For this analysis, we defined a gene signature as the molecular expression profile of a minimum of three genes exhibiting the same direction of correlation. The most extended, 13-component gene signature occurred in lymphoid neoplasm diffuse large B-cell Lymphoma (DBLC), being positively correlated (rho ≥ 0.50 p-value < 0.0003) with neutrophil, NK, T-cell CD4 memory, M1, /M2 macrophages, T-cell CD8+/CD4+, and B-cell memory cells (Figure 14).

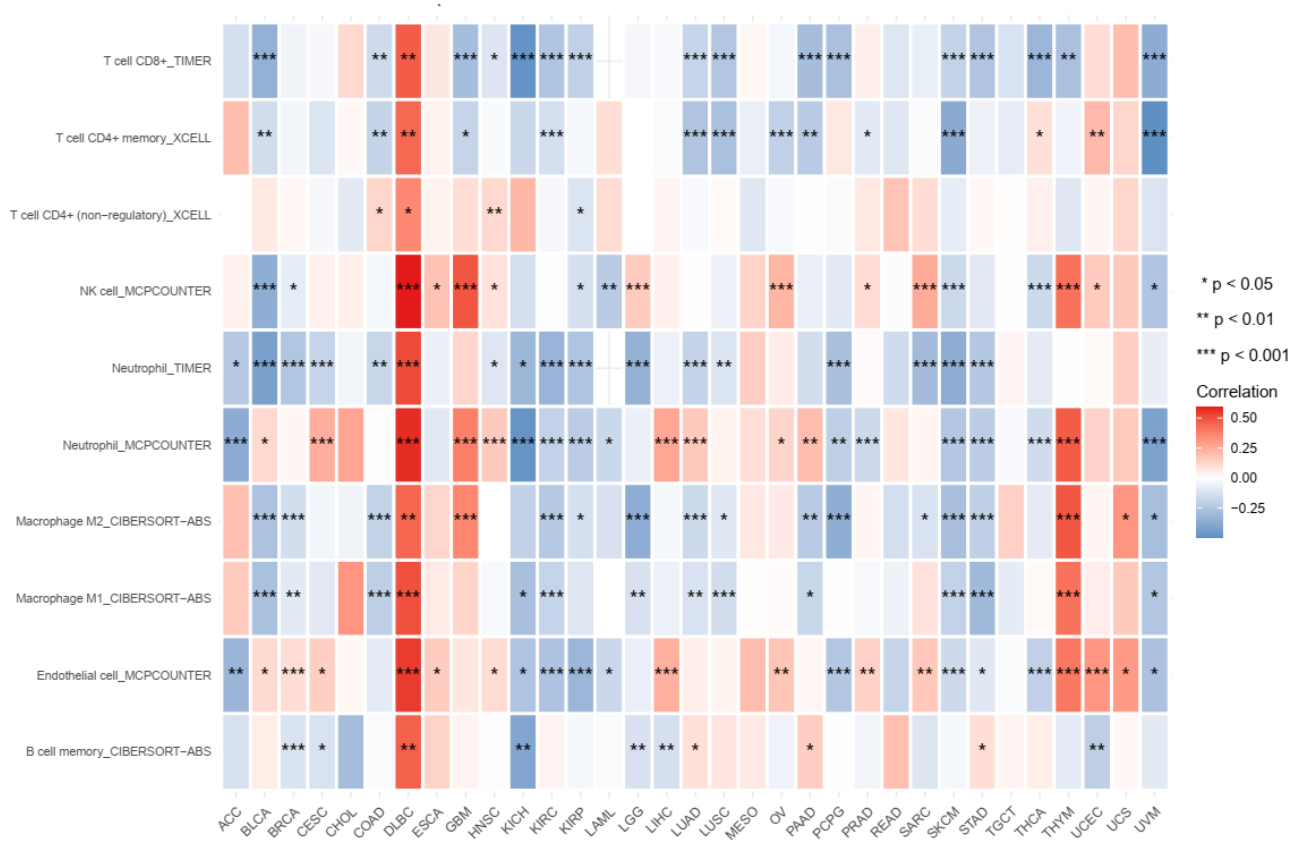


Figure 14. Heatmap of the correlations between the expression profiles of the 13-component gene signature and immune cell infiltrates. The red-blue heatmap scale shows the significance levels of Spearman correlation coefficients (***, **, *). In DBLC, he 13-component gene signature is positively correlated with infiltrates of neutrophil, NK, T-cell CD4 memory, M1/M2 macrophages, T-cell CD8+/CD4+, and B-cell memory cells (rho ≥ 0.50 p-value < 0.0003).

The second-longest signature is a 7-component gene signature (*MT-ND4L*, *MT-CYB*, *MT-ND1*, *MT-ND4*, *MT-ND5*, *MT-ND6*, *MT-ATP8*), positively correlated (rho ≥0.50, p-value ≤ 0.001) with the following cell immune infiltrates in thymoma (THYM): endothelial cell, neutrophil, myeloid dendritic cell, T cell CD4+, and macrophage M2 (Figure S10) (Table S24).

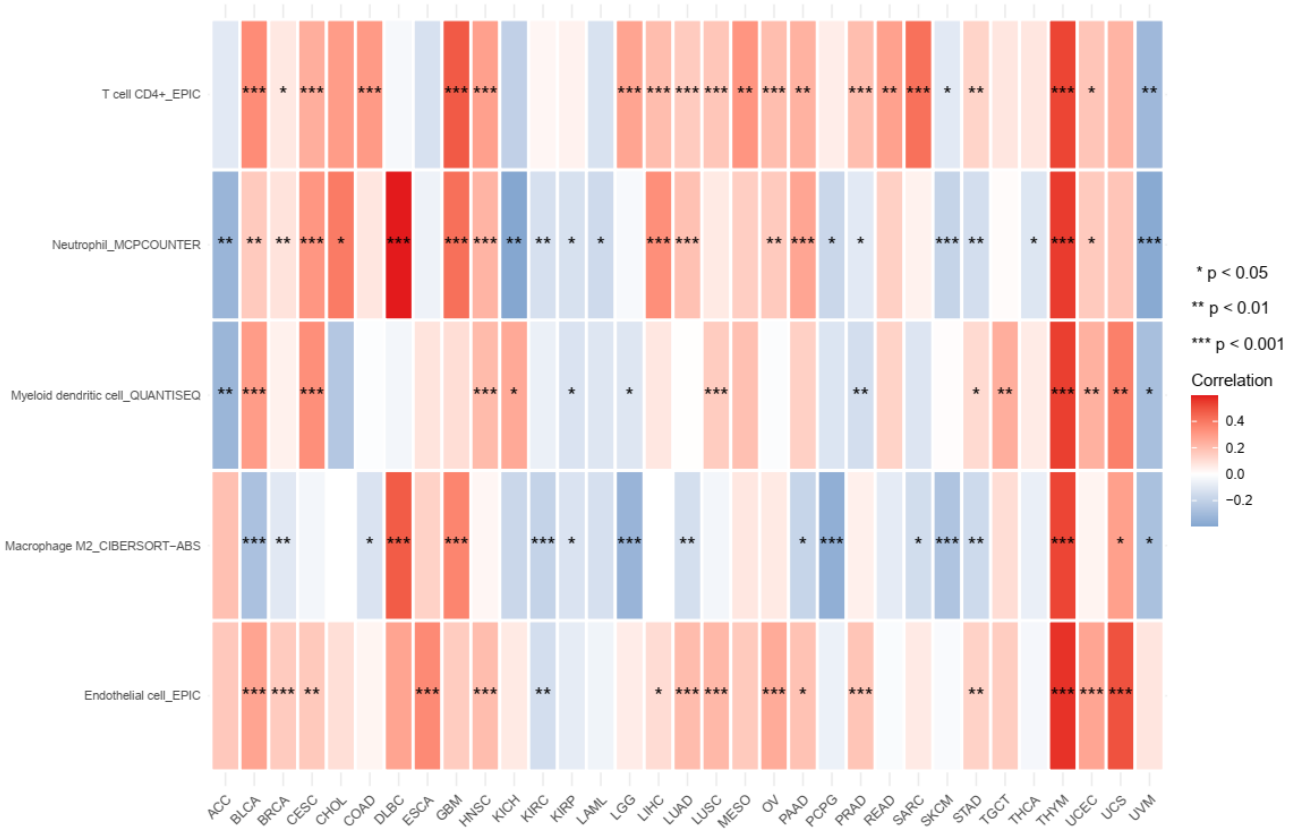


Figure S10. Heatmap of the correlations between the expression profiles of the 7-component gene signature and immune cell infiltrates. The red-blue heatmap scale shows the significance levels of Spearman correlation coefficients (***, **, *). In THYM, the 7-component gene signature is positively correlated with infiltrates of endothelial cells, neutrophils, myeloid dendritic cells, T cell CD4+, and macrophage M2 (rho ≥0.50, p-value ≤ 0.001).

In THYM, an 8-component gene signature (*MT-ATP8*, *MT-CO1*, *MT-CYB*, *MT-ND1*, *MT-ND2*, *MT-ND4*, *MT-ND4L*, *MT-ND5*) showed an inverse correlation with molecular profiles for the following cell infiltrates T cell CD4+ Th1, T cell, T cell regulatory (Tregs) (rho ≤ -0.5, p-value ≤ 3.05e-10), Figure S11 (Table S25).

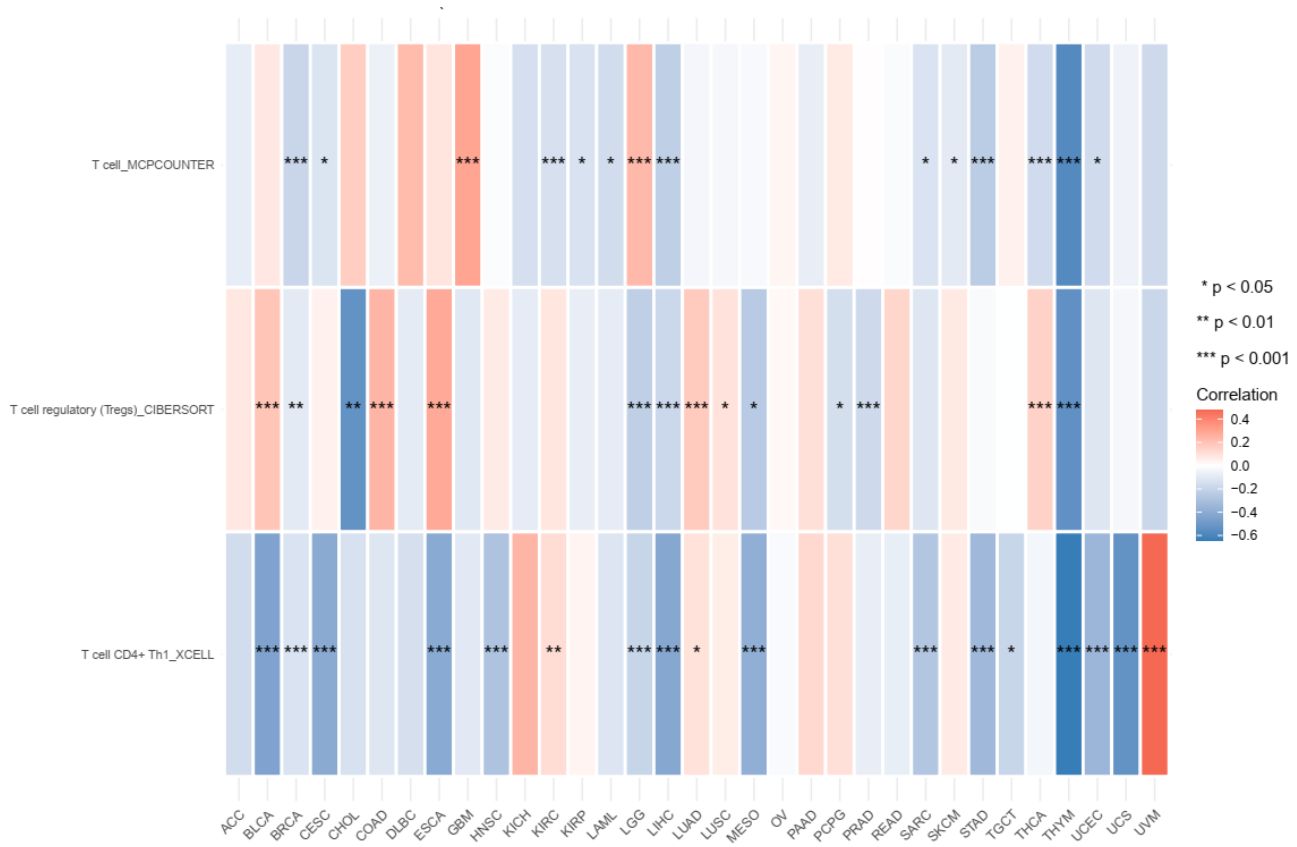


Figure S11. Heatmap of the correlations between the expression profiles of the 8-component gene signature and immune cell infiltrates. The red-blue heatmap scale shows the significance levels of Spearman correlation coefficients (***, **, *). In THYM, the 7-component gene signature is negatively correlated with infiltrates of T cell CD4+ Th1, T cell, and T cell regulatory (Tregs) (rho ≤ -0.5, p-value ≤ 3.05e-10).

We also estimated the degrees of the correlation between the mRNA expression levels for the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*) associated earlier with survival risk in LGG and the gene profiles indicative of immune cell infiltrates in 33 cancer types. The 4-component gene signature was positively correlated (rho ranging 0.06-0.6 p-value < 0.05) with 111 profiles of immune cell infiltrates and negatively correlated (rho ranging -0.56 to -0.059 p-value < 0.05) with 113 profiles of immune cell infiltrates in 33 cancer types. In LGG, the most significant positive correlation (rho > 0.3, p-value < 2.01e-12) was with monocyte, NK, and B cells (Figure 15), whereas the negative correlation (rho ≤ 0.3, p-value < 2.26E-12) were with T cell CD4+ Th2, macrophage M1 and M2, myeloid dendritic cell, neutrophil, and activated myeloid dendritic cell activated cell infiltrates (Figure 16) (Table S26).

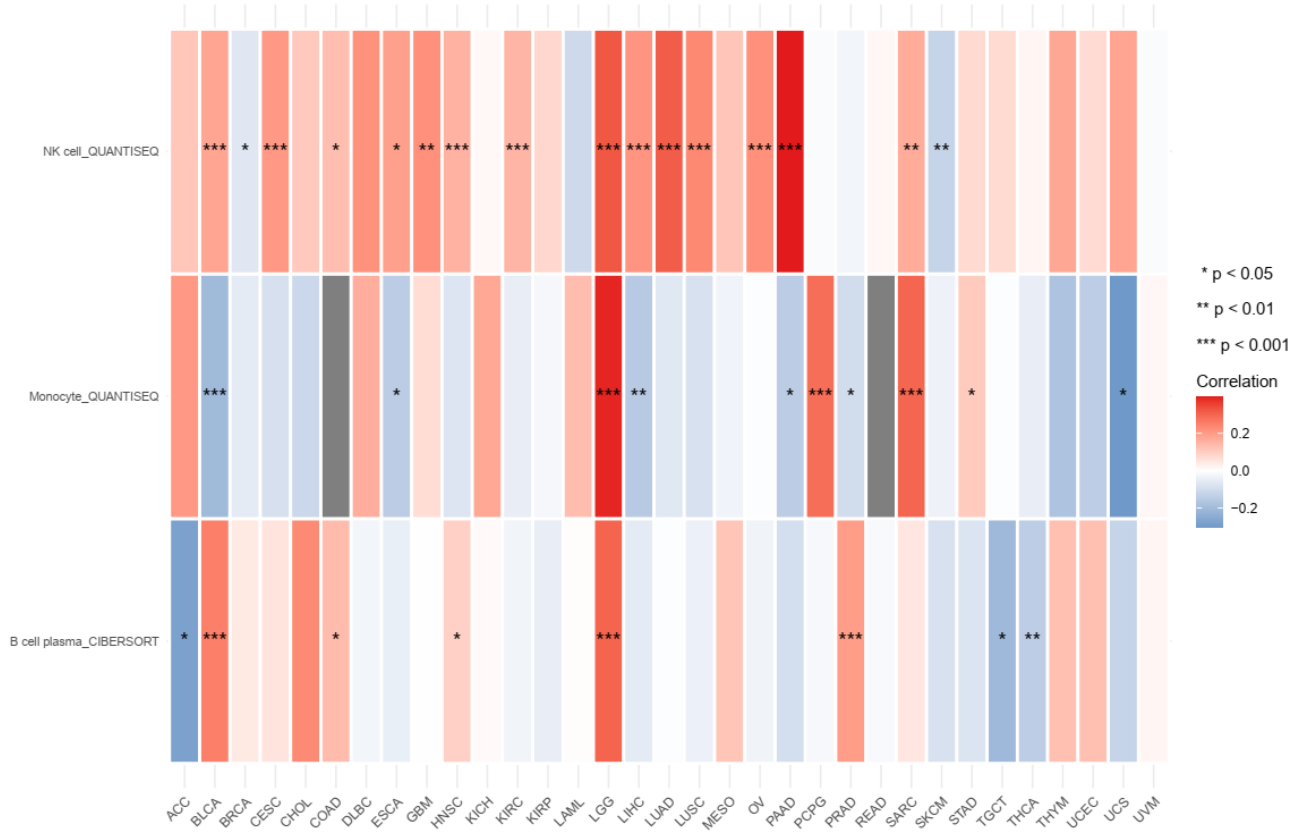


Figure 15. Heatmap of the correlations between the expression profiles of the 4-component gene signature and immune cell infiltrates. The red-blue heatmap scale shows the significance levels of Spearman correlation coefficients (***, **, *). In LGG, he 4-component gene signature is positively correlated with infiltrates of monocyte, NK, and B cells (rho > 0.3, p-value < 2.01e-12).

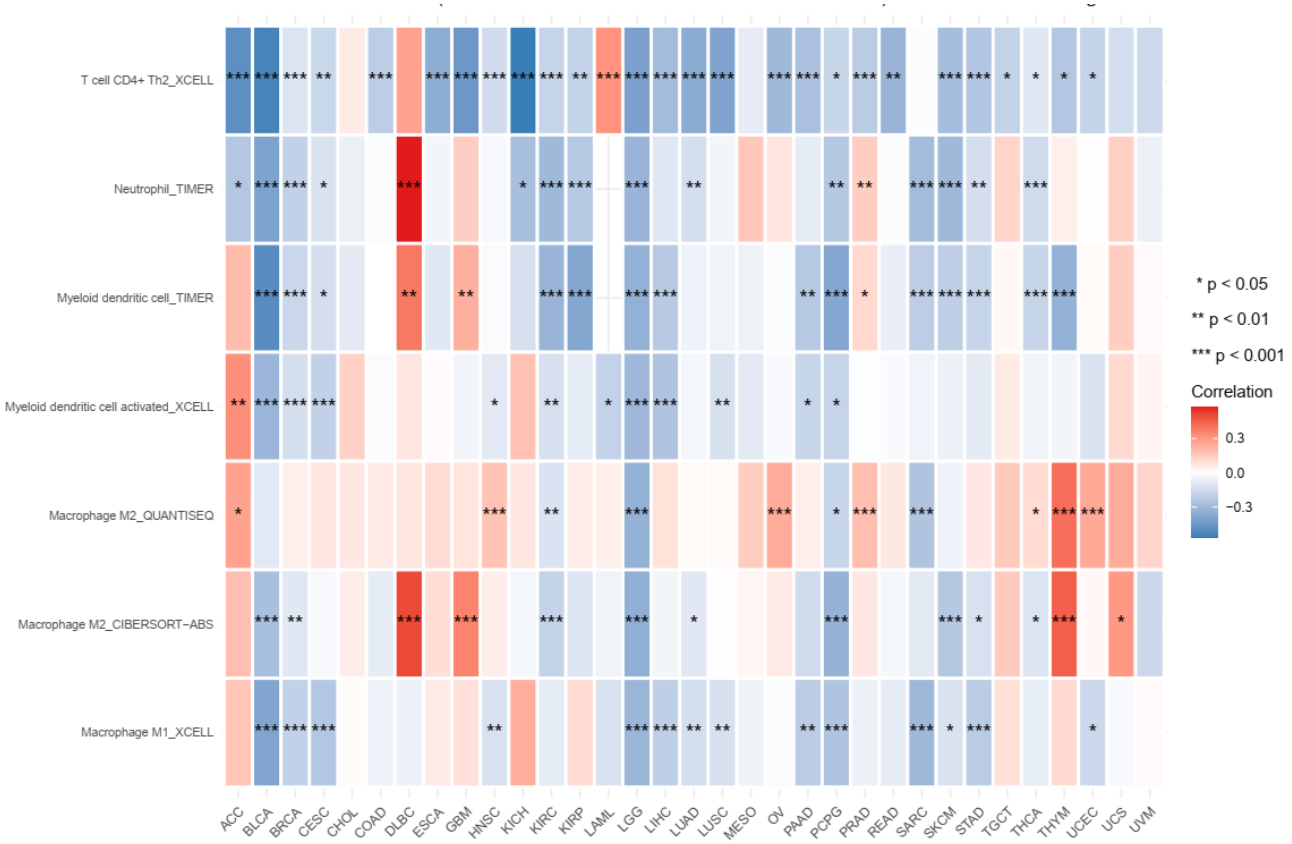


Figure 16. Heatmap of the correlations between the expression profiles of the 4-component gene signature and immune cell infiltrates. The red-blue heatmap scale shows the significance levels of Spearman correlation coefficients (***, **, *). In LGG, the 4-component gene signature negatively correlates with T cell CD4+ Th2, macrophage M1 and M2, myeloid dendritic cell, neutrophil, and activated myeloid dendritic cells cell activated cell infiltrates (rho ≤ 0.3, p-value < 2.26E-12).

Correlation Analysis between Molecular Expression Profiles and Stemness

The molecular expression profiles of the 13 mitochondrial protein-coding genes exhibited positive (rho ranging 0.08-0.55, $p < 0.045$ in 27 cancer types) and negative correlations (rho ranging -0.51 to -0.06, $p < 0.046$ in 24 cancer types) with stemness (Table S27). The most extended signatures with 13 and 12 components positively correlated with stemness in KICH (rho = 0.50 p-value = 0) and LAML (rho =0.28 p-value = 0), respectively (Figure 17) (Table S18). The 13-component signature negatively correlated with stemness in 19 cancer types, with the most significant inverse coefficient in uterine corpus endometrial carcinoma, UCEC (rho = -0.41, p =0).

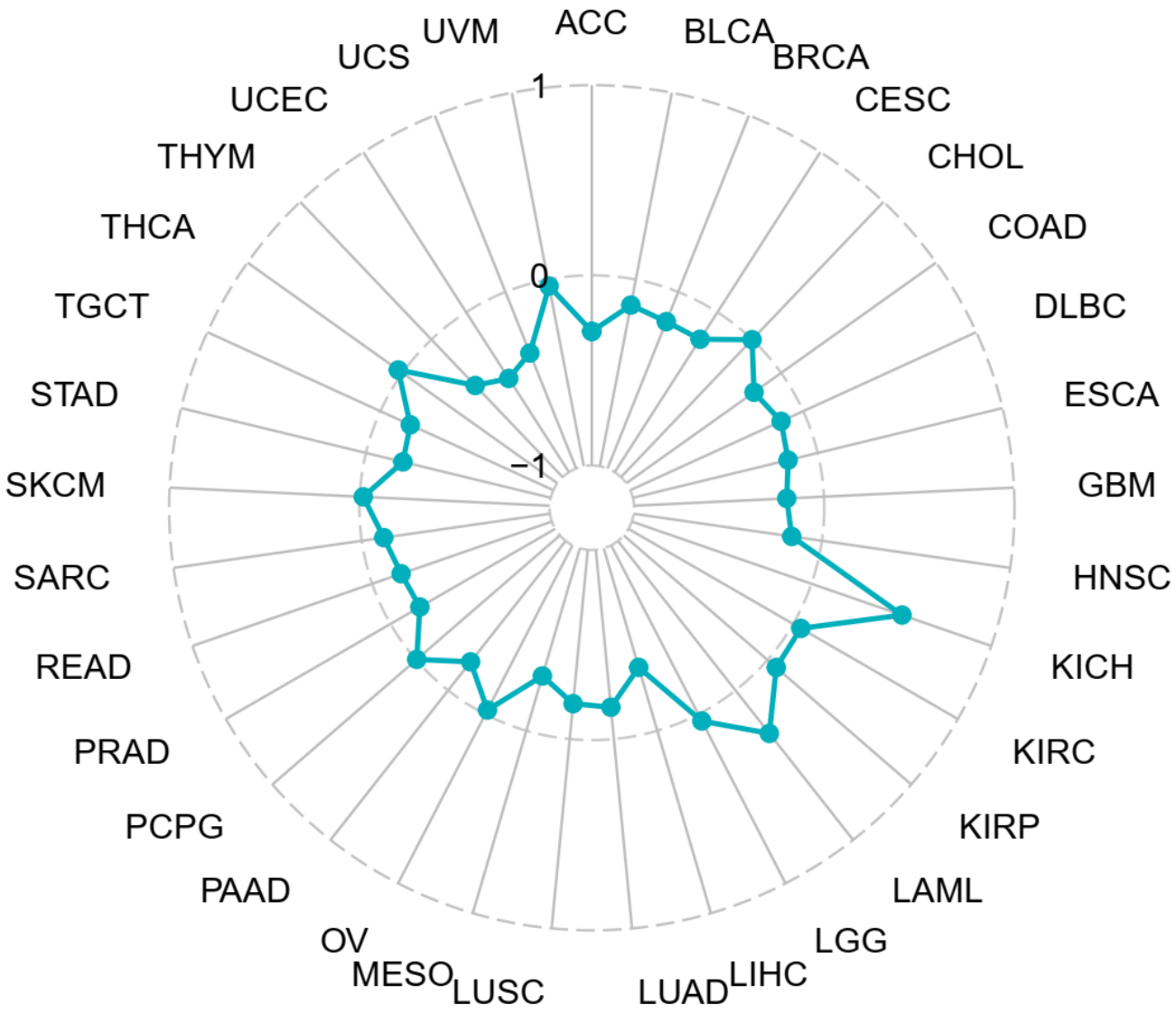


Figure 17. Radar plot for the correlation coefficients between the 13-component gene signatures and stemness in 33 cancer types. Shown is the distribution of the breadth of the Spearman correlation coefficients, where the association can be positive (+1), negative (-1), or neutral (0). The more significant positive correlation with stemness occurred in KICH and LAML.

DISCUSSION

Here, we showed differential RNA expression of the 13 mitochondrial protein-coding genes in 33 cancer types and normal tissues, where eleven genes (*MT-ND2*, *MT-ND1*, *MT-ATP8*, *MT-ATP6*, *MT-CO2*, *MT-CYB*, *MT-CO3*, *MT-ND4L*, *MT-ND4*, *MT-ND3*, *MT-CO1*) exhibited an under-expression profile in brain cancer. Moreover, the 13 mitochondrial protein-coding genes also demonstrated a sex effect in expression profiles in different tissues. A male-biased expression across ages occurred in *MT-ND5* for Artery – Aorta, Brain – Substantia Nigra, and Esophagus – Mucosa (Wilcox test, $p \leq 0.001$), and in the *MT-ND6* for Brain – Hypothalamus and Kidney – Cortex (Wilcox test, $p \leq 0.001$). A female-

biased across ages was also noticed in *MT-ATP6*, *MT-ATP8*, and *MT-CO1* for Artery – Coronary; *MT-CO1*, *MT-CO3*, and *MT-CYB* for Adipose – Visceral (Omentum); *MT-ND2* in Brain – Frontal Cortex (BA9) and Brain – Substantia Nigra; *MT-ND1* for Brain – Substantia Nigra. Interestingly, the *MT-ND5* also exhibits a biased expression across ages in Artery – Aorta, Brain – Frontal Cortex (BA9), and Colon – Transverse, but for females across different age groups. The same was noticed for *MT-ND6* in Brain – Frontal Cortex (BA9) and Brain – Substantia Nigra. The mitochondrial protein-coding genes exhibited a possible aging effect in expression profiles for the different age groups (20-49-year-old, 50-59-year-old, and 60-79-year-old) in different tissues, where the expression was higher in the 60-79-year-old group. In this group, we observed three genes (*MT-CO1*, *MT-ND5*, and *MT-ND6*, *MT-ND3*; Wilcox test, $p \leq 0.01$) with the most significant expression profile in different tissues (Artery – Coronary, Brain – Amygdala, Colon – Sigmoid, Esophagus – Muscularis, Minor Salivary Gland, and Small Intestine – Terminal Ileum). Furthermore, a possible age-group-dependent sex-biased effect was observed across the 52 tissues analyzed for the 13 mitochondrial protein-coding genes and polymorphic site expression. A female age-group-dependent sex-biased effect was noticed for ten different tissues between the three age groups (20-49-year-old, 50-59-year-old, and 60-79-year-old; Tukey test, $p \leq 0.05$), and a male age-group-dependent sex effects was observed for seven different tissues also between the three age groups (Tukey test, $p \leq 0.05$). Thus, the polymorphic site expression of 13 mitochondrial protein-coding genes exhibited a sex-age-biased between the three age groups. The difference was higher in the last group (60-79-year-old).

The RNA profiles of the 13 mitochondrial protein-coding genes also exhibited a 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, and *MT-ND6*) with under-expression in LGG (hazard ratio range 0.41-0.51, Logrank $p \leq 0.00026$, $p(HR) \leq 0.00033$). This signature was associated with survival outcomes in LGG for different categories, where the overexpression was associated with OS, DSS, and PFI (Log-rank $p \leq 0.0076$). The 4-component gene signature was not affected by sex, OS, DSS, or PFI. Only for DFI outcome, we observed a sex-specific association in female patients, with a greater prognostic factor value (Log-rank p= 0.019) compared to males (Log-rank p= 0,14). The 13 mitochondrial protein-coding genes demonstrated a protective effect for the 4-component gene signature in LGG (hazard ratio -0.17) and a risk effect on THYM – hazard ratio 0.25 and UCEC – hazard ratio 0.06. We noticed different component gene signatures for the 13 mitochondrial protein-coding genes in the immune cell infiltrates. The most extended, 13-component gene signature occurred in DBLC, positively correlated with neutrophil, NK, T-cell CD4 memory, M1 and M2 macrophages, T-cell CD8+/CD4+, and B-cell memory cells. Moreover, the 4-component gene signature's most significant positive correlation in LGG (rho > 0.3, p-value < 2.01e-12) was with monocyte, NK, and B cells, whereas the negative correlation (rho ≤ 0.3, p-value < 2.26E-12) were with T cell CD4+ Th2, macrophage M1 and M2, myeloid dendritic cell, neutrophil, and activated myeloid dendritic cell activated cell infiltrates. The molecular expression profiles of the 13 mitochondrial protein-coding genes exhibited positive and negative correlations in 24 cancer types with stemness. The most extended signatures with 13 and 12 components were positively correlated with stemness in KICH (rho = 0.50 p-value = 0) and LAML (rho =0.28 p-value = 0), respectively. The 13-component signature negatively correlated with stemness in 19 cancer types, with the most significant inverse coefficient in UCEC (rho = -0.41, p =0). The differentially expressed profiles of the 13 mitochondrial protein-coding genes in cancer have prognostic factor values in clinical studies. The correlation between the differentially expressed gene signatures and tumor stemness in LGG and immune cell infiltrates suggests that the reported signatures are proxies of oncogenic dedifferentiation.

Genome-wide screening of normal versus cancer transcription provides knowledge about the molecular information of cancer progression. It enables identifying transcriptional changes in differentially

expressed genes (DEGs) to determine their prognostic role in clinical outcomes. The gene-by-gene analysis studies in most of the 13 mitochondrial protein-coding genes demonstrated that their RNA expression is significantly suppressed in several cancer types, such as breast, colon, kidney, and liver, the exception being KICH, with all genes overexpressed (Reznik et al., 2017;Kim et al., 2020). However, these studies considered the full range of logFC values and significant FDR values. In our analysis, we used a more restrictive method accepting only magnitudes of differential expression logFC values (≥ 1.5 or ≤ -1.5) with genome-wide FDR values $\leq$ 5e-5. Thus, our restrictive methods allowed the identification of eleven genes downregulated in brain cancer (*MT-ATP6*, *MT-ATP8*, *MT-CO1*, *MT-CO2*, *MT-CO3*, *MT-CYB*, *MT-ND1*, *MT-ND2*, *MT-ND3*, *MT-ND4*, *MT-ND4L*), six in liver cancer (*MT-ATP6*, *MT-ATP8*, *MT-CYB*, *MT-ND2*, *MT-ND4*, *MT-ND4L)* and four in breast cancer (*MT-ATP8*, *MT-CYB*, *MT-ND1*, *MT-ND2*). In contrast, overexpression happens in AML (*MT-ND4*, *MT-ND5*, *MT-CO1*, *MT-ND4L*) and ovary cancer (*MT-ND6*, *MT-ND1*, *MT-CO1*).

The sex differences in gene expression have already been observed in different tissues, but nowadays, the mechanism behind the tissue-specific sex differences in gene expression is not understood (Rinn and Snyder, 2005;Ellegren and Parsch, 2007;Parsch and Ellegren, 2013). Thus, the sexual dimorphism in gene expression is still hard to describe, partly because the sex-related differences in the autosomal gene expression occur in various tissues (Mele et al., 2015). Additionally, gene expression during aging has already been studied between the sexes for several tissues (Tower, 2017). However, there are few references in the literature regarding the effects of aging on the expression of different tissues. A few reports demonstrated a low level of gene expression in older tissues, such as muscle, epicardial adipose tissue, and brain (Welle et al., 2003;Berchtold et al., 2008;Iacobellis, 2021). Thus, we wished to investigate whether there are sex-specific differences in the expression of the 13 mitochondrial protein-coding genes and if there is a difference in the expression of these genes between the tissues at different ages. Here, in males, two genes exhibited a male-biased expression in different tissues (*MT-ND5* for Artery – Aorta, Brain – Substantia Nigra, and Esophagus – Mucosa, and in the *MT-ND6* for Brain – Hypothalamus and Kidney – Cortex). Other nine genes showed a female-biased expression in different tissues *(MT-ATP6*, *MT-ATP8*, and *MT-CO1* for Artery – Coronary, *MT-CO1*, *MT-CO3* and *MT-CYB* for Adipose – Visceral (Omentum); *MT-ND2* for Brain – Frontal Cortex (BA9) and Brain – Substantia Nigra; *MT-ND1* for Brain – Substantia Nigra). *MT-ND5* also exhibits a biased expression in Artery – Aorta for females and in other tissues (Brain – Frontal Cortex (BA9) and Colon – Transverse). In agreement, Tower has reported that gene expression during aging has already been studied between the sexes for several tissues, such liver, which exhibits high sex-specific tissue expression with changes during aging (Tower, 2017).

Considering the evidence of a differential expression with age in coding genes (Frenk and Houseley, 2018), we decided to investigate possible differences in the expression of our mitochondrial protein-coding genes between 52 different tissues in the three age groups. Thus, we observe thirty-four tissues (n=34) have significantly different expression profiles (Wilcox test, $p \leq 0.05$) in 20-49-year-old and 60-79-year-old groups for the 13 mitochondrial protein-coding genes. However, when we look more specifically at the tissue-specific expression for the different ages, we observe significant differences in different tissues' expression. The artery – Coronary tissue exhibited more significant expression in the 20-49-year-old group (Wilcox test, $p \leq 0.001$) than other tissues in this group. For the 50-59-year-old group (Wilcox test, $p \leq 0.01$), the most significant differences in tissue expression occurred in the Brain – Cerebellar Hemisphere, Adipose – Visceral Omentum, and Bladder. This difference in expression of tissues also occurred in the group 60-79-year-old for Colon -Sigmoid, Artery – Coronary, Adipose – Visceral Omentum, and Minor Salivary Gland (Wilcox test, $p \leq 0.05$). In accordance, Frenk and colleagues discuss that the aging effect in gene expression may occur for reasons not necessarily

equivalent to programmed aging. Thus, gene expression changes may be reactive to physiological or cellular changes (Frenk and Houseley, 2018). Other studies also demonstrated that the most association of aging transcriptome is a consistent change of mitochondrial protein mRNAs, which was also observed in humans (de Magalhaes et al., 2009;van den Akker et al., 2014;Peters et al., 2015). No other evidence was found about the age-biased or sex-biased effect in humans' 13 mitochondrial protein-coding genes.

Although most aging-related changes are correlated with declining mitochondrial function and ROS overproduction, oxidative damage, and accumulation of mtDNA mutations in somatic tissues, it's necessary to emphasize the relationship between aging and defects of a wide range of proteins encoded by mtDNA (Poyton and McEwen, 1996;Scarpulla, 2008). In this context, we investigated the sex-dependent polymorphic site differences for mitochondrial protein-coding genes in different tissues. In this way, we observe a possible age-group-dependent sex effects expression in ten different tissues for females and seven for males between the three tested age groups. In agreement, another study reports evidence that aging occurs differently among males and females as expected from the phenotypic differences associated with aging (Mele et al., 2015;Kim et al., 2016;Gershoni and Pietrokovski, 2017;Naqvi et al., 2019). Other studies still report that breast tissue has the most sexual differentiation (Mele et al., 2015;Gershoni and Pietrokovski, 2017;Lopes-Ramos et al., 2020), but interestingly, in our study, we notice a sex-biased gene expression for breast tissue related to age groups. Thus, human aging phenotypes may be not only sex-specific but, in a more complex way, age-group-dependent.

In the last ten years, the sequencing and bioinformatic approaches expanded the possibility of accessing and handling gene signature data. This improves the analysis of survival outcomes for patients with several cancers (Rao et al., 2011;Gao et al., 2019;Li et al., 2019;Zhang et al., 2020;Pawar et al., 2022). Thus, investigating gene signature expression from mitochondrial protein-coding genes related to cancer survival outcomes was possible, which led us to perform additional analysis of these approaches in this study. In this context, Schopf and colleagues performed an analysis of gene-expression signatures of tumor sample transcriptomes that allowed the identification of several tumors associated with a shorter survival time (Schopf et al., 2020). Here we investigate whether 13 mitochondrial protein-coding genes exhibit a gene signature for survival outcomes in 33 different cancer types across donors of the TCGA cohort and whether the sex-specific differences are risk factors in cancer. We identified a novel 4-component mitochondrial protein-coding gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*) whose molecular expression profile has a prognostic factor value for survival outcomes in LGG. Xiao and colleagues demonstrate that other gene signatures, such as CD44-related genes, with high-risk expression in LGG, are associated with greater survival rates in OS and disease-free survival (Log-rank $p< 0.0001$) (Xiao et al., 2020b). Here, we observe the 4-component gene signature exhibiting an over-expression profile in OS and DSS for LGG with elevated survival rates in the high-risk group (Log-rank $p < 0.0001$). In contrast, Xiao and colleagues showed that patients in the high-risk group had significantly poorer survival results (Log-rank $p= 0.000012$) in not mitochondrial gene signatures for LGG (Xiao et al., 2020a). Other studies have already demonstrated a better prognosis in survival results for the low-risk group using gene signatures (Log-rank $p <0.0001$) (Chen et al., 2021;Guo et al., 2021). Additionally, the low-risk group related to the 4-component gene signature exhibits a poor prognosis compared to the high-risk group in OS, DSS, and PFI survival outcomes.

In a study restricted to lung squamous cell carcinoma (LUSC) and lung adenocarcinoma (LUAD) (Li et al., 2018), the two major subtypes of lung cancer, the 13 mitochondrial respiratory genes were reported to be downregulated in tumor tissues compared with matched control tissues. The under-expression of *MT-ND5* or *MT-ND6* genes was associated with overall survival outcomes and tumor

progression in LUAD and LUSC. Our gene-by-gene analysis confirmed the above associations with overall survival outcomes in LUAD (*MT-ND6* Log-rank test statistics = 10.04, p-value = 0.001529; *MT-ND5* Log-rank = 4.186, p-value = 0.04076), when the samples were clustered into two expression groups (Figures S10-S11), as well in LUSC, both *MT-ND6* and *MT-ND5* (Figure S18 and S25) was not significant for the overall survival patients (*MT-ND6* Log-rank p-value = 0.4; *MT-ND5* Log-rank p-value = 0.19). However, we extended the association with three other survival outcomes: DSS, DFI, and PFI (Figure S10-S25).

Identifying gene signatures as risk effect factors may be relevant for proposing personalized treatment of several cancers, such as hepatocellular carcinoma, colon cancer, and gliomas (Li et al., 2020;Liu and Li, 2021;Zhao et al., 2021). We aimed to identify a correlation between protective or risk effects associated with survival outcome for the 13 mitochondrial protein-coding genes in 33 different cancer types by using the TCGA cohort. Concerning the protective effect associated with survival outcome in this study, we found and 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*) with the association in six different cancer types (ACC, KICH, KIRP, LIHC, LGG, PAAD; hazard ratio ranging -0.08 to -032). Regarding the risk effects, two tissues exhibited associations (THYM and UCEC; hazard ratio ranging from 0.06 to 0.25). To LGG, Zhang and colleagues showed a protective/risk effect using gene signatures (non- mitochondrial) correlation (Zhang et al., 2019). Accordingly, autophagy gene signature also exhibited a risk effect factor for LGG patients (Lin and Lin, 2021). Equally, Wang and colleagues demonstrate evidence of risk signatures genes (Wang et al., 2022). In the same way, other gene signatures, extracellular matrix related, exhibited a risk effect for gliomas in the TCGA cohort but also demonstrated an association with immune infiltrates (Liu and Li, 2021). In contrast, we observed a protective effect factor in LGG patients to four mitochondrial components gene signature, which is not shown in the literature now.

Studying gene signatures associated with immune infiltrates may provide a potential clinical implication for cancer treatment and prognosis (Iglesia et al., 2016;Yan et al., 2020). Thus, we use the 13 mitochondrial protein-coding genes in the context of gene signature to correlate with immune cell infiltrates in 33 cancer types from the TCGA cohort. Studies already show evidence about gene signatures, such as autophagy genes related, strongly associated with six immune cell infiltrates (macrophages M0 and M1, Neutrophils, T cells CD8, T cells follicular helper, T cells reg) in LGG (Quan et al., 2021). Additionally, other gene signatures, such as immune-related genes, demonstrate positive and negative correlations with several immune cell infiltrates in LGG (macrophages, dendritic cells, B cells, and CD4 T cells) (Pan et al., 2021). In the same way, our study showed positive and negative correlations with immune infiltrates in LGG, with monocyte, NK, and B cells, T cell CD4+, macrophage (M1 and M2), and myeloid dendritic cells.

In this way, our mitochondrial protein-coding gene signature positively correlates with M1 and M2 macrophages from immune cells infiltrated in LGG. Similarly, in DLBC the 13-component gene signature show a positive correlation for six cells including neutrophils. In agreement, M2 macrophage has been confirmed to promote immunosuppression and proliferation of LGG; on the other hand, T cells and natural killer (NK) cells can be associated with poorer OS and DSS in LGG, both in non-mitochondrial gene signature (Ye et al., 2021;Zhu et al., 2022). Concerning DLBC prognosis, the tumor-associated neutrophils show a poor prognostic, in the same way, Manfroi and colleagues demonstrated that in DLBC patients with a higher ratio of neutrophils had a worse prognosis, both in non-mitochondrial gene signature (Manfroi et al., 2018;Mu et al., 2018). In other studies, it is also possible to observe T-lymphocyte signatures impacting the prognosis of patients with DBLC,

demonstrating the importance of survival gene signatures in aggressive B-cell lymphomas (Ansell et al., 2001;Keane et al., 2013;Keane et al., 2015).

Using gene signatures to study tumor stem cells has provided new insights about stemness for measuring the degree of oncogenic dedifferentiation. A correlation between stemness signatures and unfavorable outcome for some cancers, including gliomas, was observed in the literature for the TCGA cohort (Malta et al., 2018). In agreement, there is evidence of gene signatures, such as the 5-mRNA signature, that correlate with stemness in melanoma and hepatocellular carcinoma, which exhibit a negative correlation for these cancers (Cai et al., 2021;Zhang et al., 2021). Furthermore, cancer stemness has been demonstrated to be associated with kidney cancers, such as KIRC, but not for KICH, and negatively correlates with immune infiltrates (Xiao et al., 2021). Additionally, other studies demonstrated a stemness-related gene signature expression in cancers such as hepatocellular carcinoma, pancreatic ductal adenocarcinoma, and endometrial cancer as a novel prognostic marker for survival (Hong et al., 2021;Huang et al., 2021;Xu et al., 2021). In this study, we report a significant positive correlation between the expression profiles of the 13 mitochondrial gene signatures and stemness in KICH and LAML, as well as a negative correlation in THYM and UCEC. Thus, identifying gene signatures with prognostic models based on stemness provides a powerful tool for cancer treatment.

CONCLUDING REMARKS

LGG downregulates mitochondrial protein-coding genes. The downregulated 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*) exhibits good OS, DSS, and PFI predictive value, and DFI and PFI survival were female-biased. The expression of *MT-ND5* and *MT-ND6* is female-biased in brain tissues. Age-dependent sex bias occurs in brain tissues, with higher gene expression in the 50-59 and 60-79-year-old populations. Distinct immune cell infiltrates were associated with mitochondrial protein-coding gene signature in LGG. We found a stemness-related gene signature for all 13 mtDNA genes in KICH. This study identifies mitochondrial prognostic genes.

WEB RESOURCES

The URLs for data presented herein are as follows:
UCSC Genome Browser, https://genome.ucsc.edu/
UCSCXenaShiny, https://cran.r-project.org/web/packages/UCSCXenaShiny/index.html
GTEx Portal, https://www.gtexportal.org/
GEPIA2, http://gepia2.cancer-pku.cn/#index
MITOMAP, https://mitomap.org/foswiki/bin/view/MITOMAP/WebHome
R software package, http://www.R-project.org

SUPPLEMENTARY TABLES

SUPPLEMENTARY FIGURES

Supplementary Figure 1. The landscape of Mitochondrial Protein-Coding Genes and Polymorphic Sites distribution Across all Samples and Tissues.

Supplementary Figure 2. Kaplan-Meier plot of overall survival outcome in low-grade glioma. The curves denote the contribution of the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*) to the overall survival outcome in patients from the low (blue – downregulated) and high (red – upregulated) expression groups. The censoring number refers to patients who did not suffer the outcome of interest during the specified study period. The overexpression of the signature has a more significant prognostic factor value.

Supplementary Figure 3. Kaplan-Meier plot of overall survival outcome in low-grade glioma. The curves denote the contribution of the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*), in males, to the overall survival outcome in patients from the low (blue – downregulated) and high (red – upregulated) expression groups. The censoring number refers to patients who did not suffer the outcome of interest during the specified study period. The overexpression of the signature has a more significant prognostic factor value.

Supplementary Figure 4. Kaplan-Meier plot of disease-specific survival outcome in low-grade glioma. The curves denote the contribution of the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*), in females, to the disease-specific survival outcome in patients from the low (blue – downregulated) and high (red – upregulated) expression groups. The censoring number refers to patients who did not suffer the outcome of interest during the specified study period. The overexpression of the signature has a more significant prognostic factor value.

Supplementary Figure 5. Kaplan-Meier plot of disease-specific survival outcome in low-grade glioma. The curves denote the contribution of the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*), in males, to the disease-specific survival outcome in patients from the low (blue – downregulated) and high (red – upregulated) expression groups. The censoring number refers to patients who did not suffer the outcome of interest during the specified study period. The overexpression of the signature has a more significant prognostic factor value.

Supplementary Figure 6. Kaplan-Meier plot of progression-free-interval outcome in low-grade glioma. The curves denote the contribution of the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*), in females, to the progression-free-interval outcome in patients from the low (blue – downregulated) and high (red – upregulated) expression groups. The censoring number refers to patients who did not suffer the outcome of interest during the specified study period. The overexpression of the signature has a more significant prognostic factor value.

Supplementary Figure 7. Kaplan-Meier plot of progression-free-interval outcome in low-grade glioma. The curves denote the contribution of the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*), in males, to the progression-free-interval outcome in patients from the low (blue – downregulated) and high (red – upregulated) expression groups. The censoring number refers to patients who did not suffer the outcome of interest during the specified study period. The overexpression of the signature has a more significant prognostic factor value.

Supplementary Figure 8. Kaplan-Meier plot of disease-free interval outcome in low-grade glioma in females. The curves denote the contribution of the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*) to the disease-free interval outcome in female patients from the low (blue – downregulated) and high (red – upregulated) expression groups. The censoring number refers to patients who did not suffer the outcome of interest during the specified study period, and the downregulation of the signature has a more significant prognostic factor value.

Supplementary Figure 9. Kaplan-Meier plot of disease-free interval outcome in low-grade glioma. The curves denote the contribution of the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*), in males, to the disease-free interval outcome in patients from the low (blue – downregulated) and high (red – upregulated) expression groups. The censoring number refers to patients who did not suffer the outcome of interest during the specified study period. The overexpression of the signature has a more significant prognostic factor value.

Supplementary Figure 19. Heatmap of the correlations between the expression profiles of the 7-component gene signature and immune cell infiltrates. The red-blue heatmap scale shows the significance levels of Spearman correlation coefficients (***, **, *). In THYM, the 7-component gene signature is positively correlated with infiltrates of endothelial cells, neutrophils, myeloid dendritic cells, T cell CD4+, and macrophage M2 (rho $\geq$0.50, p-value $\leq$ 0.001).

Supplementary Figure 11. Heatmap of the correlations between the expression profiles of the 8-component gene signature and immune cell infiltrates. The red-blue heatmap scale shows the significance levels of Spearman correlation coefficients (***, **, *). In THYM, the 7-component gene signature is negatively correlated with infiltrates of T cell CD4+ Th1, T cell, and T cell regulatory (Tregs) (rho $\leq$ -0.5, p-value $\leq$ 3.05e-10).

DRIVE ACCESS FOR SUPPLEMENTARY DATA

https://drive.google.com/drive/folders/1JggbN51c0xpeUNrDJugP4yCqersz_F0_?usp=sharing

CONFLICT OF INTEREST

The authors declare that the research was conducted without any commercial or financial relationships that could be construed as potential conflicts of interest.

FUNDING

This work was supported by grants from the Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro – FAPERJ (BR) (http://www.faperj.br/) [grant number E26/010.001036/2015 to EM-A], and from the Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (BR) (http://cnpq.br/) [grant numbers 301034/2012-5 and 308780/2015-9 to EM-A]. ATS received a graduate fellowship from FAPERJ/UENF, and FBM received a postdoctoral fellowship from CAPES/PADCT. The agencies had no role in the study design, data collection, analysis, publication decision, or manuscript preparation.

1030 # Acknowledgments

1031 The authors wish to thank the members of the Medina-Acosta laboratory for their insights and
1032 productive lab brainstorming meetings.

Figure Legends

Figure 1. Balloon plot of the differential expressed mitochondrial protein-coding genes in normal versus cancer tissues. Balloon size refers to the average expression (norm_counts); blue color intensity refers to the Log2FC. Overexpression (+); underexpression (-). X-axis: the 13 mitochondrial protein-coding genes; Y-axis, FDR (-log10). The tissues are indicated in each box.

Figure 2. Heatmap of significant mitochondrial protein-coding gene expression across different ages for each tissue in males. The expression of all mitochondrial protein-coding genes was measured using RPKM (reads per kilobase million), and the color scale denotes the level of significance of gene expression in each tissue. The different ages consisted of three groups (20-49 years; 50-59 years, and 60-79 years). The significance level was measured based on the Tukey test and the FDR adjustment, Benjamin-Hochberg.

Figure 3. Heatmap of significant mitochondrial protein-coding gene expression across different ages for each tissue in females. The expression of all mitochondrial protein-coding genes was measured using RPKM (reads per kilobase million), and the color scale denotes the level of significance of gene expression in each tissue. The different ages consisted of three groups (20-49 years; 50-59 years, and 60-79 years). The significance level was measured based on the Tukey test and the FDR adjustment, Benjamin-Hochberg.

Figure 4. Heatmap of significant mitochondrial protein-coding gene expression between sexes in 20-49-year-old group. All mitochondrial protein-coding genes were measured between the sexes (female and male) using RPKM (reads per kilobase million) across different tissues, and the color scale denotes the level of significance of gene expression in each tissue. The sex-specific tissue samples were excluded, and the significance level was measured based on the Wilcox test and the FDR adjustment, Benjamin-Hochberg.

Figure 5. Heatmap of significant mitochondrial protein-coding gene expression between sexes in 50-59-year-old group. All mitochondrial protein-coding genes were measured between the sexes (female and male) using RPKM (reads per kilobase million) across different tissues, and the color scale denotes the level of significance of gene expression in each tissue. The sex-specific tissue samples were excluded, and the significance level was measured based on the Wilcox test and the FDR adjustment, Benjamin-Hochberg.

Figure 6. Heatmap of significant mitochondrial protein-coding gene expression between sexes in 60-79-year-old group. All mitochondrial protein-coding genes were measured between the sexes (female and male) using RPKM (reads per kilobase million) across different tissues, and the color scale denotes the level of significance of gene expression in each tissue. The sex-specific tissue samples were excluded, and the significance level was measured based on the Wilcox test and the FDR adjustment, Benjamin-Hochberg.

Figure 7. Boxplot of mitochondrial polymorphic site expression in females across tissues at different age groups. The polymorphic site expression at the mitochondrial protein-coding genes was measured using RPKM (reads per kilobase million) in three age groups, 20-49 years (red), 50-59 years (green), and 60-79 years (blue). The Tukey test and the FDR adjustment, Benjamini-Hochberg assessed significant differences between age groups.

Figure 8. Boxplot of mitochondrial polymorphic site expression in males across tissues at different age groups. The polymorphic site expression at the mitochondrial protein-coding genes was measured using RPKM (reads per kilobase million) in three age groups, 20-49 years (red), 50-59 years (green), and 60-79 years (blue). The Tukey test and the FDR adjustment, Benjamini-Hochberg assessed significant differences between age groups.

FIGURE 9. Overall survival heatmap for the 13 mitochondrial protein-coding genes. Squares denote overall survival associated with overexpression (red) or under-expression (blue). Highlighted frames indicate significant associations with overall survival outcomes based on hazard ratios (logarithmic scale log10). Y-axis: mitochondrial genes (Ensemble transcripts and UCSCXena names), X-axis: pan-cancer types. The under-expression profiles of the MT-CO1, MT-CO2, MT-ND5, and MT-ND6 genes have a significant overall survival risk in LGG (hazard ratio range 0.41-0.51, Logrank p ≤0.00026, p(HR) ≤0.00033), while the under-expression of the MT-CO3 gene is a hazard risk in PAAD (hazard ratio 0.45, p(HR)=0.00022, Logrank p=0.00017).

FIGURE 10. Kaplan-Meier plot of overall survival outcome in low-grade glioma. The curves denote the contribution of the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*) to the overall survival outcome in patients from the low (blue – downregulated) and high (red – upregulated) expression groups. The censoring number refers to patients who did not suffer the outcome of interest during the specified study period. The overexpression of the signature has a more significant prognostic factor value.

FIGURE 11. Kaplan-Meier plot of disease-specific outcome in low-grade glioma. The curves denote the contribution of the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*) to the disease-specific outcome in patients from the low (blue – downregulated) and high (red – upregulated) expression groups. The censoring number refers to patients who did not suffer the outcome of interest during the specified study period. The overexpression of the signature has a more significant prognostic factor value.

FIGURE 12. Kaplan-Meier plot of progression-free-interval outcome in low-grade glioma. The curves denote the contribution of the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*) to the progression-free-interval outcome in patients from the low (blue – downregulated) and high (red – upregulated) expression groups. The censoring number refers to patients who did not suffer the outcome of interest during the specified study period. The overexpression of the signature has a more significant prognostic factor value.

FIGURE 13. Boxplot of the molecular profile of mitochondrial protein-coding genes for protective or risk effects based on hazard ratio. Cox model analysis for the 4-component gene signature (*MT-CO1*, *MT-CO2*, *MT-ND5*, *MT-ND6*) *in 33* cancer types. The signature is protective (red) in *LGG*, *ACC*, *KICH*, *KIRP*, *LIHC*, *and PAAD* cancers (p-value < 0.0113) or risk (blue) in *THYM and UCEC* cancers (p-value < 0.0465).

FIGURE 14. Heatmap of the correlations between the expression profiles of the 13-component gene signature and immune cell infiltrates. The red-blue heatmap scale shows the Spearman

correlation coefficients' significance levels (***, **, *). In DBLC, the 13-component gene signature is positively correlated with infiltrates of neutrophil, NK, T-cell CD4 memory, M1/M2 macrophages, T-cell CD8+/CD4+, and B-cell memory cells (rho ≥ 0.50 p-value < 0.0003).

FIGURE 15. Heatmap of the correlations between the expression profiles of the 4-component gene signature and immune cell infiltrates. The red-blue heatmap scale shows the Spearman correlation coefficients' significance levels (***, **, *). In LGG, the 4-component gene signature is positively correlated with monocyte, NK, and B cell infiltrates (rho > 0.3, p-value < 2.01e-12).

FIGURE 16. Heatmap of the correlations between the expression profiles of the 4-component gene signature and immune cell infiltrates. The red-blue heatmap scale shows the Spearman correlation coefficients' significance levels (***, **, *). In LGG, the 4-component gene signature is negatively correlated with T cell CD4+ Th2, macrophage M1 and M2, myeloid dendritic cell, neutrophil, and activated myeloid dendritic cell activated cell infiltrates (rho ≤ 0.3, p-value < 2.26E-12).

FIGURE 17. Radar plot for the correlation coefficients between the 13-component gene signatures and stemness in 33 cancer types. Shown is the distribution of the breadth of the Spearman correlation coefficients, where the association can be positive (+1), negative (-1), or neutral (0). The more significant positive correlation with stemness occurred in KICH and LAML.

REFERENCES

Allaire, J.J.B., Ma (2012). RStudio: integrated development environment for R. 770**,** 165-171.

Ansell, S.M., Stenson, M., Habermann, T.M., Jelinek, D.F., and Witzig, T.E. (2001). Cd4+ T-cell immune response to large B-cell non-Hodgkin's lymphoma predicts patient outcome. *J Clin Oncol* 19**,** 720-726.

Berchtold, N.C., Cribbs, D.H., Coleman, P.D., Rogers, J., Head, E., Kim, R., Beach, T., Miller, C., Troncoso, J., Trojanowski, J.Q., Zielke, H.R., and Cotman, C.W. (2008). Gene expression changes in the course of normal brain aging are sexually dimorphic. *Proc Natl Acad Sci U S A* 105**,** 15605-15610.

Breton, S., Milani, L., Ghiselli, F., Guerra, D., Stewart, D.T., and Passamonti, M. (2014). A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs. *Trends Genet* 30**,** 555-564.

Cai, J.L., Zhu, G.Q., Du, J.X., Wang, B., Wan, J.L., Xiao, K., and Dai, Z. (2021). Identification and validation of a new gene signature predicting prognosis of hepatocellular carcinoma patients by network analysis of stemness indices. *Expert Rev Gastroenterol Hepatol* 15**,** 699-709.

Calloway, C.D., Reynolds, R.L., Herrin, G.L., Jr., and Anderson, W.W. (2000). The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. *Am J Hum Genet* 66**,** 1384-1397.

Capt, C., Passamonti, M., and Breton, S. (2016). The human mitochondrial genome may code for more than 13 proteins. *Mitochondrial DNA Part A* 27**,** 3098-3101.

Chen, J., Li, Y., Han, X., Pan, Y., and Qian, X. (2021). An autophagic gene-based signature to predict the survival of patients with low-grade gliomas. *Cancer Med* 10**,** 1848-1859.

Da Silva Francisco Junior, R., Dos Santos Ferreira, C., Santos, E.S.J.C., Terra Machado, D., Cortes Martins, Y., Ramos, V., Simoes Carnivali, G., Garcia, A.B., and Medina-Acosta, E. (2019). Pervasive Inter-Individual Variation in Allele-Specific Expression in Monozygotic Twins. *Front Genet* 10**,** 1178.

De Magalhaes, J.P., Curado, J., and Church, G.M. (2009). Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics* 25**,** 875-881.

Droge, W. (2002). Free radicals in the physiological control of cell function. *Physiol Rev* 82**,** 47-95.

Ellegren, H., and Parsch, J. (2007). The evolution of sex-biased genes and sex-biased gene expression. *Nat Rev Genet* 8**,** 689-698.

Faure, E., Delaye, L., Tribolo, S., Levasseur, A., Seligmann, H., and Barthelemy, R.M. (2011). Probable presence of an ubiquitous cryptic mitochondrial gene on the antisense strand of the cytochrome oxidase I gene. *Biol Direct* 6**,** 56.

Foster, L.J., De Hoog, C.L., Zhang, Y., Zhang, Y., Xie, X., Mootha, V.K., and Mann, M. (2006). A mammalian organelle map by protein correlation profiling. *Cell* 125**,** 187-199.

Frenk, S., and Houseley, J. (2018). Gene expression hallmarks of cellular ageing. *Biogerontology* 19**,** 547-566.

Gao, Z., Zhang, D., Duan, Y., Yan, L., Fan, Y., Fang, Z., and Liu, Z. (2019). A five-gene signature predicts overall survival of patients with papillary renal cell carcinoma. *PLOS ONE* 14**,** e0211491.

Gemmell, N.J., Metcalf, V.J., and Allendorf, F.W. (2004). Mother's curse: the effect of mtDNA on individual fitness and population viability. *Trends Ecol Evol* 19**,** 238-244.

Geromel, V., Kadhom, N., Cebalos-Picot, I., Ouari, O., Polidori, A., Munnich, A., Rotig, A., and Rustin, P. (2001). Superoxide-induced massive apoptosis in cultured skin fibroblasts

harboring the neurogenic ataxia retinitis pigmentosa (NARP) mutation in the ATPase-6 gene of the mitochondrial DNA. *Hum Mol Genet* 10**,** 1221-1228.

Gershoni, M., and Pietrokovski, S. (2017). The landscape of sex-differential transcriptome and its consequent selection in human adults. *BMC Biol* 15**,** 7.

Guo, J.C., Wei, Q.S., Dong, L., Fang, S.S., Li, F., and Zhao, Y. (2021). Prognostic Value of an Autophagy-Related Five-Gene Signature for Lower-Grade Glioma Patients. *Front Oncol* 11**,** 644443.

Gustafsson, C.M., Falkenberg, M., and Larsson, N.G. (2016). Maintenance and Expression of Mammalian Mitochondrial DNA. *Annu Rev Biochem* 85**,** 133-160.

Harman, D. (1992a). Free radical theory of aging. *Mutat Res* 275**,** 257-266.

Harman, D. (1992b). Free radical theory of aging: history. *EXS* 62**,** 1-10.

Harman, D. (1993). Free radical involvement in aging. Pathophysiology and therapeutic implications. *Drugs Aging* 3**,** 60-80.

He, Y., Wu, J., Dressman, D.C., Iacobuzio-Donahue, C., Markowitz, S.D., Velculescu, V.E., Diaz, L.A., Jr., Kinzler, K.W., Vogelstein, B., and Papadopoulos, N. (2010). Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. *Nature* 464**,** 610-614.

Hong, L., Zhou, Y., Xie, X., Wu, W., Shi, C., Lin, H., and Shi, Z. (2021). A stemness-based eleven-gene signature correlates with the clinical outcome of hepatocellular carcinoma. *BMC Cancer* 21**,** 716.

Huang, X.Y., Qin, W.T., Su, Q.S., Qiu, C.C., Liu, R.C., Xie, S.S., Hu, Y., and Zhu, S.Y. (2021). A New Stemness-Related Prognostic Model for Predicting the Prognosis in Pancreatic Ductal Adenocarcinoma. *Biomed Res Int* 2021**,** 6669570.

Iacobellis, G. (2021). Aging Effects on Epicardial Adipose Tissue. *Front Aging* 2**,** 666260.

Iglesia, M.D., Parker, J.S., Hoadley, K.A., Serody, J.S., Perou, C.M., and Vincent, B.G. (2016). Genomic Analysis of Immune Cell Infiltrates Across 11 Tumor Types. *J Natl Cancer Inst* 108.

Kassam, I., Qi, T., Lloyd-Jones, L., Holloway, A., Jan Bonder, M., Henders, A.K., Martin, N.G., Powell, J.E., Franke, L., Montgomery, G.W., Visscher, P.M., and Mcrae, A.F. (2016). Evidence for mitochondrial genetic control of autosomal gene expression. *Hum Mol Genet* 25**,** 5332-5338.

Keane, C., Gill, D., Vari, F., Cross, D., Griffiths, L., and Gandhi, M. (2013). CD4(+) tumor infiltrating lymphocytes are prognostic and independent of R-IPI in patients with DLBCL receiving R-CHOP chemo-immunotherapy. *Am J Hematol* 88**,** 273-276.

Keane, C., Vari, F., Hertzberg, M., Cao, K.A., Green, M.R., Han, E., Seymour, J.F., Hicks, R.J., Gill, D., Crooks, P., Gould, C., Jones, K., Griffiths, L.R., Talaulikar, D., Jain, S., Tobin, J., and Gandhi, M.K. (2015). Ratios of T-cell immune effectors and checkpoint molecules as prognostic biomarkers in diffuse large B-cell lymphoma: a population-based study. *Lancet Haematol* 2**,** e445-455.

Kim, G.E., Kim, N.I., Lee, J.S., Park, M.H., and Kang, K. (2020). Differentially Expressed Genes in Matched Normal, Cancer, and Lymph Node Metastases Predict Clinical Outcomes in Patients With Breast Cancer. *Appl Immunohistochem Mol Morphol* 28**,** 111-122.

Kim, S., Myers, L., Ravussin, E., Cherry, K.E., and Jazwinski, S.M. (2016). Single nucleotide polymorphisms linked to mitochondrial uncoupling protein genes UCP2 and UCP3 affect mitochondrial metabolism and healthy aging in female nonagenarians. *Biogerontology* 17**,** 725-736.

Kirches, E., Michael, M., Warich-Kirches, M., Schneider, T., Weis, S., Krause, G., Mawrin, C., and Dietzmann, K. (2001). Heterogeneous tissue distribution of a mitochondrial DNA

polymorphism in heteroplasmic subjects without mitochondrial disorders. *J Med Genet* 38**,** 312-317.

Kirkwood, T.B.a.A., S.N. (2000). Why do we age? *Nature (Lond)* 408.

Kopsidas, G., Zhang, C., Yarovaya, N., Kovalenko, S., Graves, S., Richardson, M., and Linnane, A.W. (2002). Stochastic mitochondrial DNA changes: bioenergy decline in type I skeletal muscle fibres correlates with a decline in the amount of amplifiable full-length mtDNA. *Biogerontology* 3**,** 29-36.

Larsson Ng, W.J., Wilhelmsson H, Oldfors a, Rustin P, Et Al. (1998). Mitochondrial transcription factor A is necessary for mtDNA maintenance and embryogenesis in mice. *Nat. Genet* 18.

Lee, C., Yen, K., and Cohen, P. (2013). Humanin: a harbinger of mitochondrial-derived peptides? *Trends Endocrinol Metab* 24**,** 222-228.

Li, N., Zhao, J., Ma, Y., Roy, B., Liu, R., Kristiansen, K., and Gao, Q. (2018). Dissecting the expression landscape of mitochondrial genes in lung squamous cell carcinoma and lung adenocarcinoma. *Oncol Lett* 16**,** 3992-4000.

Li, W., Lu, J., Ma, Z., Zhao, J., and Liu, J. (2019). An Integrated Model Based on a Six-Gene Signature Predicts Overall Survival in Patients With Hepatocellular Carcinoma. *Front Genet* 10**,** 1323.

Li, X., Wen, D., Li, X., Yao, C., Chong, W., and Chen, H. (2020). Identification of an Immune Signature Predicting Prognosis Risk and Lymphocyte Infiltration in Colon Cancer. *Front Immunol* 11**,** 1678.

Lin, J.Z., and Lin, N. (2021). A risk signature of three autophagy-related genes for predicting lower grade glioma survival is associated with tumor immune microenvironment. *Genomics* 113**,** 767-777.

Linnane Aw, M.S., Ozawa T, and Tanaka M. (1989). Mitochondrial DNA mutations as an important contributor to ageing and degenerative diseases. *Lancet* 1.

Liu, J., and Li, G. (2021). Identification and validation of a risk signature based on extracellular matrix-related genes in gliomas. *Medicine (Baltimore)* 100**,** e25603.

Lopes-Ramos, C.M., Chen, C.Y., Kuijjer, M.L., Paulson, J.N., Sonawane, A.R., Fagny, M., Platig, J., Glass, K., Quackenbush, J., and Demeo, D.L. (2020). Sex Differences in Gene Expression and Regulatory Networks across 29 Human Tissues. *Cell Rep* 31**,** 107795.

Malta, T.M., Sokolov, A., Gentles, A.J., Burzykowski, T., Poisson, L., Weinstein, J.N., Kaminska, B., Huelsken, J., Omberg, L., Gevaert, O., Colaprico, A., Czerwinska, P., Mazurek, S., Mishra, L., Heyn, H., Krasnitz, A., Godwin, A.K., Lazar, A.J., Cancer Genome Atlas Research, N., Stuart, J.M., Hoadley, K.A., Laird, P.W., Noushmehr, H., and Wiznerowicz, M. (2018). Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell* 173**,** 338-354 e315.

Manfroi, B., Moreaux, J., Righini, C., Ghiringhelli, F., Sturm, N., and Huard, B. (2018). Tumor-associated neutrophils correlate with poor prognosis in diffuse large B-cell lymphoma patients. *Blood Cancer J* 8**,** 66.

Mele, M., Ferreira, P.G., Reverter, F., Deluca, D.S., Monlong, J., Sammeth, M., Young, T.R., Goldmann, J.M., Pervouchine, D.D., Sullivan, T.J., Johnson, R., Segre, A.V., Djebali, S., Niarchou, A., Consortium, G.T., Wright, F.A., Lappalainen, T., Calvo, M., Getz, G., Dermitzakis, E.T., Ardlie, K.G., and Guigo, R. (2015). Human genomics. The human transcriptome across tissues and individuals. *Science* 348**,** 660-665.

Miquel, J., Economos, A.C., Fleming, J., and Johnson, J.E., Jr. (1980). Mitochondrial role in cell aging. *Exp Gerontol* 15**,** 575-591.

Mu, S., Ai, L., Fan, F., Qin, Y., Sun, C., and Hu, Y. (2018). Prognostic role of neutrophil-to-lymphocyte ratio in diffuse large B cell lymphoma patients: an updated dose-response meta-analysis. *Cancer Cell Int* 18**,** 119.

Muller-Hocker, J. (1989). Cytochrome-c-oxidase deficient cardiomyocytes in the human heart--an age-related phenomenon. A histochemical ultracytochemical study. *Am J Pathol* 134**,** 1167-1173.

Nagley, P., and Wei, Y.H. (1998). Ageing and mammalian mitochondrial genetics. *Trends in Genetics* 14**,** 513-517.

Naqvi, S., Godfrey, A.K., Hughes, J.F., Goodheart, M.L., Mitchell, R.N., and Page, D.C. (2019). Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. *Science* 365.

Ozawa, T. (1997). Genetic and functional changes in mitochondria associated with aging. *Physiol Rev* 77**,** 425-464.

Pan, X., Wang, Z., Liu, F., Zou, F., Xie, Q., Guo, Y., and Shen, L. (2021). A novel tailored immune gene pairs signature for overall survival prediction in lower-grade gliomas. *Transl Oncol* 14**,** 101109.

Papa, S. (1996). Mitochondrial oxidative phosphorylation changes in the lifespan. Molecular aspects and physiopathological implications. *Biochim Biophys Acta*.

Parsch, J., and Ellegren, H. (2013). The evolutionary causes and consequences of sex-biased gene expression. *Nat Rev Genet* 14**,** 83-87.

Pawar, A., Chowdhury, O.R., Chauhan, R., Talole, S., and Bhattacharjee, A. (2022). Identification of key gene signatures for the overall survival of ovarian cancer. *Journal of Ovarian Research* 15**,** 12.

Peters, M.J., Joehanes, R., Pilling, L.C., Schurmann, C., Conneely, K.N., Powell, J., Reinmaa, E., Sutphin, G.L., Zhernakova, A., Schramm, K., Wilson, Y.A., Kobes, S., Tukiainen, T., Consortium, N.U., Ramos, Y.F., Goring, H.H., Fornage, M., Liu, Y., Gharib, S.A., Stranger, B.E., De Jager, P.L., Aviv, A., Levy, D., Murabito, J.M., Munson, P.J., Huan, T., Hofman, A., Uitterlinden, A.G., Rivadeneira, F., Van Rooij, J., Stolk, L., Broer, L., Verbiest, M.M., Jhamai, M., Arp, P., Metspalu, A., Tserel, L., Milani, L., Samani, N.J., Peterson, P., Kasela, S., Codd, V., Peters, A., Ward-Caviness, C.K., Herder, C., Waldenberger, M., Roden, M., Singmann, P., Zeilinger, S., Illig, T., Homuth, G., Grabe, H.J., Volzke, H., Steil, L., Kocher, T., Murray, A., Melzer, D., Yaghootkar, H., Bandinelli, S., Moses, E.K., Kent, J.W., Curran, J.E., Johnson, M.P., Williams-Blangero, S., Westra, H.J., Mcrae, A.F., Smith, J.A., Kardia, S.L., Hovatta, I., Perola, M., Ripatti, S., Salomaa, V., Henders, A.K., Martin, N.G., Smith, A.K., Mehta, D., Binder, E.B., Nylocks, K.M., Kennedy, E.M., Klengel, T., Ding, J., Suchy-Dicey, A.M., Enquobahrie, D.A., Brody, J., Rotter, J.I., Chen, Y.D., Houwing-Duistermaat, J., Kloppenburg, M., Slagboom, P.E., Helmer, Q., Den Hollander, W., Bean, S., Raj, T., Bakhshi, N., Wang, Q.P., Oyston, L.J., Psaty, B.M., Tracy, R.P., Montgomery, G.W., Turner, S.T., et al. (2015). The transcriptional landscape of age in human peripheral blood. *Nat Commun* 6**,** 8570.

Poyton, R.O., and Mcewen, J.E. (1996). Crosstalk between nuclear and mitochondrial genomes. *Annual Review of Biochemistry* 65**,** 563-607.

Quan, Q., Xiong, X., Wu, S., and Yu, M. (2021). Identification of Autophagy-Related Prognostic Signature and Analysis of Immune Cell Infiltration in Low-Grade Gliomas. *Biomed Res Int* 2021**,** 7918693.

Rao, S., Welsh, L., Cunningham, D., Te-Poele, R.H., Benson, M., Norman, A., Saffery, C., Giddings, I., Workman, P., and Clarke, P.A. (2011). Correlation of overall survival with gene

expression profiles in a prospective study of resectable esophageal cancer. *Clin Colorectal Cancer* 10**,** 48-56.

Reznik, E., Wang, Q., La, K., Schultz, N., and Sander, C. (2017). Mitochondrial respiratory gene expression is suppressed in many cancers. *Elife* 6.

Richter, C. (1995). Oxidative damage to mitochondrial DNA and its relationship to ageing. *Int J Biochem Cell Biol* 27**,** 647-653.

Rinn, J.L., and Snyder, M. (2005). Sexual dimorphism in mammalian gene expression. *Trends Genet* 21**,** 298-305.

Santos, C., Sierra, B., Alvarez, L., Ramos, A., Fernandez, E., Nogues, R., and Aluja, M.P. (2008). Frequency and pattern of heteroplasmy in the control region of human mitochondrial DNA. *J Mol Evol* 67**,** 191-200.

Scarpulla, R.C. (2008). Transcriptional paradigms in mammalian mitochondrial biogenesis and function. *Physiol Rev* 88**,** 611-638.

Schapira, A.H. (2006). Mitochondrial disease. *Lancet* 368**,** 70-82.

Schopf, B., Weissensteiner, H., Schafer, G., Fazzini, F., Charoentong, P., Naschberger, A., Rupp, B., Fendt, L., Bukur, V., Giese, I., Sorn, P., Sant'anna-Silva, A.C., Iglesias-Gonzalez, J., Sahin, U., Kronenberg, F., Gnaiger, E., and Klocker, H. (2020). OXPHOS remodeling in high-grade prostate cancer involves mtDNA mutations and increased succinate oxidation. *Nat Commun* 11**,** 1487.

Shokolenko, I., Venediktova, N., Bochkareva, A., Wilson, G.L., and Alexeyev, M.F. (2009). Oxidative stress induces degradation of mitochondrial DNA. *Nucleic Acids Res* 37**,** 2539-2548.

Srinivasan, S., Guha, M., Kashina, A., and Avadhani, N.G. (2017). Mitochondrial dysfunction and mitochondrial dynamics-The cancer connection. *Biochim Biophys Acta Bioenerg* 1858**,** 602-614.

Stenton, S.L., and Prokisch, H. (2020). Genetics of mitochondrial diseases: Identifying mutations to help diagnosis. *EBioMedicine* 56**,** 102784.

Stewart, J.B., and Chinnery, P.F. (2021). Extreme heterogeneity of human mitochondrial DNA from organelles to populations. *Nat Rev Genet* 22**,** 106-118.

Taylor, R.W., and Turnbull, D.M. (2005). Mitochondrial DNA mutations in human disease. *Nat Rev Genet* 6**,** 389-402.

Tower, J. (2017). Sex-Specific Gene Expression and Life Span Regulation. *Trends Endocrinol Metab* 28**,** 735-747.

Van Den Akker, E.B., Passtoors, W.M., Jansen, R., Van Zwet, E.W., Goeman, J.J., Hulsman, M., Emilsson, V., Perola, M., Willemsen, G., Penninx, B.W., Heijmans, B.T., Maier, A.B., Boomsma, D.I., Kok, J.N., Slagboom, P.E., Reinders, M.J., and Beekman, M. (2014). Meta-analysis on blood transcriptomic studies identifies consistently coexpressed protein-protein interaction modules as robust markers of human aging. *Aging Cell* 13**,** 216-225.

Van Oven, M., and Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30**,** E386-394.

Wallace, D.C. (1992). Diseases of the mitochondrial DNA. *Annu Rev Biochem* 61**,** 1175-1212.

Wang, J., Ma, X., and Ma, J. (2022). Identification of Four Enhancer-Associated Genes as Risk Signature for Diffuse Glioma Patients. *J Mol Neurosci* 72**,** 410-419.

Welle, S., Brooks, A.I., Delehanty, J.M., Needler, N., and Thornton, C.A. (2003). Gene expression profile of aging in human muscle. *Physiol Genomics* 14**,** 149-159.

Xiao, K., Liu, Q., Peng, G., Su, J., Qin, C.Y., and Wang, X.Y. (2020a). Identification and validation of a three-gene signature as a candidate prognostic biomarker for lower grade glioma. *PeerJ* 8**,** e8312.

Xiao, L., Zou, G., Cheng, R., Wang, P., Ma, K., Cao, H., Zhou, W., Jin, X., Xu, Z., Huang, Y., Lin, X., Nie, H., and Jiang, Q. (2021). Alternative splicing associated with cancer stemness in kidney renal clear cell carcinoma. *BMC Cancer* 21**,** 703.

Xiao, Y., Cui, G., Ren, X., Hao, J., Zhang, Y., Yang, X., Wang, Z., Zhu, X., Wang, H., Hao, C., and Duan, H. (2020b). A Novel Four-Gene Signature Associated With Immune Checkpoint for Predicting Prognosis in Lower-Grade Glioma. *Front Oncol* 10**,** 605737.

Xu, H., Zou, R., Liu, J., and Zhu, L. (2021). A Risk Signature with Nine Stemness Index-Associated Genes for Predicting Survival of Patients with Uterine Corpus Endometrial Carcinoma. *J Oncol* 2021**,** 6653247.

Yan, S., Fang, J., Chen, Y., Xie, Y., Zhang, S., Zhu, X., and Fang, F. (2020). Comprehensive analysis of prognostic gene signatures based on immune infiltration of ovarian cancer. *BMC Cancer* 20**,** 1205.

Ye, W., Luo, C., Liu, F., Liu, Z., and Chen, F. (2021). CD96 Correlates With Immune Infiltration and Impacts Patient Prognosis: A Pan-Cancer Analysis. *Front Oncol* 11**,** 634617.

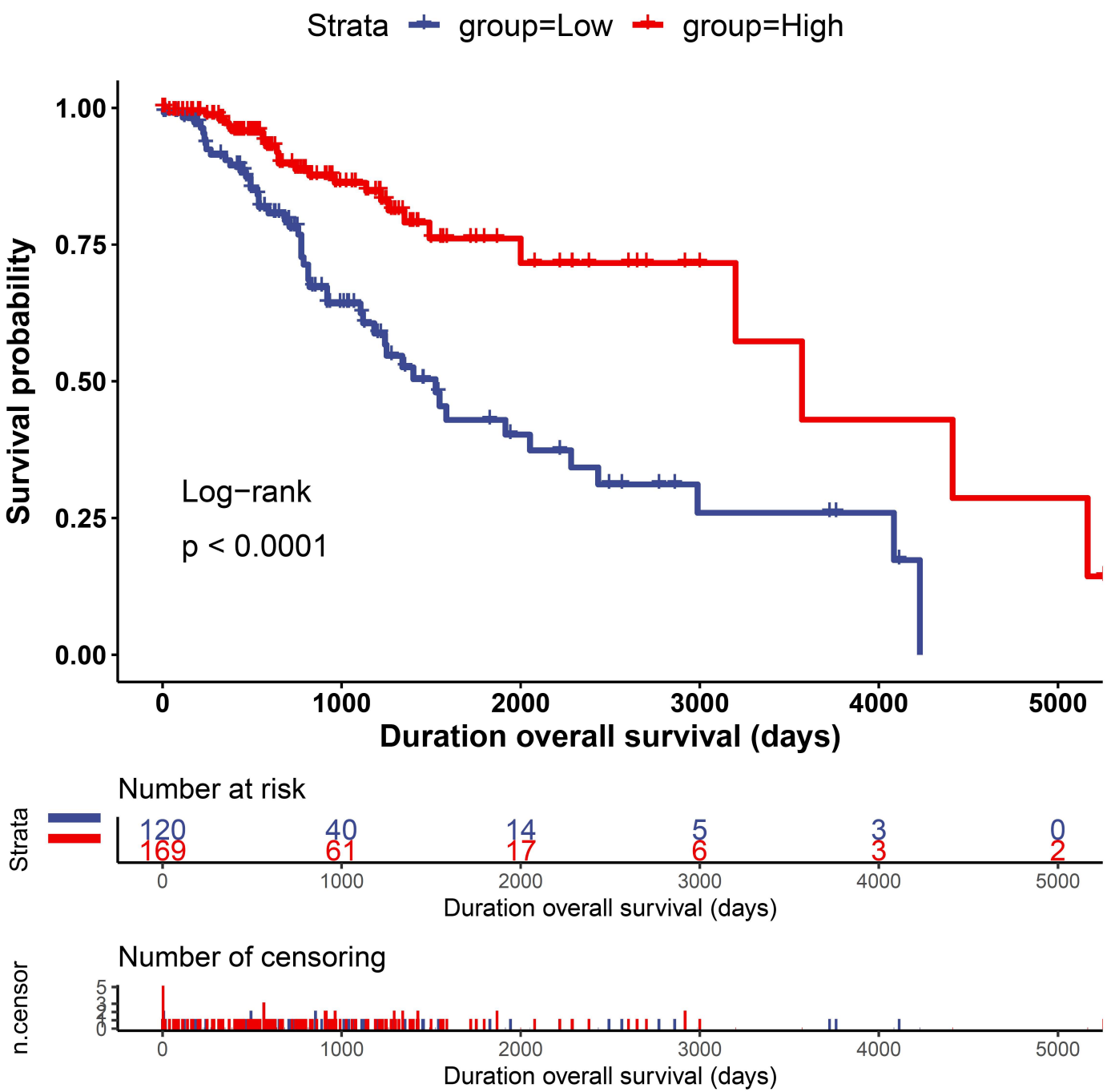
Zhang, G.H., Zhong, Q.Y., Gou, X.X., Fan, E.X., Shuai, Y., Wu, M.N., and Yue, G.J. (2019). Seven genes for the prognostic prediction in patients with glioma. *Clin Transl Oncol* 21**,** 1327-1335.

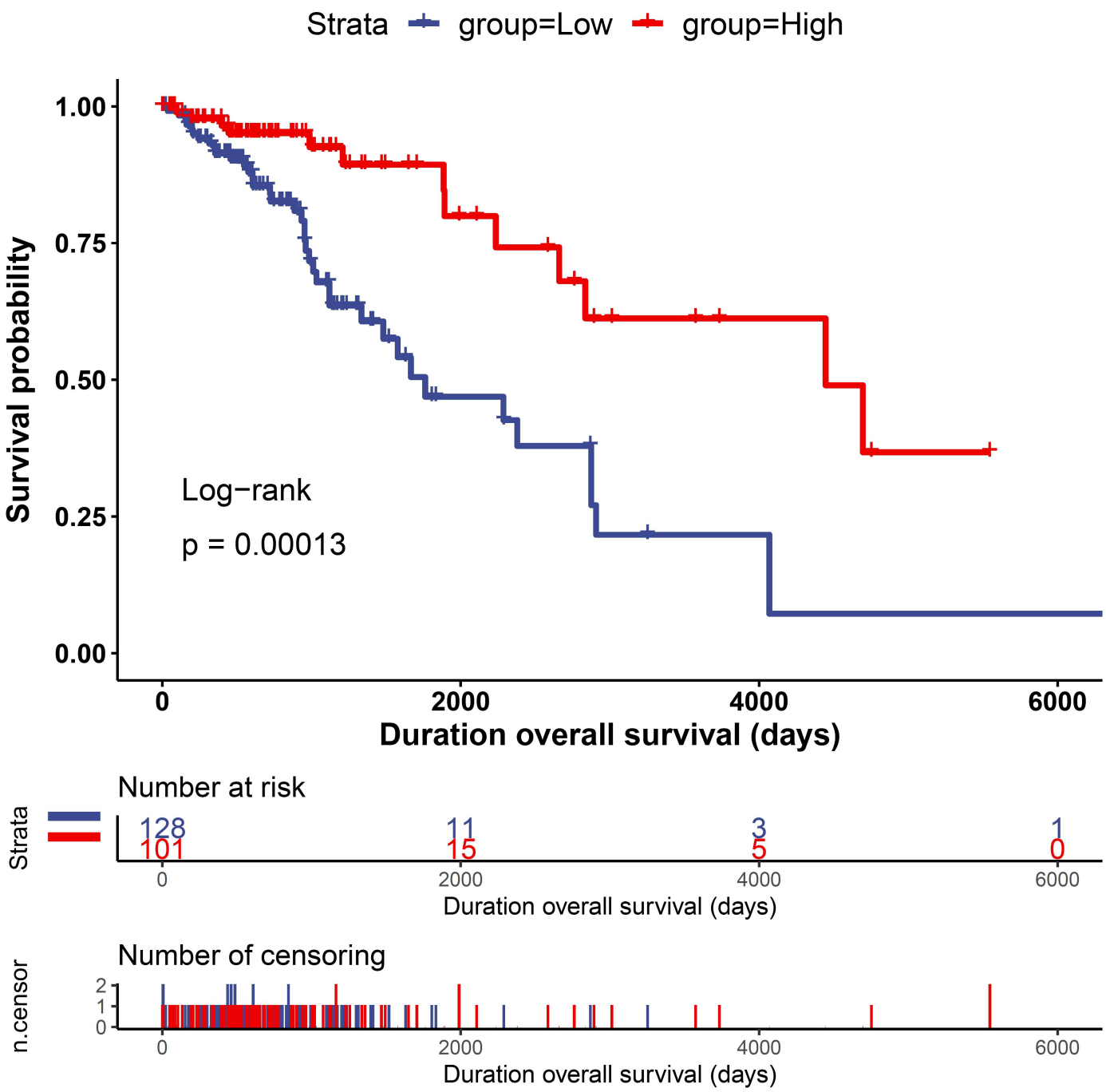
Zhang, M., Wang, X., Chen, X., Zhang, Q., and Hong, J. (2020). Novel Immune-Related Gene Signature for Risk Stratification and Prognosis of Survival in Lower-Grade Glioma. *Front Genet* 11**,** 363.

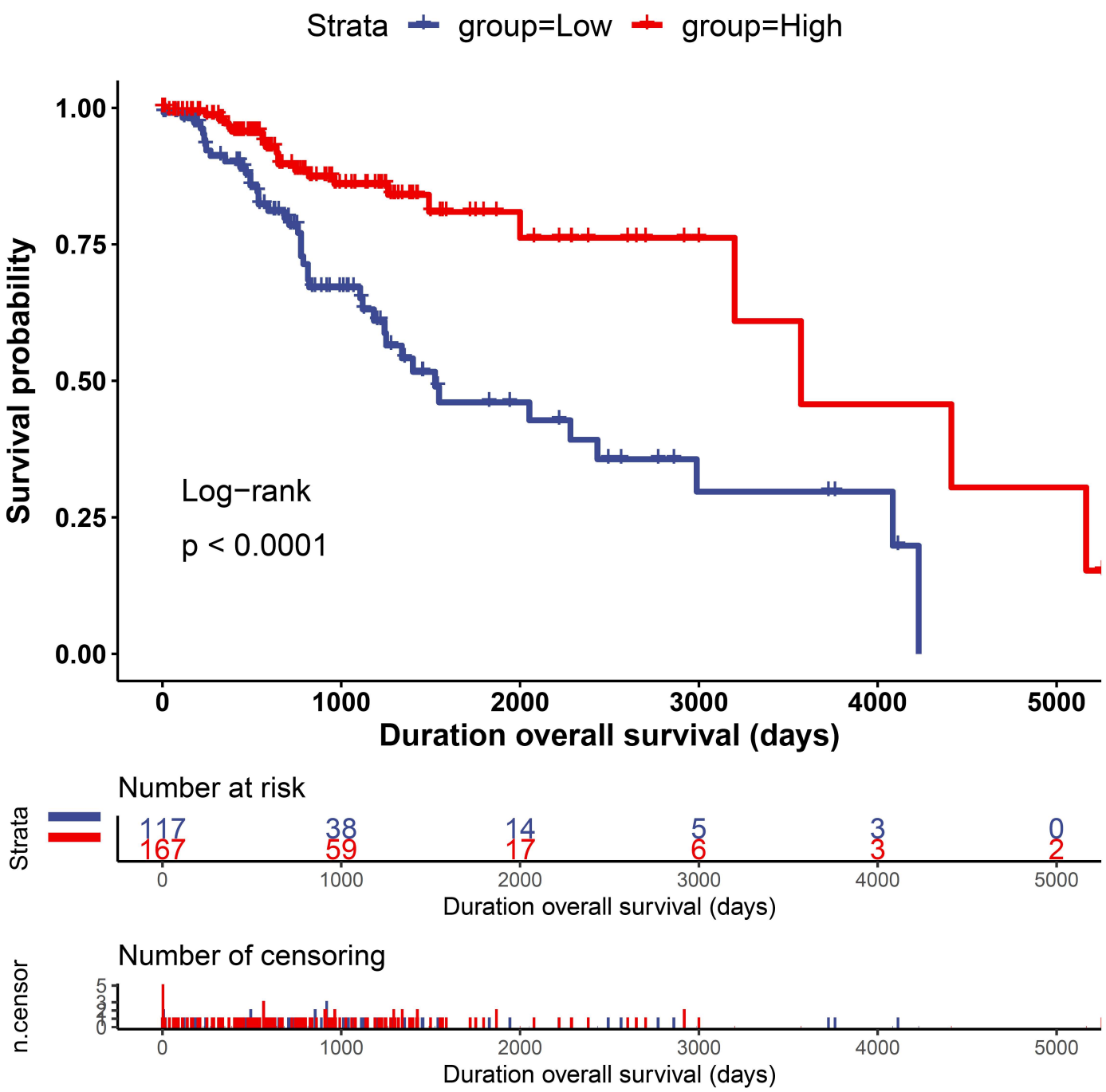
Zhang, Y., Peng, J., Du, H., Zhang, N., and Fang, X. (2021). Identification and Validation of Immune- and Stemness-Related Prognostic Signature of Melanoma. *Front Cell Dev Biol* 9**,** 755284.

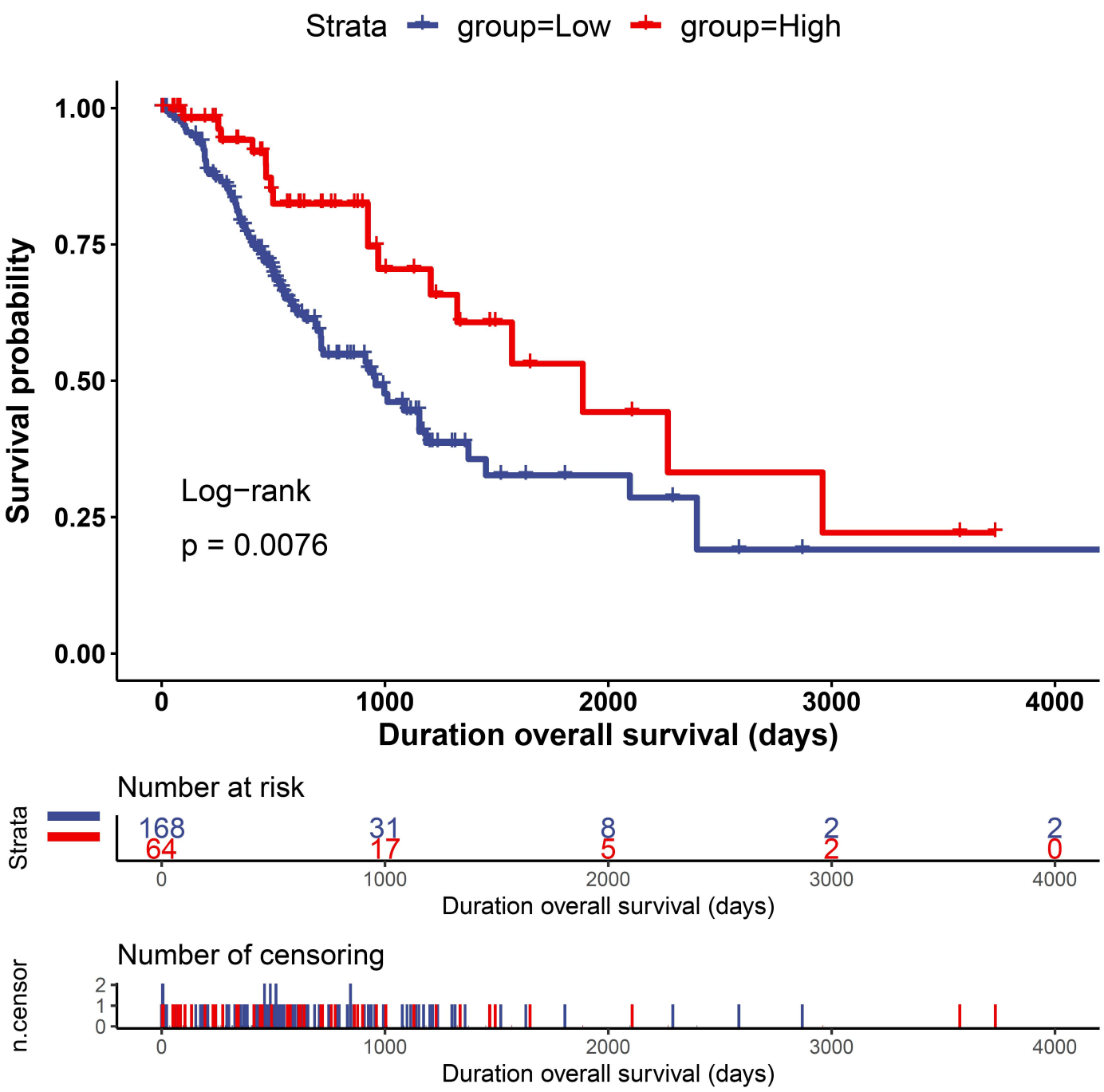
Zhao, Y., Zhang, J., Wang, S., Jiang, Q., and Xu, K. (2021). Identification and Validation of a Nine-Gene Amino Acid Metabolism-Related Risk Signature in HCC. *Front Cell Dev Biol* 9**,** 731790.

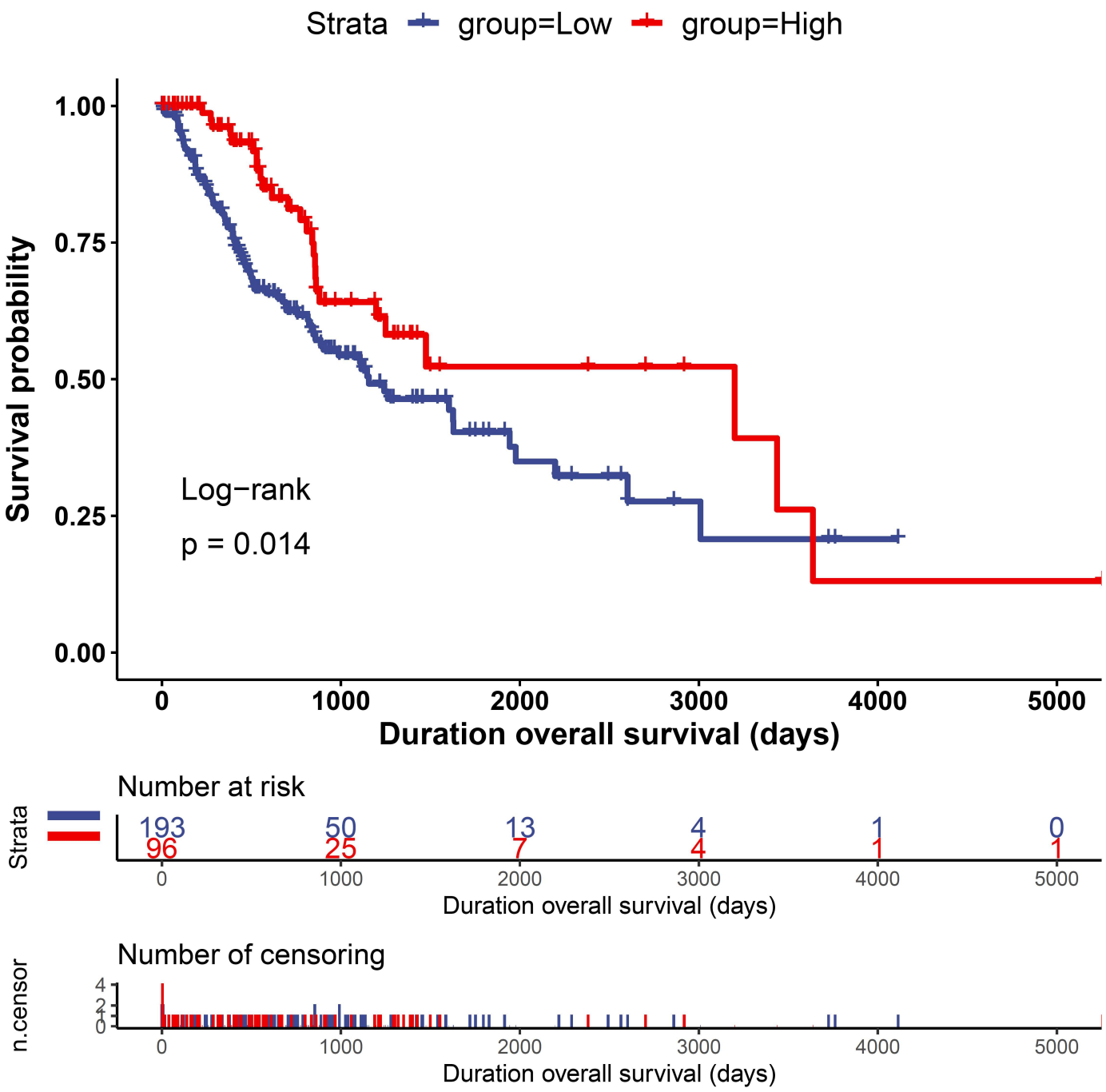
Zhu, Y., Song, Z., Wang, Z., and Chen, G. (2022). Protective Prognostic Biomarkers Negatively Correlated with Macrophage M2 Infiltration in Low-Grade Glioma. *J Oncol* 2022**,** 3623591.

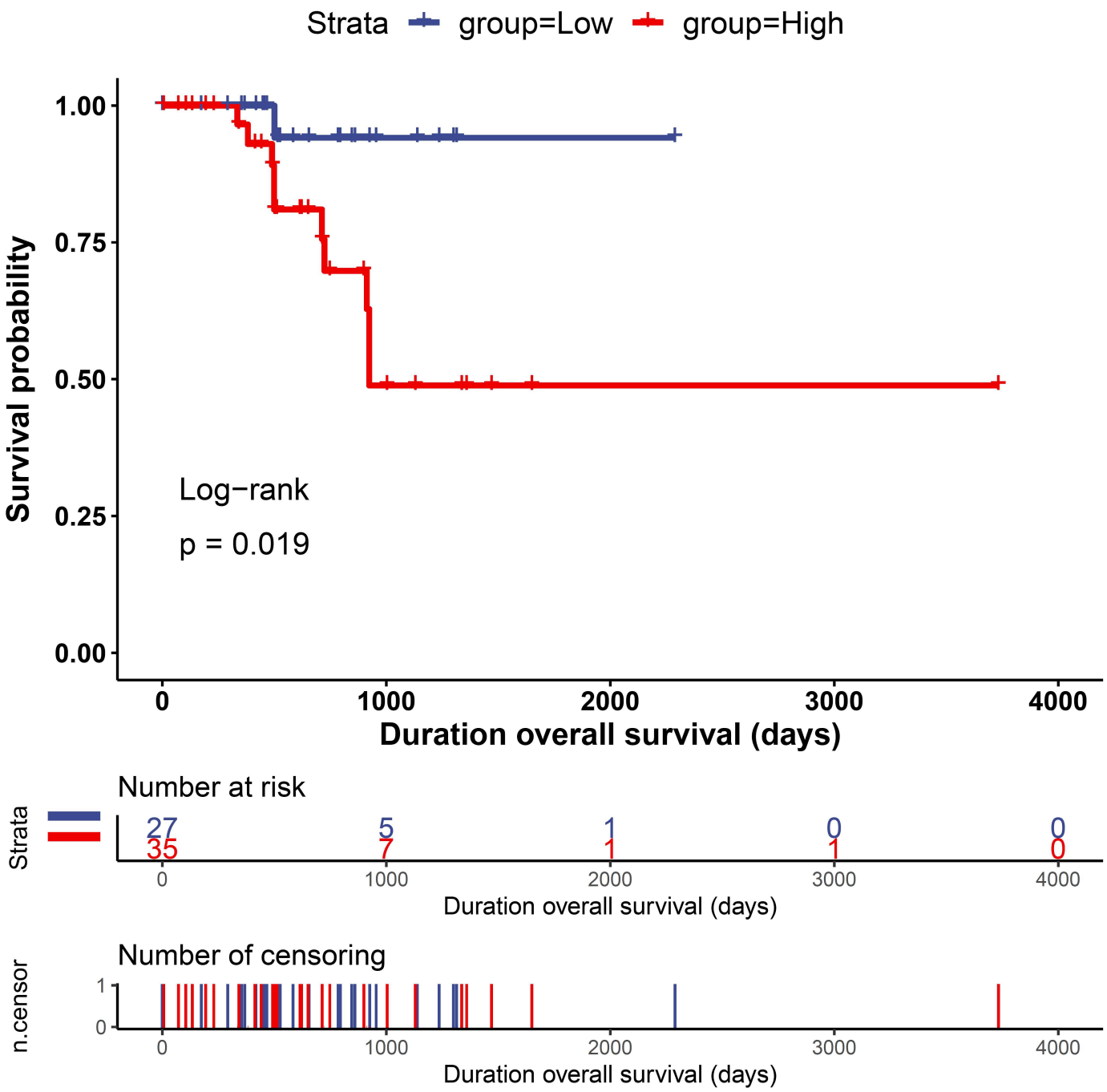

Strata
group=Low
group=High
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank
p < 0.0001
0
1000
2000
3000
4000
5000
Duration overall survival (days)
Number at risk
Strata
120
169
40
61
14
17
5
6
3
3
0
2
Duration overall survival (days)
Number of censoring
n.censor
5
3
1
0
Duration overall survival (days)

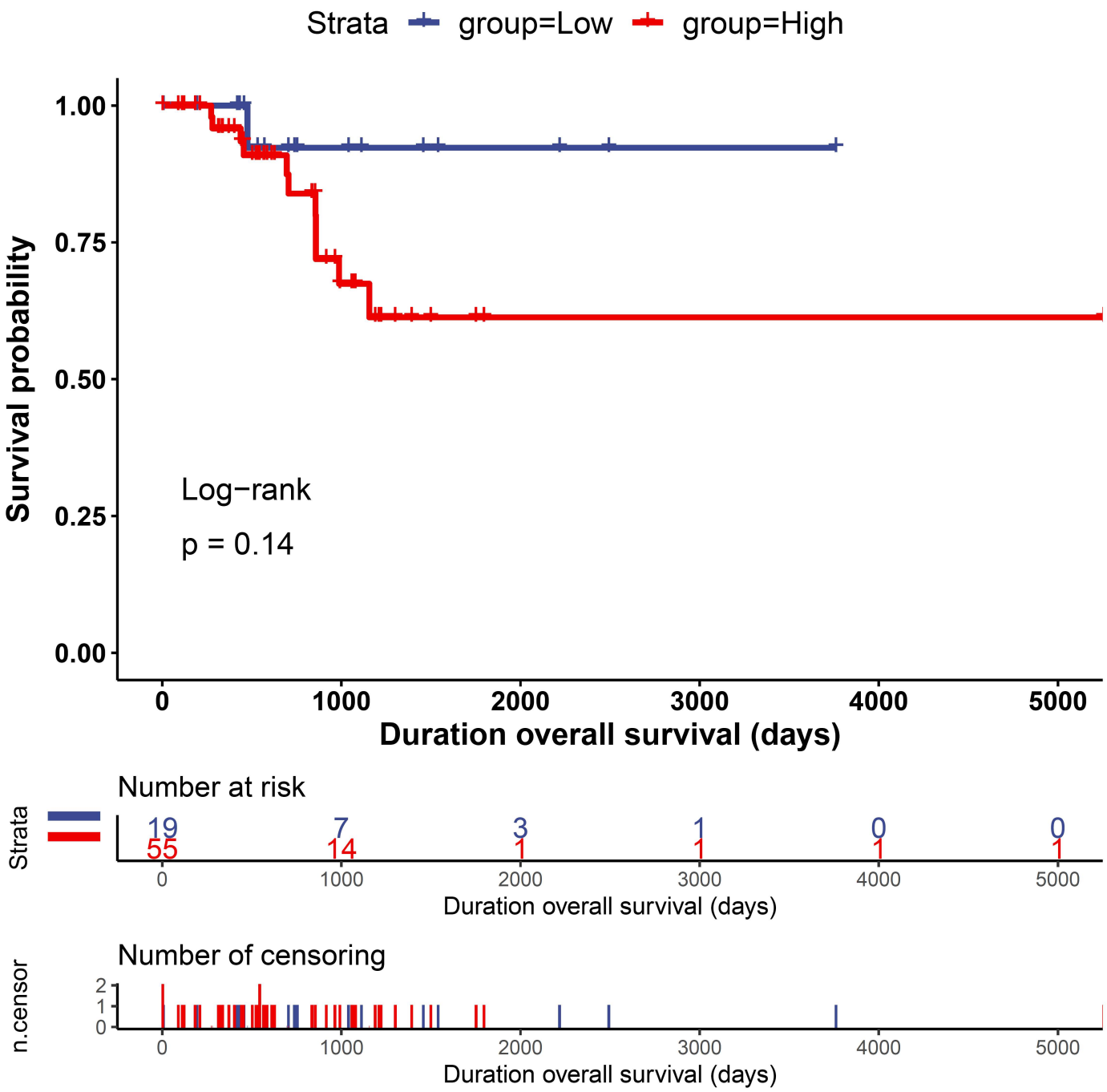

Strata
group=Low
group=High
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank
p = 0.00013
0
2000
4000
6000
Duration overall survival (days)
Number at risk
Strata
128
101
11
15
3
5
1
0
0
2000
4000
6000
Duration overall survival (days)
Number of censoring
n.censor
2
1
0
0
2000
4000
6000
Duration overall survival (days)

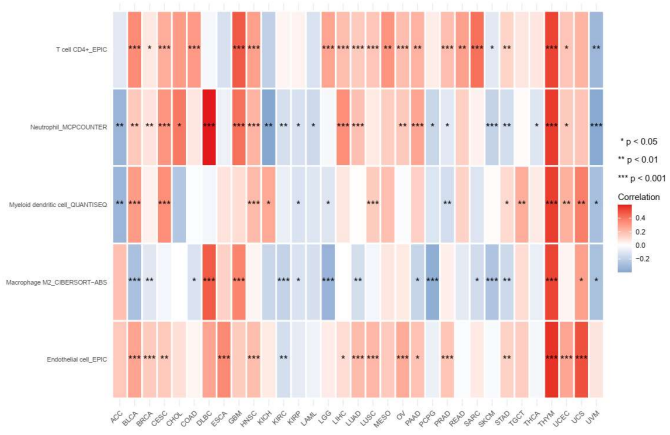

Strata
group=Low
group=High
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank
p < 0.0001
0
1000
2000
3000
4000
5000
Duration overall survival (days)
Number at risk
Strata
117
167
38
59
14
17
5
6
3
3
0
2
0
1000
2000
3000
4000
5000
Duration overall survival (days)
Number of censoring
n.censor
5
3
2
1
0
0
1000
2000
3000
4000
5000
Duration overall survival (days)

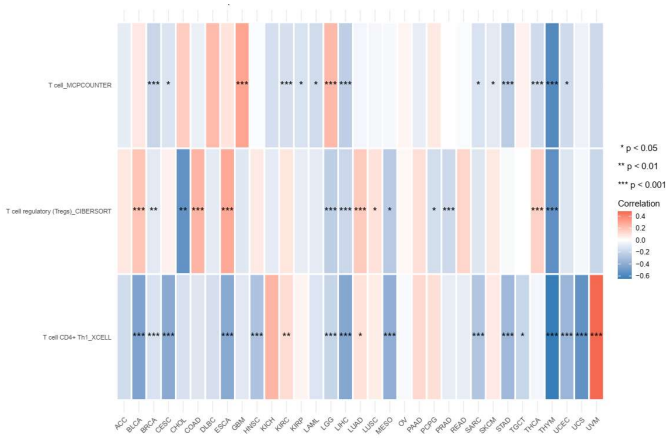

Strata
group=Low
group=High
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank
p = 0.0076
0
1000
2000
3000
4000
Duration overall survival (days)
Number at risk
Strata
168
64
31
17
8
5
2
2
2
0
0
1000
2000
3000
4000
Duration overall survival (days)
Number of censoring
n.censor
2
1
0
0
1000
2000
3000
4000
Duration overall survival (days)

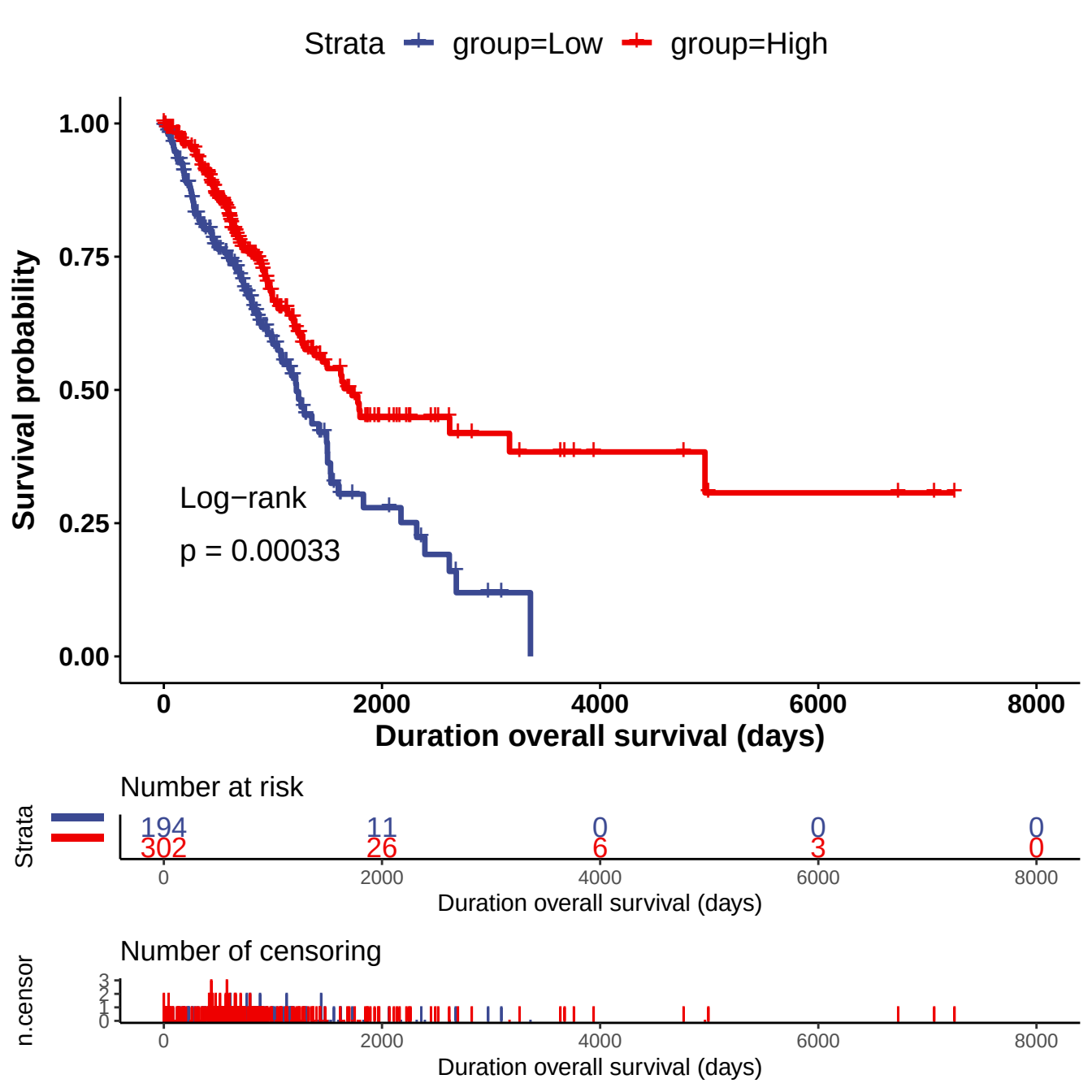

Strata
group=Low
group=High
Survival probability
1.00
0.75
0.50
0.25
0.00
Log−rank
p = 0.014
0
1000
2000
3000
4000
5000
Duration overall survival (days)
Number at risk
Strata
193
50
13
4
1
0
96
25
7
4
1
1
0
1000
2000
3000
4000
5000
Duration overall survival (days)
Number of censoring
n.censor
4
2
0
0
1000
2000
3000
4000
5000
Duration overall survival (days)

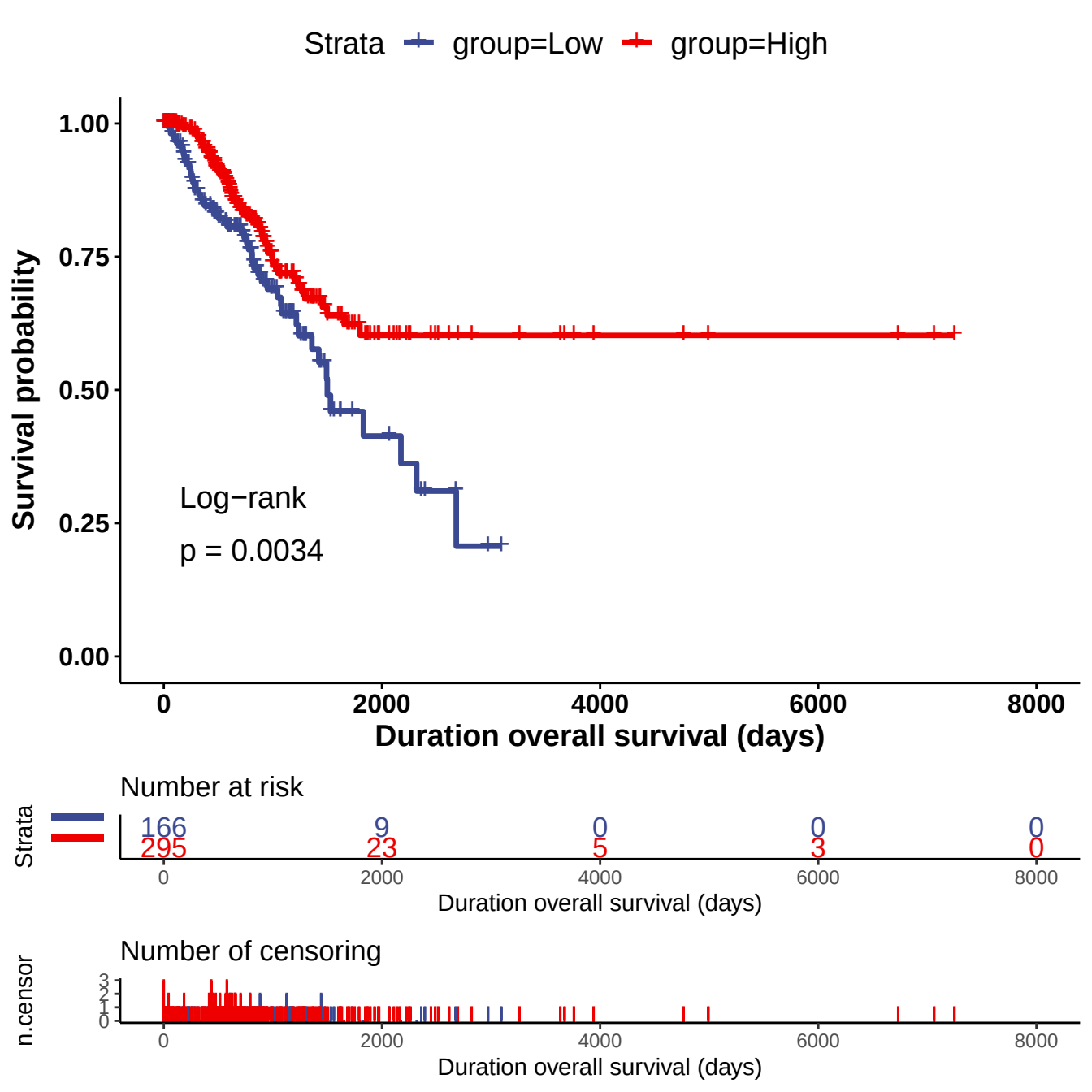

Strata
group=Low
group=High
Survival probability
1.00
0.75
0.50
0.25
0.00
Log−rank
p = 0.019
0
1000
2000
3000
4000
Duration overall survival (days)
Number at risk
Strata
27
35
5
7
1
1
0
1
0
0
Duration overall survival (days)
Number of censoring
n.censor
1
0
Duration overall survival (days)

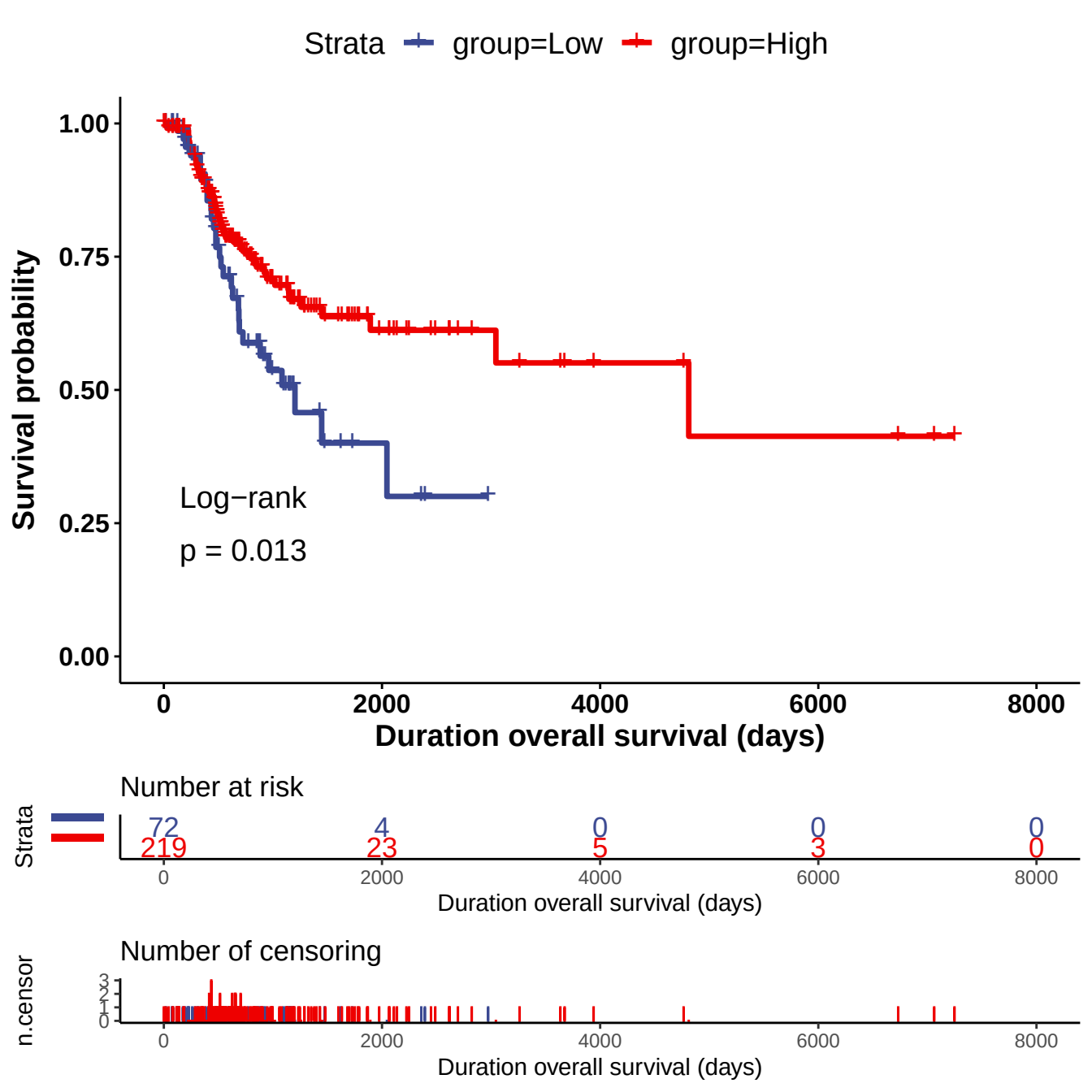

Strata
group=Low
group=High
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank
p = 0.14
0
1000
2000
3000
4000
5000
Duration overall survival (days)
Number at risk
Strata
19
55
7
14
3
1
1
1
0
1
0
1
0
1000
2000
3000
4000
5000
Duration overall survival (days)
Number of censoring
n.censor
2
1
0
0
1000
2000
3000
4000
5000
Duration overall survival (days)

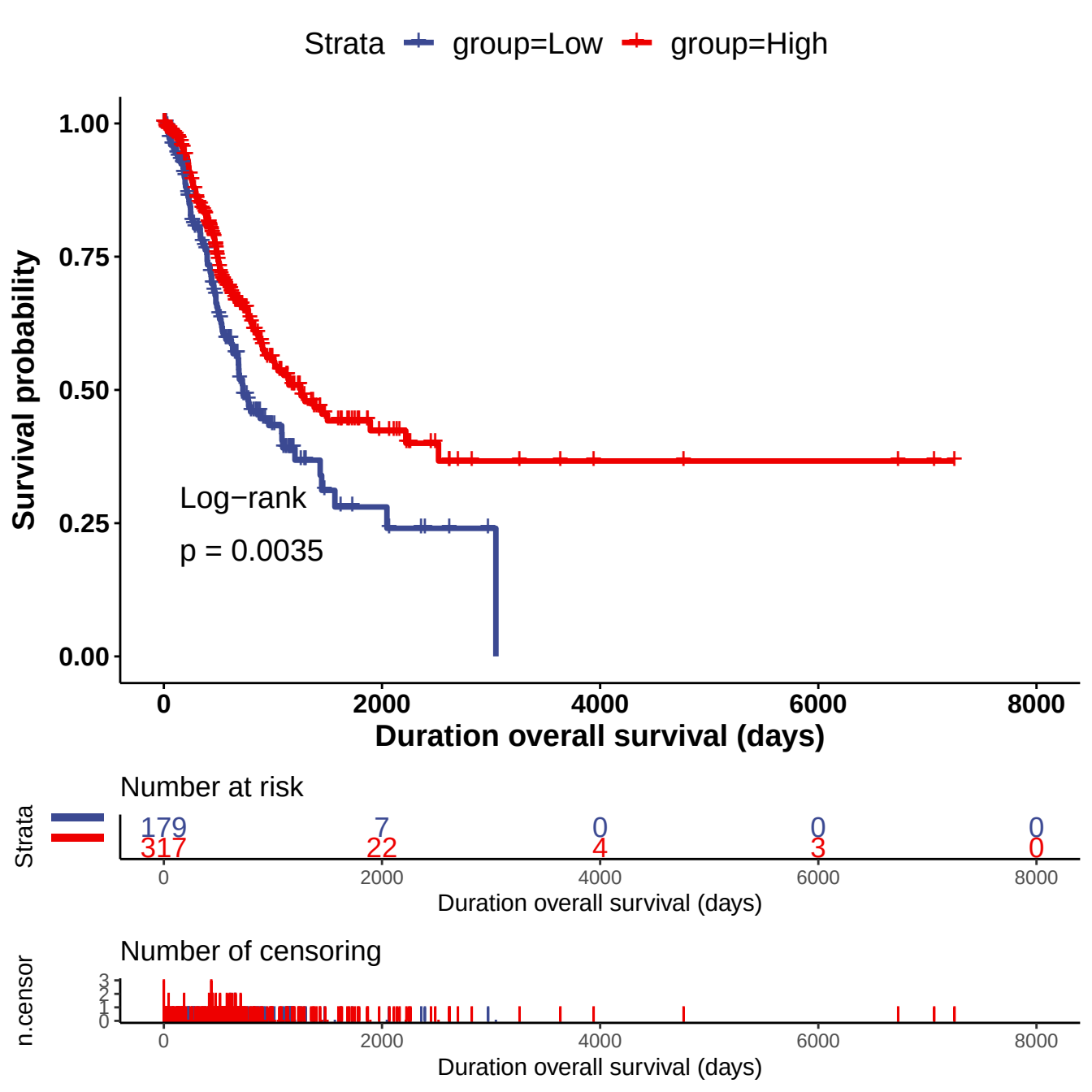

T cell CD4+_EPIC
Neutrophil_MCPCOUNTER
Myeloid dendritic cell_QUANTISEQ
Macrophage M2_CIBERSORT−ABS
Endothelial cell_EPIC
ACC
BLCA
BRCA
CESC
CHOL
COAD
DLBC
ESCA
GBM
HNSC
KICH
KIRC
KIRP
LAML
LGG
LIHC
LUAD
LUSC
MESO
OV
PAAD
PCPG
PRAD
READ
SARC
SKCM
STAD
TGCT
THCA
THYM
UCEC
UCS
UVM
* p < 0.05
** p < 0.01
*** p < 0.001
Correlation
0.4
0.2
0.0
−0.2

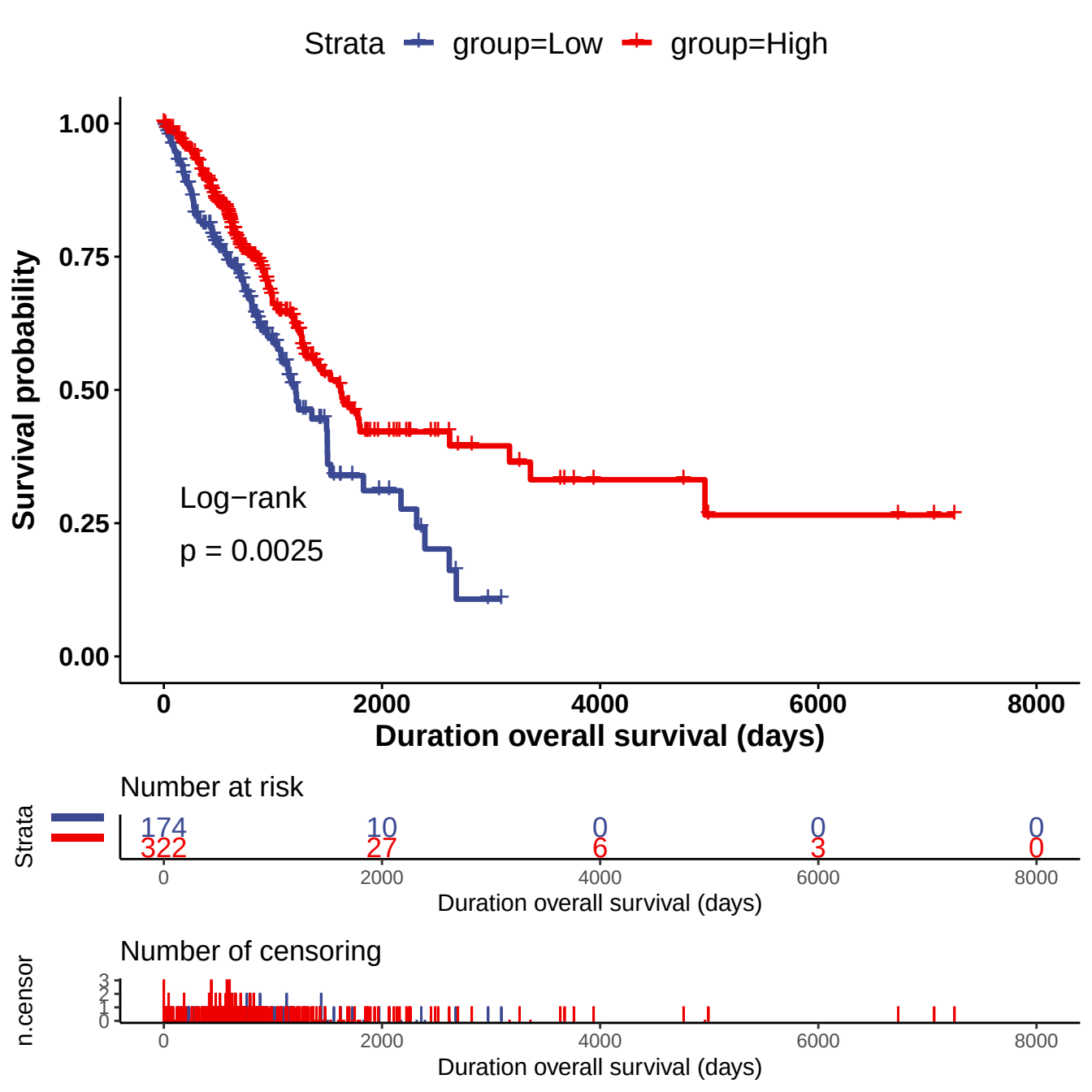

T cell_MCPCOUNTER
T cell regulatory (Tregs)_CIBERSORT
T cell CD4+ Th1_XCELL
ACC BLCA BRCA CESC CHOL COAD DLBC ESCA GBM HNSC KICH KIRC KIRP LAML LGG LIHC LUAD LUSC MESO OV PAAD PCPG PRAD READ SARC SKCM STAD TGCT THCA THYM UCEC UCS UVM
* p < 0.05
** p < 0.01
*** p < 0.001
Correlation
0.4
0.2
0.0
−0.2
−0.4
−0.6

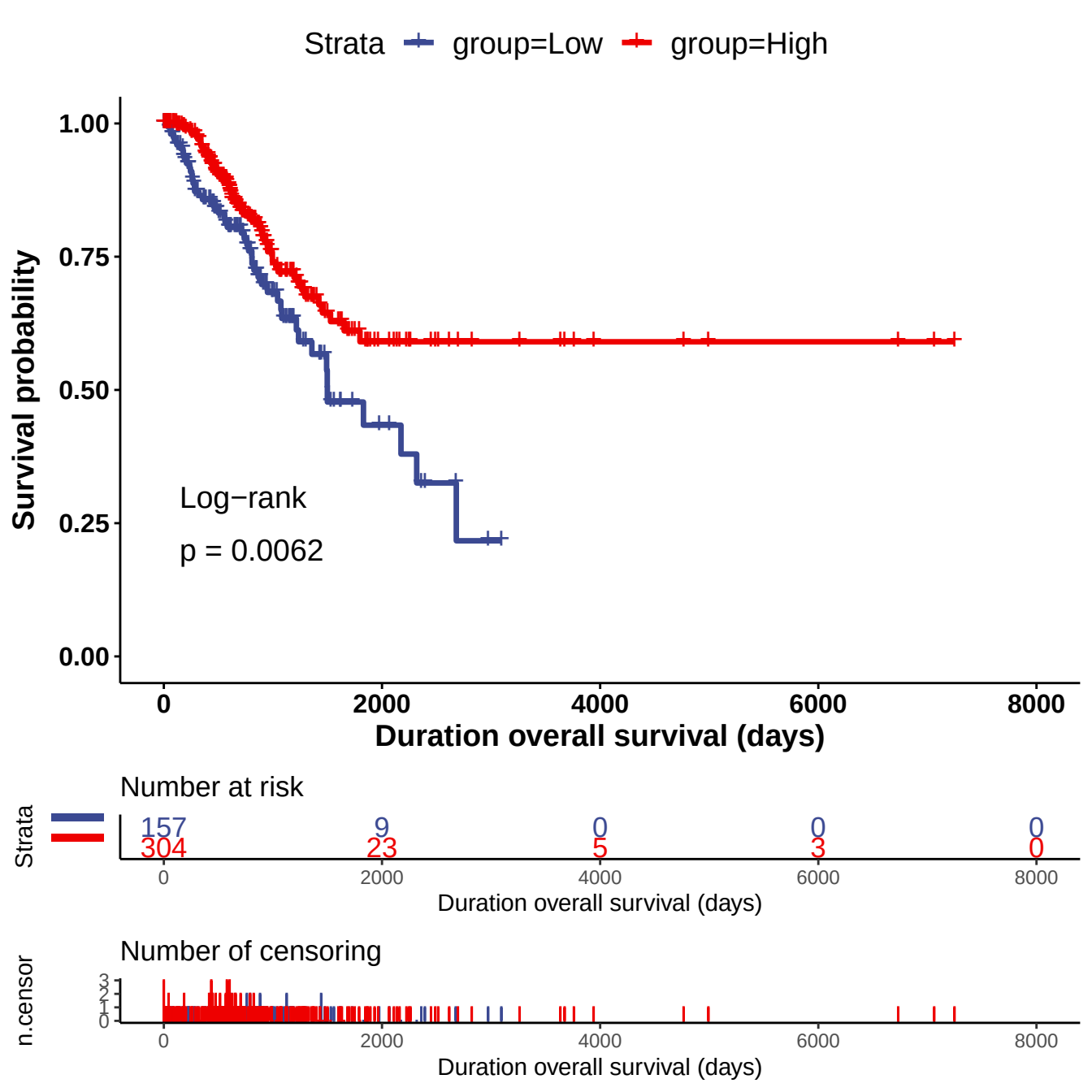

Strata
group=Low
group=High
Survival probability
1.00
0.75
0.50
0.25
0.00
Log−rank
p = 0.00033
0
2000
4000
6000
8000
Duration overall survival (days)
Number at risk
Strata
194
11
0
0
0
302
26
6
3
0
0
2000
4000
6000
8000
Duration overall survival (days)
Number of censoring
n.censor
3
2
1
0
0
2000
4000
6000
8000
Duration overall survival (days)

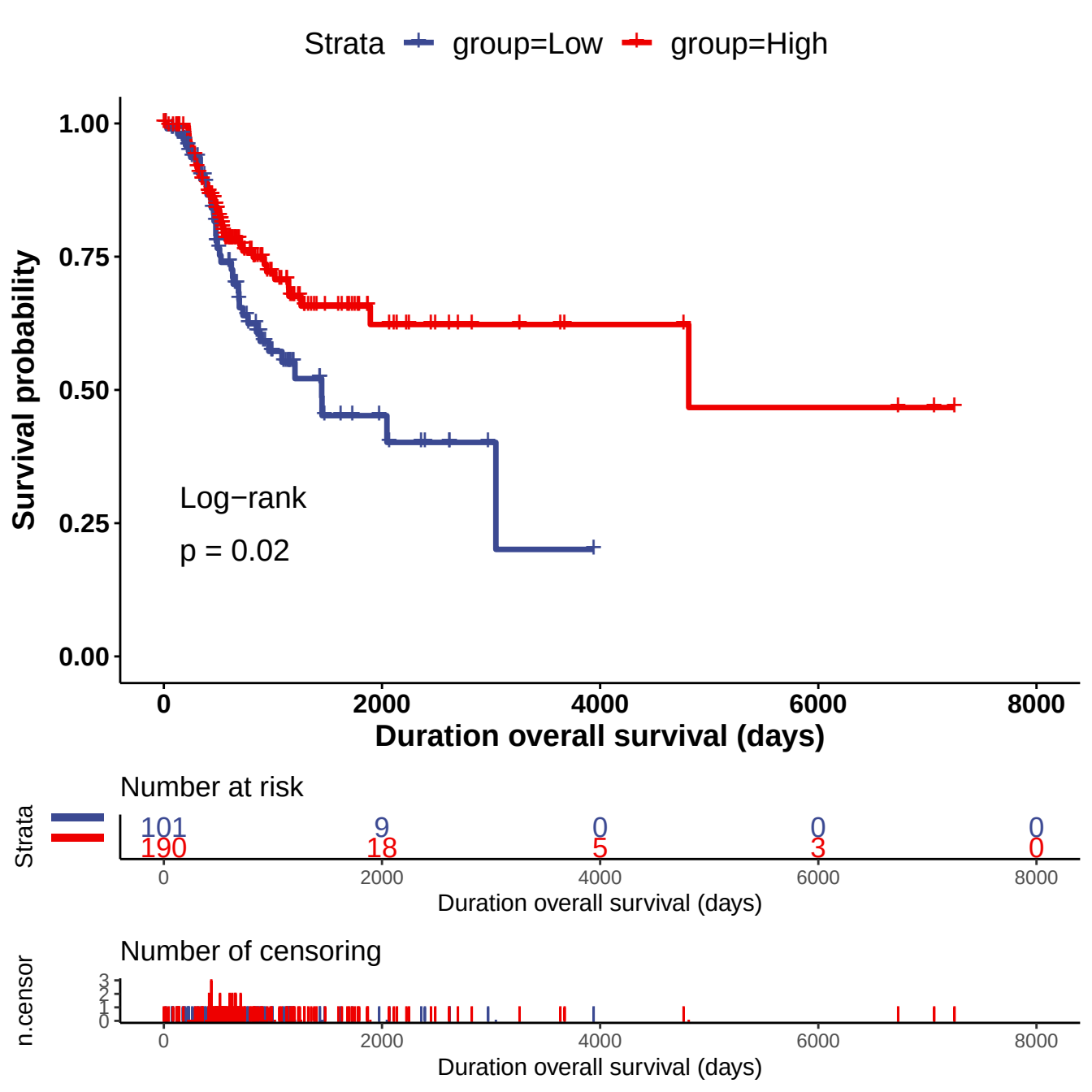

Strata
group=Low
group=High
Survival probability
1.00
0.75
0.50
0.25
0.00
Log−rank
p = 0.0034
0
2000
4000
6000
8000
Duration overall survival (days)
Number at risk
Strata
166
295
9
23
0
5
0
3
0
0
0
2000
4000
6000
8000
Duration overall survival (days)
Number of censoring
n.censor
3
2
1
0
0
2000
4000
6000
8000
Duration overall survival (days)

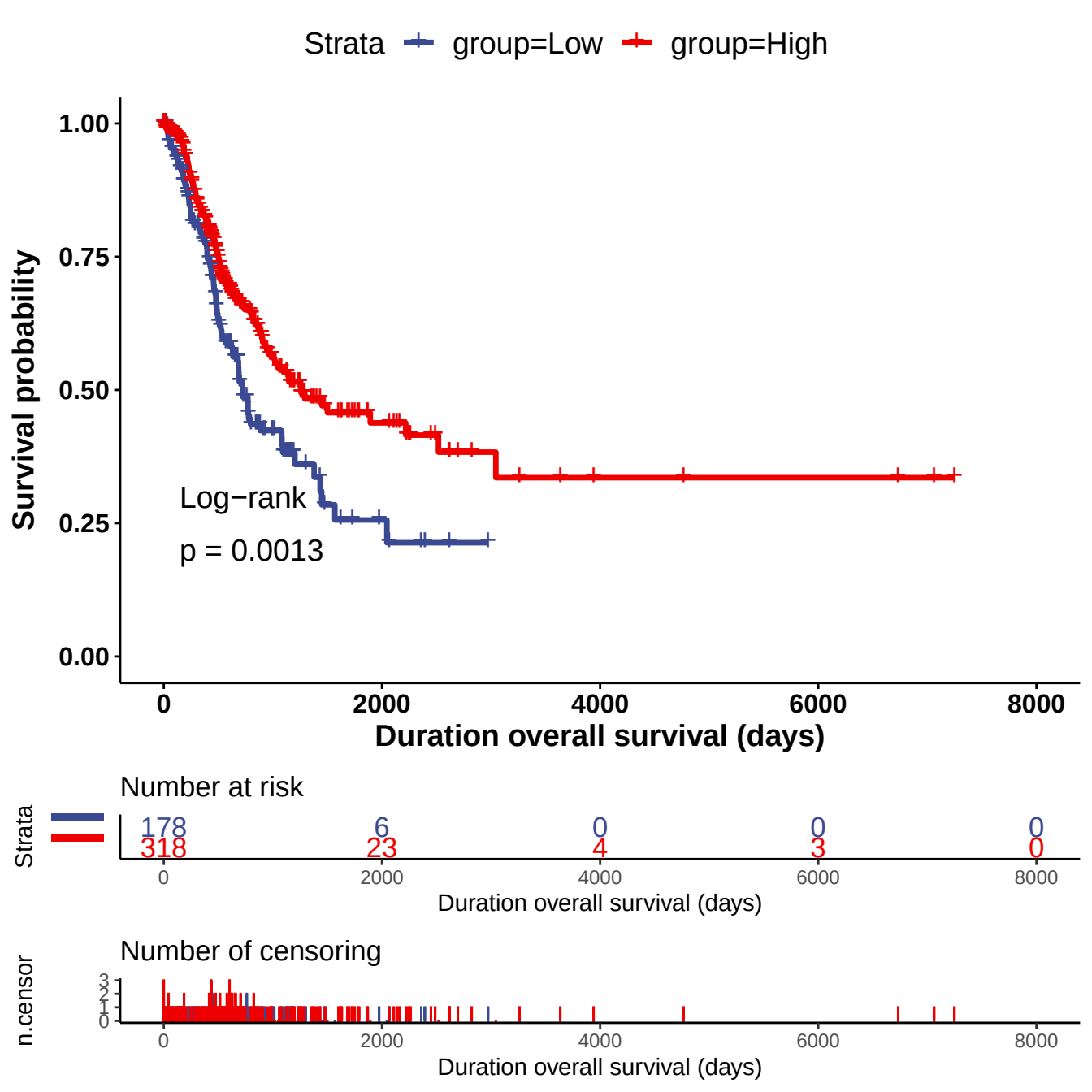

Strata
group=Low
group=High
Survival probability
1.00
0.75
0.50
0.25
0.00
Log−rank
p = 0.013
0
2000
4000
6000
8000
Duration overall survival (days)
Number at risk
Strata
72
219
4
23
0
5
0
3
0
0
0
2000
4000
6000
8000
Duration overall survival (days)
Number of censoring
n.censor
3
2
1
0
0
2000
4000
6000
8000
Duration overall survival (days)

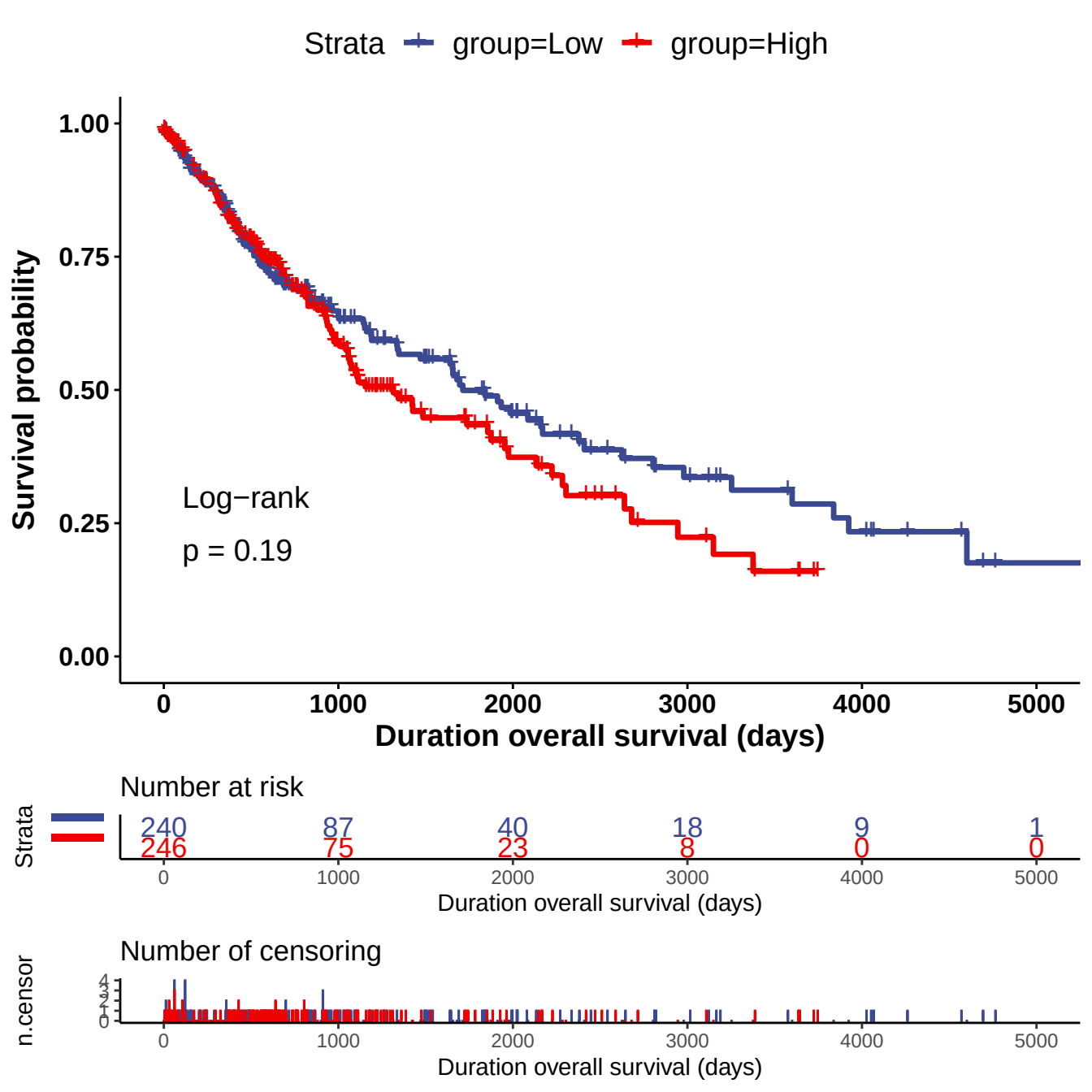

Strata
group=Low
group=High
Survival probability
1.00
0.75
0.50
0.25
0.00
Log−rank
p = 0.0035
0
2000
4000
6000
8000
Duration overall survival (days)
Number at risk
Strata
179
317
7
22
0
4
0
3
0
0
0
2000
4000
6000
8000
Duration overall survival (days)
Number of censoring
n.censor
3
2
1
0
0
2000
4000
6000
8000
Duration overall survival (days)

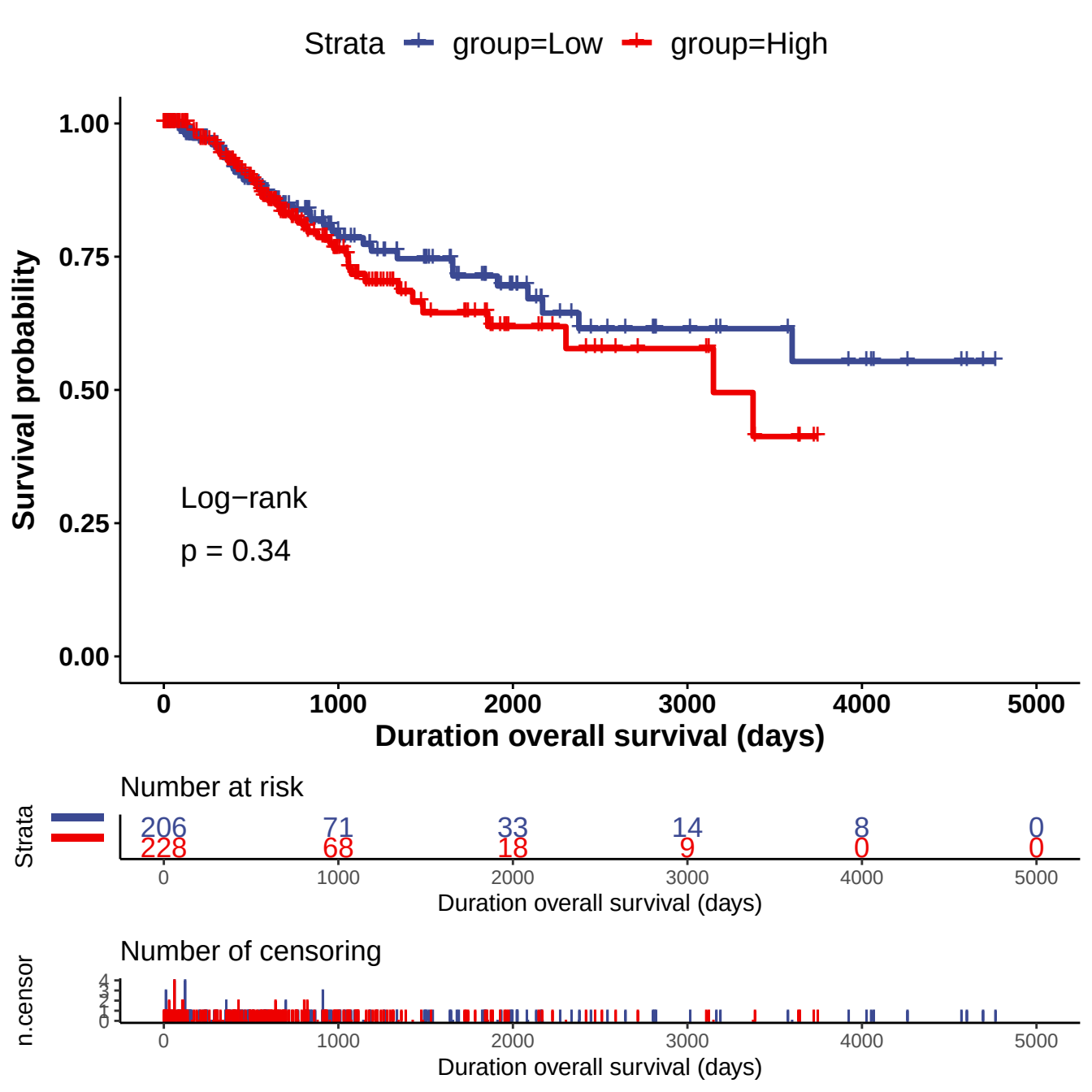

Strata
group=Low
group=High
Survival probability
1.00
0.75
0.50
0.25
0.00
Log−rank
p = 0.0025
0
2000
4000
6000
8000
Duration overall survival (days)
Number at risk
Strata
174
322
10
27
0
6
0
3
0
0
0
2000
4000
6000
8000
Duration overall survival (days)
Number of censoring
n.censor
3
2
1
0
0
2000
4000
6000
8000
Duration overall survival (days)

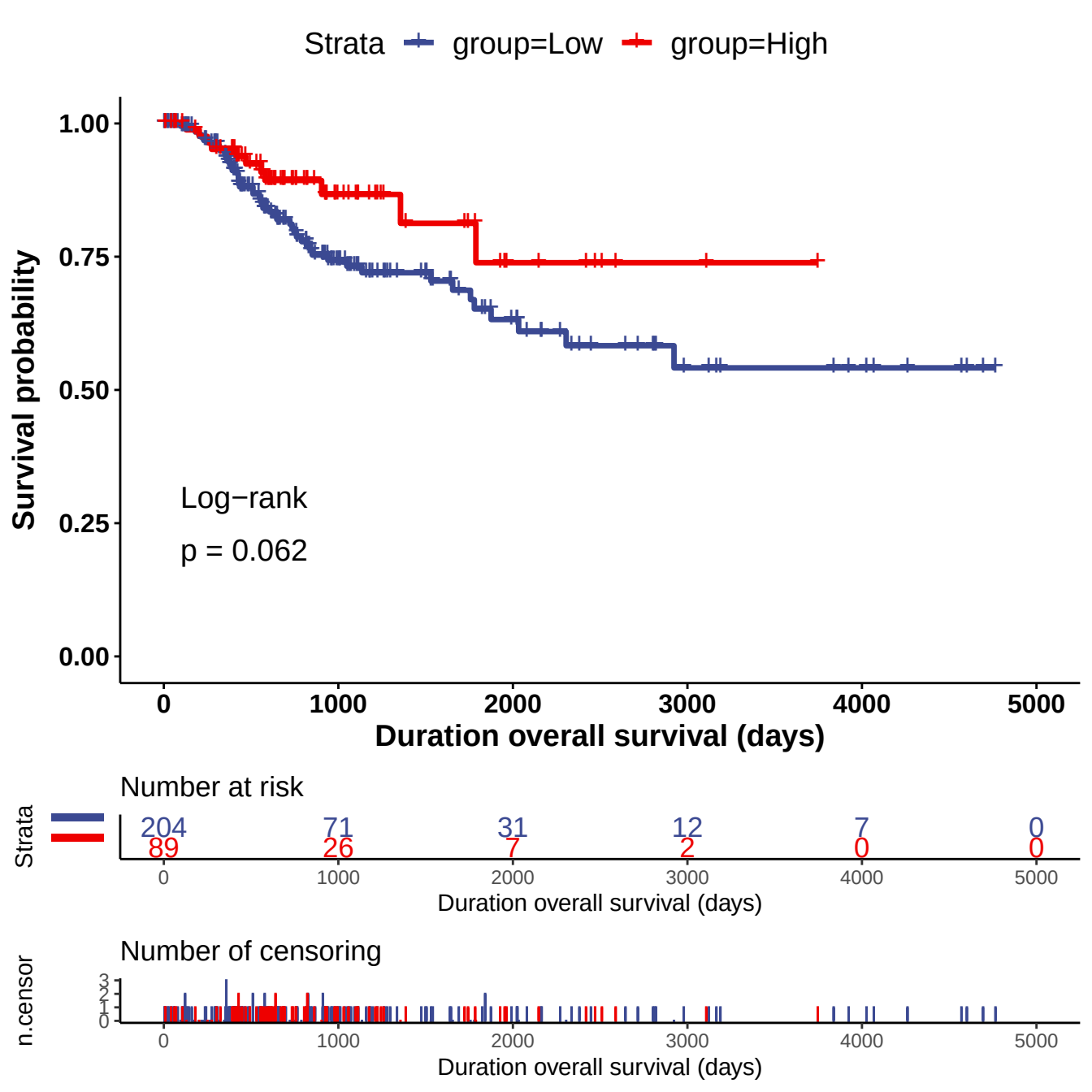

Strata
group=Low
group=High
Survival probability
1.00
0.75
0.50
0.25
0.00
Log−rank
p = 0.0062
0
2000
4000
6000
8000
Duration overall survival (days)
Number at risk
Strata
157
304
9
23
0
5
0
3
0
0
0
2000
4000
6000
8000
Duration overall survival (days)
Number of censoring
n.censor
3
2
1
0
0
2000
4000
6000
8000
Duration overall survival (days)

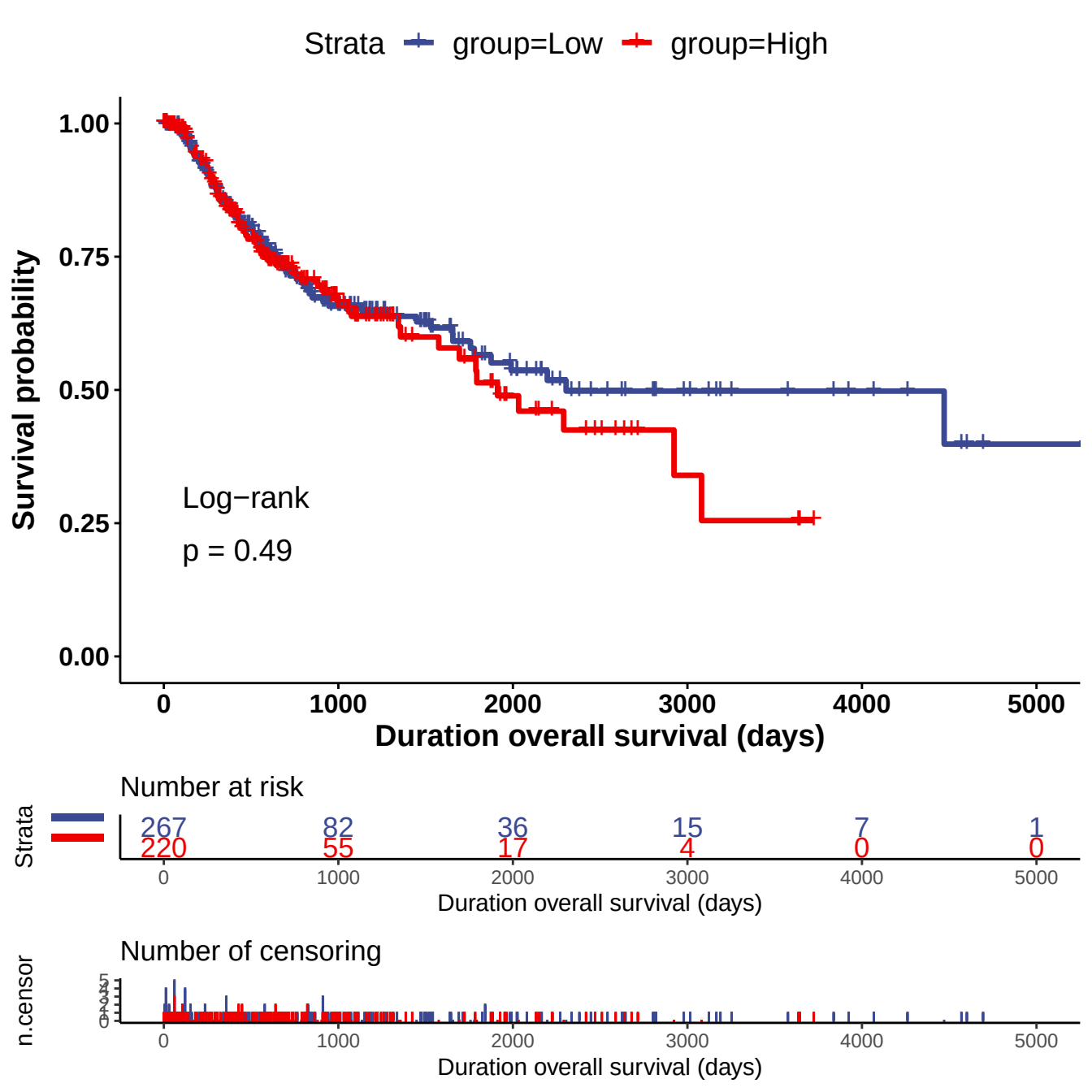

Strata
group=Low
group=High
Survival probability
1.00
0.75
0.50
0.25
0.00
Log−rank
p = 0.02
0
2000
4000
6000
8000
Duration overall survival (days)
Number at risk
Strata
101
190
9
18
0
5
0
3
0
0
0
2000
4000
6000
8000
Duration overall survival (days)
Number of censoring
n.censor
3
2
1
0
0
2000
4000
6000
8000
Duration overall survival (days)

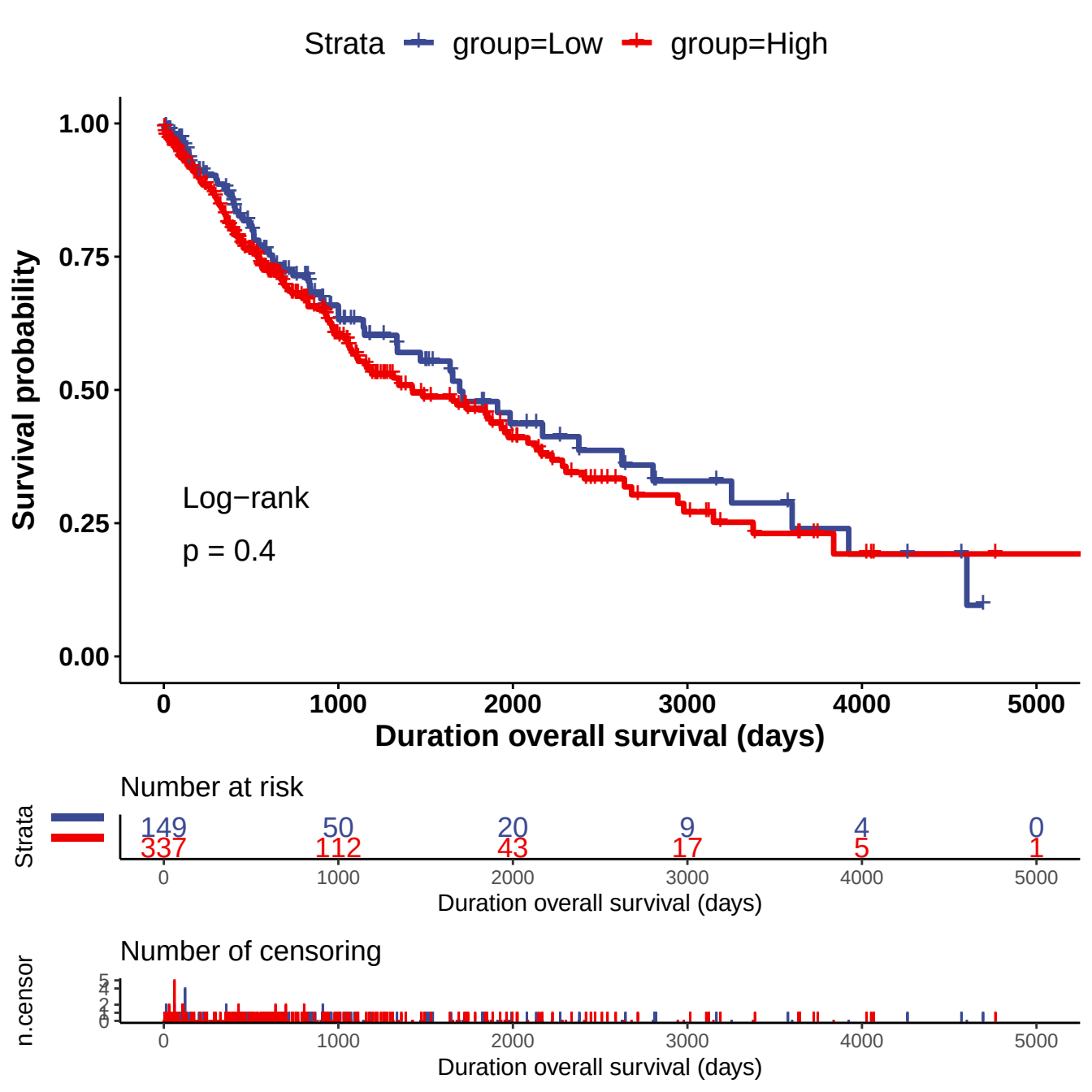

Strata
group=Low
group=High
Survival probability
1.00
0.75
0.50
0.25
0.00
Log−rank
p = 0.0013
0
2000
4000
6000
8000
Duration overall survival (days)
Number at risk
Strata
178
6
0
0
0
318
23
4
3
0
0
2000
4000
6000
8000
Duration overall survival (days)
Number of censoring
n.censor
3
2
1
0
0
2000
4000
6000
8000
Duration overall survival (days)

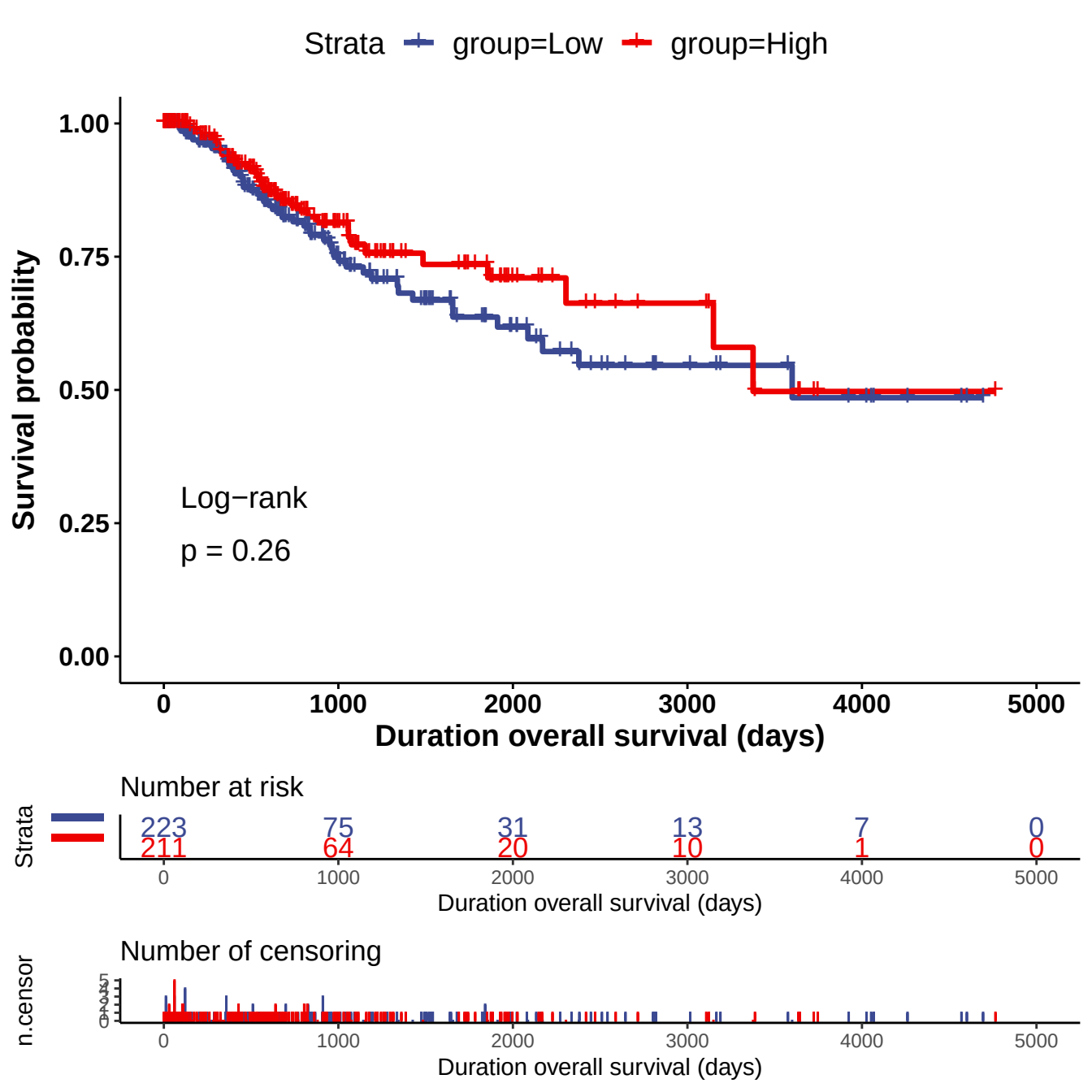

Strata
group=Low
group=High
Survival probability
1.00
0.75
0.50
0.25
0.00
Log−rank
p = 0.19
0
1000
2000
3000
4000
5000
Duration overall survival (days)
Number at risk
Strata
240
246
87
75
40
23
18
8
9
0
1
0
0
1000
2000
3000
4000
5000
Duration overall survival (days)
Number of censoring
n.censor
4
3
2
1
0
0
1000
2000
3000
4000
5000
Duration overall survival (days)

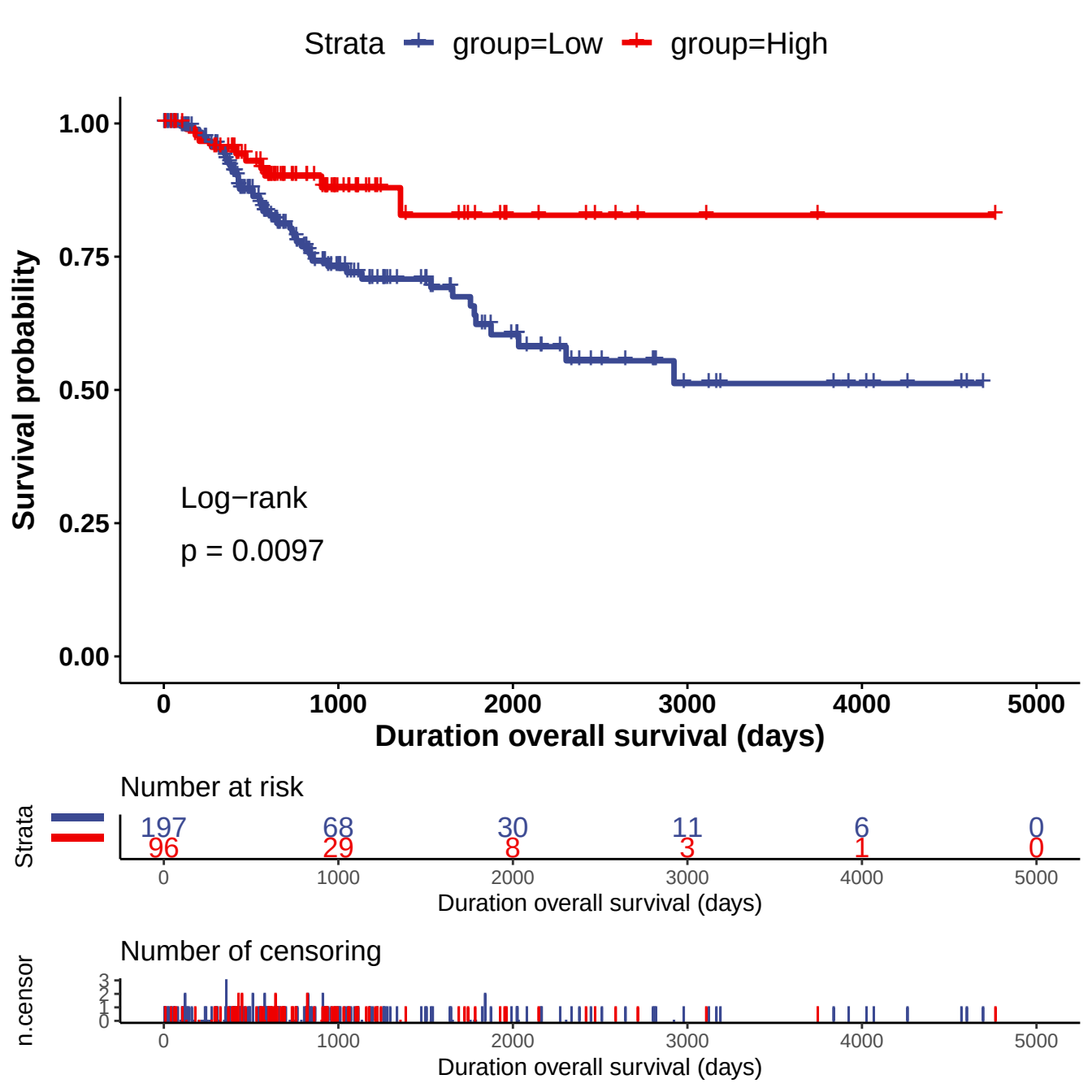

Strata
group=Low
group=High
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank
p = 0.34
0
1000
2000
3000
4000
5000
Duration overall survival (days)
Number at risk
Strata
206
71
33
14
8
0
228
68
18
9
0
0
0
1000
2000
3000
4000
5000
Duration overall survival (days)
Number of censoring
n.censor
4
3
2
1
0
0
1000
2000
3000
4000
5000
Duration overall survival (days)

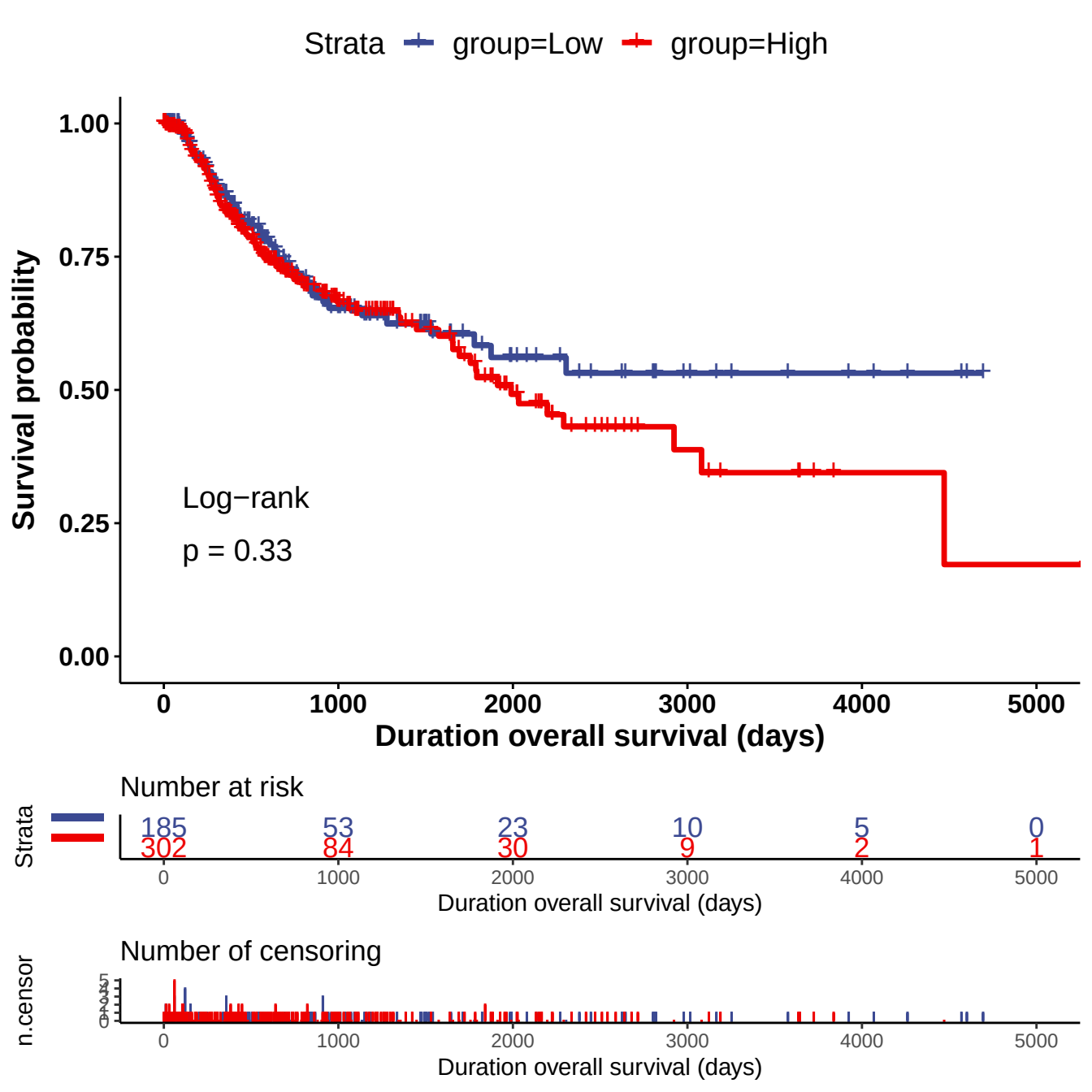

Strata
group=Low
group=High
Survival probability
1.00
0.75
0.50
0.25
0.00
Log−rank
p = 0.062
0
1000
2000
3000
4000
5000
Duration overall survival (days)
Number at risk
Strata
204
71
31
12
7
0
89
26
7
2
0
0
0
1000
2000
3000
4000
5000
Duration overall survival (days)
Number of censoring
n.censor
3
2
1
0
0
1000
2000
3000
4000
5000
Duration overall survival (days)

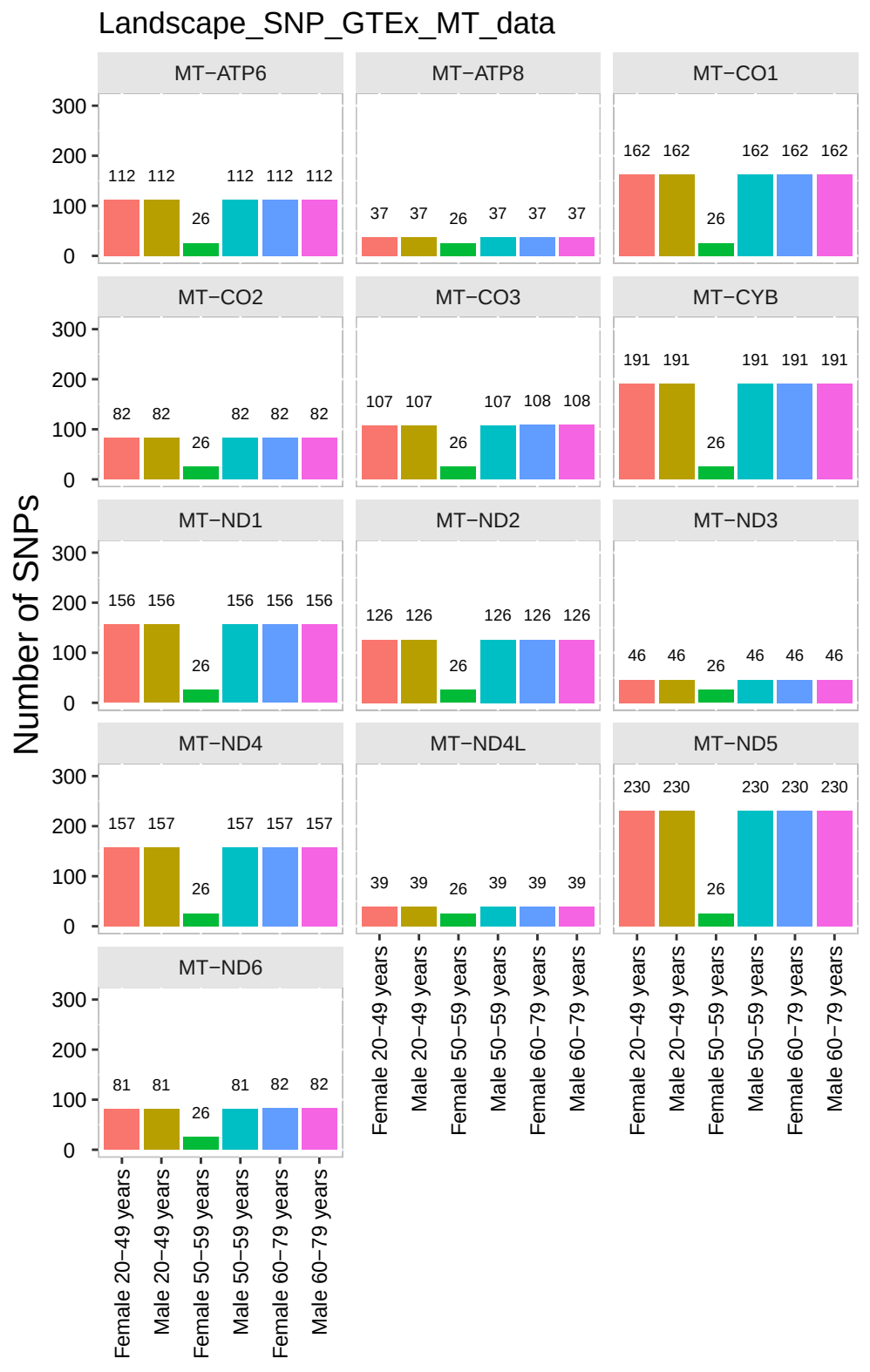

Strata
group=Low
group=High
Survival probability
1.00
0.75
0.50
0.25
0.00
Log−rank
p = 0.49
0
1000
2000
3000
4000
5000
Duration overall survival (days)
Number at risk
Strata
267
220
82
55
36
17
15
4
7
0
1
0
0
1000
2000
3000
4000
5000
Duration overall survival (days)
Number of censoring
n.censor
0
1000
2000
3000
4000
5000
Duration overall survival (days)

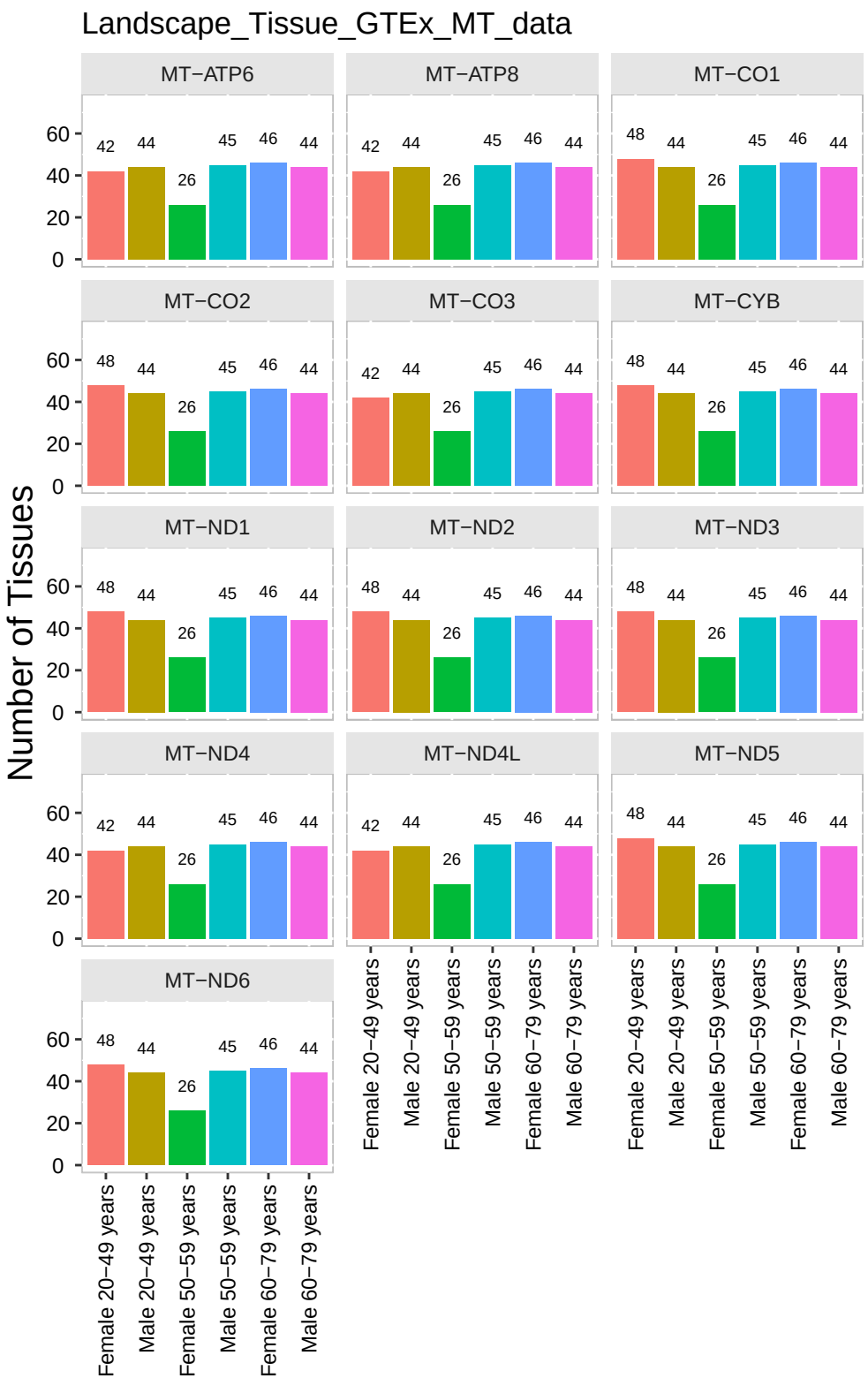

Strata
group=Low
group=High
Survival probability
1.00
0.75
0.50
0.25
0.00
Log−rank
p = 0.4
0
1000
2000
3000
4000
5000
Duration overall survival (days)
Number at risk
Strata
149
337
50
112
20
43
9
17
4
5
0
1
0
1000
2000
3000
4000
5000
Duration overall survival (days)
Number of censoring
n.censor
0
1
2
3
4
5
0
1000
2000
3000
4000
5000
Duration overall survival (days)

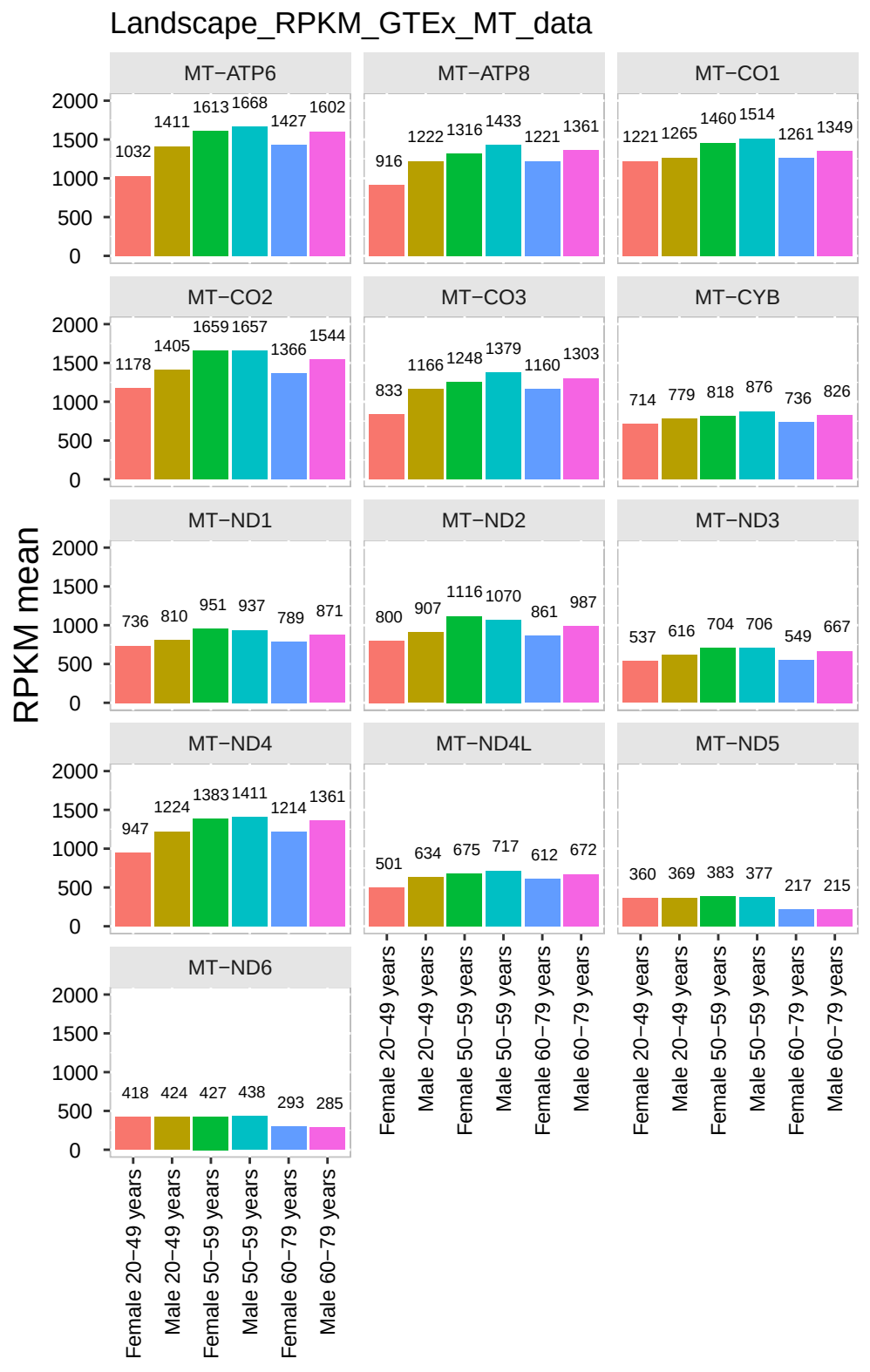

Strata
group=Low
group=High
Survival probability
1.00
0.75
0.50
0.25
0.00
Log−rank
p = 0.26
0
1000
2000
3000
4000
5000
Duration overall survival (days)
Number at risk
Strata
223
211
75
64
31
20
13
10
7
1
0
0
0
1000
2000
3000
4000
5000
Duration overall survival (days)
Number of censoring
n.censor
0
1
2
3
4
5
0
1000
2000
3000
4000
5000
Duration overall survival (days)

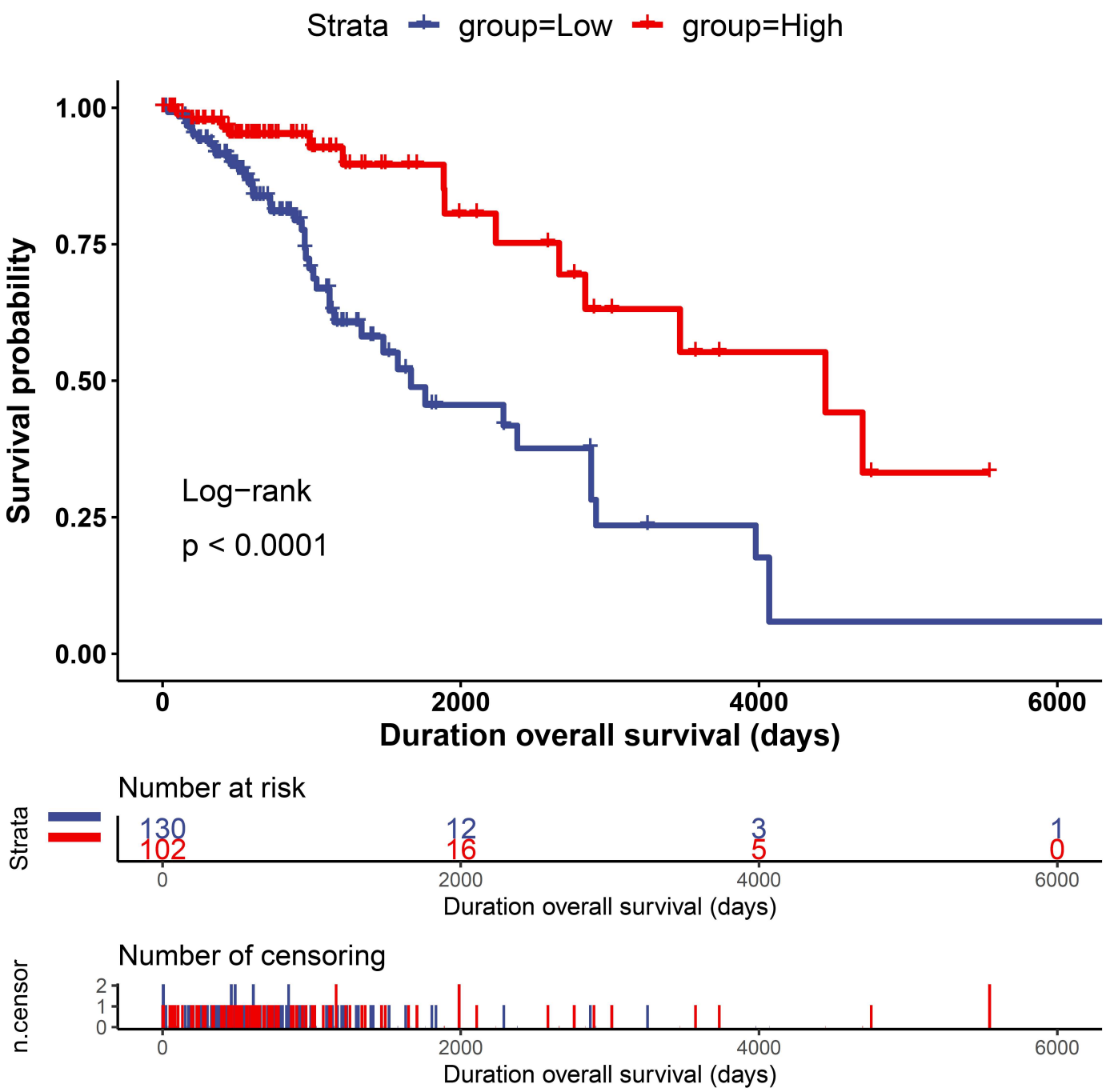

Strata
group=Low
group=High
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank
p = 0.0097
0
1000
2000
3000
4000
5000
Duration overall survival (days)
Number at risk
Strata
197
96
68
29
30
8
11
3
6
1
0
0
0
1000
2000
3000
4000
5000
Duration overall survival (days)
Number of censoring
n.censor
3
2
1
0
0
1000
2000
3000
4000
5000
Duration overall survival (days)


Strata
group=Low
group=High
Survival probability
1.00
0.75
0.50
0.25
0.00
Log−rank
p = 0.33
0
1000
2000
3000
4000
5000
Duration overall survival (days)
Number at risk
Strata
185
302
53
84
23
30
10
9
5
2
0
1
0
1000
2000
3000
4000
5000
Duration overall survival (days)
Number of censoring
n.censor
0
1
2
3
4
5
0
1000
2000
3000
4000
5000
Duration overall survival (days)


Landscape_SNP_GTEx_MT_data
MT−ATP6
MT−ATP8
MT−CO1
MT−CO2
MT−CO3
MT−CYB
MT−ND1
MT−ND2
MT−ND3
MT−ND4
MT−ND4L
MT−ND5
MT−ND6
Number of SNPs
0
100
200
300
112 112 26 112 112 112
37 37 26 37 37 37
162 162 26 162 162 162
82 82 26 82 82 82
107 107 26 107 108 108
191 191 26 191 191 191
156 156 26 156 156 156
126 126 26 126 126 126
46 46 26 46 46 46
157 157 26 157 157 157
39 39 26 39 39 39
230 230 26 230 230 230
81 81 26 81 82 82
Female 20−49 years
Male 20−49 years
Female 50−59 years
Male 50−59 years
Female 60−79 years
Male 60−79 years


Landscape_Tissue_GTEx_MT_data
MT−ATP6
MT−ATP8
MT−CO1
MT−CO2
MT−CO3
MT−CYB
MT−ND1
MT−ND2
MT−ND3
MT−ND4
MT−ND4L
MT−ND5
MT−ND6
Number of Tissues
0
20
40
60
42
44
26
45
46
44
48
Female 20−49 years
Male 20−49 years
Female 50−59 years
Male 50−59 years
Female 60−79 years
Male 60−79 years


Landscape_RPKM_GTEx_MT_data
RPKM mean
MT−ATP6
1032 1411 1613 1668 1427 1602
MT−ATP8
916 1222 1316 1433 1221 1361
MT−CO1
1221 1265 1460 1514 1261 1349
MT−CO2
1178 1405 1659 1657 1366 1544
MT−CO3
833 1166 1248 1379 1160 1303
MT−CYB
714 779 818 876 736 826
MT−ND1
736 810 951 937 789 871
MT−ND2
800 907 1116 1070 861 987
MT−ND3
537 616 704 706 549 667
MT−ND4
947 1224 1383 1411 1214 1361
MT−ND4L
501 634 675 717 612 672
MT−ND5
360 369 383 377 217 215
MT−ND6
418 424 427 438 293 285
0
500
1000
1500
2000
Female 20−49 years
Male 20−49 years
Female 50−59 years
Male 50−59 years
Female 60−79 years
Male 60−79 years


Strata
group=Low
group=High
1.00
0.75
0.50
0.25
0.00
Survival probability
Log−rank
p < 0.0001
0
2000
4000
6000
Duration overall survival (days)
Number at risk
Strata
130
102
12
16
3
5
1
0
0
2000
4000
6000
Duration overall survival (days)
Number of censoring
n.censor
2
1
0
0
2000
4000
6000
Duration overall survival (days)