1	Molecular Feature-Based Classification of Retroperitoneal Liposarcoma:
2	A Prospective Cohort Study
3	
4	Mengmeng Xiao ^{1,3†} , Xiangji Li ^{2,3†} , Fanqin Bu ^{2†} , Shixiang Ma ³ , Xiaohan Yang ² , Jun Chen ³ , Yu
5	Zhao ² , Ferdinando Cananzi ⁴ , Chenghua Luo ^{1,3*} , Li Min ^{2*}
6	
7	¹ Department of General Surgery, Peking University People's Hospital, 100044 Beijing, P. R.
8	China
9	² Department of Gastroenterology, Beijing Friendship Hospital, Capital Medical University,
10	State Key Laboratory for Digestive Health, National Clinical Research Center for Digestive
11	Disease, Beijing Digestive Disease Center, Beijing Key Laboratory for Precancerous Lesion of
12	Digestive Disease, 100050 Beijing, P. R. China
13	³ Department of Retroperitoneal Tumor Surgery, Peking University International Hospital,
14	102206 Beijing, P. R. China
15	⁴ Department of Biomedical Sciences, Humanitas University, 20089, Milan, Italy
16	[†] These authors contributed equally to this work
17	Running title: Molecular feature-based classification of RPLS
18	Corresponding author:
19	Prof. Li Min, PhD, Email: minli@ccmu.edu.cn
20	Prof. Chenghua Luo, MD, PhD, Email: pkuihlch@163.com
21	
22	

23 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.

27 Methods: RNA sequencing was performed on a training cohort of 88 RPLS patients to identify 28 dysregulated genes and pathways using clusterprofiler. The GSVA algorithm was utilized to 29 assess signaling pathways levels in each sample, and unsupervised clustering was employed to 30 distinguish RPLS subtypes. Differentially expressed genes (DEGs) between RPLS subtypes were 31 identified to construct a simplified dichotomous clustering via nonnegative matrix factorization. 32 The feasibility of this classification was validated in a separate validation cohort (n=241) using 33 immunohistochemistry (IHC) from the Retroperitoneal SArcoma Registry (RESAR). The study 34 is registered with ClinicalTrials.gov under number NCT03838718.

35 **Results:** Cell cycle, DNA damage & repair, and Metabolism were identified as the most aberrant 36 biological processes in RPLS, enabling the division of RPLS patients into two distinct subtypes 37 with unique molecular signatures, tumor microenvironment, clinical features and outcomes 38 (overall survival, OS and disease-free survival, DFS). A simplified RPLS classification based on 39 representative biomarkers (LEP and PTTG1) demonstrated high accuracy (AUC>0.99), with 40 patients classified as LEP+ and PTTG1- showing lower aggressive pathological composition 41 ratio and fewer surgery times, along with better OS (HR=0.41, P<0.001) and DFS (HR=0.60, 42 *P*=0.005).

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

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

48 **INTRODUCTION**

49 Retroperitoneal liposarcoma (RPLS) is a soft tissue sarcoma (STS) originating in the 50 retroperitoneum with an insidious onset. Traditional surgical resection has been regarded as a 51 primary and curable treatment strategy of RPLS for the past fifty years (Ecker et al., 2016). 52 However, the anatomical complexity and biological properties of sarcoma brought great 53 difficulty to achieve microscopically margin-negative resection, leading to a high postoperative 54 recurrence rate in RPLS patients. During the past decade, scientists tried to improve the 55 patients postoperative survival of RPLS by personalized surgical resection and 56 neoadjuvant/adjuvant therapies, but the effect was not satisfactory (Littau et al., 2020; Gronchi

57 et al., 2015; Gronchi et al., 2009; Gronchi et al., 2013; Pisters et al., 2009).

58 Recently, precision medicine greatly enriched the therapeutic approaches and reformed the 59 clinical decision-making chain of tumor diagnosis and treatment, prolonging the median survival 60 of main tumor types 2-10 times (Kam et al., 2021; Zeng et al., 2022; Alifrangis et al., 2019; 61 *Frese et al.*, 2021). Biomarker-based patient stratification and targeted therapy together make up 62 the kernel of precision medicine, which is intrinsically based on the molecular profiling of 63 cancers. However, our knowledge of the molecular features of RPLS is limited, and few 64 clinically applicable molecular biomarkers and targeted drugs are available for RPLS treatment. 65 Only sporadic molecules such as CDK4 (*Pilotti et al., 2000*), MDM2 (*Binh et al., 2005*), 66 AURK4 (Yen et al., 2019) and CCNDBP1 (Yang et al., 2021) have been reported as prognostic 67 and diagnostic biomarkers, but these biomarkers were poorly represented and verified. Therefore, 68 it is crucial to reveal the molecular landscape of RPLS and explore a feasible classification for its 69 diagnosis and treatment.

70 Here, we conducted a comprehensive inbestigation into the molecular characteristics of 71RPLS through the delineation of the largest gene expression landscape ever assembled for this 72 rare disease entity. By identifying both RPLS-specific genes and prognostic biomarkers, we 73 unveiled their intricate relationships with clinical parameters. Our findings revealed the existence 74 of two distinct molecular subtypes within all RPLS patients, characterized by diverse 75 pathological compositions, enriched signaling pathways, and varying clinical outcomes. This 76 highlights the limitations of relying solely on traditional pathological classification for surgical 77 decision-making in certain cases where patients exhibit favorable histological features but poor 78 prognoses. Emphasizing the pivotal role of molecular subtyping in guiding individualized 79 treatment strategies and enhancing patient management. To facilitate practical application in 80 clinical settings, we developed a simplified RPLS classification system based on key biomarkers 81 (LEP and PTTG1) representative of each subtype. Notably, this classification scheme was 82 validated in a larger cohort of RPLS patients through immunohistochemistry assays (Figure 1), 83 laying the groundwork for precise surgical interventions guided by molecular insights in the 84 realm of RPLS treatment.



85

86 Figure 1. Flow diagram of exploring RPLS dichotomous classification

87

88 MATERIALS AND METHODS

89 Patients and tissue specimens

90 Patients who diagnosed with RPLS amenable to surgical resection were eligible for the 91 study. The RPLS histology was confirmed according to the WHO criteria done on biopsy or 92 surgical specimen by dedicated sarcoma pathologist. The exclusion criteria included the age<18 93 years; serious psychiatric disease that precludes informed consent or limits compliance; 94 impossibility to ensure adequate follow-up. Tumor specimens from 88 RPLS patients (Training 95 cohort 1, Table S1; Training cohort 2, Table S2) and another cohort of 241 RPLS patients 96 (Validation cohort, Table S3) were obtained from our local Hospital. These cohorts are sourced 97 from Retroperitoneal SArcoma Registry (RESAR, NCT03838718). All the patients underwent

98 curative resection from January 2015 to May 2019. RPLS tissue specimens were snap-frozen in 99 liquid nitrogen within 1 h and then stored in a -80 °C refrigerator before use. Clinical information 100 was collected from the medical records, and no patient had undergone previous chemotherapy or 101 radiotherapy. Overall survival (OS) was defined as the interval between the latest surgery and 102 death from tumors or between the latest surgery and the last observation taken for surviving 103 patients. Disease-free survival (DFS) was defined as the interval between the latest surgery and 104 diagnosis of relapse or death. Informed consent for surgical procedures and specimen collection 105 were obtained from each patient. This study has been reported in line with the REMARK criteria 106 (McShane et al., 2005).

107 **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

121 NovaSeq 6000 platform, and single-end reads were generated (Novogene Bioinformatic
122 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.

128 *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".

132 Identification of prognostic genes

133 Cox univariate regression analysis was used to screen the prognostic genes of RPLS. 134 Results of P < 0.05 was considered significant. This process was conducted with the R package 135 "survival".

136 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 (<u>www.gsea-msigdb.org</u>) 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.

144 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.

149 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 tdistributed 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".

156 Consensus clustering with nonnegative matrix factorization (NMF)

157 Nonnegative matrix factorization (NMF) was used to perform RPLS subtyping. Specifically, 158 NMF was applied to gene expression matrix *A* which contained gene sets of major signaling 159 pathways and prognostic genes. Matrix *A* was factorized into 2 nonnegative matrices *W* and *H*. 160 Repeated factorization of matrix *A* was performed and its outputs were aggregated to obtain 161 consensus clustering of RPLS samples. The optimal number of subtypes was selected according 162 to cophenetic, dispersion, and silhouette coefficients. This consensus clustering was conducted 163 with the R package "NMF".

164 *Construction of machine learning models*

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.

168 LR is a statistical method that establishes a relationship between asset of independent 169 variables and a binary outcome. It calculates the probability of an event occurring based on the 170input features derived from biomarkers relevant to surgical patient. SVM is a supervised learning 171algorithm that categorizes data points by identifying the optimal hyperplane that separates 172distinct classes within a high-dimensional space. This approach effectively maps biomarker data 173 into a multidimensional space to facilitate accurate classification of patient outcomes. RF is an 174ensemble learning technique that generated multiple decision trees during training and 175aggregates the results to make predictions. By leveraging this method, we can enhance predictive 176accuracy by mitigating overfitting and increasing model robustness when analyzing biomarker-177driven 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.

184 *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 $15\min \times 3$ in dimethylbenzene and routine

188 hydration, the tissues were soaked in phosphate buffer saline (PBS) for 10min and then 189 performed high-pressure antigen retrieval (Tris-EDTA, PH=9.0) for 2.5min. After being treated 190 with a 3% endogenous catalase blocker (ZSBIO, PV-6000) for 10min, the tissues were incubated 191 in goat serum (ZSBIO, ZLI-9022) for the blocking of nonspecific reaction and then incubated 192 with primary antibody (LEP=1:300 and PTTG1=1:300) at 4°C overnight. The next day, tissues 193 were washed and incubated with goat anti-rabbit secondary antibody (ZSBIO, PV-9000) for 1h 194 at room temperature, then washed and stained with DAB reagents (ZSBIO, ZLI-9018). Then 195 hematoxylin staining, 1% hydrochloric acid alcohol differentiation, ammonia water anti-blue, 196 and neutral gum sealing.

197 The IHC results were evaluated by pathologists, the staining extent was scored as 0-100%. 198 The intensity score was defined as negative, low-expression, medium-expression, and high-199 expression, which were documented as 0, 1, 2, and 3 respectively. The final scores were 200 calculated by the formula: *IHC score* = *Staining extent score* × *Staining intensity score*.

201 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 Oneway 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.

208

209 **RESULTS**

- 210 Baseline characteristics were shown in Table 1. Of 329 RPLS patients, 88 in training cohort and
- 211 241 in validation cohort. No statistically significant differences were found in the age, sex,
- 212 pathology, surgery times, tumor size, and multilocation between the two cohorts (P>0.05). The
- 213 IHC score of LEP and PTTG1 were 1.62 (0.820) and 0.830 (0.75) in validation cohort.
- 214 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)	0.074
Pathology			
WDLS	29 (36.25)	75 (31.12)	
DDLS	48 (60.00)	144 (59.75)	0.078
MLS and PLS	3 (3.75)	5 (2.07)	0.078
NR	0 (0)	17 (7.06)	_
Surgery times ^{††}			
0-1	50 (62.50)	143 (59.34)	
2-3	21 (26.25)	73 (30.29)	0.834
4-7	9 (11.25)	24 (9.96)	0.834
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)	
>18 cm	39 (48.75)	92 (38.17)	0.095
NR	1 (1.25)	14 (5.81)	

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

²¹⁵

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

216 ^{*‡*}*The data is shown as Mean (SD); other data is shown as Number (%)*

- 217 *^{††}The definition of surgical times is the sum of current admission surgery and previous surgical resection*
- 218 DDLS, Dedifferentiated liposarcoma; LEP, Leptin; MDM2, Mouse double minute 2; MLS, Myxoid
- 219 liposarcoma; NA, Not applicable; NR, Not reported; PLS, Pleomorphic liposarcoma; PTTG1, Pituitary tumor
- 220 transforming gene 1; WDLS, Well-differentiated liposarcoma

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

222 To reveal the general molecular features of RPLS compared to noncancerous adipose 223 tissues, we first recruited 8 RPLS patients and collected paired tumor and normal tissues for 224 differentially expressed gene (DEG) analysis. A total of 1354 DEGs, 554 upregulated and 800 225 downregulated, were identified (Figure 2A-B). To assess the underlying pathways of RPLS, 226 GSEA analyses were performed for those DEGs. We found that proliferation-associated 227 pathways, such as Mitotic spindle, E2F target, G2/M checkpoint and Separation of sister 228 chromatids, were mainly enriched in tumors; while metabolism-related pathways, such as Bile 229 acid metabolism, Heme and fatty acid metabolism and Integration of energy metabolism, were 230 enriched in normal controls (Figure 2C).



231

232 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.

236

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-

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

244 **RPLS** subgroups based on molecular features show different clinical outcomes

245 To evaluate heterogeneous molecular clustering characteristics in RPLS, ssGSEA emerged 246 as a widely adopted method for computing the enrichment level of specific biological signaling 247 pathways for each sample based on gene expression data. This aids in gaining insights into the 248 overall activity level of signaling pathways. Here, we scored each sample on the dysregulated 249 pathways by ssGSEA and divided RPLS patients into two subgroups (Figure 3A). Subgroup 1 250 (G1) showed better OS and DFS compared to subgroup 2 (G2) (Figure 3B-C), G1 displayed 251 elevated ssGSEA scores associated with Metabolism, whereas G2 exhibited heightened ssGSEA 252 scores linked on Cell cycle and DNA damage and repair (Figure 3D). These features suggested 253that effective monitoring of the prognosis of RPLS patients can be achieved based on the 254 activation status of specific pathways. We also evaluated the clinical features and immune 255infiltration levels



256



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

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

268 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).

275 To develop a simplified RPLS clustering based on DEGs, we adopted NMF and tSNE for a 276 re-classification of those patients. The results showed RPLS patients were also divided into two 277 clusters (Figure 4A and Figure S1F-G). We then annotated the samples of two clusters by 278 ssGSEA and found Cluster1 (C1) was related to metabolic processes, and Cluster2 (C2) was 279 mainly related to the processes of Cell cycle and DNA damage and repair (Figure S2A). Also, 280 C1 showed better OS and DFS, lower pathological composition ratio and MDM2 expression, and 281 fewer surgery times (Figure 4B-F). Lower Ki67 expression and larger tumor size were observed 282 in C2 (Figure S2B-C). Interestingly, the biological annotations of the C1/C2 classification were 283 greatly consistent with G1/G2. Therefore, a simplified RPLS classification strategy derived from 284 RPLS subgroups was provisionally established.



285

286 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.

290 Development of a dichotomous RPLS classification model

291 For NMF classification of RPLS patients, LEP and PTTG1 were identified as representative 292 biomarkers of C1 and C2, respectively (Figure 5A). We aimed to replicate the RPLS 293 classification of C1 and C2 by integrating these two biomarkers with the assistance of machine 294 learning algorithms, and this two-gene panel achieved promising results (Logistic, AUC=0.995; 295 SVM, AUC=0.997; RF, AUC=1.000; Figure 5B). Also, a linear negative correlation between 296 LEP and PTTG1 expression was detected (Figure 5C). Considering the enhanced interpretability 297 and generalization of linear models, we adopted the results of Logistic regression for subsequent 298 analysis (Risk values=2.182×PTTG1-2.204×LEP). The patients marked as high-risk (Cluster_H) 299 exhibited worse OS and DFS than those marked as low-risk (Cluster_L) (Figure 5D-E).

300 Dysregulated pathways, such as DNA repair and Cell cycle regulation, were enriched in 301 Cluster_H (Figure 5F and Figure S2D), and Cluster_H presented more aggressive pathological 302 composition ratio, higher MDM2 levels and marginally increased in surgery times than 303 Cluster L (Figure 5G-I). Similarly, the Cluster H showed higher Ki67 level and smaller tumor 304 sizes (Figure S2E-F). Moreover, a Sankey diagram was drawn to show the correlation among 305 G1/G2, C1/C2, and Cluster_L/H. Cluster_L/H were well-matched to C1/C2 and G1/G2, 306 suggesting LEP and PTTG1 were promising biomarkers for a dichotomous RPLS classification 307 (Figure 5J).



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

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

315 correlation among G1/G2, C1/C2, and Cluster_C1/C2.

316 Validation of the dichotomous RPLS classification in another 241 RPLS patients

317 To validate LEP and PTTG1 as biomarkers for a dichotomous RPLS classification, we 318 performed IHC staining of two biomarkers in validation cohort. The representative images of 319 LEP and PTTG1 with different expression levels were shown in Figure 6A-B. The IHC scores 320 were integrated with the previously fitted coefficients to evaluate the prognosis of RPLS patients 321 (Risk values= $2.182 \times PTTG1_{IHC}$ - $2.204 \times LEP_{IHC}$). The cutoff value of validation cohort is the 322 median of risk value. The high-risk group had worse OS and DFS (Figure 6C-D), along with 323 more surgery times and more aggressive pathological composition ratio (Figure 7A-B), but the 324 difference of tumor size was not observed between the two groups (Figure 7C). Then we 325 constructed visual nomograms for a precise survival prediction of RPLS patients by combining 326 the risk score with clinical features. The predictive abilities of the 1-, 2-, and 3-year OS (Figure 327 7D-F) and DFS (Figure S3A-C) were 0.743-0.788. Together, we proposed a simple and 328 clinically applicable molecular classification strategy for RPLS patients.

329

330 **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

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

335 G2



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

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

340 had more aggressive molecular features and worse clinical outcomes compared to G1, in 341 accordance with previously reported tumor classification (Lindskrog et al., 2021; Yu et al., 2021; 342 Zhang et al., 2022). In fact, Demicco et al., 2017 has integrated SCNA and DNA methylation 343 divide dedifferentiated liposarcoma into two subtypes (S1 and S2), the unfavorable cluster was 344 characterized as JUN amplified (An oncogene that promotes proliferation and metastasis) and 345 lower inferred fraction of immature dendritic cells. However, the patients in *Demicco et al.*, 2017 346 were of complex origin (mixed limbs, trunk and retroperitoneum), providing limited guidance for 347 RPLS molecular classification. Here we reported the first clinically applicable RPLS molecular 348 classification based on RNA sequencing and IHC validation cohorts.

349 To facilitate the clinical application, we constructed a simplified RPLS molecular 350 classification derived from the original Cell cycle/Metabolism subgroups. By NMF algorithm, 351 we identified LEP and PTTG1 as representative biomarkers for each subtype. A model based on 352 IHC staining of LEP and PTTG1 successfully approximated the original dichotomous RPLS 353 classification in biological features and survival outcomes. LEP is an important regulator of basal 354 metabolism and food intake, which is considered a linkage between metabolism and the immune 355 system (Jiménez-Cortegana et al., 2021). Although LEP-based targeting therapies have not yet 356 been fully applied, LEP has already been identified as a potent metabolic reprogramming agent 357 to support antitumor responses in aggressive melanomas (Waldman et al., 2020; De la et al., 358 2019; Rivadeneira et al., 2019). In addition, LEP improves the immunotherapeutic effects by 359 regulating innate and adaptive immune responses via increasing the cytotoxicity of NK cells 360 (Vera et al., 2018), stimulating the proliferation of T/B cells (Vera et al., 2018; Bernotiene et al., 361 2006), and activating DC cells (Hu et al., 2019). Those reported roles of LEP provided a good 362 mechanism explanation on the features of metabolism pathway-enriched, better prognosis, higher

363 TME level of Metabolism subgroup (LEP+). In contrast, PTTG1 acts as a regulator of sister 364 chromatid separation during cell division under physiological conditions (Zou et al., 1999), 365 which is closely linked to genetic instability, aneuploidy, tumor progression, invasion, and 366 metastasis (Heaney et al., 2000; Ramaswamy et al., 2003; Kim et al., 2005; Yu et al., 2003; 367 Teveroni et al., 2021; Romero et al., 2001). PTTG1 also regulates the cell cycle and the 368 transactivation of growth factors as an initiator and promoter of tumorigenesis (Zou et al., 1999; 369 Mora-Santos et al., 2013; McCabe et al., 2002; Ishikawa et al 2001; Hamid et al., 2005). 370 Overexpressing PTTG1 was correlated with worse prognosis in tumors, such as ovarian cancer 371 (Parte et al., 2019), cervical cancer (Guo et al., 2019), renal cell carcinoma (Tian et al., 2022), 372 and colorectal cancer (Heaney et al., 2000). Therefore, the biological functions of PTTG1 373 provided a good mechanism explanation of the pathway-enriched of cell cycle/DNA damage & 374 repair-associated, worse prognosis, and more aggressive pathological composition ratio the Cell 375 cycle subgroup (PTTG1+).

376

377 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.



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

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

392 Abbreviations

- 393 RPLS Retroperitoneal liposarcoma 394 DEGs Differentially expressed genes 395 IHC Immunohistochemistry 396 STS Soft tissue sarcoma 397 OS Overall survival 398 DFS Disease-free survival 399 OMIX Open archive for miscellaneous data 400 CNCB China national center for bioinformation 401 **GSEA** Gene set enrichment analysis 402 GO Gene ontology 403 KEGG Kyoto encyclopedia of gene and genomes 404 t-SNE t-distributed stochastic neighbor embedding 405 DBSCAN Density-based spatial clustering of applications with noise 406 NMF Nonnegative matrix factorization 407 LR Logistic regression 408 Support vector machine SVM 409 RF Random forest 410 AUC Area under curve 411
- 412

413 Authors' contributors

414 Mengmeng Xiao: Methodology, Formal analysis, Data curation, Writing-original draft, Writing-415 review & editing. Xiang Ji Li: Methodology, Formal analysis, Investigation, Data curation, 416 Writing-original draft. Fanqin Bu: Conceptualization, Methodology, Formal analysis, 417 Investigation, Data curation, Writing-review & editing. Shixiang Ma: Methodology, Formal 418 analysis. Xiaohan Yang: Methodology, Investigation. Jun Chen: Formal analysis. Yu Zhao: 419 Investigation. Ferdinando Cananzi: Investigation. Chenghua Luo: Conceptualization, 420 Methodology, Writing-review & editing, Supervision. Li Min: Conceptualization, Resources, 421 Writing-review & editing, Supervision.

422

423 **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.

428

429 **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.

434

435 **Research registration unique identifying number (UIN)**

- 436 1. Name of the registry: ClinicalTrials.gov
- 437 2. Unique identifying number or registration ID: NCT03838718
- 438 3. Hyperlink to your specific registration: <u>https://clinicaltrials.gov</u>
- 439
- 440 **Consent for publication**
- 441 All authors have read and approved the manuscript and agree with submission to *eLife*.

442

443 **Data availability**

444 Data supporting the conclusions of this article are presented within the article and its

- 445 supplementary files.
- 446

447 **Competing interests**

- 448 The authors declare that they have no competing interests.
- 449

450 **REFERENCES**

- 451 Ecker BL, Peters MG, McMillan MT, Sinnamon AJ, Zhang PJ, Fraker DL, Levin WP, Roses
- 452 RE, Karakousis GC. Preoperative radiotherapy in the management of retroperitoneal liposarcoma.
- 453 Br J Surg. 2016; 103(13): 1839-1846. doi: 10.1002/bjs.10305
- 454 Littau MJ, Kulshrestha S, Bunn C, Agnew S, Sweigert P, Luchette FA, Baker MS. The
- 455 importance of the margin of resection and radiotherapy in retroperitoneal liposarcoma. Am J
- 456 Surg. 2021;221(3):554-560. doi:10.1016/j.amjsurg.2020.11.041
- 457 Gronchi A, Miceli R, Allard MA, Callegaro D, Le Péchoux C, Fiore M, Honoré C, Sanfilippo R,
- 458 Coppola S, Stacchiotti S, Terrier P, Casali PG, Cesne AL, Mariani L, Colombo C, Bonvalot S.
- 459 Personalizing the approach to retroperitoneal soft tissue sarcoma: histology-specific patterns of
- 460 failure and postrelapse outcome after primary extended resection. Ann Surg Oncol.
- 461 2015;22(5):1447-1454. doi:10.1245/s10434-014-4130-7
- 462 Gronchi A, Lo Vullo S, Fiore M, Mussi C, Stacchiotti S, Collini P, Lozza L, Pennacchioli E,
 463 Mariani L, Casali PG. Aggressive surgical policies in a retrospectively reviewed single464 institution case series of retroperitoneal soft tissue sarcoma patients. *J Clin Oncol.*465 2009;27(1):24-30. doi:10.1200/JCO.2008.17.8871
- 466 Gronchi A, Pollock RE. Quality of local treatment or biology of the tumor: which are the trump
- 467 cards for loco-regional control of retroperitoneal sarcoma?. Ann Surg Oncol. 2013;20(7):2111-
- 468 2113. doi:10.1245/s10434-013-2971-0
- 469 **Pisters PW**. Resection of some -- but not all -- clinically uninvolved adjacent viscera as part of
- 470 surgery for retroperitoneal soft tissue sarcomas. J Clin Oncol. 2009;27(1):6-8.
- 471 doi:10.1200/JCO.2008.18.7138

- 472 Kam AE, Masood A, Shroff RT. Current and emerging therapies for advanced biliary tract
 473 cancers. *Lancet Gastroenterol Hepatol.* 2021;6(11):956-969. doi:10.1016/S2468474 1253(21)00171-0
- 475 Zeng Y, Jin RU. Molecular pathogenesis, targeted therapies, and future perspectives for gastric
- 476 cancer. Semin Cancer Biol. 2022;86(Pt 3):566-582. doi:10.1016/j.semcancer.2021.12.004
- 477 Alifrangis C, McGovern U, Freeman A, Powles T, Linch M. Molecular and histopathology
- 478 directed therapy for advanced bladder cancer. *Nat Rev Urol.* 2019;16(8):465-483.
 479 doi:10.1038/s41585-019-0208-0
- 480 Frese KK, Simpson KL, Dive C. Small cell lung cancer enters the era of precision medicine.
- 481 *Cancer Cell*. 2021;39(3):297-299. doi:10.1016/j.ccell.2021.02.002
- 482 Pilotti S, Della Torre G, Mezzelani A, Tamborini E, Azzarelli A, Sozzi G, Pierotti MA. The 483 expression of MDM2/CDK4 gene product in the differential diagnosis of well differentiated 484 liposarcoma and large deep-seated lipoma. Br JCancer. 2000;82(7):1271-1275. 485 doi:10.1054/bjoc.1999.1090
- 486 Binh MB, Sastre-Garau X, Guillou L, Pinieux G, Terrier P, Lagacé R, Aurias A, Hostein I, 487 Coindre JM. MDM2 and CDK4 immunostainings are useful adjuncts in diagnosing well-488 differentiated and dedifferentiated liposarcoma subtypes: a comparative analysis of 559 soft 489 Am J Surg Pathol. tissue neoplasms with genetic data. 2005;29(10):1340-1347. 490 doi:10.1097/01.pas.0000170343.09562.39
- 491 Yen CC, Chen SC, Hung GY, Wu PK, Chua WY, Lin YC, Yen CH, Chen YC, Wang JY, Yang
- 492 MH, Chao Y, Chang MC, Chen WM. Expression profile-driven discovery of AURKA as a
- 493 treatment target for liposarcoma. Int J Oncol. 2019;55(4):938-948. doi:10.3892/ijo.2019.4861

- 494 Yang L, Wu Z, Sun W, Luo P, Chen S, Chen Y, Yan W, Li Y, Wang C. CCNDBP1, a
- 495 Prognostic Marker Regulated by DNA Methylation, Inhibits Aggressive Behavior in
- 496 Dedifferentiated Liposarcoma via Repressing Epithelial Mesenchymal Transition. Front Oncol.

497 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.
- 501 Guo L, Chen G, Zhang W, Zhou L, Xiao T, Di X, Wang Y, Feng L, Zhang K. A high-risk
- 502 luminal A dominant breast cancer subtype with increased mobility. Breast Cancer Res Treat.
- 503 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-
- 505 S in retroperitoneal liposarcoma via increasing ROS levels. *Cell Commun Signal*.
 506 2022;20(1):128. doi:10.1186/s12964-022-00873-9
- 507 Lindskrog SV, Prip F, Lamy P, Taber A, Groeneveld CS, Birkenkamp-Demtröder K, Jensen JB,
- 508 Strandgaard T, Nordentoft I, Christensen E, Sokac M, Birkbak NJ, Maretty L, Hermann GG,
- 509 Petersen AC, Weyerer V, Grimm MO, Horstmann M, Sjödahl G, Höglund M, Steiniche T,
- 510 Mogensen K, Reyniès A, Nawroth R, Jordan B, Lin X, Dragicevic D, Ward DG, Goel A, Hurst
- 511 C, Kessel K, Maurer T, Meeks JJ, DeGraff DJ, Bryan RT, Knowles MA, Simic T, Hartmann A,
- 512 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
- 514 *Commun.* 2021;12(1):2301.
- 515 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

517 Cell Cycle Machinery Overcomes Resistance to Oxaliplatin in Colorectal Cancer. *Adv Sci*518 (*Weinh*). 2021;8(23):e2100759.

519 **Zhang F**, Zhang Q, Zhu J, Yao B, Ma C, Qiao N, He S, Ye Z, Wang Y, Han R, Feng J, Wang Y,

520 Qin Z, Ma Z, Li K, Zhang Y, Tian S, Chen Z, Tan S, Wu Y, Ran P, Wang Y, Ding C, Zhao Y.

521 Integrated proteogenomic characterization across major histological types of pituitary 522 neuroendocrine tumors. *Cell Res.* 2022;32(12):1047-1067.

523CancerGenomeAtlasResearchNetwork.Electronicaddress:524elizabeth.demicco@sinaihealthsystem.ca;CancerGenomeAtlasResearchNetwork.

525 Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. *Cell*.

526 2017;171(4):950-965.e28. doi:10.1016/j.cell.2017.10.014

527 Jiménez-Cortegana C, López-Saavedra A, Sánchez-Jiménez F, Pérez-Pérez A, Castiñeiras J,

528 Virizuela-Echaburu JA, Cruz-Merino L, Sánchez-Margalet V. Leptin, Both Bad and Good Actor

529 in Cancer. *Biomolecules*. 2021;11(6):913. doi:10.3390/biom11060913

530 Waldman AD, Fritz JM, Lenardo MJ. A guide to cancer immunotherapy: from T cell basic

531 science to clinical practice. *Nat Rev Immunol.* 2020;20(11):651-668. doi:10.1038/s41577-020-

532 **0306-5**

533 De la Cruz-Merino L, Palazón-Carrión N, Henao-Carrasco F, Nogales-Fernández E, Álamo-de

⁵³⁴ 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-

536 018-1907-3

537 **Rivadeneira DB**, DePeaux K, Wang Y, Kulkarni A, Tabib T, Menk AV, Sampath P, Lafyatis R,

538 Ferris RL, Sarkar SN, Thorne SH, Delgoffe GM. Oncolytic Viruses Engineered to Enforce

- 539 Leptin Expression Reprogram Tumor-Infiltrating T Cell Metabolism and Promote Tumor
- 540 Clearance. *Immunity*. 2019;51(3):548-560.e4. doi:10.1016/j.immuni.2019.07.003
- 541 Vera F, Pino J, Campos-Cabaleiro V, Ruiz-Fernández C, Mera A, Gonzalez-Gay MA, Gómez R,
- 542 Gualillo O. Obesity, Fat Mass and Immune System: Role for Leptin. *Front Physiol.* 2018;9:640.
- 543 doi: 10.3389/fphys.2018.00640
- 544 **Bernotiene E**, Palmer G, Gabay C. The role of leptin in innate and adaptive immune responses.
- 545 Arthritis Res Ther. 2006;8(5):217. doi:10.1186/ar2004
- 546 **Hu W**, Wang G, Huang D, Sui M, Xu Y. Cancer Immunotherapy Based on Natural Killer Cells:
- 547 Current Progress and New Opportunities. *Front Immunol.* 2019;10:1205. 548 doi:10.3389/fimmu.2019.01205
- 549 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-
- 551 422. doi:10.1126/science.285.5426.418
- 552 Heaney AP, Singson R, McCabe CJ, Nelson V, Nakashima M, Melmed S. Expression of
- 553 pituitary-tumour transforming gene in colorectal tumours. *Lancet*. 2000;355(9205):716-719.
 554 doi:10.1016/S0140-6736(99)10238-1
- - **Ramaswamy S**, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary
 - 556 solid tumors. Nat Genet. 2003;33(1):49-54. doi:10.1038/ng1060
 - 557 Kim D, Pemberton H, Stratford AL, Buelaert K, Watkinson JC, Lopes V, Franklyn JA, McCabe
 - 558 CJ. Pituitary tumour transforming gene (PTTG) induces genetic instability in thyroid cells.
 - 559 Oncogene. 2005;24(30):4861-4866. doi:10.1038/sj.onc.1208659

- 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
- 563 Teveroni E, Di Nicuolo F, Bianchetti G, Epstein AL, Grande G, Maulucci G, Spirito MD,
- 564 Pontecorvi A, Milardi D, Mancini F. Nuclear Localization of PTTG1 Promotes Migration and
- 565 Invasion of Seminoma Tumor through Activation of MMP-2. Cancers (Basel). 2021;13(2):212.
- 566 doi:10.3390/cancers13020212
- 567 Romero F, Multon MC, Ramos-Morales F, Domínguez A, Bernal JA, Pintor-Toro JA, Tortolero
- 568 M. Human securin, hPTTG, is associated with Ku heterodimer, the regulatory subunit of the
- 569 DNA-dependent protein kinase. *Nucleic Acids Res.* 2001;29(6):1300-1307.
 570 doi:10.1093/nar/29.6.1300
- 571 Mora-Santos M, Castilla C, Herrero-Ruiz J, Giráldez S, Limón-Mortés MC, Sáez C, Japón MÁ,
- 572 Tortolero M, Romero F. A single mutation in Securin induces chromosomal instability and
- 573 enhances cell invasion. Eur J Cancer. 2013;49(2):500-510. doi:10.1016/j.ejca.2012.06.024
- 574 McCabe CJ, Boelaert K, Tannahill LA, Heaney AP, Stratford AL, Khaira JS, Hussain S,
- 575 Sheppard MC, Franklyn JA, Gittoes NJL. Vascular endothelial growth factor, its receptor
- 576 KDR/Flk-1, and pituitary tumor transforming gene in pituitary tumors. J Clin Endocrinol Metab.
- 577 2002;87(9):4238-4244. doi:10.1210/jc.2002-020309
- 578 Ishikawa H, Heaney AP, Yu R, Horwitz GA, Melmed S. Human pituitary tumor-transforming
- 579 gene induces angiogenesis. J Clin Endocrinol Metab. 2001;86(2):867-874.
 580 doi:10.1210/jcem.86.2.7184

- 581 **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.
- 583 doi:10.1186/1476-4598-4-3
- 584 Parte S, Virant-Klun I, Patankar M, Batra SK, Straughn A, Kakar SS. PTTG1: a Unique
- 585 Regulator of Stem/Cancer Stem Cells in the Ovary and Ovarian Cancer. Stem Cell Rev Rep.
- 586 2019;15(6):866-879. doi:10.1007/s12015-019-09911-5
- 587 Guo XC, Li L, Gao ZH, Zhou HW, Li J, Wang QQ. The long non-coding RNA PTTG3P
- 588 promotes growth and metastasis of cervical cancer through PTTG1. Aging (Albany NY).
- 589 2019;11(5):1333-1341. doi:10.18632/aging.101830
- 590 Tian X, Xu WH, Xu FJ, Li H, Anwaier A, Wang HK, Wan FN, Zhu Y, Cao DL, Zhu YP, Shi
- 591 GH, Qu YY, Zhang HL, Ye DW. Identification of prognostic biomarkers in papillary renal cell
- 592 carcinoma and PTTG1 may serve as a biomarker for predicting immunotherapy response. Ann
- 593 *Med.* 2022;54(1):211-226.
- 594

595 **Table 1. Baseline characteristics of training cohort and validation cohort**

597 Figure Legend

598 Figure 1. Flow diagram of exploring RPLS dichotomous classification

599

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

601 Volcano plot of the DEGs in 8 normal vs 8 RPLS tissues (A). Venn diagram showed shared

602 genes between DEGs and prognostic genes (B). GSEA analysis of RPLS tumors, including

603 HALLMARK gene sets and REACTOME gene sets (C). Circular plots of the prognostic genes

- 604 in GO, KEGG, and enrichWP (**D**).
- 605

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

608 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**).

610 Histograms revealed the difference of pathological composition ratio (E), surgery times (F), and

611 MDM2 (G) in G1 and G2. Violin plot of the microenvironmental scores in G1 and G2 (H).

612

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

614 NMF for a re-classification of training cohort 1 (C1 and C2) (A). Survival cures of OS (B) and

615 DFS (C) in C1 and C2. Histograms revealed the difference in pathological composition ratio (D),

616 MDM2 (E), and surgery times (F) in C1 and C2.

617

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

619 **RPLS clusters**

620	Heatmap of biomarkers identified (LEP and PTTG1) in C1 and C2 (A). ROC curves of the
621	machine learning models to identify C1 and C2 (B). Correlation between LEP and PTTG1
622	expression (C). Survival curves of OS (D) and DFS (E) in Cluster_C1 (low risk) and Cluster_C2
623	(high risk) groups. GSEA of HALLMARK gene sets in Cluster_C1 and Cluster_C2 (F).
624	Histograms revealed the difference of pathological composition ratio (G), MDM2 level (H), and
625	surgery times (I) in Cluster_C1 and Cluster_C2. Sankey diagram indicated the correlation among
626	G1/G2, C1/C2, and Cluster_C1/C2 (J).
627	
628	Figure 6. Validation of the RPLS dichotomous classification in another 241 RPLS cohort
629	Representative IHC staining images of LEP (A) and PTTG1 (B). Survival curves of OS (C) and

630 DFS (**D**) in high-risk and low-risk groups.

631

632 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**).

639 Supplementary Information

- 640 **Table S1.** The detailed clinicopathological characteristics of the training cohort 1
- 641 **Table S2.** The RNA-seq data of the training cohort 2
- 642 **Table S3.** The detailed clinicopathological characteristics of the validation cohort
- 643 **Table S4.** The RNA-seq data of the training cohort 1
- 644

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

- 646 The difference of Ki67 (A) and tumor size (B) in G1 and G2. Volcano plot of the DEGs (G1 vs
- 647 G2) (C). Venn diagram showed shared genes between DEGs and prognostic genes (D). Bubble

648 plot of the DEGs enrichment (E). NMF for a re-classification of training cohort 1 (F). tSNE

649 exhibits the RPLS clusters (C1 and C2) (G).

650

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.

657

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
- 660 differentiation (A). ROC curves of 1-, 2-, and 3-year DFS in validation cohort (B). Calibration
- 661 curves of predicting 1-, 2-, and 3-year DFS (C).











Cell cycle

Oocyte meiosis

Fanconi anemia pathway

p53 signaling pathway

Cellular senescence

Mismatch repair

DNA replication

Spliceosome

PPAR signaling pathway

am		8.0	
	StdM	Mb466	
	log EC	7.00	
	logro	2-500	bre
• do	wnregulated 🗕 upregulated		
		decreasing	increasing
ID	Descrip	otion	
VP4016	DNA IR-damage and ce	llular respons	se via ATR
WP179	Cell c	ycle	
VP5102	Familial partial lipe	odystrophy (F	PLD)
VP4946	DNA repair pathy	vays, full netw	vork
WP466	DNA rep	lication	
WP45	G1 to S cell of	ycle control	
VP4240	Regulation of sister	chromatid se	eparation
NP1971	Integrated car	ncer pathway	shon
WP531	DNA misma	atch repair	
WP707	DNA damag	e resnonse	

HALLMARK

Mitotic spindle

REACTOME Bile acid/salt metabolism

Separation of sister chromatids

WP4016





D





1.0

MDM2 expression 70 0.0 8.0 8.0 8.0

0.2

0.0



Surgery times











С

D

LEP (leptin)





В

PTTG1 (securin)







0.4 0.6 0.8 Nomogram-predicted OS (%) 1.0

0.0

0.5

0.4

0.2

Α

0.03

0.05

0.07

0.1

0.2

0.3

	Training cohort (N=80) [†]	Validation cohort (N=241)	P value
Age (v)	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)	1
4-7	9 (11.25)	24 (9.96)	1
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			1
Yes	52 (65.00)	153 (63.49)	0.081
No	28 (35.00)	74 (30.71)	-
NR	0 (0)	14 (5.80)	
MDM2 score			1
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	$1.62 (0.82)^{\ddagger}$	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	$0.83 (0.75)^{\ddagger}$	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.

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