medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Molecular Feature-Based Classification of Retroperitoneal Liposarcoma: A Prospective Cohort Study

Mengmeng Xiao[1,3†], Xiangji Li[2,3†], Fanqin Bu[2†], Shixiang Ma[3], Xiaohan Yang[2], Jun Chen[3], Yu Zhao[2], Ferdinando Cananzi[4], Chenghua Luo[1,3*], Li Min[2*]

[1]*Department of General Surgery, Peking University People's Hospital, 100044 Beijing, P. R. China*

[2]*Department of Gastroenterology, Beijing Friendship Hospital, Capital Medical University, State Key Laboratory for Digestive Health, National Clinical Research Center for Digestive Disease, Beijing Digestive Disease Center, Beijing Key Laboratory for Precancerous Lesion of Digestive Disease, 100050 Beijing, P. R. China*

[3]*Department of Retroperitoneal Tumor Surgery, Peking University International Hospital, 102206 Beijing, P. R. China*

[4]*Department of Biomedical Sciences, Humanitas University, 20089, Milan, Italy*

[†]***These authors contributed equally to this work***

Running title: Molecular feature-based classification of RPLS

Corresponding author:

Prof. Li Min, PhD, Email: minli@ccmu.edu.cn

Prof. Chenghua Luo, MD, PhD, Email: pkuihlch@163.com

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

ABSTRACT

Background: Retroperitoneal liposarcoma (RPLS) is a critical malignant disease with various clinical outcomes. However, the molecular heterogeneity of RPLS was poorly elucidated, and few biomarkers were proposed to monitor its progression.

Methods: RNA sequencing was performed on a training cohort of 88 RPLS patients to identify dysregulated genes and pathways using clusterprofiler. The GSVA algorithm was utilized to assess signaling pathways levels in each sample, and unsupervised clustering was employed to distinguish RPLS subtypes. Differentially expressed genes (DEGs) between RPLS subtypes were identified to construct a simplified dichotomous clustering via nonnegative matrix factorization. The feasibility of this classification was validated in a separate validation cohort (n=241) using immunohistochemistry (IHC) from the Retroperitoneal SArcoma Registry (RESAR). The study is registered with ClinicalTrials.gov under number NCT03838718.

Results: Cell cycle, DNA damage & repair, and Metabolism were identified as the most aberrant biological processes in RPLS, enabling the division of RPLS patients into two distinct subtypes with unique molecular signatures, tumor microenvironment, clinical features and outcomes (overall survival, OS and disease-free survival, DFS). A simplified RPLS classification based on representative biomarkers (LEP and PTTG1) demonstrated high accuracy (AUC>0.99), with patients classified as LEP+ and PTTG1- showing lower aggressive pathological composition ratio and fewer surgery times, along with better OS (HR=0.41, $P<0.001$) and DFS (HR=0.60, $P=0.005$).

Conclusions: Our study provided an ever-largest gene expression landscape of RPLS and established an IHC-based molecular classification that was clinically relevant and cost-effective for guiding treatment decisions.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Keywords: LEP; PTTG1; Retroperitoneal liposarcoma (RPLS); Molecular classification

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

INTRODUCTION

Retroperitoneal liposarcoma (RPLS) is a soft tissue sarcoma (STS) originating in the retroperitoneum with an insidious onset. Traditional surgical resection has been regarded as a primary and curable treatment strategy of RPLS for the past fifty years (***Ecker et al., 2016***). However, the anatomical complexity and biological properties of sarcoma brought great difficulty to achieve microscopically margin-negative resection, leading to a high postoperative recurrence rate in RPLS patients. During the past decade, scientists tried to improve the postoperative survival of RPLS patients by personalized surgical resection and neoadjuvant/adjuvant therapies, but the effect was not satisfactory (***Littau et al., 2020; Gronchi et al., 2015; Gronchi et al., 2009; Gronchi et al., 2013; Pisters et al., 2009***).

Recently, precision medicine greatly enriched the therapeutic approaches and reformed the clinical decision-making chain of tumor diagnosis and treatment, prolonging the median survival of main tumor types 2-10 times (***Kam et al., 2021; Zeng et al., 2022; Alifrangis et al., 2019; Frese et al., 2021***). Biomarker-based patient stratification and targeted therapy together make up the kernel of precision medicine, which is intrinsically based on the molecular profiling of cancers. However, our knowledge of the molecular features of RPLS is limited, and few clinically applicable molecular biomarkers and targeted drugs are available for RPLS treatment. Only sporadic molecules such as CDK4 (***Pilotti et al., 2000***), MDM2 (***Binh et al., 2005***), AURK4 (***Yen et al., 2019***) and CCNDBP1 (***Yang et al., 2021***) have been reported as prognostic and diagnostic biomarkers, but these biomarkers were poorly represented and verified. Therefore, it is crucial to reveal the molecular landscape of RPLS and explore a feasible classification for its diagnosis and treatment.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Here, we conducted a comprehensive inbestigation into the molecular characteristics of RPLS through the delineation of the largest gene expression landscape ever assembled for this rare disease entity. By identifying both RPLS-specific genes and prognostic biomarkers, we unveiled their intricate relationships with clinical parameters. Our findings revealed the existence of two distinct molecular subtypes within all RPLS patients, characterized by diverse pathological compositions, enriched signaling pathways, and varying clinical outcomes. This highlights the limitations of relying solely on traditional pathological classification for surgical decision-making in certain cases where patients exhibit favorable histological features but poor prognoses. Emphasizing the pivotal role of molecular subtyping in guiding individualized treatment strategies and enhancing patient management. To facilitate practical application in clinical settings, we developed a simplified RPLS classification system based on key biomarkers (LEP and PTTG1) representative of each subtype. Notably, this classification scheme was validated in a larger cohort of RPLS patients through immunohistochemistry assays (Figure 1), laying the groundwork for precise surgical interventions guided by molecular insights in the realm of RPLS treatment.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

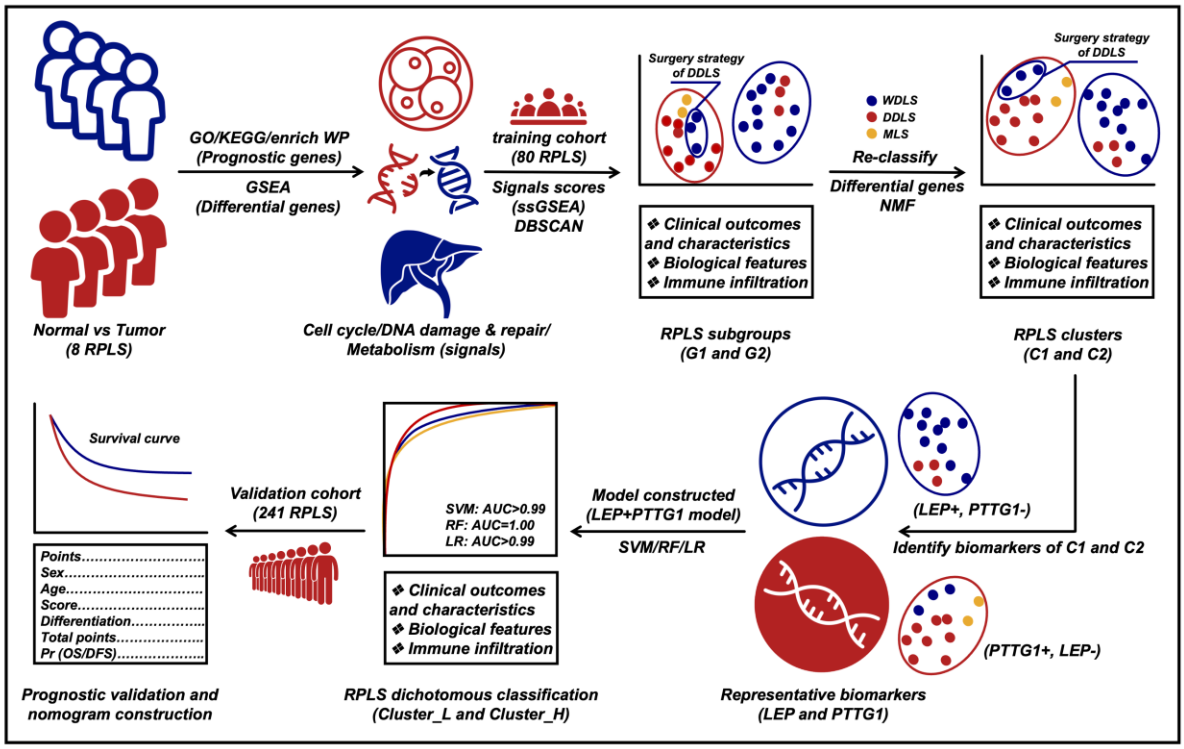


Figure 1. Flow diagram of exploring RPLS dichotomous classification

MATERIALS AND METHODS

Patients and tissue specimens

Patients who diagnosed with RPLS amenable to surgical resection were eligible for the study. The RPLS histology was confirmed according to the WHO criteria done on biopsy or surgical specimen by dedicated sarcoma pathologist. The exclusion criteria included the age<18 years; serious psychiatric disease that precludes informed consent or limits compliance; impossibility to ensure adequate follow-up. Tumor specimens from 88 RPLS patients (Training cohort 1, Table S1; Training cohort 2, Table S2) and another cohort of 241 RPLS patients (Validation cohort, Table S3) were obtained from our local Hospital. These cohorts are sourced from Retroperitoneal SArcoma Registry (RESAR, NCT03838718). All the patients underwent

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

curative resection from January 2015 to May 2019. RPLS tissue specimens were snap-frozen in liquid nitrogen within 1 h and then stored in a -80 ℃ refrigerator before use. Clinical information was collected from the medical records, and no patient had undergone previous chemotherapy or radiotherapy. Overall survival (OS) was defined as the interval between the latest surgery and death from tumors or between the latest surgery and the last observation taken for surviving patients. Disease-free survival (DFS) was defined as the interval between the latest surgery and diagnosis of relapse or death. Informed consent for surgical procedures and specimen collection were obtained from each patient. This study has been reported in line with the REMARK criteria (***McShane et al., 2005***).

RNA sequencing, primary data processing, and analysis

Total RNA was extracted from Training cohort 1 (Table S4) and Training cohort 2 (Table S2) using TRIzol Reagent (Invitrogen). RNA degradation and contamination were monitored with 1% agarose gel. RNA purity was checked by the NanoPhotometer spectrophotometer (IMPLEN, Los Angeles, CA, USA). RNA concentration was measured using the Qubit RNA Assay Kit with the Qubit 2.0 Fluorometer (Life Technologies, CA, USA). RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 System (Agilent Technologies, CA, USA).

A total amount of 3-5 ug RNA per sample was used as input material for the RNA library. Sequencing libraries were generated using NEBNext® Multiplex Small RNA Library Prep Set for Illumina® (NEB, USA) following the manufacturer's recommendations and index codes were added to attribute sequences to each sample. The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq SR Cluster Kit v3-cBot-HS (Illumia). After cluster generation, the strand-specific cDNA were sequenced on an Illumina

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

NovaSeq 6000 platform, and single-end reads were generated (Novogene Bioinformatic Technology, Beijing, China).

FPKMs of mRNAs and non-coding RNAs in each sample were calculated by Cuffdiff (v2.1.1). FPKMs were calculated based on the length of the fragments and read counts mapped to this fragment. These sequencing data have been deposited at the Open Archive for Miscellaneous Data (OMIX) database of China National Center for Bioinformation (CNCB) under the accession number OMIX002786.

Identification of differential genes

Gene difference analysis was performed to determine the differential genes (DEGs). An adjusted *FDR*<0.05 and |log2FC|>0.585 was considered significant. This process was conducted with the R package "limma".

Identification of prognostic genes

Cox univariate regression analysis was used to screen the prognostic genes of RPLS. Results of *P*<0.05 was considered significant. This process was conducted with the R package "survival".

Gene set enrichment analysis (GSEA) and immune infiltrate analysis

GSEA was performed in the tumor and normal groups to explore the biological signaling pathways. Pathway annotation files were downloaded from the msigdb (www.gsea-msigdb.org) platform. This process was conducted by the GSEA R package to elucidate the representative HALLMARK and REACTOME pathways enriched in RPLS patients. Immunocyte infiltration (immune score and stromal score) was measured by the Estimation of STromal and Immune cells in Malignant Tumor tissues using Expression data (ESTIMATE) algorithm. This process was completed via the "estimate" R package.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Functional annotation

Functional enrichment analyses were performed to elucidate the possible biological processes and signaling pathways of the prognostic genes. Gene ontology (GO) and Kyoto Encyclopedia of Gene and Genomes (KEGG) analyses were conducted by R package "clusterprofiler", and the false discovery rate<0.05 was considered significantly enriched.

Consensus clustering with t-distributed stochastic neighbor embedding (t-SNE)

After evaluated the relative abundance level of related pathways, the Euclidean distance was calculated between any two samples and condensed into two-dimensional points using t-distributed stochastic neighbor embedding (t-SNE) (***Guo et al., 2019***) and subsequently visualized automatically with the density-based spatial clustering of applications with noise (DBSCAN) algorithm. This consensus clustering was conducted with the R packages "Rtsne" and "dbscan".

Consensus clustering with nonnegative matrix factorization (NMF)

Nonnegative matrix factorization (NMF) was used to perform RPLS subtyping. Specifically, NMF was applied to gene expression matrix *A* which contained gene sets of major signaling pathways and prognostic genes. Matrix *A* was factorized into 2 nonnegative matrices *W* and *H*. Repeated factorization of matrix *A* was performed and its outputs were aggregated to obtain consensus clustering of RPLS samples. The optimal number of subtypes was selected according to cophenetic, dispersion, and silhouette coefficients. This consensus clustering was conducted with the R package "NMF".

Construction of machine learning models

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Machine learning models based on biomarkers were constructed by logistic regression (LR), support vector machine (SVM), and random forest (RF). These models were specifically tailored to analyze biomarker data in order to predict clinical outcomes in surgical patients.

LR is a statistical method that establishes a relationship between asset of independent variables and a binary outcome. It calculates the probability of an event occurring based on the input features derived from biomarkers relevant to surgical patient. SVM is a supervised learning algorithm that categorizes data points by identifying the optimal hyperplane that separates distinct classes within a high-dimensional space. This approach effectively maps biomarker data into a multidimensional space to facilitate accurate classification of patient outcomes. RF is an ensemble learning technique that generated multiple decision trees during training and aggregates the results to make predictions. By leveraging this method, we can enhance predictive accuracy by mitigating overfitting and increasing model robustness when analyzing biomarker-driven patient data.

The performance of these machine learning models was assessed using the area under the curve (AUC) metric. A higher AUC value indicates superior discriminatory power of model in distinguishing different clinical outcomes. An AUC value closer to 1.0 signifies strong predictive capability, while 0.5 indicates no discriminatory ability at all. By evaluating the AUC values generated by LR, SVM, and RF models, clinicians can identify which algorithm yields the most reliable predictions based on biomarker profiles for surgical patients.

Immunohistochemistry

The protocol was performed as previously described (***Li et al., 2022***). In brief, the LEP and PTTG1 antibodies for immunohistochemistry were purchased from Proteintech (Cat No: bs-0409R and bs-1881R). With deparaffinization for 15min × 3 in dimethylbenzene and routine

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

hydration, the tissues were soaked in phosphate buffer saline (PBS) for 10min and then performed high-pressure antigen retrieval (Tris-EDTA, PH=9.0) for 2.5min. After being treated with a 3% endogenous catalase blocker (ZSBIO, PV-6000) for 10min, the tissues were incubated in goat serum (ZSBIO, ZLI-9022) for the blocking of nonspecific reaction and then incubated with primary antibody (LEP=1:300 and PTTG1=1:300) at 4℃ overnight. The next day, tissues were washed and incubated with goat anti-rabbit secondary antibody (ZSBIO, PV-9000) for 1h at room temperature, then washed and stained with DAB reagents (ZSBIO, ZLI-9018). Then hematoxylin staining, 1% hydrochloric acid alcohol differentiation, ammonia water anti-blue, and neutral gum sealing.

The IHC results were evaluated by pathologists, the staining extent was scored as 0-100%. The intensity score was defined as negative, low-expression, medium-expression, and high-expression, which were documented as 0, 1, 2, and 3 respectively. The final scores were calculated by the formula: *IHC score = Staining extent score × Staining intensity score*.

Statistical Methods

R software (version 4.1.3) was used in this study. For quantitative variables, differences between the two groups and among multiple groups were analyzed by Wilcoxon's test and One-way analysis of variance (ANOVA), respectively. For categoric variables, groups were compared by use of Chi-square test. Survival curves were determined by Log-rank test. The clinicopathological features and levels of immune infiltration were conducted by Wilcoxon's test. A difference of $P<0.05$ indicated statistical significance unless specified otherwise.

RESULTS

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Baseline characteristics were shown in Table 1. Of 329 RPLS patients, 88 in training cohort and 241 in validation cohort. No statistically significant differences were found in the age, sex, pathology, surgery times, tumor size, and multilocation between the two cohorts (P>0.05). The IHC score of LEP and PTTG1 were 1.62 (0.820) and 0.830 (0.75) in validation cohort.

Table 1 Baseline characteristics of training cohort and validation cohort

	Training cohort (N=80)[†]	**Validation cohort (N=241)**	***P* value**
Age (y)	56.34 (11.14)[‡]	55.11 (10.80)[‡]	0.384
Sex			
Male	37 (46.25)	118 (48.96)	0.674
Female	43 (53.75)	123 (51.04)	
Pathology			
WDLS	29 (36.25)	75 (31.12)	0.078
DDLS	48 (60.00)	144 (59.75)	
MLS and PLS	3 (3.75)	5 (2.07)	
NR	0 (0)	17 (7.06)	
Surgery times[††]			
0-1	50 (62.50)	143 (59.34)	0.834
2-3	21 (26.25)	73 (30.29)	
4-7	9 (11.25)	24 (9.96)	
NR	0 (0)	1 (0.41)	
Tumor size			
All	18.65 (8.70)[‡]	16.90 (7.94)[‡]	0.101
<18 cm	40 (50.00)	135 (56.02)	0.095
>18 cm	39 (48.75)	92 (38.17)	
NR	1 (1.25)	14 (5.81)	

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Multilocation			
Yes	52 (65.00)	153 (63.49)	0.081
No	28 (35.00)	74 (30.71)	
NR	0 (0)	14 (5.80)	
MDM2 score			
0	11 (13.75)	NA	NA
1	12 (15.00)	NA	
2	39 (48.75)	NA	
3	5 (6.25)	NA	
4	9 (11.25)	NA	
NR	4 (5.00)	NA	
LEP score	NA	1.62 (0.82)[‡]	NA
LEP strength			
0	NA	10 (4.15)	NA
1	NA	39 (16.18)	
2	NA	74 (30.71)	
3	NA	115 (47.72)	
NR	NA	3 (1.24)	
PTTG1 score	NA	0.83 (0.75)[‡]	NA
PTTG1 strength			
0	NA	38 (15.77)	NA
1	NA	100 (41.49)	
2	NA	61 (25.31)	
3	NA	39 (16.18)	
NR	NA	3 (1.25)	

[‡]*Clinical information missing in 8 RPLS patients (Training cohort 2)*

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

‡The data is shown as Mean (SD); other data is shown as Number (%)

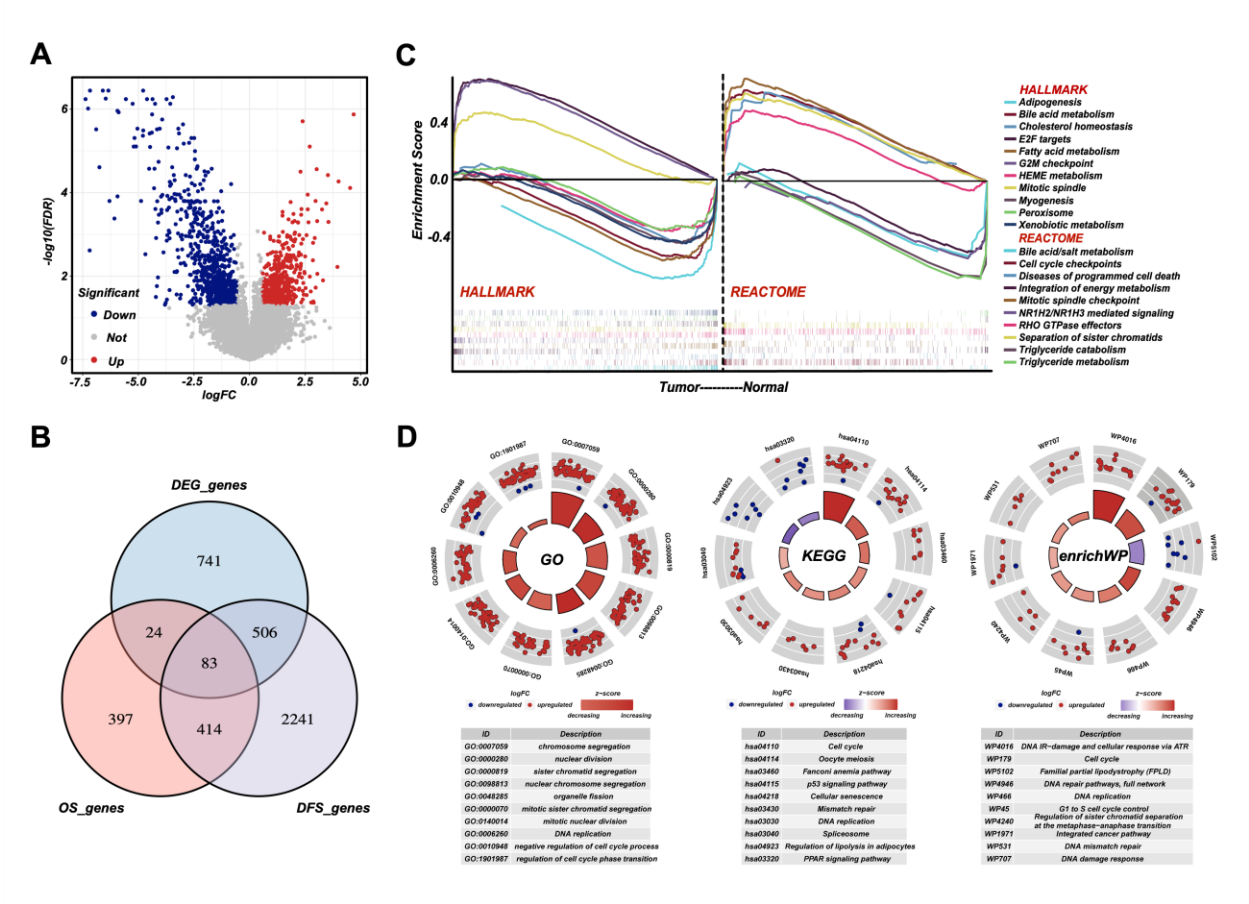
††The definition of surgical times is the sum of current admission surgery and previous surgical resection

DDLS, Dedifferentiated liposarcoma; LEP, Leptin; MDM2, Mouse double minute 2; MLS, Myxoid liposarcoma; NA, Not applicable; NR, Not reported; PLS, Pleomorphic liposarcoma; PTTG1, Pituitary tumor transforming gene 1; WDLS, Well-differentiated liposarcoma

Cell cycle, DNA damage and repair, and Metabolism are dysregulated in RPLS

To reveal the general molecular features of RPLS compared to noncancerous adipose tissues, we first recruited 8 RPLS patients and collected paired tumor and normal tissues for differentially expressed gene (DEG) analysis. A total of 1354 DEGs, 554 upregulated and 800 downregulated, were identified (Figure 2A-B). To assess the underlying pathways of RPLS, GSEA analyses were performed for those DEGs. We found that proliferation-associated pathways, such as Mitotic spindle, E2F target, G2/M checkpoint and Separation of sister chromatids, were mainly enriched in tumors; while metabolism-related pathways, such as Bile acid metabolism, Heme and fatty acid metabolism and Integration of energy metabolism, were enriched in normal controls (Figure 2C).

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .



ID	Description
GO:0007059	chromosome segregation
GO:0000280	nuclear division
GO:0000819	sister chromatid segregation
GO:0098813	nuclear chromosome segregation
GO:0048285	organelle fission
GO:0000070	mitotic sister chromatid segregation
GO:0140014	mitotic nuclear division
GO:0006260	DNA replication
GO:0010948	negative regulation of cell cycle process
GO:1901987	regulation of cell cycle phase transition

ID	Description
hsa04110	Cell cycle
hsa04114	Oocyte meiosis
hsa03460	Fanconi anemia pathway
hsa04115	p53 signaling pathway
hsa04218	Cellular senescence
hsa03430	Mismatch repair
hsa03030	DNA replication
hsa03040	Spliceosome
hsa04923	Regulation of lipolysis in adipocytes
hsa03320	PPAR signaling pathway

ID	Description
WP4016	DNA IR-damage and cellular response via ATR
WP179	Cell cycle
WP5102	Familial partial lipodystrophy (FPLD)
WP4946	DNA repair pathways, full network
WP466	DNA replication
WP45	G1 to S cell cycle control
WP4240	Regulation of sister chromatid separation at the metaphase-anaphase transition
WP1971	Integrated cancer pathway
WP531	DNA mismatch repair
WP707	DNA damage response

Figure 2. Cell cycle, DNA damage and repair, and metabolism are dysregulated in RPLS

A. *Volcano plot of the DEGs in 8 normal vs 8 RPLS tissues.* ***B.*** *Venn diagram showed shared genes between DEGs and prognostic genes.* ***C.*** *GSEA analysis of RPLS tumors, including HALLMARK gene sets and REACTOME gene sets.* ***D.*** *Circular plots of the prognostic genes in GO, KEGG, and enrichWP.*

Then, we collected another 80 samples to investigate the molecular heterogeneity of RPLS. Gene expression profiles showed 918 and 3244 genes associated with overall survival (OS) and disease-free survival (DFS), respectively. Among 497 candidate genes associated with both OS and DFS, 83 of them also overlapped with DEGs (Figure 2B). Functional annotation (GO, KEGG, and enrichWP) demonstrated that Cell cycle, DNA damage and repair, and Metabolism-

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

related pathways were significantly enriched (Figure 2D), suggesting these signaling pathways were dysregulated in RPLS.

RPLS subgroups based on molecular features show different clinical outcomes

To evaluate heterogeneous molecular clustering characteristics in RPLS, ssGSEA emerged as a widely adopted method for computing the enrichment level of specific biological signaling pathways for each sample based on gene expression data. This aids in gaining insights into the overall activity level of signaling pathways. Here, we scored each sample on the dysregulated pathways by ssGSEA and divided RPLS patients into two subgroups (Figure 3A). Subgroup 1 (G1) showed better OS and DFS compared to subgroup 2 (G2) (Figure 3B-C), G1 displayed elevated ssGSEA scores associated with Metabolism, whereas G2 exhibited heightened ssGSEA scores linked on Cell cycle and DNA damage and repair (Figure 3D). These features suggested that effective monitoring of the prognosis of RPLS patients can be achieved based on the activation status of specific pathways. We also evaluated the clinical features and immune infiltration levels

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

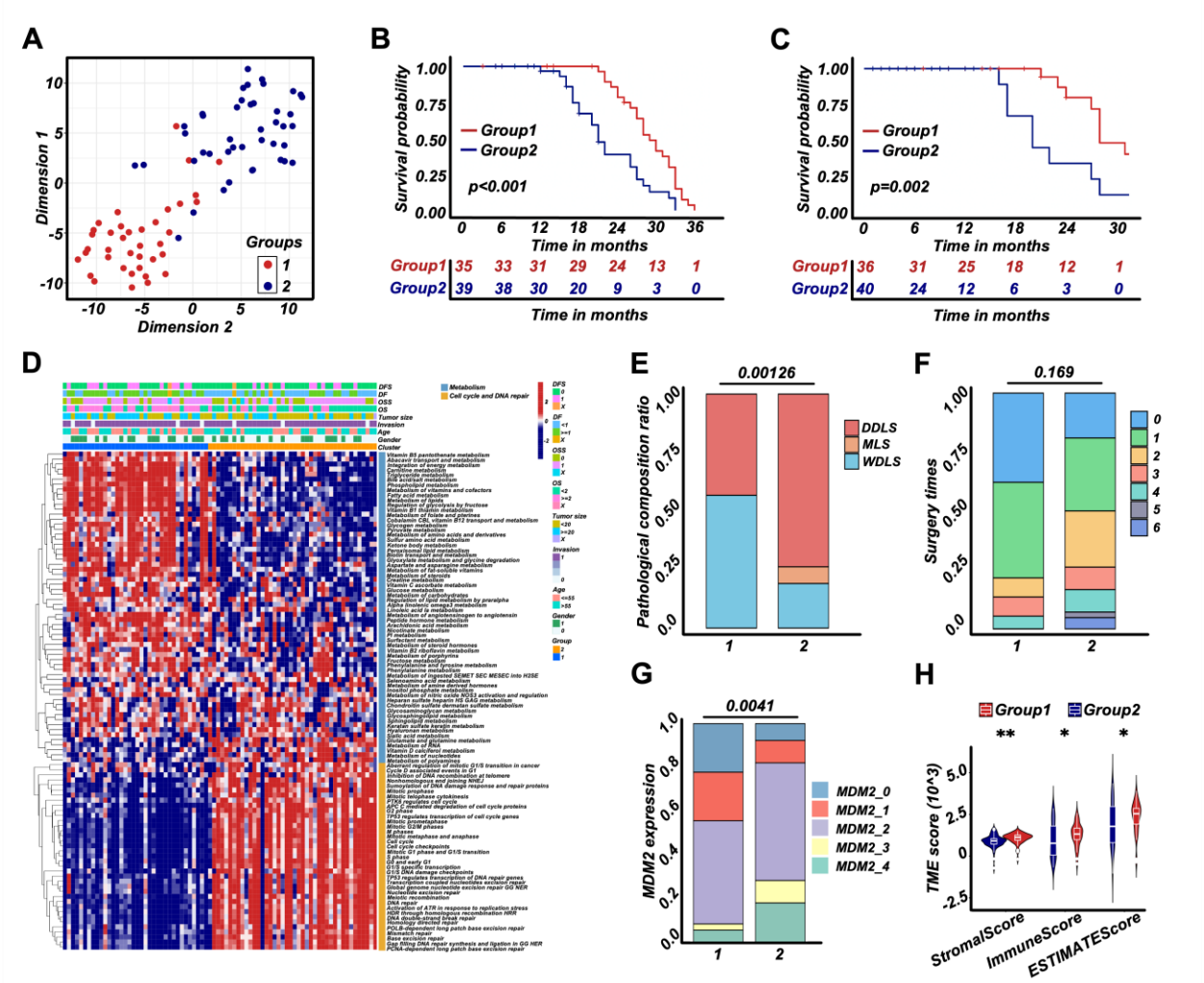


Figure 3. RPLS subgroups (G1 and G2) based on cell cycle, DNA damage and repair, and metabolism

A. *tSNE exhibited the subgroups (G1 and G2) of RPLS.* ***B-C.*** *Survival cures of OS (B) and DFS (C) in G1 and G2.* ***D.*** *The hierarchical clustering heatmap of dysregulated pathways in G1 and G2.* ***E-G.*** *Histograms revealed the difference of pathological composition ratio (E), surgery times (F), and MDM2 (G) in G1 and G2.* ***H.*** *Violin plot of the microenvironmental scores in G1 and G2.*

of those samples. The results showed that G1 had lower aggressive pathological composition ratio (Figure 3E), MDM2 (Figure 3G) and Ki67 (Figure S1A) expression, larger tumor size (Figure S1B), and higher tumor microenvironment (TME) level compared to G2 (Figure 3H). Surgery times for G1 were also tended to decrease (Figure 3F). Taken together, the above restuls

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

indicated that RPLS subgroups based on molecular features showed distinct clinical features and clinical outcomes.

A simplified RPLS classification strategy derived from RPLS G1/G2 subgroups

To explore representative biomarkers for different RPLS subgroups, we performed a DEG analysis between G1 and G2. There were 1258 genes downregulated, among which 112 of them indicated good prognosis (protective genes). Correspondingly, 754 genes were upregulated, and 28 of them indicated poor prognosis (aggressive genes) (Figure S1C-D). Enrichment analysis suggested that those DEGs were also associated with cell cycle regulation and metabolism (Figure S1E), which was consistent with previous results (Figure 2C).

To develop a simplified RPLS clustering based on DEGs, we adopted NMF and tSNE for a re-classification of those patients. The results showed RPLS patients were also divided into two clusters (Figure 4A and Figure S1F-G). We then annotated the samples of two clusters by ssGSEA and found Cluster1 (C1) was related to metabolic processes, and Cluster2 (C2) was mainly related to the processes of Cell cycle and DNA damage and repair (Figure S2A). Also, C1 showed better OS and DFS, lower pathological composition ratio and MDM2 expression, and fewer surgery times (Figure 4B-F). Lower Ki67 expression and larger tumor size were observed in C2 (Figure S2B-C). Interestingly, the biological annotations of the C1/C2 classification were greatly consistent with G1/G2. Therefore, a simplified RPLS classification strategy derived from RPLS subgroups was provisionally established.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

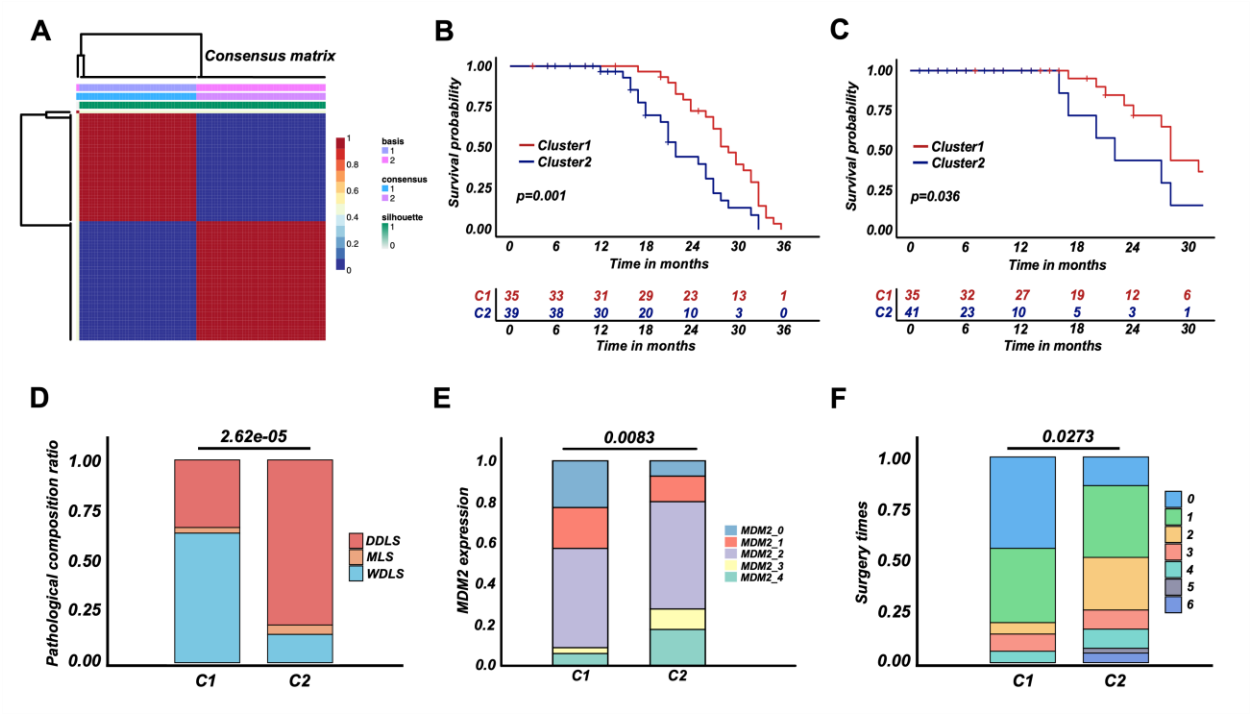


Figure 4. RPLS classification strategy (C1 and C2) derived from RPLS subgroups

A. *NMF for a re-classification of training cohort 1 (C1 and C2) (F).* ***B-C.*** *Survival cures of OS (B) and DFS (C) in C1 and C2.* ***D-F.*** *Histograms revealed the difference in pathological composition ratio (D), MDM2 (E), and surgery times (F) in C1 and C2.*

Development of a dichotomous RPLS classification model

For NMF classification of RPLS patients, LEP and PTTG1 were identified as representative biomarkers of C1 and C2, respectively (Figure 5A). We aimed to replicate the RPLS classification of C1 and C2 by integrating these two biomarkers with the assistance of machine learning algorithms, and this two-gene panel achieved promising results (Logistic, AUC=0.995; SVM, AUC=0.997; RF, AUC=1.000; Figure 5B). Also, a linear negative correlation between LEP and PTTG1 expression was detected (Figure 5C). Considering the enhanced interpretability and generalization of linear models, we adopted the results of Logistic regression for subsequent analysis (Risk values=2.182×PTTG1-2.204×LEP). The patients marked as high-risk (Cluster_H) exhibited worse OS and DFS than those marked as low-risk (Cluster_L) (Figure 5D-E).

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Dysregulated pathways, such as DNA repair and Cell cycle regulation, were enriched in Cluster_H (Figure 5F and Figure S2D), and Cluster_H presented more aggressive pathological composition ratio, higher MDM2 levels and marginally increased in surgery times than Cluster_L (Figure 5G-I). Similarly, the Cluster_H showed higher Ki67 level and smaller tumor sizes (Figure S2E-F). Moreover, a Sankey diagram was drawn to show the correlation among G1/G2, C1/C2, and Cluster_L/H. Cluster_L/H were well-matched to C1/C2 and G1/G2, suggesting LEP and PTTG1 were promising biomarkers for a dichotomous RPLS classification (Figure 5J).

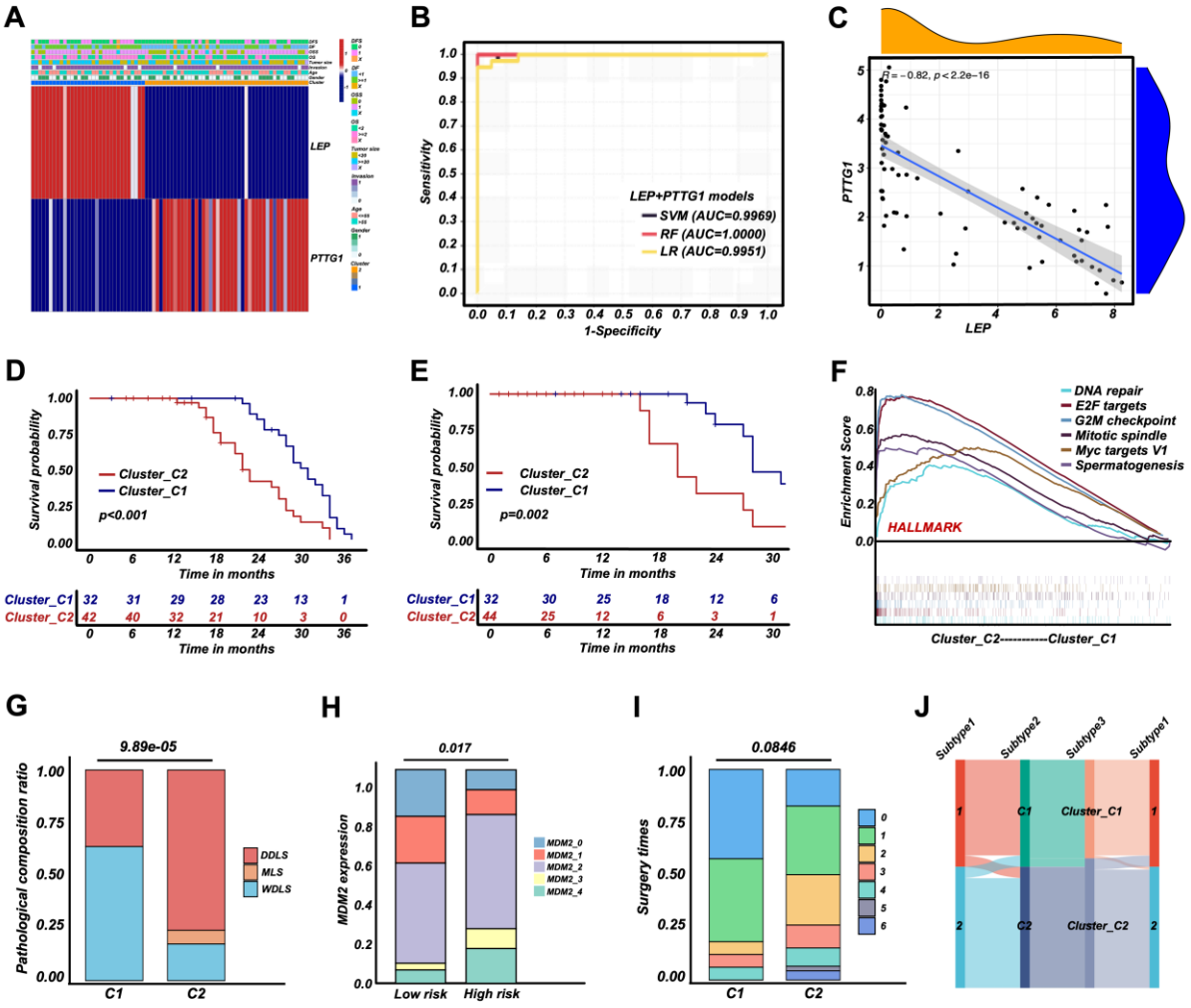


Figure 5. RPLS dichotomous classification (Cluster_C1 and Cluster_C2) derived from RPLS clusters

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

A. *Heatmap of biomarkers identified (LEP and PTTG1) in C1 and C2.* ***B.*** *ROC curves of the machine learning models to identify C1 and C2.* ***C.*** *Correlation between LEP and PTTG1 expression.* ***D-E*** *Survival curves of OS (D) and DFS (E) in Cluster_C1 (low risk) and Cluster_C2 (high risk) groups.* ***F.*** *GSEA of HALLMARK gene sets in Cluster_C1 and Cluster_C2.* ***G-I.*** *Histograms revealed the difference of pathological composition ratio (G), MDM2 level (H), and surgery times (I) in Cluster_C1 and Cluster_C2.* ***J.*** *Sankey diagram indicated the correlation among G1/G2, C1/C2, and Cluster_C1/C2.*

Validation of the dichotomous RPLS classification in another 241 RPLS patients

To validate LEP and PTTG1 as biomarkers for a dichotomous RPLS classification, we performed IHC staining of two biomarkers in validation cohort. The representative images of LEP and PTTG1 with different expression levels were shown in Figure 6A-B. The IHC scores were integrated with the previously fitted coefficients to evaluate the prognosis of RPLS patients (Risk values=$2.182\times PTTG1_{IHC}-2.204\times LEP_{IHC}$). The cutoff value of validation cohort is the median of risk value. The high-risk group had worse OS and DFS (Figure 6C-D), along with more surgery times and more aggressive pathological composition ratio (Figure 7A-B), but the difference of tumor size was not observed between the two groups (Figure 7C). Then we constructed visual nomograms for a precise survival prediction of RPLS patients by combining the risk score with clinical features. The predictive abilities of the 1-, 2-, and 3-year OS (Figure 7D-F) and DFS (Figure S3A-C) were 0.743-0.788. Together, we proposed a simple and clinically applicable molecular classification strategy for RPLS patients.

DISCUSSION

Here we divided RPLS patients into two subgroups based on Cell cycle, DNA damage & repair, and Metabolism-related pathways. G1 was annotated as Metabolism-active, which exhibited high ssGSEA scores on Metabolism-associated pathways, while G2 showed high

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

ssGSEA scores on Cell cycle and DNA damage and repair with a high Ki67 and MDM2 level. G2

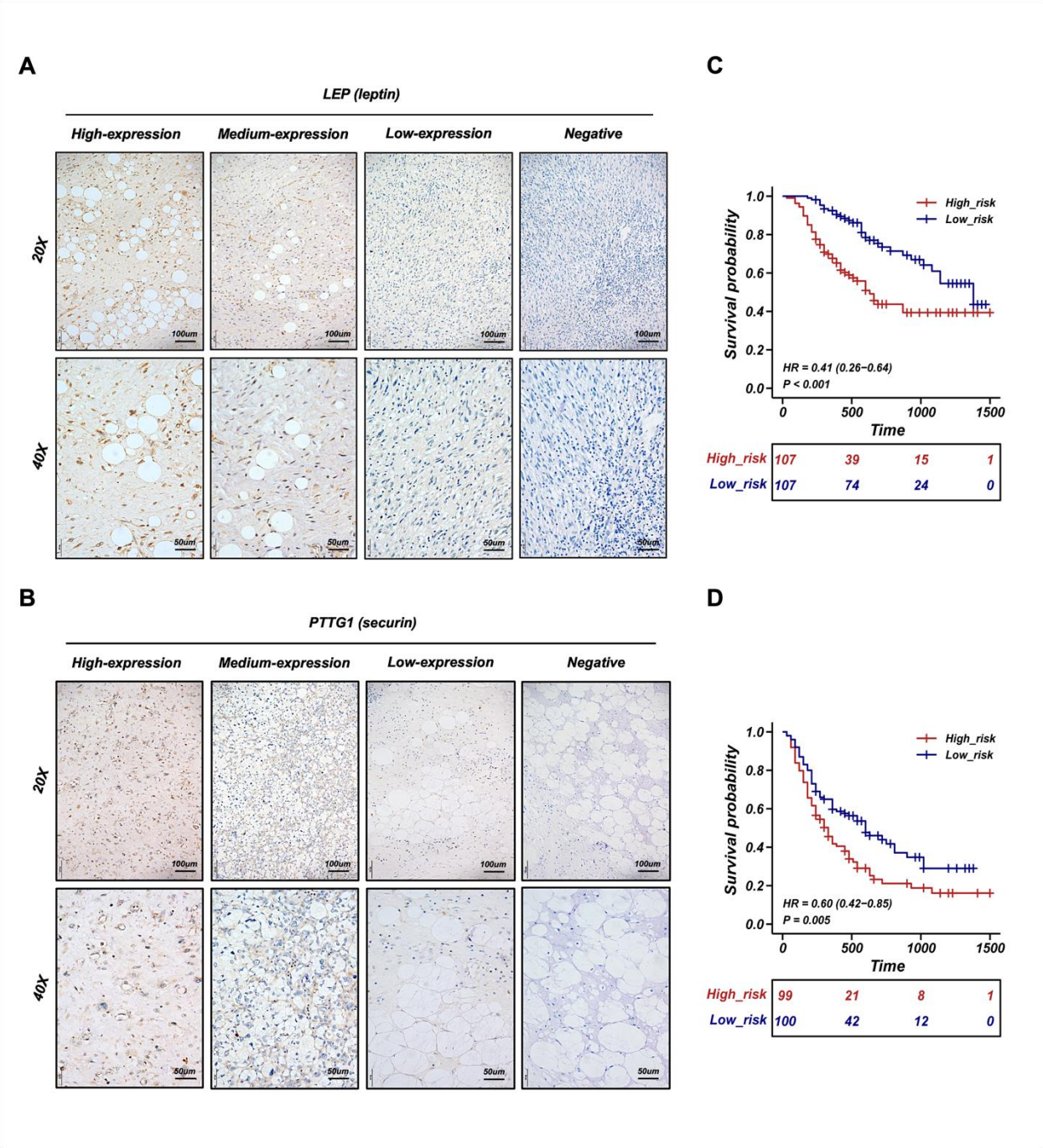


Figure 6. Validation of the RPLS dichotomous classification in another 241 RPLS cohort

A-B. *Representative IHC staining images of LEP (A) and PTTG1 (B).* ***C-D.*** *Survival curves of OS (C) and DFS (D) in high-risk and low-risk groups.*

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

had more aggressive molecular features and worse clinical outcomes compared to G1, in accordance with previously reported tumor classification (***Lindskrog et al., 2021; Yu et al., 2021; Zhang et al., 2022***). In fact, ***Demicco et al., 2017*** has integrated SCNA and DNA methylation divide dedifferentiated liposarcoma into two subtypes (S1 and S2), the unfavorable cluster was characterized as JUN amplified (An oncogene that promotes proliferation and metastasis) and lower inferred fraction of immature dendritic cells. However, the patients in ***Demicco et al., 2017*** were of complex origin (mixed limbs, trunk and retroperitoneum), providing limited guidance for RPLS molecular classification. Here we reported the first clinically applicable RPLS molecular classification based on RNA sequencing and IHC validation cohorts.

To facilitate the clinical application, we constructed a simplified RPLS molecular classification derived from the original Cell cycle/Metabolism subgroups. By NMF algorithm, we identified LEP and PTTG1 as representative biomarkers for each subtype. A model based on IHC staining of LEP and PTTG1 successfully approximated the original dichotomous RPLS classification in biological features and survival outcomes. LEP is an important regulator of basal metabolism and food intake, which is considered a linkage between metabolism and the immune system (***Jiménez-Cortegana et al., 2021***). Although LEP-based targeting therapies have not yet been fully applied, LEP has already been identified as a potent metabolic reprogramming agent to support antitumor responses in aggressive melanomas (***Waldman et al., 2020; De la et al., 2019; Rivadeneira et al., 2019***). In addition, LEP improves the immunotherapeutic effects by regulating innate and adaptive immune responses via increasing the cytotoxicity of NK cells (***Vera et al., 2018***), stimulating the proliferation of T/B cells (***Vera et al., 2018; Bernotiene et al., 2006***), and activating DC cells (***Hu et al., 2019***). Those reported roles of LEP provided a good mechanism explanation on the features of metabolism pathway-enriched, better prognosis, higher

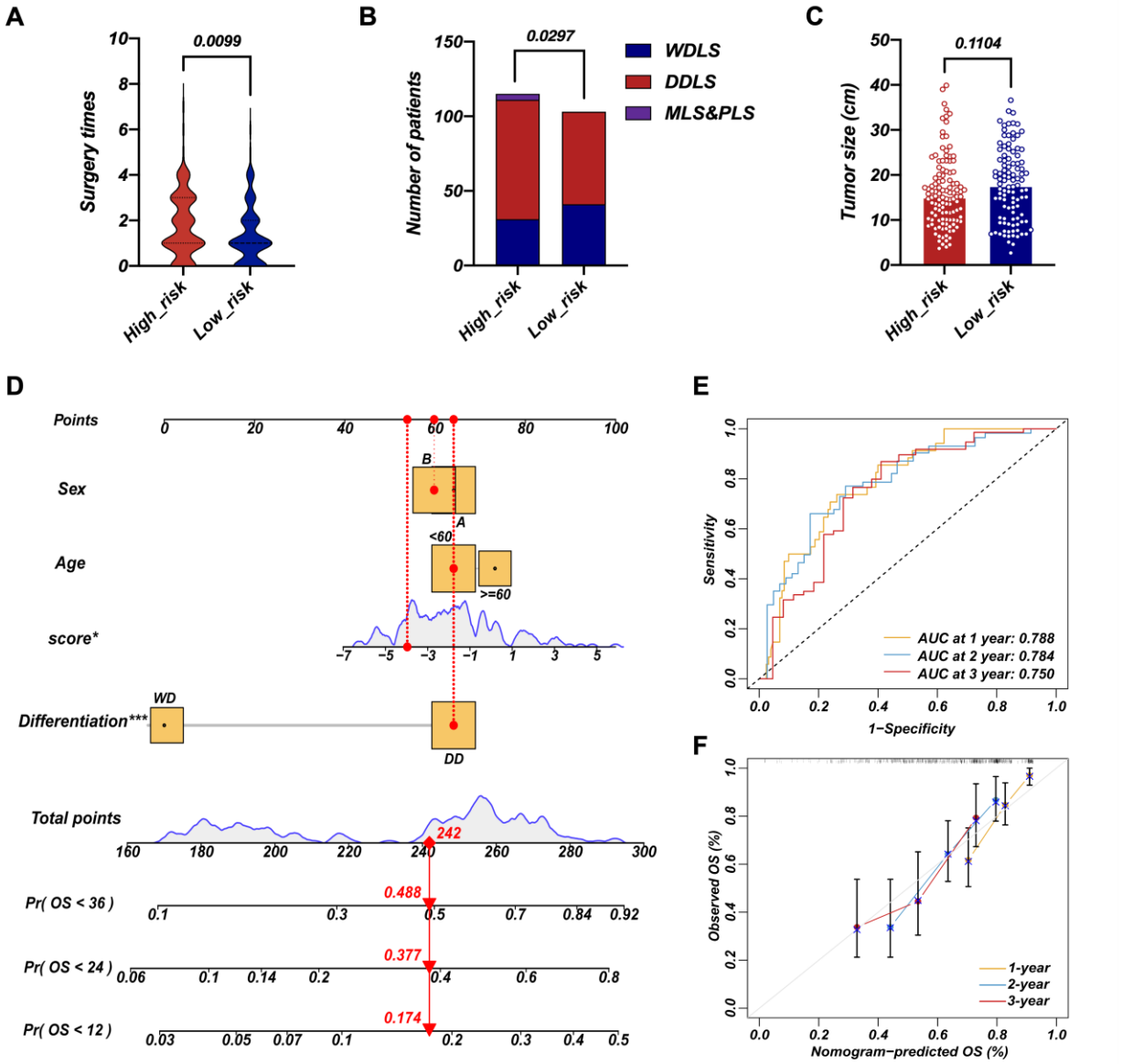
medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

TME level of Metabolism subgroup (LEP+). In contrast, PTTG1 acts as a regulator of sister chromatid separation during cell division under physiological conditions (Zou et al., 1999), which is closely linked to genetic instability, aneuploidy, tumor progression, invasion, and metastasis (***Heaney et al., 2000; Ramaswamy et al., 2003; Kim et al., 2005; Yu et al., 2003; Teveroni et al., 2021; Romero et al., 2001***). PTTG1 also regulates the cell cycle and the transactivation of growth factors as an initiator and promoter of tumorigenesis (***Zou et al., 1999; Mora-Santos et al., 2013; McCabe et al., 2002; Ishikawa et al 2001; Hamid et al., 2005***). Overexpressing PTTG1 was correlated with worse prognosis in tumors, such as ovarian cancer (***Parte et al., 2019***), cervical cancer (***Guo et al., 2019***), renal cell carcinoma (***Tian et al., 2022***), and colorectal cancer (***Heaney et al., 2000***). Therefore, the biological functions of PTTG1 provided a good mechanism explanation of the pathway-enriched of cell cycle/DNA damage & repair-associated, worse prognosis, and more aggressive pathological composition ratio the Cell cycle subgroup (PTTG1+).

CONCLUSION

Our study presented a comprehensive gene expression landscape of RPLS, revealing distinct molecular features. Through categorizing RPLS into Metabolism and Cell Cycle subtypes and identifying key biomarkers LEP and PTTG1, we established a dichotomous classification system verified by IHC assays. This innovative approach enables precise guidance for surgeons in adjusting treatment strategies for patients with histologically favorable but prognostically challenging RPLS cases, thereby advancing the implementation of precision medicine in guiding surgical interventions for RPLS.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .



386 ***Figure 7. Survival nomogram of LEP+PTTG1 model in validation cohort***

387 ***A-C.*** *The difference of surgery times (A), pathological composition (B), and surgery times (C) in high-risk and*
388 *low-risk groups.* ***D.*** *Nomograms for OS was developed in REASR cohort with four factors: sex, age, risk score,*
389 *and differentiation.* ***E.*** *ROC curves of 1-, 2-, and 3-year OS in validation cohort.* ***F.*** *Calibration curves of*
390 *predicting 1-, 2-, and 3-year OS in validation cohort.*

391

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Abbreviations

RPLS	Retroperitoneal liposarcoma
DEGs	Differentially expressed genes
IHC	Immunohistochemistry
STS	Soft tissue sarcoma
OS	Overall survival
DFS	Disease-free survival
OMIX	Open archive for miscellaneous data
CNCB	China national center for bioinformation
GSEA	Gene set enrichment analysis
GO	Gene ontology
KEGG	Kyoto encyclopedia of gene and genomes
t-SNE	t-distributed stochastic neighbor embedding
DBSCAN	Density-based spatial clustering of applications with noise
NMF	Nonnegative matrix factorization
LR	Logistic regression
SVM	Support vector machine
RF	Random forest
AUC	Area under curve

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Authors' contributors

Mengmeng Xiao: Methodology, Formal analysis, Data curation, Writing-original draft, Writing-review & editing. Xiang Ji Li: Methodology, Formal analysis, Investigation, Data curation, Writing-original draft. Fanqin Bu: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing-review & editing. Shixiang Ma: Methodology, Formal analysis. Xiaohan Yang: Methodology, Investigation. Jun Chen: Formal analysis. Yu Zhao: Investigation. Ferdinando Cananzi: Investigation. Chenghua Luo: Conceptualization, Methodology, Writing-review & editing, Supervision. Li Min: Conceptualization, Resources, Writing-review & editing, Supervision.

Ethics approval and consent to participate

Specimens of RPLS were obtained from Peking University International Hospital. The study protocol was approved by the Ethics Committee of Peking University International Hospital, Peking University Health Science Center (WA2020RW29) and conducted in accordance with Helsinki Declaration. All patients signed the informed consent.

Finding information

This work was supported by grants from the Beijing Municipal Science and Technology Project (Z191100006619081), National Natural Science Foundation of China (82073390), and Young Elite Scientists Sponsorship Program (2023QNRC001). The study sponsors had no role in the design and preparation of this manuscript.

Research registration unique identifying number (UIN)

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

1. Name of the registry: ClinicalTrials.gov

2. Unique identifying number or registration ID: NCT03838718

3. Hyperlink to your specific registration: https://clinicaltrials.gov

Consent for publication

All authors have read and approved the manuscript and agree with submission to ***eLife***.

Data availability

Data supporting the conclusions of this article are presented within the article and its supplementary files.

Competing interests

The authors declare that they have no competing interests.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

REFERENCES

Ecker BL, Peters MG, McMillan MT, Sinnamon AJ, Zhang PJ, Fraker DL, Levin WP, Roses RE, Karakousis GC. Preoperative radiotherapy in the management of retroperitoneal liposarcoma. *Br J Surg*. 2016; 103(13): 1839-1846. doi: 10.1002/bjs.10305

Littau MJ, Kulshrestha S, Bunn C, Agnew S, Sweigert P, Luchette FA, Baker MS. The importance of the margin of resection and radiotherapy in retroperitoneal liposarcoma. *Am J Surg*. 2021;221(3):554-560. doi:10.1016/j.amjsurg.2020.11.041

Gronchi A, Miceli R, Allard MA, Callegaro D, Le Péchoux C, Fiore M, Honoré C, Sanfilippo R, Coppola S, Stacchiotti S, Terrier P, Casali PG, Cesne AL, Mariani L, Colombo C, Bonvalot S. Personalizing the approach to retroperitoneal soft tissue sarcoma: histology-specific patterns of failure and postrelapse outcome after primary extended resection. *Ann Surg Oncol*. 2015;22(5):1447-1454. doi:10.1245/s10434-014-4130-7

Gronchi A, Lo Vullo S, Fiore M, Mussi C, Stacchiotti S, Collini P, Lozza L, Pennacchioli E, Mariani L, Casali PG. Aggressive surgical policies in a retrospectively reviewed single-institution case series of retroperitoneal soft tissue sarcoma patients. *J Clin Oncol*. 2009;27(1):24-30. doi:10.1200/JCO.2008.17.8871

Gronchi A, Pollock RE. Quality of local treatment or biology of the tumor: which are the trump cards for loco-regional control of retroperitoneal sarcoma?. *Ann Surg Oncol*. 2013;20(7):2111-2113. doi:10.1245/s10434-013-2971-0

Pisters PW. Resection of some -- but not all -- clinically uninvolved adjacent viscera as part of surgery for retroperitoneal soft tissue sarcomas. *J Clin Oncol*. 2009;27(1):6-8. doi:10.1200/JCO.2008.18.7138

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Kam AE, Masood A, Shroff RT. Current and emerging therapies for advanced biliary tract cancers. *Lancet Gastroenterol Hepatol*. 2021;6(11):956-969. doi:10.1016/S2468-1253(21)00171-0

Zeng Y, Jin RU. Molecular pathogenesis, targeted therapies, and future perspectives for gastric cancer. *Semin Cancer Biol*. 2022;86(Pt 3):566-582. doi:10.1016/j.semcancer.2021.12.004

Alifrangis C, McGovern U, Freeman A, Powles T, Linch M. Molecular and histopathology directed therapy for advanced bladder cancer. *Nat Rev Urol*. 2019;16(8):465-483. doi:10.1038/s41585-019-0208-0

Frese KK, Simpson KL, Dive C. Small cell lung cancer enters the era of precision medicine. *Cancer Cell*. 2021;39(3):297-299. doi:10.1016/j.ccell.2021.02.002

Pilotti S, Della Torre G, Mezzelani A, Tamborini E, Azzarelli A, Sozzi G, Pierotti MA. The expression of MDM2/CDK4 gene product in the differential diagnosis of well differentiated liposarcoma and large deep-seated lipoma. *Br J Cancer*. 2000;82(7):1271-1275. doi:10.1054/bjoc.1999.1090

Binh MB, Sastre-Garau X, Guillou L, Pinieux G, Terrier P, Lagacé R, Aurias A, Hostein I, Coindre JM. MDM2 and CDK4 immunostainings are useful adjuncts in diagnosing well-differentiated and dedifferentiated liposarcoma subtypes: a comparative analysis of 559 soft tissue neoplasms with genetic data. *Am J Surg Pathol*. 2005;29(10):1340-1347. doi:10.1097/01.pas.0000170343.09562.39

Yen CC, Chen SC, Hung GY, Wu PK, Chua WY, Lin YC, Yen CH, Chen YC, Wang JY, Yang MH, Chao Y, Chang MC, Chen WM. Expression profile-driven discovery of AURKA as a treatment target for liposarcoma. *Int J Oncol*. 2019;55(4):938-948. doi:10.3892/ijo.2019.4861

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Yang L, Wu Z, Sun W, Luo P, Chen S, Chen Y, Yan W, Li Y, Wang C. CCNDBP1, a Prognostic Marker Regulated by DNA Methylation, Inhibits Aggressive Behavior in Dedifferentiated Liposarcoma *via* Repressing Epithelial Mesenchymal Transition. *Front Oncol*. 2021;11:687012. doi:10.3389/fonc.2021.687012

McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer*. 2005;93(4):387-391.

Guo L, Chen G, Zhang W, Zhou L, Xiao T, Di X, Wang Y, Feng L, Zhang K. A high-risk luminal A dominant breast cancer subtype with increased mobility. *Breast Cancer Res Treat*. 2019;175(2):459-472. doi:10.1007/s10549-019-05135-w

Li X, Bu F, Ma S, Cananzi F, Zhao Y, Xiao M, Min L, Luo C. The Janus-faced role of TRPM2-S in retroperitoneal liposarcoma via increasing ROS levels. *Cell Commun Signal*. 2022;20(1):128. doi:10.1186/s12964-022-00873-9

Lindskrog SV, Prip F, Lamy P, Taber A, Groeneveld CS, Birkenkamp-Demtröder K, Jensen JB, Strandgaard T, Nordentoft I, Christensen E, Sokac M, Birkbak NJ, Maretty L, Hermann GG, Petersen AC, Weyerer V, Grimm MO, Horstmann M, Sjödahl G, Höglund M, Steiniche T, Mogensen K, Reyniès A, Nawroth R, Jordan B, Lin X, Dragicevic D, Ward DG, Goel A, Hurst C, Kessel K, Maurer T, Meeks JJ, DeGraff DJ, Bryan RT, Knowles MA, Simic T, Hartmann A, Zwarthoff EC, Malmström PU, Malats N, Real FX, Dyrskjøt L. An integrated multi-omics analysis identifies prognostic molecular subtypes of non-muscle-invasive bladder cancer. *Nat Commun*. 2021;12(1):2301.

Yu Z, Deng P, Chen Y, Liu S, Chen J, Yang Z, Chen J, Fan X, Wang P, Cai Z, Wang Y, Hu P, Lin D, Xiao R, Zou Y, Huang Y, Yu Q, Lan P, Tan J, Wu X. Inhibition of the PLK1-Coupled

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Cell Cycle Machinery Overcomes Resistance to Oxaliplatin in Colorectal Cancer. *Adv Sci (Weinh)*. 2021;8(23):e2100759.

Zhang F, Zhang Q, Zhu J, Yao B, Ma C, Qiao N, He S, Ye Z, Wang Y, Han R, Feng J, Wang Y, Qin Z, Ma Z, Li K, Zhang Y, Tian S, Chen Z, Tan S, Wu Y, Ran P, Wang Y, Ding C, Zhao Y. Integrated proteogenomic characterization across major histological types of pituitary neuroendocrine tumors. *Cell Res*. 2022;32(12):1047-1067.

Cancer Genome Atlas Research Network. Electronic address: elizabeth.demicco@sinaihealthsystem.ca; Cancer Genome Atlas Research Network. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. *Cell*. 2017;171(4):950-965.e28. doi:10.1016/j.cell.2017.10.014

Jiménez-Cortegana C, López-Saavedra A, Sánchez-Jiménez F, Pérez-Pérez A, Castiñeiras J, Virizuela-Echaburu JA, Cruz-Merino L, Sánchez-Margalet V. Leptin, Both Bad and Good Actor in Cancer. *Biomolecules*. 2021;11(6):913. doi:10.3390/biom11060913

Waldman AD, Fritz JM, Lenardo MJ. A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nat Rev Immunol*. 2020;20(11):651-668. doi:10.1038/s41577-020-0306-5

De la Cruz-Merino L, Palazón-Carrión N, Henao-Carrasco F, Nogales-Fernández E, Álamo-de Ia Gala A, Vallejo-Benítez A, Chiesa M, Sánchez-Margalet V. New horizons in breast cancer: the promise of immunotherapy. *Clin Transl Oncol*. 2019;21(2):117-125. doi:10.1007/s12094-018-1907-3

Rivadeneira DB, DePeaux K, Wang Y, Kulkarni A, Tabib T, Menk AV, Sampath P, Lafyatis R, Ferris RL, Sarkar SN, Thorne SH, Delgoffe GM. Oncolytic Viruses Engineered to Enforce

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Leptin Expression Reprogram Tumor-Infiltrating T Cell Metabolism and Promote Tumor Clearance. *Immunity*. 2019;51(3):548-560.e4. doi:10.1016/j.immuni.2019.07.003

Vera F, Pino J, Campos-Cabaleiro V, Ruiz-Fernández C, Mera A, Gonzalez-Gay MA, Gómez R, Gualillo O. Obesity, Fat Mass and Immune System: Role for Leptin. *Front Physiol*. 2018;9:640. doi: 10.3389/fphys.2018.00640

Bernotiene E, Palmer G, Gabay C. The role of leptin in innate and adaptive immune responses. *Arthritis Res Ther*. 2006;8(5):217. doi:10.1186/ar2004

Hu W, Wang G, Huang D, Sui M, Xu Y. Cancer Immunotherapy Based on Natural Killer Cells: Current Progress and New Opportunities. *Front Immunol*. 2019;10:1205. doi:10.3389/fimmu.2019.01205

Zou H, McGarry TJ, Bernal T, Kirschner MW. Identification of a vertebrate sister-chromatid separation inhibitor involved in transformation and tumorigenesis. *Science*. 1999;285(5426):418-422. doi:10.1126/science.285.5426.418

Heaney AP, Singson R, McCabe CJ, Nelson V, Nakashima M, Melmed S. Expression of pituitary-tumour transforming gene in colorectal tumours. *Lancet*. 2000;355(9205):716-719. doi:10.1016/S0140-6736(99)10238-1

Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. *Nat Genet*. 2003;33(1):49-54. doi:10.1038/ng1060

Kim D, Pemberton H, Stratford AL, Buelaert K, Watkinson JC, Lopes V, Franklyn JA, McCabe CJ. Pituitary tumour transforming gene (PTTG) induces genetic instability in thyroid cells. *Oncogene*. 2005;24(30):4861-4866. doi:10.1038/sj.onc.1208659

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Yu R, Lu W, Chen J, McCabe CJ, Melmed S. Overexpressed pituitary tumor-transforming gene causes aneuploidy in live human cells. *Endocrinology*. 2003;144(11):4991-4998. doi:10.1210/en.2003-0305

Teveroni E, Di Nicuolo F, Bianchetti G, Epstein AL, Grande G, Maulucci G, Spirito MD, Pontecorvi A, Milardi D, Mancini F. Nuclear Localization of PTTG1 Promotes Migration and Invasion of Seminoma Tumor through Activation of MMP-2. *Cancers (Basel)*. 2021;13(2):212. doi:10.3390/cancers13020212

Romero F, Multon MC, Ramos-Morales F, Domínguez A, Bernal JA, Pintor-Toro JA, Tortolero M. Human securin, hPTTG, is associated with Ku heterodimer, the regulatory subunit of the DNA-dependent protein kinase. *Nucleic Acids Res*. 2001;29(6):1300-1307. doi:10.1093/nar/29.6.1300

Mora-Santos M, Castilla C, Herrero-Ruiz J, Giráldez S, Limón-Mortés MC, Sáez C, Japón MÁ, Tortolero M, Romero F. A single mutation in Securin induces chromosomal instability and enhances cell invasion. *Eur J Cancer*. 2013;49(2):500-510. doi:10.1016/j.ejca.2012.06.024

McCabe CJ, Boelaert K, Tannahill LA, Heaney AP, Stratford AL, Khaira JS, Hussain S, Sheppard MC, Franklyn JA, Gittoes NJL. Vascular endothelial growth factor, its receptor KDR/Flk-1, and pituitary tumor transforming gene in pituitary tumors. *J Clin Endocrinol Metab*. 2002;87(9):4238-4244. doi:10.1210/jc.2002-020309

Ishikawa H, Heaney AP, Yu R, Horwitz GA, Melmed S. Human pituitary tumor-transforming gene induces angiogenesis. *J Clin Endocrinol Metab*. 2001;86(2):867-874. doi:10.1210/jcem.86.2.7184

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Hamid T, Malik MT, Kakar SS. Ectopic expression of PTTG1/securin promotes tumorigenesis in human embryonic kidney cells. *Mol Cancer*. 2005;4(1):3. Published 2005 Jan 13. doi:10.1186/1476-4598-4-3

Parte S, Virant-Klun I, Patankar M, Batra SK, Straughn A, Kakar SS. PTTG1: a Unique Regulator of Stem/Cancer Stem Cells in the Ovary and Ovarian Cancer. *Stem Cell Rev Rep*. 2019;15(6):866-879. doi:10.1007/s12015-019-09911-5

Guo XC, Li L, Gao ZH, Zhou HW, Li J, Wang QQ. The long non-coding RNA PTTG3P promotes growth and metastasis of cervical cancer through PTTG1. *Aging (Albany NY)*. 2019;11(5):1333-1341. doi:10.18632/aging.101830

Tian X, Xu WH, Xu FJ, Li H, Anwaier A, Wang HK, Wan FN, Zhu Y, Cao DL, Zhu YP, Shi GH, Qu YY, Zhang HL, Ye DW. Identification of prognostic biomarkers in papillary renal cell carcinoma and PTTG1 may serve as a biomarker for predicting immunotherapy response. *Ann Med*. 2022;54(1):211-226.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Table 1. Baseline characteristics of training cohort and validation cohort

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Figure Legend

Figure 1. Flow diagram of exploring RPLS dichotomous classification

Figure 2. Cell cycle, DNA damage and repair, and metabolism are dysregulated in RPLS

Volcano plot of the DEGs in 8 normal vs 8 RPLS tissues **(A)**. Venn diagram showed shared genes between DEGs and prognostic genes **(B)**. GSEA analysis of RPLS tumors, including HALLMARK gene sets and REACTOME gene sets **(C)**. Circular plots of the prognostic genes in GO, KEGG, and enrichWP **(D)**.

Figure 3. RPLS subgroups (G1 and G2) based on cell cycle, DNA damage and repair, and metabolism

tSNE exhibited the subgroups (G1 and G2) of RPLS **(A)**. Survival cures of OS **(B)** and DFS **(C)** in G1 and G2. The hierarchical clustering heatmap of dysregulated pathways in G1 and G2 **(D)**. Histograms revealed the difference of pathological composition ratio **(E)**, surgery times **(F)**, and MDM2 **(G)** in G1 and G2. Violin plot of the microenvironmental scores in G1 and G2 **(H)**.

Figure 4. RPLS classification strategy (C1 and C2) derived from RPLS subgroups

NMF for a re-classification of training cohort 1 (C1 and C2) **(A)**. Survival cures of OS **(B)** and DFS **(C)** in C1 and C2. Histograms revealed the difference in pathological composition ratio **(D)**, MDM2 **(E)**, and surgery times **(F)** in C1 and C2.

Figure 5. RPLS dichotomous classification (Cluster_C1 and Cluster_C2) derived from RPLS clusters

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Heatmap of biomarkers identified (LEP and PTTG1) in C1 and C2 **(A)**. ROC curves of the machine learning models to identify C1 and C2 **(B)**. Correlation between LEP and PTTG1 expression **(C)**. Survival curves of OS **(D)** and DFS **(E)** in Cluster_C1 (low risk) and Cluster_C2 (high risk) groups. GSEA of HALLMARK gene sets in Cluster_C1 and Cluster_C2 **(F)**. Histograms revealed the difference of pathological composition ratio **(G)**, MDM2 level **(H)**, and surgery times **(I)** in Cluster_C1 and Cluster_C2. Sankey diagram indicated the correlation among G1/G2, C1/C2, and Cluster_C1/C2 **(J)**.

Figure 6. Validation of the RPLS dichotomous classification in another 241 RPLS cohort

Representative IHC staining images of LEP **(A)** and PTTG1 **(B)**. Survival curves of OS **(C)** and DFS **(D)** in high-risk and low-risk groups.

Figure 7. Survival nomogram of LEP+PTTG1 model in validation cohort

The difference of surgery times **(A)**, pathological composition **(B)**, and surgery times **(C)** in high-risk and low-risk groups. Nomograms for OS was developed in REASR cohort with four factors: sex, age, risk score, and differentiation **(D)**. ROC curves of 1-, 2-, and 3-year OS in validation cohort **(E)**. Calibration curves of predicting 1-, 2-, and 3-year OS in validation cohort **(F)**.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Supplementary Information

Table S1. The detailed clinicopathological characteristics of the training cohort 1

Table S2. The RNA-seq data of the training cohort 2

Table S3. The detailed clinicopathological characteristics of the validation cohort

Table S4. The RNA-seq data of the training cohort 1

Figure S1. Clinical features and re-classification of RPLS subgroups (G1 and G2)

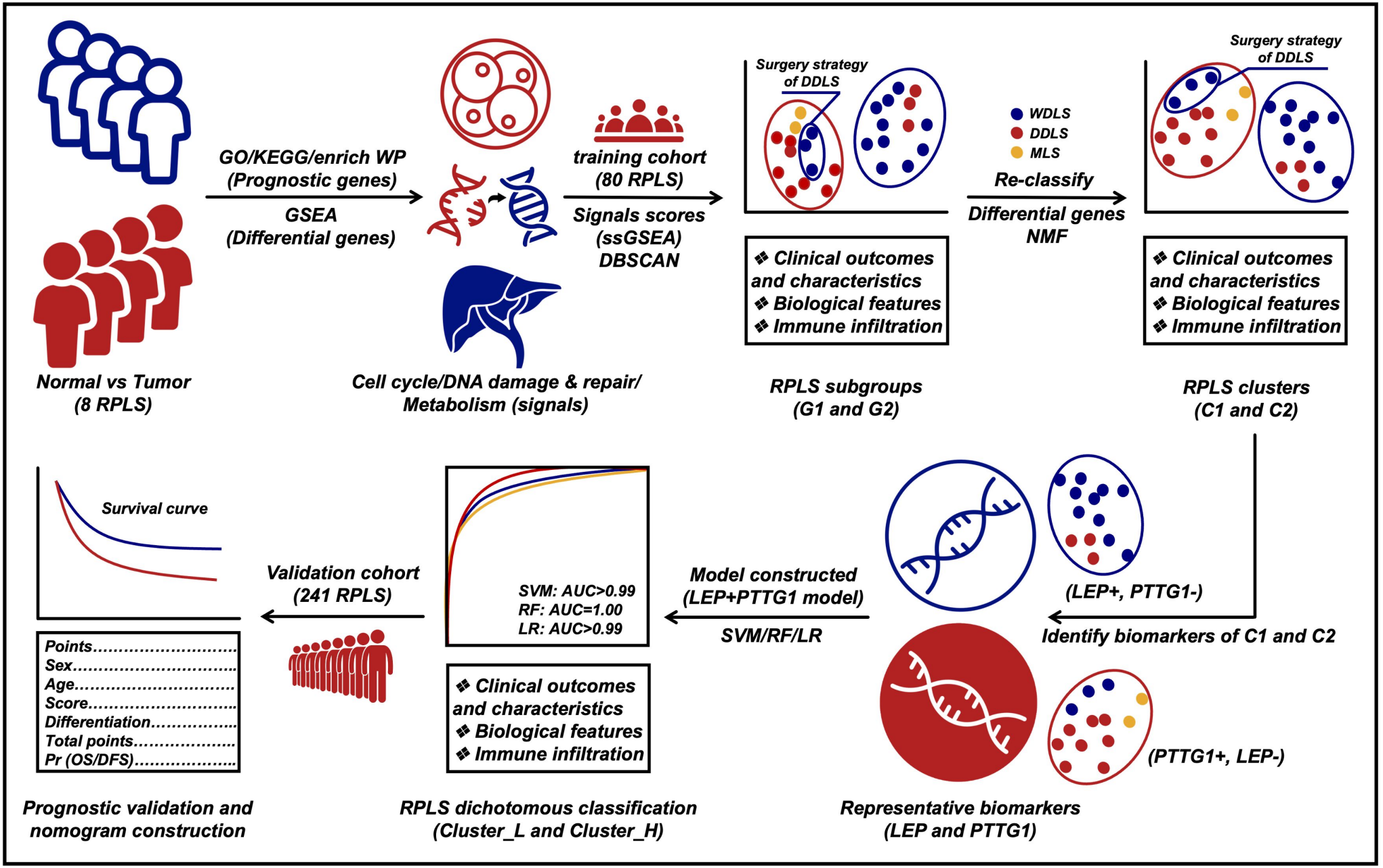
The difference of Ki67 **(A)** and tumor size **(B)** in G1 and G2. Volcano plot of the DEGs (G1 vs G2) **(C)**. Venn diagram showed shared genes between DEGs and prognostic genes **(D)**. Bubble plot of the DEGs enrichment **(E)**. NMF for a re-classification of training cohort 1 **(F)**. tSNE exhibits the RPLS clusters (C1 and C2) **(G)**.

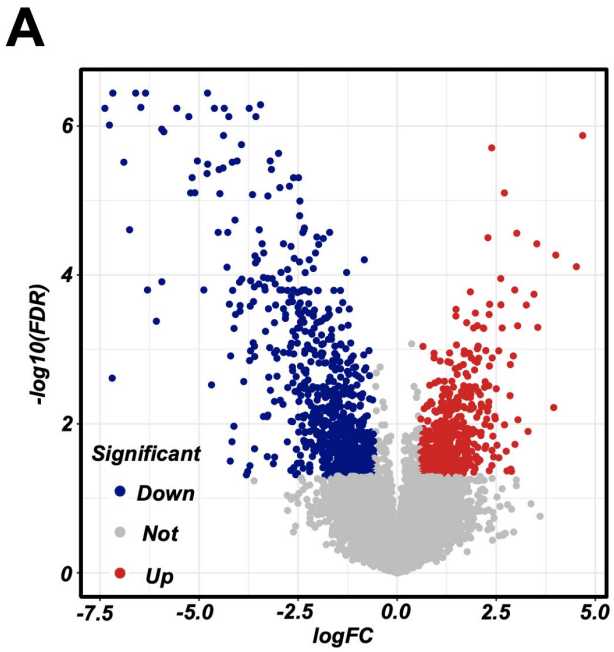
Figure S2. Dysregulated pathways and clinical features of RPLS clusters and high-/low-risk groups

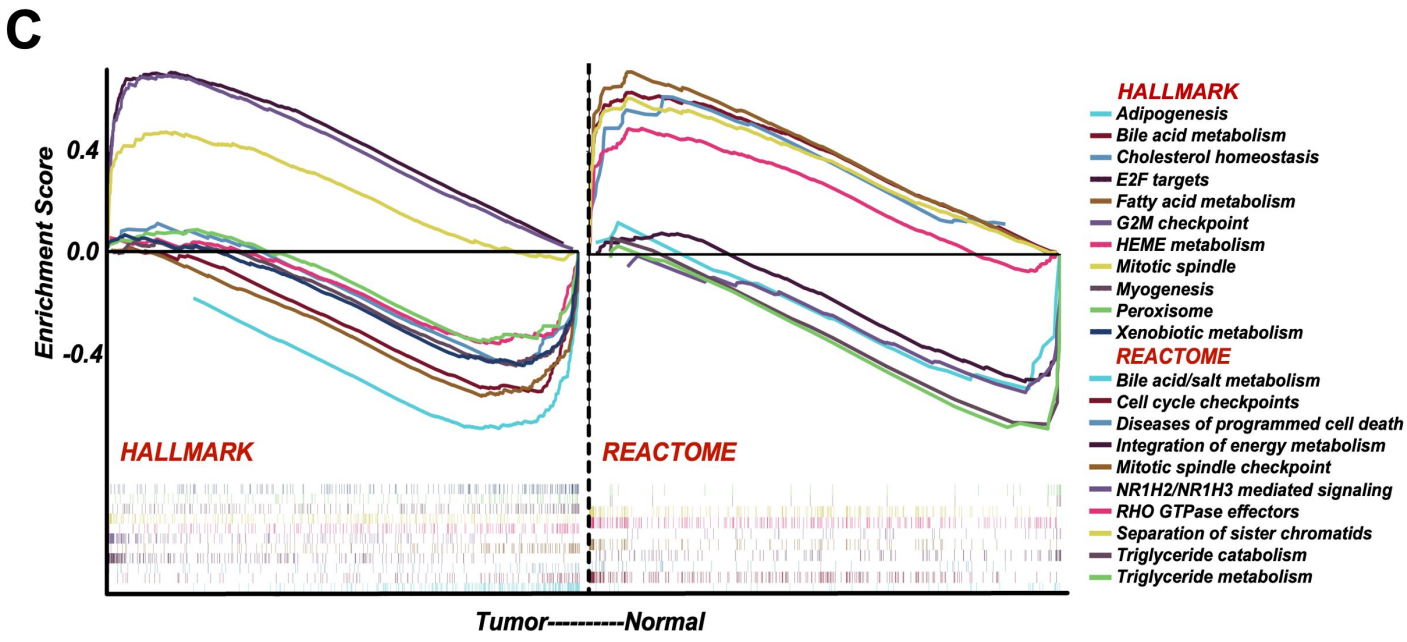
The hierarchical clustering heatmap of dysregulated pathways in C1 and C2 **(A)**. The difference of Ki67 **(B)** and tumor size **(C)** in C1 and C2. The hierarchical clustering heatmap of dysregulated pathways in high- and low-risk groups **(D)**. The difference of tumor size **(E)** and Ki67 **(F)** in high- and low-risk groups.

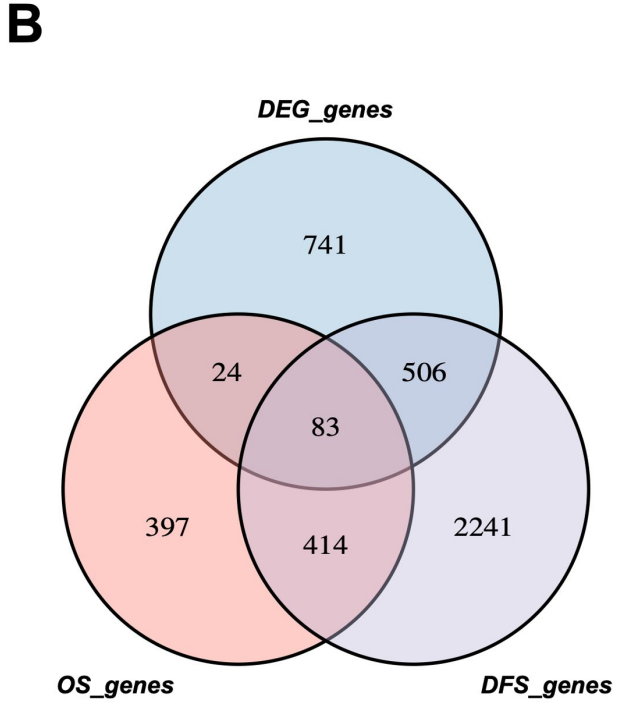
Figure S3. Survival nomogram of LEP+PTTG1 model in validation cohort

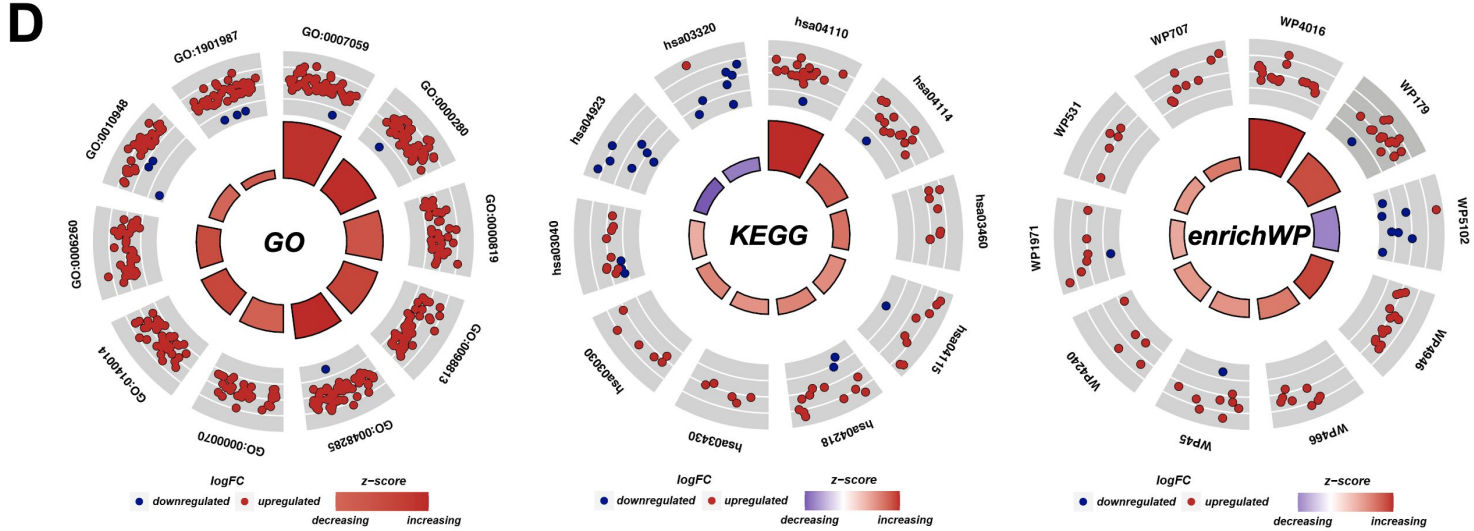
Nomograms for DFS was developed in REASR cohort with four factors: sex, age, risk score, and differentiation **(A)**. ROC curves of 1-, 2-, and 3-year DFS in validation cohort **(B)**. Calibration curves of predicting 1-, 2-, and 3-year DFS **(C)**.


GO/KEGG/enrich WP
(Prognostic genes)
GSEA
(Differential genes)
training cohort
(80 RPLS)
Signals scores
(ssGSEA)
DBSCAN
Surgery strategy of DDLS
WDLS
DDLS
MLS
Re-classify
Differential genes
NMF
Surgery strategy of DDLS
❖ Clinical outcomes and characteristics
❖ Biological features
❖ Immune infiltration
❖ Clinical outcomes and characteristics
❖ Biological features
❖ Immune infiltration
Normal vs Tumor
(8 RPLS)
Cell cycle/DNA damage & repair/
Metabolism (signals)
RPLS subgroups
(G1 and G2)
RPLS clusters
(C1 and C2)
Survival curve
Validation cohort
(241 RPLS)
SVM: AUC>0.99
RF: AUC=1.00
LR: AUC>0.99
Model constructed
(LEP+PTTG1 model)
SVM/RF/LR
(LEP+, PTTG1-)
Identify biomarkers of C1 and C2
Points
Sex
Age
Score
Differentiation
Total points
Pr (OS/DFS)
❖ Clinical outcomes and characteristics
❖ Biological features
❖ Immune infiltration
(PTTG1+, LEP-)
Prognostic validation and
nomogram construction
RPLS dichotomous classification
(Cluster_L and Cluster_H)
Representative biomarkers
(LEP and PTTG1)





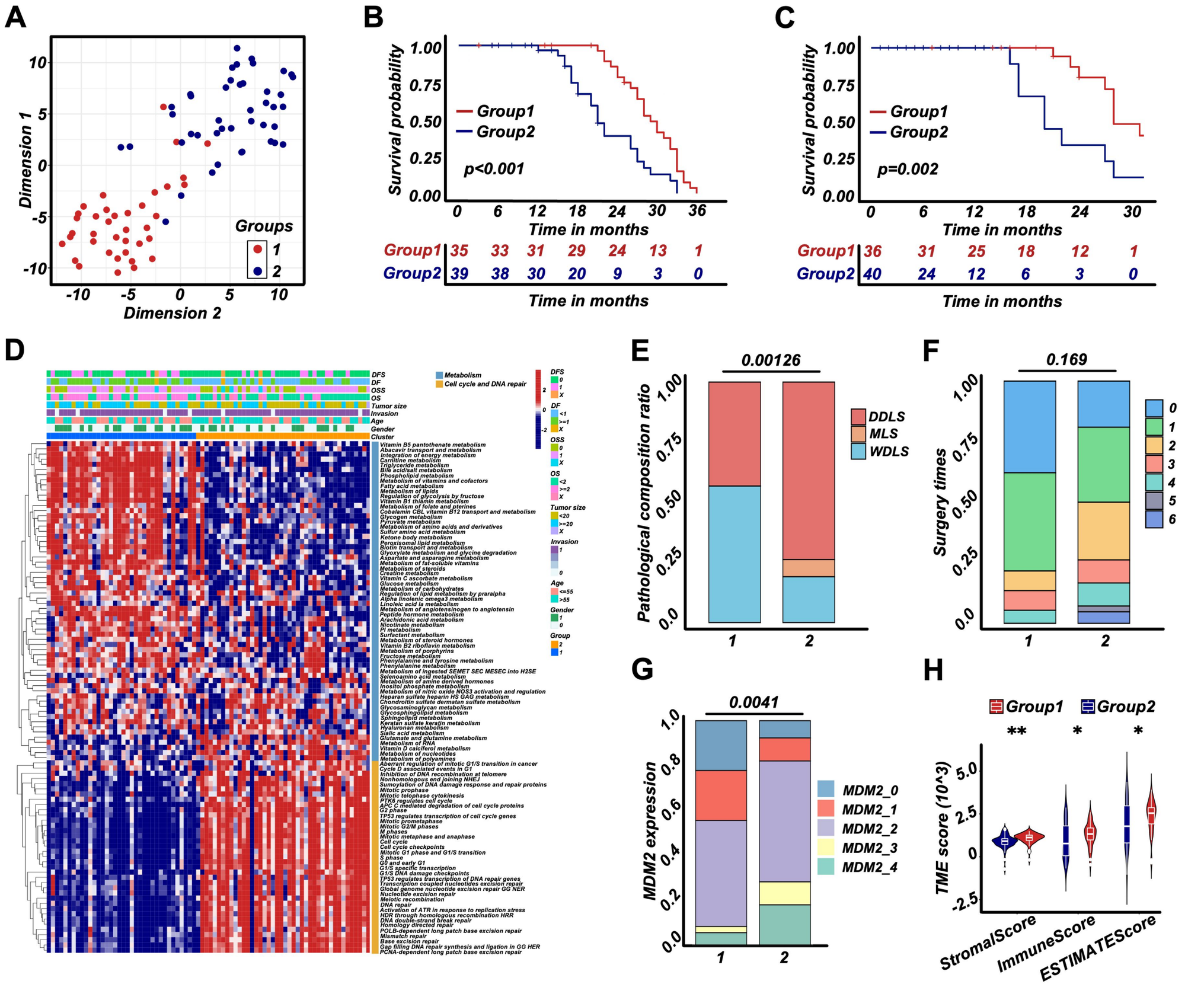


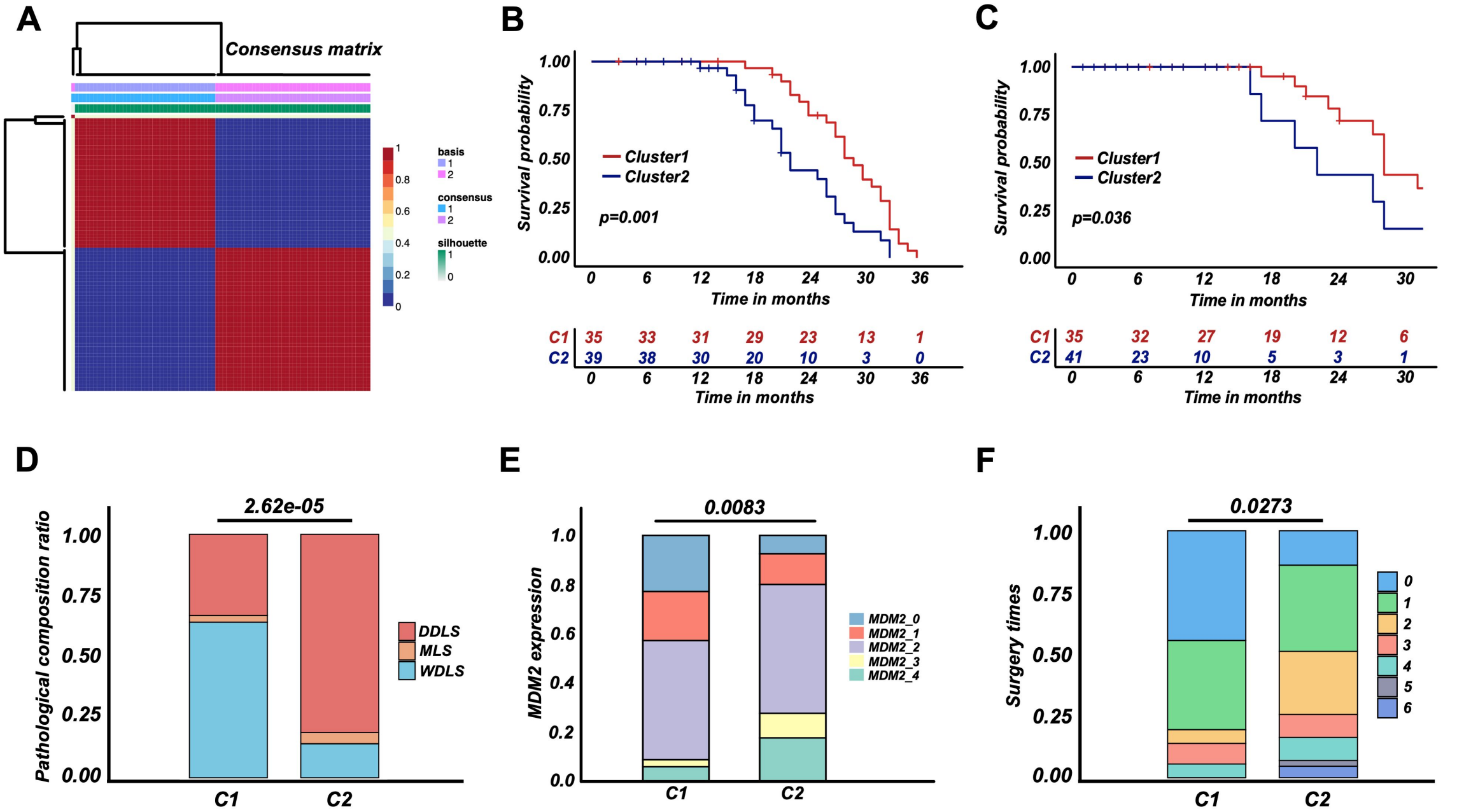


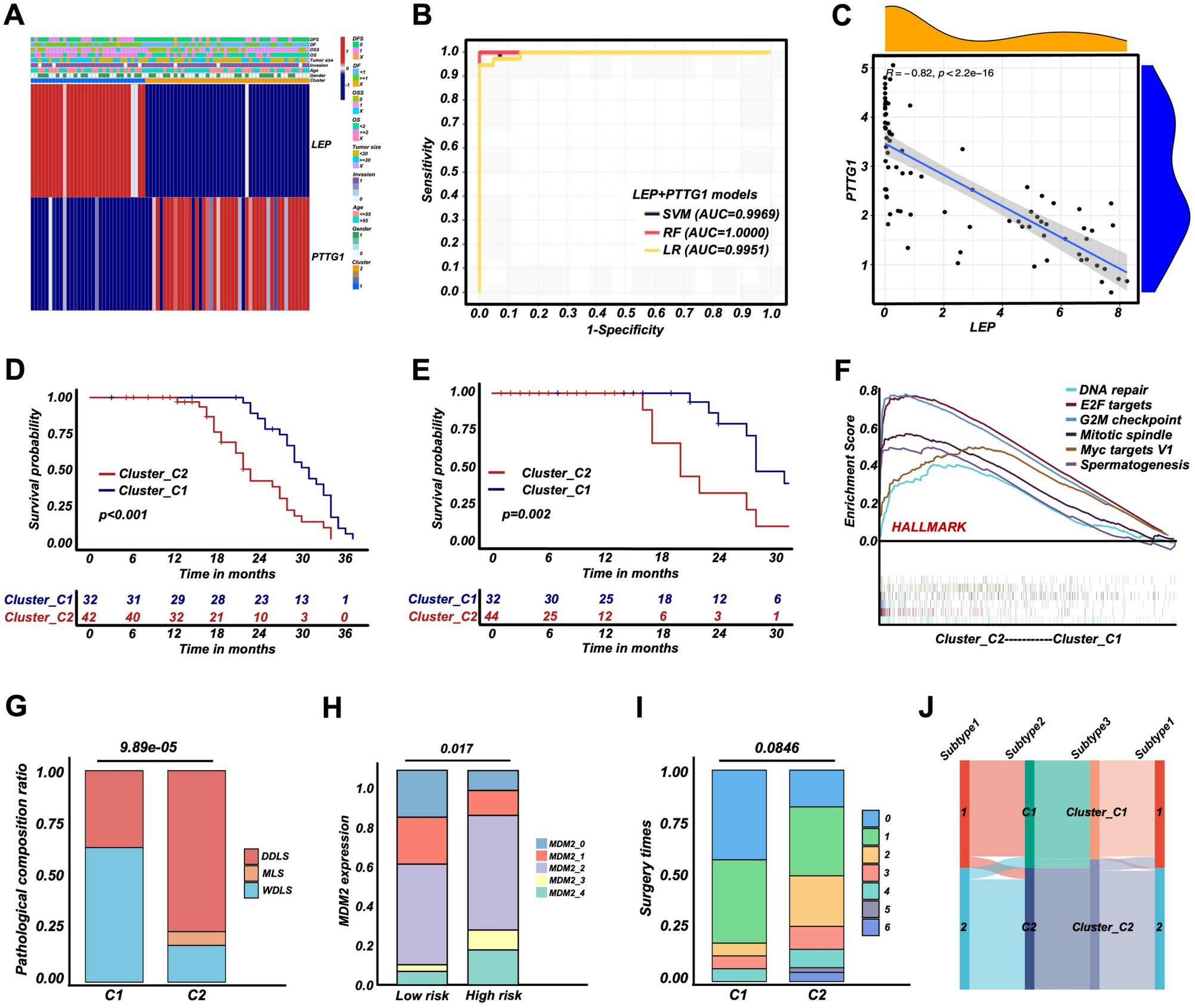
ID	Description
GO:0007059	chromosome segregation
GO:0000280	nuclear division
GO:0000819	sister chromatid segregation
GO:0098813	nuclear chromosome segregation
GO:0048285	organelle fission
GO:0000070	mitotic sister chromatid segregation
GO:0140014	mitotic nuclear division
GO:0006260	DNA replication
GO:0010948	negative regulation of cell cycle process
GO:1901987	regulation of cell cycle phase transition

ID	Description
hsa04110	Cell cycle
hsa04114	Oocyte meiosis
hsa03460	Fanconi anemia pathway
hsa04115	p53 signaling pathway
hsa04218	Cellular senescence
hsa03430	Mismatch repair
hsa03030	DNA replication
hsa03040	Spliceosome
hsa04923	Regulation of lipolysis in adipocytes
hsa03320	PPAR signaling pathway

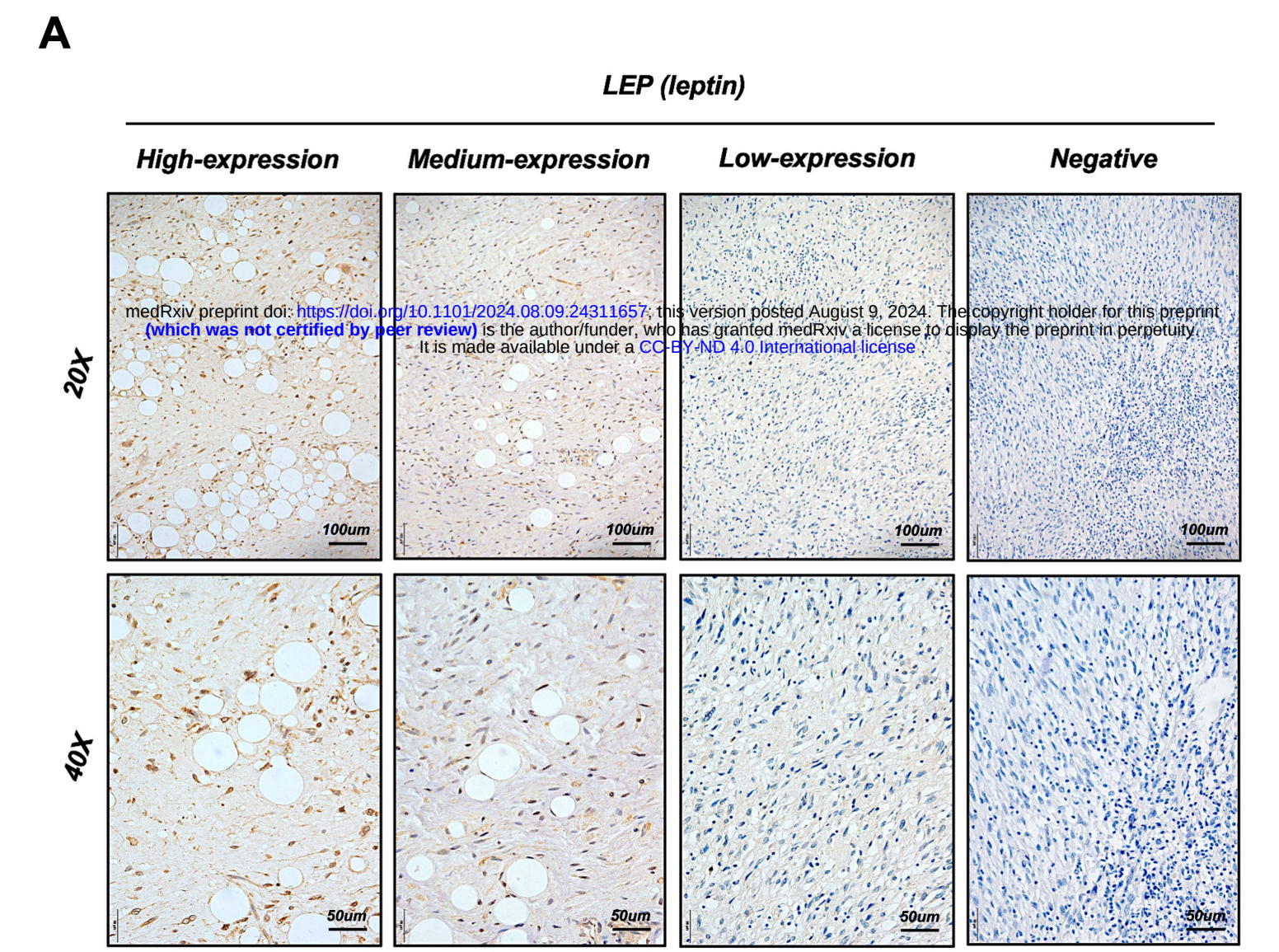
ID	Description
WP4016	DNA IR−damage and cellular response via ATR
WP179	Cell cycle
WP5102	Familial partial lipodystrophy (FPLD)
WP4946	DNA repair pathways, full network
WP466	DNA replication
WP45	G1 to S cell cycle control
WP4240	Regulation of sister chromatid separation at the metaphase−anaphase transition
WP1971	Integrated cancer pathway
WP531	DNA mismatch repair
WP707	DNA damage response

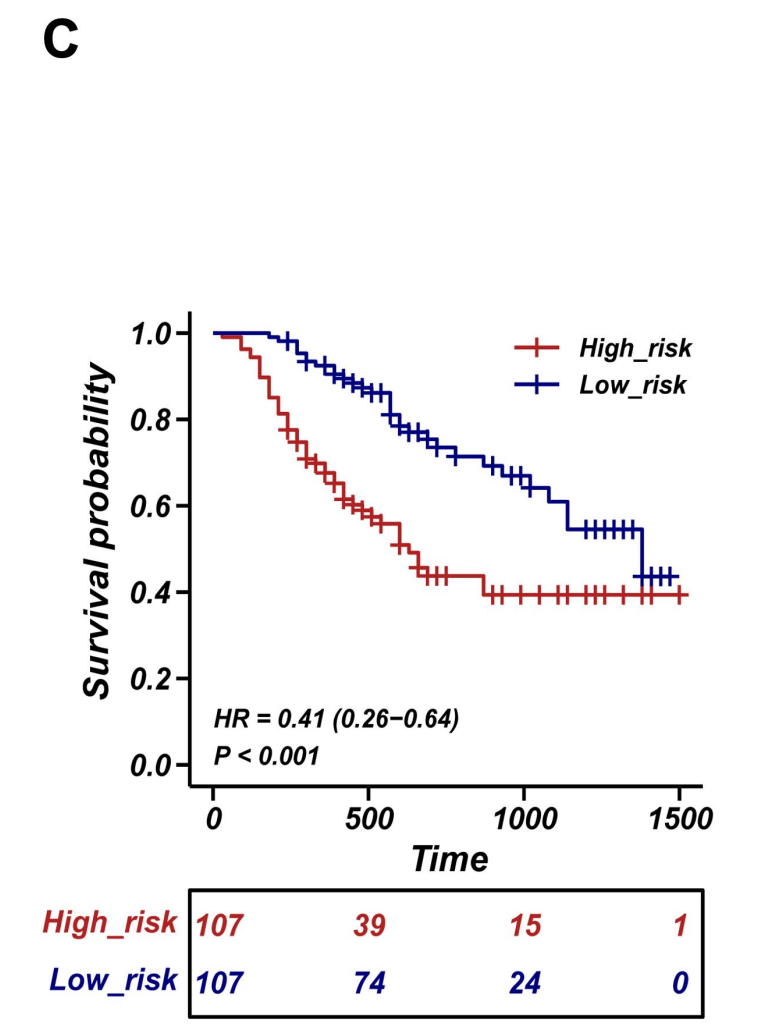

A
Dimension 1
Dimension 2
Groups
1
2
B
Survival probability
Group1
Group2
p<0.001
Time in months
Group1 35 33 31 29 24 13 1
Group2 39 38 30 20 9 3 0
C
Survival probability
Group1
Group2
p=0.002
Time in months
Group1 36 31 25 18 12 1
Group2 40 24 12 6 3 0
D
DFS
DF
OSS
OS
Tumor size
Invasion
Age
Gender
Cluster
Metabolism
Cell cycle and DNA repair
E
0.00126
Pathological composition ratio
DDLS
MLS
WDLS
F
0.169
Surgery times
G
0.0041
MDM2 expression
MDM2_0
MDM2_1
MDM2_2
MDM2_3
MDM2_4
H
Group1
Group2
TME score (10^3)
StromalScore
ImmuneScore
ESTIMATEScore

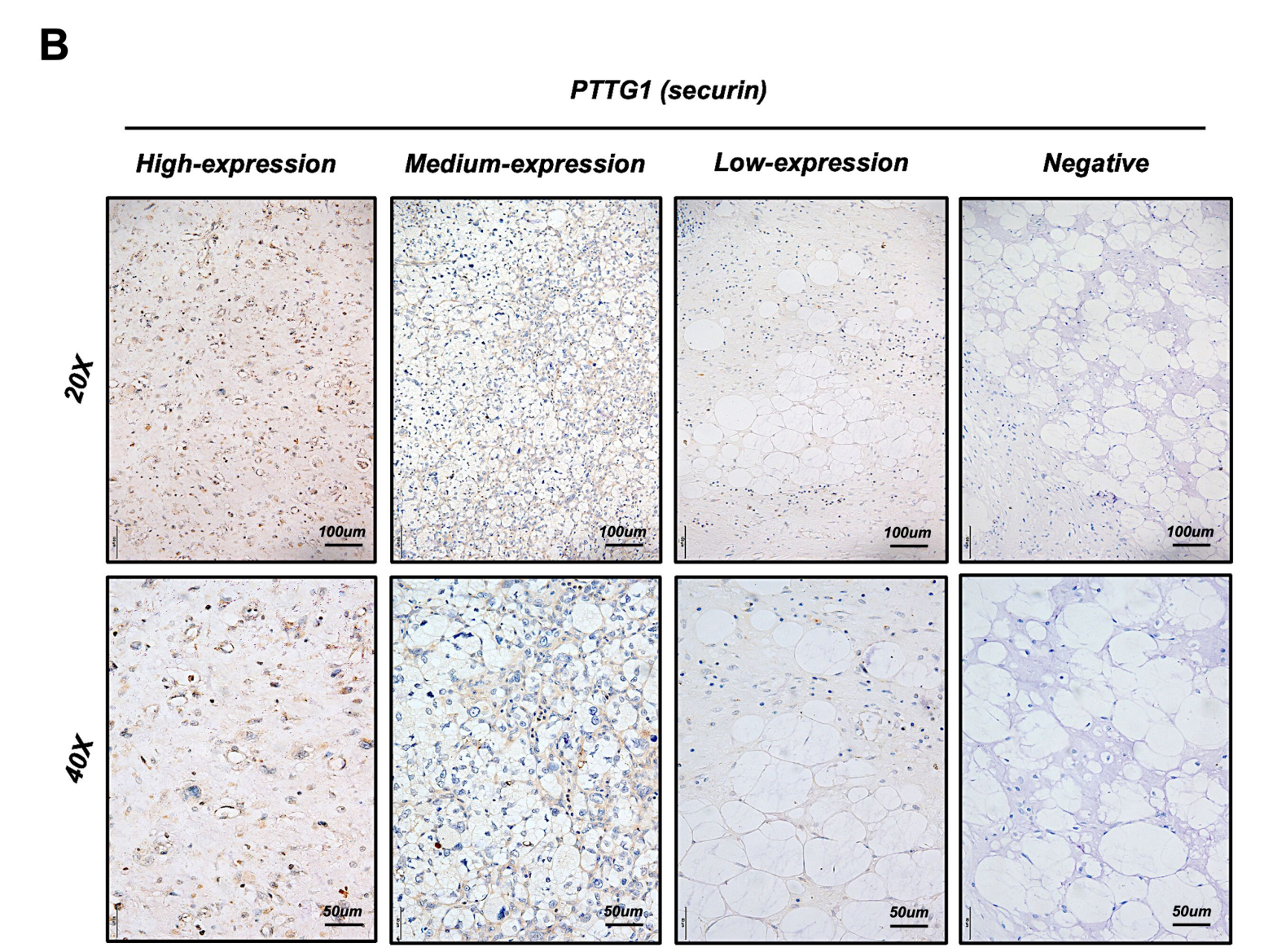

A
Consensus matrix
basis
1
2
consensus
1
2
silhouette
1
0
1
0.8
0.6
0.4
0.2
0
B
Survival probability
1.00
0.75
0.50
0.25
0.00
Cluster1
Cluster2
p=0.001
0 6 12 18 24 30 36
Time in months
C1 35 33 31 29 23 13 1
C2 39 38 30 20 10 3 0
0 6 12 18 24 30 36
Time in months
C
Survival probability
1.00
0.75
0.50
0.25
0.00
Cluster1
Cluster2
p=0.036
0 6 12 18 24 30
Time in months
C1 35 32 27 19 12 6
C2 41 23 10 5 3 1
0 6 12 18 24 30
Time in months
D
2.62e-05
Pathological composition ratio
1.00
0.75
0.50
0.25
0.00
DDLS
MLS
WDLS
C1
C2
E
0.0083
MDM2 expression
1.0
0.8
0.6
0.4
0.2
0.0
MDM2_0
MDM2_1
MDM2_2
MDM2_3
MDM2_4
C1
C2
F
0.0273
Surgery times
1.00
0.75
0.50
0.25
0.00
0
1
2
3
4
5
6
C1
C2

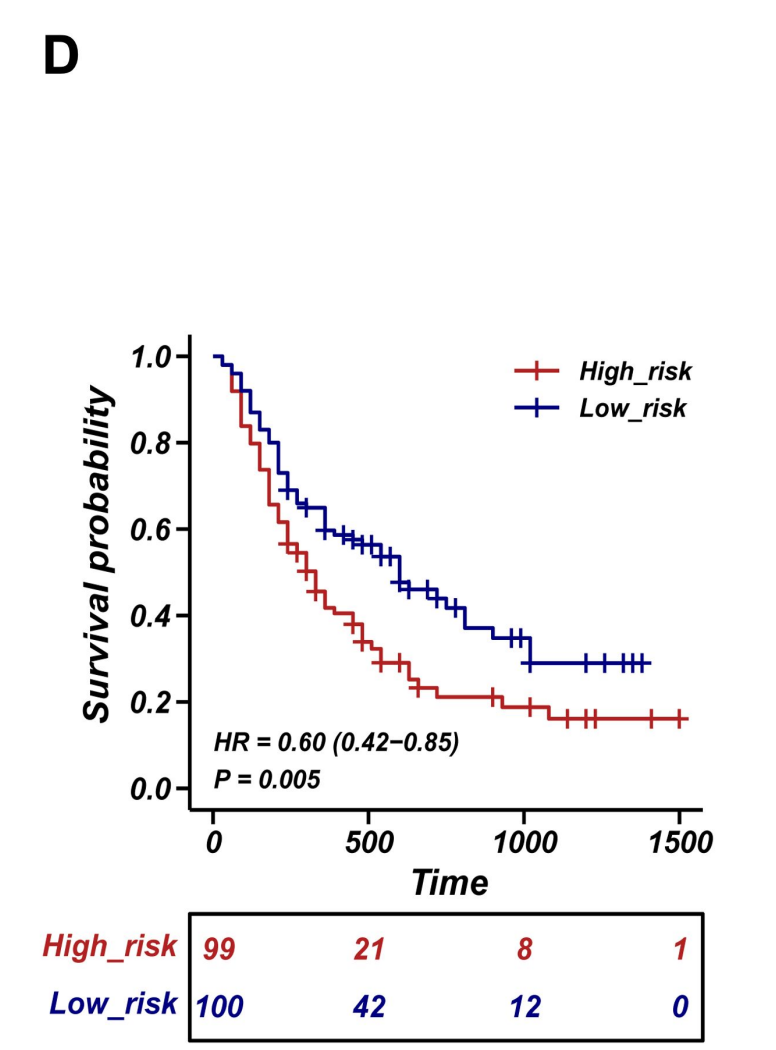

A
DFS
DF
OSS
OS
Tumor size
Invasion
Age
Gender
Cluster
LEP
PTTG1
B
Sensitivity
1-Specificity
LEP+PTTG1 models
SVM (AUC=0.9969)
RF (AUC=1.0000)
LR (AUC=0.9951)
C
R = −0.82, p < 2.2e−16
PTTG1
LEP
D
Survival probability
Cluster_C2
Cluster_C1
p<0.001
Time in months
Cluster_C1 32 31 29 28 23 13 1
Cluster_C2 42 40 32 21 10 3 0
E
Survival probability
Cluster_C2
Cluster_C1
p=0.002
Time in months
Cluster_C1 32 30 25 18 12 6
Cluster_C2 44 25 12 6 3 1
F
Enrichment Score
DNA repair
E2F targets
G2M checkpoint
Mitotic spindle
Myc targets V1
Spermatogenesis
HALLMARK
Cluster_C2----------Cluster_C1
G
9.89e-05
Pathological composition ratio
DDLS
MLS
WDLS
C1
C2
H
0.017
MDM2 expression
MDM2_0
MDM2_1
MDM2_2
MDM2_3
MDM2_4
Low risk
High risk
I
0.0846
Surgery times
C1
C2
J
Subtype1
Subtype2
Subtype3
Subtype1
C1
Cluster_C1
C2
Cluster_C2

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

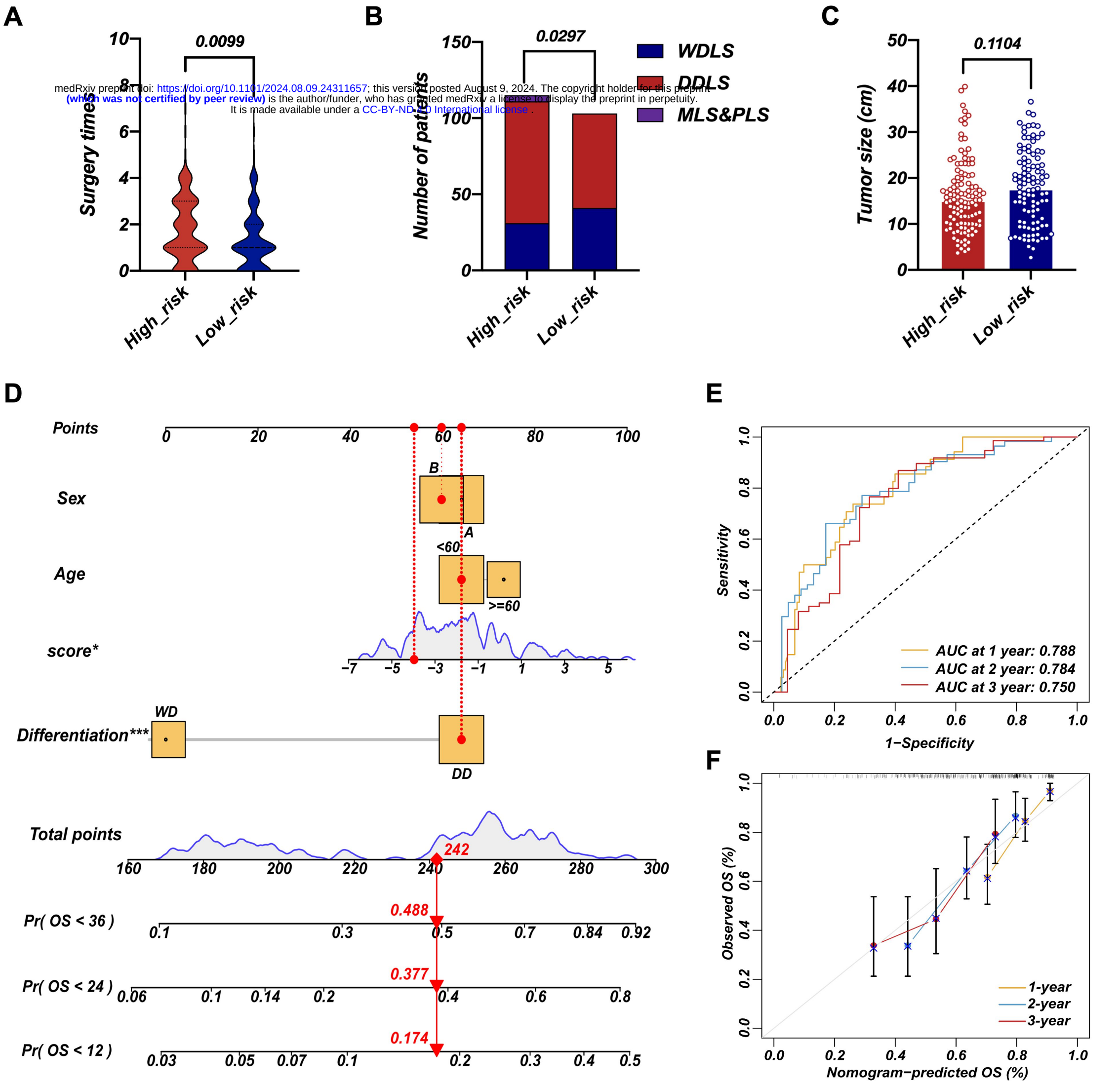








medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND-4.0 International license .



medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Table 1 Baseline characteristics of training cohort and validation cohort

	Training cohort (N=80)[†]	**Validation cohort (N=241)**	***P* value**
Age (y)	56.34 (11.14)[‡]	55.11 (10.80)[‡]	0.384
Sex			
Male	37 (46.25)	118 (48.96)	0.674
Female	43 (53.75)	123 (51.04)	
Pathology			
WDLS	29 (36.25)	75 (31.12)	0.078
DDLS	48 (60.00)	144 (59.75)	
MLS and PLS	3 (3.75)	5 (2.07)	
NR	0 (0)	17 (7.06)	
Surgery times[††]			
0-1	50 (62.50)	143 (59.34)	0.834
2-3	21 (26.25)	73 (30.29)	
4-7	9 (11.25)	24 (9.96)	
NR	0 (0)	1 (0.41)	
Tumor size			
All	18.65 (8.70)[‡]	16.90 (7.94)[‡]	0.101
<18 cm	40 (50.00)	135 (56.02)	0.095
>18 cm	39 (48.75)	92 (38.17)	
NR	1 (1.25)	14 (5.81)	
Multilocation			
Yes	52 (65.00)	153 (63.49)	0.081
No	28 (35.00)	74 (30.71)	
NR	0 (0)	14 (5.80)	
MDM2 score			
0	11 (13.75)	NA	NA
1	12 (15.00)	NA	
2	39 (48.75)	NA	
3	5 (6.25)	NA	
4	9 (11.25)	NA	
NR	4 (5.00)	NA	
LEP score	NA	1.62 (0.82)[‡]	NA
LEP strength			
0	NA	10 (4.15)	NA
1	NA	39 (16.18)	
2	NA	74 (30.71)	
3	NA	115 (47.72)	
NR	NA	3 (1.24)	
PTTG1 score	NA	0.83 (0.75)[‡]	NA
PTTG1 strength			
0	NA	38 (15.77)	NA

medRxiv preprint doi: https://doi.org/10.1101/2024.08.09.24311657; this version posted August 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND-4.0 International license .

1	NA	100 (41.49)	
2	NA	61 (25.31)	
3	NA	39 (16.18)	
NR	NA	3 (1.25)	

†Clinical information missing in 8 RPLS patients (Training cohort 2)

‡The data is shown as Mean (SD); other data is shown as Number (%)

††The definition of surgical times is the sum of current admission surgery and previous surgical resection

DDLS, Dedifferentiated liposarcoma; LEP, Leptin; MDM2, Mouse double minute 2; MLS, Myxoid liposarcoma; NA, Not applicable; NR, Not reported; PLS, Pleomorphic liposarcoma; PTTG1, Pituitary tumor transforming gene 1; WDLS, Well-differentiated liposarcoma