Proteogenomic profiling of soft tissue leiomyosarcoma reveals distinct molecular subtypes with divergent outcomes and therapeutic vulnerabilities

Atsushi Tanaka[1,2,3]*, Makiko Ogawa[1,3], Yusuke Otani[1], Ronald C. Hendrickson[4,5,6], Zhuoning Li[7], Narasimhan P. Agaram[8], David S. Klimstra[3,9], Julia Y. Wang[10], and Michael H. A. Roehrl[1,3,11,12]*

[1]Department of Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA

[2]Broad Institute of MIT and Harvard, Cambridge, MA, USA

[3]Former address: Department of Pathology and Laboratory Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[4]Department of Biochemistry and Molecular Biology, Miller School of Medicine, University of Miami, Miami, FL, USA

[5]Sylvester Comprehensive Cancer Center, Miller School of Medicine, University of Miami, Miami, FL, USA

[6]Former address: Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[7]Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[8]Department of Pathology and Laboratory Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[9]Department of Pathology, Yale University School of Medicine, New Haven, CT, USA

[10]Curandis, Boston, MA, USA

[11]Wyss Institute at Harvard, Boston, MA, USA

[12]Lead contact

*Correspondence to: atanaka@bidmc.harvard.edu, michael_roehrl@bidmc.harvard.edu

Running title: Proteogenomics of leimyosarcoma

Summary

Soft tissue leiomyosarcoma (STLMS) is an aggressive malignancy lacking validated molecular subclassification and effective targeted treatments. We performed comprehensive proteogenomic analysis of primary and metastatic STLMS to uncover biological traits and therapeutic weaknesses. Integrative proteomic and phosphoproteomic analyses using non-negative matrix factorization identified three subtypes. Subtype P1 shows genomic stability, low proliferation, and enrichment of FGFR2 and PDK signaling pathways. Subtype P2 exhibits chromosomal instability, inflammatory programs, activation of CDK-AURKA/B-mTOR/ERK kinome with *IGF1R*/*PDGFRA* gene alterations, and poorest survival outcomes. Subtype P3 is highly proliferative, with E2F/DNA-repair programs, elevated NCOR1, and shift towards non-homologous end joining with upregulation of PARP1. Homologous recombination deficiency (HRD) analysis distinguishes HRD-low P1 from HRD-high P2/P3. Paired analyses suggest HRD increases in metastases within P3. Immune profiling shows P2 as immunosuppressive, characterized by LGALS9 and M2 macrophages. Our proteogenomic analyses provide a molecular landscape of LMS, revealing biological insights, patient outcome stratification, and therapeutic targets.

Introduction

Soft tissue sarcomas (STS) are a diverse group of mesenchymal malignancies that include over 80 histological subtypes[1]. Leiomyosarcoma (LMS) accounts for 10-20% of all STS[2]. Soft tissue leiomyosarcoma (STLMS), i.e., extra-uterine LMS, develops mainly in the retroperitoneum, abdomen/pelvis, trunk, and extremities, with the first two being more common[3]. Clinical management of localized disease depends on anatomical site and tumor grade[4,5]. Typical STLMS shows spindle-shaped cells that are set in long intersecting fascicles, diffuse hypercellularity with high-grade histology, and the expression of one or more myogenic marker proteins. Despite multidisciplinary management, cure rates for STLMS remain low. Half of patients experience tumor relapse after surgery[6-8]. Patients with advanced or metastatic STLMS have poor outcomes and limited treatment options[9,10]. To improve outcomes, there are urgent needs to develop better classification and prognostic stratification of STLMS patients for the development of efficient therapeutics.

Several genomic and epigenomic studies of pan-STS reported enrichment of PIK3/AKT pathway alterations and deletions of tumor suppressor genes, including *TP53* (9% deep and 60% shallow deletions), *RB1* (14% deep, 78% shallow), and *PTEN* (13% deep, 68% shallow)[11-13]. Multiomic analyses showed that LMS is molecularly distinct from other STS, and that STLMS differs from uterine LMS in hormone response and DNA damage response pathways. Integrated clustering of STLMS based on WES, epigenome, and whole transcriptome data revealed two subtypes: one with short survival and frequent PI3K/AKT/MTOR pathway alterations, the other with longer survival and minimal PI3K/AKT/MTOR alterations. While MTOR inhibitors like everolimus or temsirolimus show some clinical efficacy in LMS[14,15], their efficacy is reduced by indirect AKT upregulation.

Proteins are the “machines of life” carrying out complex cellular and biochemical functions and thus serve as cancer treatment targets[16]. mRNA abundance poorly predicts protein amounts[17,18]. Proteomics complements genomic research by providing accurate assessment of tumor biology and patient stratification that is inaccessible to nucleic acid-based approaches. Two proteomic analyses of pan-STS were recently published, although the use of formalin-fixed, paraffin-embedded samples limits proteomic studies and largely precludes analyses of post-translational modifications, such as phosphorylation[19,20]. Here, we present the first integrative proteogenomic study of STLMS using freshly frozen tissue and the inclusion of phosphoproteomics to discover alterations of the kinome.

To gain a better understanding of STLMS tumor biology with focus on associations between genomic alterations and both total proteome and phosphoproteome, we analyzed 72

primary site and metastatic STLMS samples, including 18 paired primary-metastasis lesions. Additionally, we incorporated our own independent validation cohort of 222 STLMS FFPE samples as tissue microarrays.

Results

Unsupervised clustering of proteome identifies 3 proteomic subtypes with clinical relevance

We obtained 72 frozen tissues of primary and metastatic STLMS (cohort 1) for liquid chromatography-mass spectrometry (LC-MS) analysis (Figure 1A). For independent cohort 2, we obtained FFPE blocks of 222 STLMS cases and constructed tissue microarray blocks (Figure 1A). Patient demographics are summarized in Table S1. LC-MS analysis of global proteomes quantified 10,249 proteins in total and 7,325 proteins in at least 30% of samples (Table S2A). For phosphoproteomics, the analysis quantified 6,542 phosphorylated proteins in total and 4,584 in at least 30% of samples (Figure S1A, Table S2B). Batch correction removed TMT multiplexing effects from the data (Figure S1B). Using non-negative matrix factorization (NMF) on filtered global proteome and phosphoproteome data, we identified three distinct proteome-based subtypes (P1, P2, and P3) of STLMS (Figures 1BC, S1CD). Survival analyses showed significant differences in overall survival time (OS) and recurrence-free survival time (RFS) between subtypes, with P2 showing the shortest patient survival. We validated these findings using a TCGA STLMS dataset, where P2/P3 subtypes showed significantly shorter recurrence-free survival and overlapped with the TCGA STLMS iCluster1 subtype (Fig. S1EF). Analysis of clinical attributes (Figure 1D) revealed no significant differences in tumor size, age, tumor status (primary lesion or metastatic lesion), gender, TNM stage, or prior treatment between proteomic subtypes. However, metastatic samples were enriched in P2/P3 subtypes compared to P1 ($p=0.0308$). Primary tumor site distribution varied significantly among subtypes ($p=0.01735$), with trunk samples predominantly in P1 and pelvis samples in P2. FNCLCC Grade 1 tumors were only seen in P1, while FNCLCC Grade 3 tumors were only seen in P3.

Differences of proteogenomic landscape and transcription regulatory networks between proteome-based subtypes

We evaluated genomic alterations in 69 samples by MSK-IMPACT targeted sequencing. P2/3 subtypes showed higher ploidy, whole genome duplication, and chromosomal instability versus P1 (Figure 2AB, findings supported by analysis of the TCGA STLMS cohort (Figure S2AB). Genomically, *TP53* (68%) and *RB1* (41%) were most frequently altered, followed by *NCOR1*

114 (35%), *MAP2K4* (33%), and *FLCN* (28%) (Figure 2C). The GENIE STLMS cohort (v15.0,
115 n=774)[21] showed similar frequencies. P2/P3 showed frequent genomic alterations in *NCOR1*,
116 *MAP2K4*, *FLCN*, and *IGF1R*. P2 showed commonly alterations in oncogenic drivers like *IGF1R*
117 or *PDGFRA* and SWI/SNIF complex genes, associated with poor outcomes[22-25]. Somatic copy
118 number alteration (SCNA) analysis revealed more arm-level amplifications in P2/P3 than in P1,
119 while P3 showed higher deletion frequencies (Figure 2DE). Each subtype showed unique focal
120 amplification/deletion patterns, with P1 showing fewer focal peaks (Figure 2F). The 17p12 peak,
121 containing *NCOR1* and *MAP2K4*, was shared between P2 and P3. *NCOR1* and *MAP2K4*
122 alterations occurred frequently together (p=1.03E-14, Fisher's exact test). We analyzed gene
123 overlap in recurrent focal peaks, finding 1,818, 1,125, and 3,556 involved genes respectively
124 (Figure 2G). Gene ontology analyses showed P1 focal deletion peaks enriched in copper/zinc
125 ion processes, P2 focal peaks in extracellular matrix processes, and P3 focal peaks in T-cell
126 receptor signaling (Figure S2C). KEGG analysis showed P2/P3 deletion peaks enriched in
127 immune-related pathways, while only P3 showed deletions involving in PI3K-Akt signaling
128 (Figure S2D).
129 Pairwise differential expression analyses between subtypes (P1 vs. P2/P3, P2 vs. P1/P3,
130 P3 vs. P1/P2) revealed 2,277 dysregulated proteins in P1 vs. others (207 upregulated >2-fold,
131 211 downregulated >2-fold), 2,503 in P2 vs. others (282 upregulated >2-fold, 286 downregulated
132 >2-fold), and 2,652 in P3 vs. others (106 upregulated >2-fold, 377 downregulated >2-fold) with
133 q-values <0.05 (Figure S2E) (Table S3ABC). KEGG and Hallmark analyses identified cancer-
134 related pathways in each subtype (Figure 2H). P1 showed low proliferation and enriched
135 metabolic processes. P2 showed enrichment of epithelial-mesenchymal transition (EMT), NF-kB
136 pathway, inflammatory responses, and KRAS signaling, with downregulated oxidative
137 phosphorylation. P3 exhibited cell cycle activation, DNA repair, low EMT, and higher muscle
138 differentiation. Transcriptome regulatory network analysis identified 60 dysregulated regulons
139 across proteome subtypes (Figure 2I). The regulons formed 3 clusters. Cluster 1 showed high
140 activity in the P2 subtype with poor outcome (Figure 1B), with MYC enrichment confirmed by IHC
141 (Figure 2J). Cluster 2 showed high activity in P3, with enriched cell cycle-related transcription
142 factors. DNMT1 enrichment in P3 suggests methylation differences. PAX6, a tumor progression
143 factor[26-28], was upregulated in P3 and confirmed by IHC (Figure 2J). Cluster 3 showed high
144 activity in P1 and some P3 samples. MITF, regulating immune response[29], was enriched in P1
145 (Figure 2J) and validated in the TCGA STLMS dataset (Figures 2J, S2F).
146
147 **Impact of genomic alterations on proteomic subtypes and cell cycle-related components**

148 We studied cis effects of copy number alteration (CNA) on mRNA and protein. Analysis found
149 320 and 365 cis correlations of CNA-mRNA and CNA-protein respectively (FDR <0.1) (Figure
150 3A), with 131 genes overlapping (Figure 3B). GOBP over-representation analysis of cis regulated
151 genes that showed positive CNA-protein abundance correlations revealed significant enrichment
152 of cell cycle-related terms (Figure 3C). Most cell cycle components showed positive correlation
153 with the cell cycle score at CNA, mRNA, and (phosho)protein levels (Figure 3D). The P3 subtype
154 showed highest cell cycle activity, Ki67 immunolabeling (Figure 3E), and AURKA protein
155 expression (Figure 3F), an emerging cancer treatment target[30]. E2F targets were enriched in
156 P2/P3 (Figure 3G), with CDK1/2 kinase activity significantly higher than in P1 (Figure 3H).
157 Independent TCGA STLMS data validated these findings (Figure S3A-C). CRISPR-Cas9 gene
158 perturbation experiments of the DepMap project showed CDK1 and AURKA inhibition greatly
159 affected cell survival (Figure 3I), suggesting that these proteins are potential treatment targets
160 for high proliferation P3 subtype STLMS. Metastatic STLMS samples showed increased cell
161 cycle score, proliferation rate, and CDK2 kinase activity relative to primary site STLMS (Figure
162 S3DE), with a trend for elevated AURKA protein expression (Figure S3F). TMA cohort 2 IHC
163 scores validated these findings (Figure S3G). Combined cohort analyses of recurrence-free
164 survival showed poor outcomes associated with increased expression of cell cycle-related
165 components (Figure S3H), supporting AURKA and CDKs as potential therapeutic targets against
166 STLMS including metastatic lesions.
167
168 **Phosphoproteomic alterations associated with proteomic subtypes and receptor tyrosine**
169 **kinase pathways**
170 We investigated phosphoproteomic differences between STLMS subtypes focusing on receptor
171 tyrosine kinase molecules (RTK-RAS pathways). Among 69 DNA-sequenced samples, 56
172 samples (81.2%) had at least one genomic alteration in RTK-RAS pathway genes. *MAP2K4* and
173 *IGF1R* were the most frequent events. Though alteration frequencies were higher in P2/P3 vs.
174 P1, differences were not statistically significant (Figure 4A). Genomic alteration frequencies of
175 *PDGFRA* (overall frequency,14%), *FGFR2* (12%), *MET* (9%), *TSC2* (7%), *ERBB3* (6%), *FGFR4*
176 (6%), *INPP4B* (6%) differed significantly between proteomic subtypes. All these genes except
177 *FGFR2* showed enrichment in P2, while *FGFR2* was enriched in P1. Analyses of differential
178 expression at phosphoproteome level between subtypes revealed Ras protein signal
179 transduction enrichment in P2 and cell cycle term enrichment in P3 (Figure 4B), compatible with
180 RTK-RAS gene enrichment in P2 and cell cycle enrichment in P3 (as shown in Figures 2 and 3).
181 We profiled kinase activity using phosphopeptide expression data and the RoKAI app[31] to

182 generate kinase-substrate networks (Figure 4CD). CSNK1A1 and PDK1-4 were enriched in P1.
183 AURKA activity was downregulated in P1, matching AURKA IHC assessment (Figure 3F). CDKs
184 (CDK1,2,5,6), AURKB, mTOR, and MAPK1 (ERK2) were enriched in P2. CDKs (CDK2,7),
185 AURKB, ROCK1, and BRAF were enriched in P3. AKT1/3 showed enrichment in P3 versus
186 P1/P2, though not statistically significant. P1 had a distinct kinase activity profile from P2/P3,
187 while P2/3 shared kinase activities such as AURKB or CDK2. This difference may derive from
188 distinct genomic alteration patterns (Figure 4A). PTM-GSEA analyses showed similar kinase
189 activity differences between STLMS subtypes (Figure S4A). AURKB kinase activity scores were
190 higher in P2/P3 versus P1 (Figure 4E). Furthermore, AURKB IHC positivity was higher in P2/P3
191 compared to P1 (Figure 4F). Consistent with this observation, AURKB mRNA expression in the
192 independent TCGA cohort was also higher in P2/P3 versus P1 (Figure S4B). The most frequent
193 AURKB phosphorylation site with activating effect is T232 based on the PhosphositePlus
194 database (https://www.phosphosite.org/homeAction). Motif analysis of AURKB-T232 predicted
195 responsible kinases for phosphorylation and inferred AURKB upstream regulators (Figure 4G).
196 Most inferred upstream kinases showed activation in P2/P3, while AURKB-activating kinases
197 were downregulated in P1.
198
199 **NCOR1 protein expression is associated with smooth muscle differentiation, cancer**
200 **stemness, and immune-suppressive phenotype**
201 Epigenetic regulators affect gene expression via chromatin structure regulation, including histone
202 protein modification, DNA methylation, TF binding, and chromatin accessibility[32]. With *NCOR1*
203 amplification frequently observed (Figure 2C), we investigated epigenetic regulator gene
204 alterations in cohort 1 (Figure 5A). *NCOR1* showed highest alteration frequency (35%), followed
205 by *SMARCA4* (13%), *BRD4* (12%), and *CARM1* (12%). *BRD4*, *NSD1*, and *SMARCA4* alterations
206 differed significantly between subtypes. *NCOR1* was amplified in over 30% of P2/P3 samples
207 versus 14.2% in P1 (Figure 5B). *NCOR1* copy number correlated with NCOR1 protein level, with
208 P3 showing a trend for highest expression. P2/3 of TCGA STLMS showed significantly higher
209 *NCOR1* amplification frequency, resulting in highest *NCOR1* mRNA expression in P3 (Figure
210 S5AB). As *NCOR1* was a top altered epigenomic regulator enriched in poor-outcome P3, we
211 investigated the impact of high vs. low NCOR1 protein expression. We divided our cohort by
212 median NCOR1 protein expression and performed a differential expression analysis with GSEA.
213 GSEA revealed muscle organ development terms enriched in the NCOR1-high group, while
214 immune and epithelial process terms were depleted (Figure 5E), suggesting NCOR1's role in
215 muscle differentiation. Smooth muscle differentiation markers trended higher in the NCOR1-high

216 group (Figure 5F), with TCGA STLMS data showing similar results (Figure S5C). The NCOR1-
217 high group showed significantly higher stemness scores (Figures 5G, S5D) and PROM1 (CD133)
218 and ALDH7A1 protein expression (Figure 5H). TCGA data showed higher *POU5F1* (Oct3/4),
219 *PROM1* (CD133), and *ALCAM* mRNA in the *NCOR1*-high group (Figure S5E). Consistent with
220 the Hippo pathway's role in stemness, we found YAP1/TEAD expression significantly higher in
221 the NCOR1-high group (Figure 5IJ), confirmed by TCGA data showing YAP1/TAZ (encoded by
222 WWTR1)/TEAD enrichment (Figure S5F). These findings suggest a role for NCOR in stemness
223 maintenance and smooth muscle differentiation regulation.
224
225 **Enrichment of homologous recombination deficiency and nonhomologous end joining in**
226 **poor-prognosis subtypes**
227 We observed high chromosomal instability (CIN) in P2/P3 (Figures 2B, S2B) with whole genome
228 duplication and DNA repair enrichment in P3 (Figure 2H). The main repair pathway for DNA
229 double strand breaks (DSB) is homologous recombination repair (HRR)[33]. Inactivation leads to
230 genomic alterations, CIN, poor prognosis, and metastasis in tumors[34-36]. In cohort 1, we
231 examined 64 HRR pathway genes from the TCGA sarcoma cohort[37] and found 38 genes altered.
232 Using allele specific copy number and variant calls, HRR genes showed frequent loss of
233 heterozygosity (LOH) (Figure 6A). HRR gene events were enriched in P2/P3 (Figure 6B). The
234 HRD (homologous recombination deficiency) score was highest in P2 (Figure 6C) and correlated
235 with CIN (Figure 6D). The TCGA dataset confirmed these findings (Figure S6A-D). We
236 additionally assessed RAD51 protein (a key recombinase for homologous recombination[38])
237 expression by IHC, finding a trend for decreased RAD51 staining in P2/P3 vs. P1 (Figure 6E),
238 also supported by TCGA STLMS RPPA data (Figure S6E). Analysis of paired samples from the
239 same patients showed increased HRD scores in metastatic sites vs, primary sites for P3 STLMS
240 (Figure 6F), aligning with CIN-metastasis association[35,36]. We analyzed genomic alteration
241 patterns of the top 5 altered HRR genes in STLMS (*BRCA2*, *RPA1*, *ATM*, *MRE11*, and *FANCA*)
242 in cohort 1. The most frequent combination affected all 5 genes simultaneously (Figure 6G). For
243 single or 2-gene combinations, *BRCA2* was predominant (total events = 17). *BRCA2* LOH tumors
244 showed higher HRD scores than non-LOH tumors (Figure 6H). Assessment of LOH effects of
245 the remaining 37 HRR genes on both mRNA and protein levels revealed decreased expression
246 in LOH-positive STLMS (Figure 6I), a finding also supported by the external TCGA STLMS
247 mRNA dataset (Figure S6F). TCGA STLMS RPPA protein data confirmed reduced BRCA2
248 protein in HRD-high and BRCA2-LOH groups (Figure S6G). BRCA2-loss by IHC was frequent in
249 P2 (Figure 6J), which showed the highest HRD score among subtypes (Figure 6C). We assessed

250 RAD50 expression in tumor tissue by IHC since RAD50 is a core protein of DSB recognition, a
251 starting point of the HRR pathway. While HRD status alone showed no statistical difference in
252 RAD50 expression in cohort 1, P2/P3 subtypes featured decreased RAD50 expression versus
253 P1 (Figure 6K), a finding also supported by external TCGA data (Figure S6H). Assessment of
254 phosphorylated H2AX (DNA double-strand break marker) showed a positivity range shift
255 between subtypes (Figure 6L), suggesting impaired DSB recognition and repair in P2/P3
256 compared to P1. We also evaluated nonhomologous end joining (NHEJ), an alternative DSB
257 repair process. We assessed both classical NHEJ driven by KU70/80 (coded by XRCC6/5) and
258 alternative NHEJ driven by PARP1[39]. While XRCC5/6 increased in P3 compared to P1, P2
259 showed no significant difference to P1 (Figure 6M). PARP1 showed increased expression in the
260 HRD-high group and the P3 subtype (Figure 6N), with P2 showing a close-to-significant increase,
261 suggesting NHEJ pathway activation in P3 and possibly P2/P3. External TCGA data confirmed
262 these results (Figure S6I). Using RAD51 and BRCA2 IHC protein expression data from cohorts
263 1 and 2, we found decreased expression of both proteins significantly associated with poor
264 patient outcome (Figure 6O).
265
266 **Tumor immune microenvironment profiling reveals possible mechanisms of immune**
267 **evasion**
268 To elucidate cancer immune evasion mechanisms, we performed integrated proteogenomic
269 analyses with IHC assessment of immune cell infiltration. Immune microenvironment status was
270 determined using the ESTIMATE ImmuneScore[40], which correlates with the CD3+ T-cell count
271 by IHC (Figure 7A). The ImmuneScore correlates with antigen presenting machinery (APM)
272 components, including MHC class I pathway members and IRF1 (Figure 7A). APM-related
273 molecules showed good concordance between genomic copy number, mRNA, and protein
274 expression (Figure 7A), indicating genomic alterations affect the immune microenvironment. The
275 P3 subtype of STLMS showed a significantly reduced ImmuneScore (Figure 7B), a finding
276 corroborated by external TCGA data (Figure S7AB). Paired sample analyses showed metastatic
277 tumors had lower ImmuneScores (Figure 7C) and downregulated immune-related processes
278 (Figure S7C). Comparing immune-cold vs. immune-hot tumors revealed 356 differentially
279 expressed proteins, with 75 upregulated and 281 downregulated in immune-cold tumors (Figure
280 7D, Table S4). GSEA confirmed downregulation of antigen-presenting pathways in immune-cold
281 tumors (Figure 7E), also supported by TCGA STLMS data (Figure S7D). Metastatic tumors
282 showed similar downregulated antigen processing pathways (Figure S7E). ERAP1, TAP1/2,
283 MHC class I complex molecules (HLA-B/C, B2M), and proteasome components were

284 downregulated in immune-cold tumors (Figure 7F). IRF1, a major transcription regulator of MHC
285 class I and APM molecules[41], showed lower expression in immune-cold tumors (Figure 7G). We
286 observed positive correlations between IRF1 and MHC class I molecules (HLA-B/C, B2M),
287 though not statistically significant. IRF1 correlated positively with proteins in antigen processing
288 (including immunoproteasome components), such as TAP1, TAP2, TAPBP, ERAP1, PSMB8,
289 and PSMB9, while catalytic subunits of proteasome (PSMB5, PSMB7) were negatively correlated
290 (Figure 7H), findings supported by TCGA transcriptome data (Figure S7F). IRF1 correlated with
291 IFNγ-JAK-STAT pathway, which regulates its expression[42], suggesting IFNγ signaling may have
292 therapeutic potential against immune-cold STLMS[43-45]. Hypoxic environments may suppress host
293 immune responses against tumors[46-48]. However, we found no significant correlation between
294 hypoxia and the STLMS ImmunsScore (Figure 7A).
295 Next, we assessed the expression of tumor immune suppressors in immune-hot tumors,
296 which may abrogate an immune response. Transcriptome analysis of cohort 1 found LGALS9
297 was upregulated in immune-hot tumors (Figure 7I). TCGA STLMS data showed high expression
298 of immune suppressive molecules including LGALS9 (Figure S7G). As LGALS9 overlapped
299 between the two cohorts, we evaluated its protein expression by IHC in cohorts 1 and 2. IHC
300 confirmed tumor cells expressed LGALS9 with a trend for higher expression in the P2 (immune-
301 hot) subtype (Figure 7J). Interestingly, LGALS9 protein expression was not detectable by IHC in
302 benign smooth muscle, such as blood vessel walls or intestinal walls. The TCGA STLMS cohort
303 showed a similar result (Figure S7H). Given LGALS9 upregulation in immune-hot tumors, we
304 assessed tumor-associated macrophages via CD68 (pan-macrophage marker) and CD163 (M2
305 macrophage marker) IHC to evaluate their contribution to an immunosuppressive
306 microenvironment. M2-type macrophages were enriched in the P2 subtype (Figure 7K). Survival
307 analyses showed high CD68 or CD163 cell counts correlated with shorter recurrence-free
308 survival (Figure 7L). These results suggest that the P2 subtype evades immune surveillance
309 through LGALS9 expression and M2 macrophage recruitment, creating an immunosuppressive
310 microenvironment (despite an immune-hot background) that is associated with poor clinical
311 outcome.
312 To concisely summarize the multi-omic findings, we generated an integrative schematic
313 of the proteome-defined subtypes of STLMS (Figure 8). The figure summarizes, for each subtype,
314 the characteristic genomic and pathway alterations, kinome alterations, DNA-repair features,
315 tumor microenvironment, prognostic associations, and hypothesis-generating drug targets.
316
317 **Discussion**

318 In this detailed proteogenomic analysis of STLMS, we identified three proteomic subtypes
319 (P1, P2, and P3) with notable clinical and molecular differences (Figure 8). These subtypes are
320 associated with patient outcomes, genomic alterations, signaling pathway activations, and
321 immune microenvironment profiles, offering insights into STLMS heterogeneity and potential
322 targeted therapies.

323 Unsupervised clustering of proteomic data identified the P2 subtype as having the worst
324 prognosis with the shortest overall survival (OS) and recurrence-free survival (RFS). Both P2
325 and P3 subtypes were significantly enriched in metastatic behavior and higher-grade tumors,
326 linking proteomic profiles to tumor aggressiveness. Validation using an external TCGA STLMS
327 dataset confirmed the clinical relevance of our proteomic subtyping. The TCGA study identified
328 only two subtypes (iClusters 1 and 2)[12] but lacked comprehensive (phospho)proteomic data,
329 focusing mainly on the AKT signaling pathway. Our integrated proteogenomic analyses,
330 however, identified three molecular subtypes, providing a more detailed characterization that
331 enhances understanding of STLMS tumor biology and heterogeneity.

332 Genomic analysis shows key differences in genomic instability between P1, P2, and P3
333 subtypes. High ploidy and whole-genome duplication in P2 and P3 indicate aggressive tumor
334 phenotype, linked to rapid evolution and poor prognosis[49]. Enhanced chromosomal instability
335 (CIN) in these subtypes may drive tumor heterogeneity and therapy resistance[50]. Higher
336 alterations in oncogenic drivers like *IGF1R* and *PDGFRA* in P2 and P3 indicate activation of
337 pathways of cell proliferation and survival[49]. Frequent alterations in SWI/SNF complex genes in
338 P2, including *SMARCA4* and *BRD4*, suggest disrupted chromatin remodeling, causing aberrant
339 gene expression and tumorigenesis[51]. The co-occurrence of *NCOR1* and *MAP2K4* gene
340 amplifications in P2 and P3 is significant; NCOR1 regulates gene expression through chromatin
341 remodeling, and MAP2K4 is vital in the MAPK/ERK pathway[49]. These amplifications contribute
342 to tumor progression by maintaining cancer stemness and activating RTK pathways, crucial for
343 cancer progression and metastasis[49].

344 Copy number alterations (CNAs) correlated with mRNA and protein expression levels,
345 particularly in genes linked to cell cycle regulation (Figure 3). The P3 subtype showed highest
346 cell cycle activity and elevated AURKA expression, revealing potential therapeutic targets such
347 as CDK1/2 and AURKA against aggressive STLMS subtypes. CRISPR/Cas9 gene perturbation
348 experiments have supported the viability of these targets, offering promising avenues for targeted
349 therapies[52,53].

350 The P1 subtype shows low proliferation and metabolic process enrichment, suggesting
351 slower-growing tumors with metabolic vulnerabilities. The P2 subtype exhibits EMT, NF-κB

signaling, immune responses, and RAS pathways, indicating an aggressive, metastatic phenotype. The P3 subtype shows cell cycle activation and DNA repair dysregulation, suggesting highly proliferative tumors susceptible to therapies targeting cell cycle checkpoints or DNA repair deficiencies (Figures 2, 3, 6, 7). Phosphoproteomic analyses have revealed subtype-specific kinome programs corresponding to biological characteristics and therapeutic hypotheses. In subtype P1, kinase activity differed from P2 and P3, favoring metabolic regulation over proliferation. CSNK1A1 and the pyruvate dehydrogenase kinase axis (PDK1-PDK4) were selectively enriched, while AURKA/B activity was diminished (Figure 4C-D). This aligns with oxidative-metabolic pathway enrichment and low proliferation in P1 and *FGFR2* alterations. These results suggest a metabolically reprogrammed, growth factor-dependent, yet cell cycle-quiescent state may be targetable with FGFR2/PDK-axis strategies (Figures 2H,4A, and 8). P2 showed an RTK-RAS/ERK-mTOR cell cycle signature. *IGF1R*/*PDGFRA* alterations were enriched, and CDK1/2/5/6, AURKB, mTOR, and MAPK1(ERK2) were highly scored, with downstream S6K/RSK families (Figure 4A-D). These signals match the EMT phenotype and poor outcomes of P2, supporting CDK1/2 or AURKB blockade with mTOR/ERK pathway inhibition (Figures 2H and 4C-D) as therapeutic targets. Subtype P3 showed a proliferative mitotic/replication kinome with CDK2/7, AURKB, ROCK1, and BRAF activity, and elevated AKT1/3 compared to P1/P2 (Figure 4C-D). Together with P3's high cell-cycle scores and AURKA protein, these data suggest sensitivity to CDK/Aurora inhibitors, while ROCK1 and BRAF enrichment implies actomyosin and MAPK dependencies (Figures 3D-H and 4C-D). AURKB activity was elevated in P2/P3 versus P1, validated by IHC, identifying AURKB as a cross-subtype mitotic driver (Figure 4E-F). AURKB-T232 motif analysis suggested upstream mitotic and checkpoint kinase activation in P2/P3, supporting AURKB-targeted therapy (Figure 4G). These kinome data define P1 as PDK-activated with metabolic/FGFR2 bias, P2 as RTK-ERK-mTOR/CDK-driven, and P3 as CDK/Aurora-centered with ROCK/MAPK inputs (Figures 4 and 8).

Transcriptional regulatory network analysis identified multiple transcription factors for each subtype (Figure 2I). These findings highlight the complexity of STLMS biology and importance of understanding subtype-specific regulatory networks for targeted therapies. Future studies should focus on validating these transcription factors as therapeutic targets and exploring their regulatory network interactions. The P2 subtype showed elevated activity of oncogenic regulators, including MYC. MYC is an oncogene that drives cell proliferation, growth, and metabolism. Its activation links to aggressive tumor phenotypes and poor prognosis[56], aligning with P2's worse outcome (Figure 1B). The P3 subtype showed enrichment for cell cycle-related transcription factors, including E2Fs, DNMT1, and PAX6, suggesting epigenetic regulation in its

386 pathogenesis.

387 NCOR1 expression in our cohort was associated with both increased smooth-muscle
388 marker expression and higher expression of progenitor-like stemness markers (PROM1 (CD133)
389 and ALDH7A1) (Figure 5). NCOR1, as a chromatin corepressor interacting with HDAC3, can
390 reshape transcriptional programs to favor a lineage-restricted progenitor state that co-expresses
391 differentiation markers while retaining self-renewal properties, which is a behavior seen in muscle
392 physiology and chromatin regulation. Additionally, the increased activity of YAP1/TAZ-TEAD in
393 tumors with high NCOR1 levels suggests a potential collaborative pathway, as YAP/TEAD
394 signaling encourages the proliferation of muscle progenitors and supports regenerative
395 processes. These combined characteristics might contribute to a differentiated yet progenitor-
396 like phenotype that is resistant to therapy in STLMS. To validate this hypothesis, functional
397 disruption and chromatin analysis will be necessary[57].

398 Our observations showed that HRD is common in both P2 and P3 subtypes, alongside
399 activation of NHEJ pathways (Figure 6)[58]. Elevated HRD scores correlated with increased
400 chromosomal instability. Loss of heterozygosity in HRR genes, notably *BRCA2*, was linked to
401 reduced HRR protein expression and unrepaired DNA double-strand breaks, as shown by higher
402 phosphorylated H2AX levels (Figure 6L). PARP1 upregulation suggests that PARP inhibitors
403 may be effective therapeutic agents in HRD-high STLMS (Figure 6N). HRD, a hallmark of
404 genomic instability, is characterized by inability to repair DNA double-strand breaks through
405 homologous recombination[59]. This deficiency leads to reliance on error-prone repair mechanisms
406 like NHEJ, resulting in chromosomal rearrangements and mutations. In sarcomas, particularly
407 uterine leiomyosarcoma, HRD is associated with aggressive tumor behavior and poor
408 prognosis[60]. Identification of HRD in two of our proteomic subtypes of STLMS points to the
409 potential for targeted therapies against STLMS exploiting this vulnerability, such as PARP
410 inhibitors.

411 Immune profiling analysis revealed that the P3 subtype and metastatic tumors exhibit
412 lower ImmuneScores and downregulation of antigen-presenting machinery (APM) components,
413 contributing to an immunosuppressive tumor microenvironment. Reduced expression of IRF1
414 and APM molecules in immunosuppressed tumors may facilitate immune evasion mechanisms.
415 Conversely, immune-hot tumors demonstrated concomitant upregulation of suppressive immune
416 checkpoint molecules, such as LGALS9, and M2 macrophage recruitment. These results suggest
417 that combinatory immunomodulatory therapies, such as LGALS9 or PD-L1 inhibition[61], could be
418 strategically tailored to the immune profile and proteomic subtype of STLMS.

419 Our study presents a large, systematic proteogenomic analysis of STLMS that also

420 includes a phosphoproteomic characterization of kinome alterations. By identifying subtype-
421 specific genomic alterations and pathway activations, the study establishes a foundational
422 molecular basis for precision diagnostics and therapeutic interventions. Aggressive STLMS
423 subtypes may benefit from therapies targeting CDK1/2 and AURKA or from exploiting HRD with
424 PARP inhibitors. Patients with immune-hot tumors expressing LGALS9 may respond to immune
425 checkpoint inhibitors. These findings highlight clinically relevant heterogeneity within STLMS and
426 suggest that future precision medicine testing should incorporate proteogenomic or surrogate
427 assays to assign molecular subtypes and guide risk stratification and treatment selection. Our
428 comprehensive proteogenomic dataset of primary and metastatic STLMS provides a valuable
429 resource for further validation of these observations and for the development of practical assays
430 to implement subtype–based management in routine care for patients with STLMS.
431

432 Materials and Methods

433 Clinical specimens and pathological data

434 We examined tissues from two cohorts of STLMS patients. Cohort 1 consists of 72 primary and
435 metastatic STLMS tissues from 54 patients, including 18 primary-metastasis-paired samples.
436 Cohort 2 consists of 222 surgically resected STLMS samples. The study was approved by the
437 Institutional Review Board (MSKCC #16-1683, BIDMC #2023P000933). Clinical data, such as
438 patient demographics, treatment history, recurrence status, clinical targeted MSK-IMPACT
439 sequencing results, or histologic grade were retrieved from medical records in an anonymized
440 fashion. Tumor content ratios for all samples were verified by board-certified pathologists.
441 Clinicopathologic features and prognosis information, including gender, age, pathological TNM
442 stage, anatomic site, are summarized in Table S1.
443

444 Targeted cancer gene sequencing

445 MSK-IMPACT targeted sequencing was conducted as previously outlined[62]. In summary,
446 genomic DNA was extracted from both tumor and adjacent normal tissues using the DNeasy
447 Tissue kit and the EZ1 Advanced XL system (Qiagen). The extracted DNA was fragmented with
448 the Covaris E200 device. Custom DNA probes were crafted for targeted sequencing of all exons
449 and selected introns across 505 genes. These probes were synthesized using the NimbleGen
450 SeqCap EZ library custom oligo system and were biotinylated. Sequencing libraries were
451 prepared following the KAPA HTP protocol (Kapa Biosystems) and the Biomek FX system
452 (Beckman Coulter), and sequencing was performed on the Illumina HiSeq 2500 to achieve high,
453 uniform coverage (>500x median coverage). Demultiplexed fastq files underwent trimming with

454 Trim Galore v0.2.5mod (https://github.com/FelixKrueger/TrimGalore) to remove adapters and
455 short reads. These files were then aligned to the human reference genome (GRCh37) using
456 BWA-MEM v0.7.5a (with arguments -M -t 6)[63], and Picard Tools v2.9
457 (https://github.com/broadinstitute/picard) was used. AddOrReplaceReadGroups was applied to
458 annotate read groups, and MarkDuplicates was used to identify PCR duplicates. Genomic
459 regions were pinpointed using FindCoveredIntervals from the GATK Tool Kit v3.3-0[64]and
460 underwent indel realignment with Assembly Based ReAligner (ABRA) v2.12[65]. GATK
461 BaseRecalibrator was employed to identify systematic errors in base quality scores. Variant
462 calling was executed in paired tumor/normal mode using MuTect v1.1.4[66]for single nucleotide
463 variants (SNV). Pindel v0.2.5a7[67]was utilized for small insertions and deletions (indels). Vardict
464 v1.5.1[68]served as the variant caller to report both SNVs and indels. The vcfs generated by
465 MuTect, Vardict, and Pindel were subsequently combined. Additionally, copy-number variants,
466 including chromosomal instability (CIS) and whole-genome doubling (WGD), were identified
467 using FACETS[69]. The resulting variants were annotated using vcf2maf v1.6.14
468 (https://github.com/mskcc/vcf2maf), which employs Ensembl's Variant Effect Predictor v86. In
469 conclusion, we successfully conducted MSK-IMPACT sequencing on 69 out of 72 samples in
470 cohort 1 (LM_010P, LM_020P, LM_043P lacked sufficient DNA for sequencing). Amplifications
471 and deep deletions were identified from the GISTIC result table named
472 "allthresholded.bygenes.txt" (part of the GISTIC output used to determine the copy number status
473 of each gene in each sample). Genes with a value of +2 were classified as amplifications, while
474 those with a value of -2 were considered deep deletions. Sequencing results are detailed in Table
475 S5.
476
477 **Allele-specific copynumber analysis**
478 FACETS analyses were conducted to ascertain genome-wide allele-specific and absolute DNA
479 copy numbers for all samples (FACETS version 0.5.6)[69]. Allele-specific copy number data were
480 utilized to estimate tumor purity and ploidy. Prior to further analyses, total copy number log ratios
481 were adjusted for ploidy and purity. Tumors exhibiting whole-genome doubling (WGD) were
482 identified as those in which more than 50% of the autosomal genome possessed a major copy
483 number ≥2, where the major copy number is defined as the number of copies of the most
484 prevalent allele present in the sample[70]. Loss of heterozygosity (LOH) was defined as the loss of
485 one allele in cohort 1. For the TCGA STLMS cohort, allele-specific copy number data were
486 obtained via the R TCGAbiolink package[71], and LOH was defined using the same criteria.
487

488 **HRD score calculation**

489 The scarHRD R package was utilized to calculate the HRD score of our cohort 1 samples[72]. This
490 metric was derived from the simple addition of three types of genetic alterations: loss of
491 heterozygosity (LOH), large-scale transitions (LST), and telomeric allelic imbalance (TAI) based
492 on allele-specific copy number data. LOH refers to the elimination of genomic segments
493 exceeding 15 Mb in size, excluding entire chromosomes. LST encompasses chromosomal
494 breakages between adjacent regions of at least 10 Mb, with breaks occurring no more than 3 Mb
495 apart. TAI quantifies the disparities in allele sequence contributions within the telomeric regions
496 of chromosomes. For the TCGA cohort, we downloaded precalculated HRD score from the prior
497 study[73].

498

499 **Chromosomal instability inference**

500 Chromosomal instability (CIN) scores of our cohort 1 and the TCGA STLMS data was computed
501 on the allele-specific copy number segments by adding the number of gains (CNV >2), losses
502 (CNV <2), and loss of heterozygosity per sample based on data from the prior study[37].

503

504 **Recurrent somatic copy number alteration detection**

505 Recurrent somatic copy number alterations (SCNAs) were identified using the Genomic
506 Identification of Significant Targets in Cancer (GISTIC, version 2.0.23)[74]to ascertain which SCNA
507 regions exhibited amplification or deletion at a frequency greater than expected by chance, with
508 a q value threshold of 0.05. The analysis employed the following parameters: Amplification
509 Threshold = 0.3, Deletion Threshold = 0.3, Cap Values = 1.5, Confidence Level = 0.99, Join
510 Segment Size = 4, Arm Level Peel Off = 1, and Sample Normalization Method = mean. Default
511 values were applied for all other parameters.

512

513 **RNA sequencing**

514 For each sample, 20-30 mg of frozen tissue was homogenized in 1 mL of TRIzol Reagent
515 (ThermoFisher catalog # 15596018), followed by phase separation induced with 200 µL of
516 chloroform. RNA was extracted from 350 µL of the aqueous phase utilizing the miRNeasy Mini
517 Kit (Qiagen catalog # 217004) on the QIAcube Connect (Qiagen) in accordance with the
518 manufacturer's protocol. Samples were eluted in 34 µL of RNase-free water. Subsequent to
519 RiboGreen quantification and quality control via the Agilent BioAnalyzer, 500 ng of total RNA with
520 RIN values ranging from 5.9 to 9.9 underwent polyA selection and TruSeq library preparation as
521 per the instructions provided by Illumina (TruSeq Stranded mRNA LT Kit, catalog # RS-122-

522 2102) with 8 cycles of PCR. Samples were barcoded and sequenced on a NovaSeq 6000 in a
523 PE100 run, employing the NovaSeq 6000 S1 Reagent Kit (200 Cycles) (Illumina). Read quality
524 was evaluated using FastQC. Trimmed raw reads were aligned to the human genome version
525 hg19 using STAR (v2.7.5a)[75]with default parameters, and gene annotation was incorporated
526 using Gencode v36, followed by the calculation of gene-level count values and TPM values via
527 the RSEM tool (version 1.3.1)[76].
528
529 **Tissue proteome extraction**
530 Aliquots of 5 mg of frozen tissue were lysed using a lysis buffer composed of 8 M urea and
531 200 mM EPPS (4-(2-Hydroxyethyl)-1-piperazinepropanesulfonic acid), at pH 8.5, supplemented
532 with protease inhibitor (complete mini EDTA-free, Roche) and phosphatase inhibitors (cocktails
533 2 and 3, Sigma). This was followed by homogenization using a Bead Ruptor. Subsequently, the
534 tissue mixture underwent sonication through 12 cycles of 1-minute sonication at 120 W power
535 (FB120, Fisher Scientific), with intermittent cooling. Post-centrifugation at 18,000 g for 10 minutes
536 at 4 °C, the supernatant containing all soluble proteins was collected. Protein concentrations
537 were quantified using BCA assays (Pierce). Pooled protein samples were prepared and stored,
538 with equal amounts of all samples combined to assess MS run quality.
539
540 **Liquid chromatography- mass spectrometry (LC-MS) proteomic analysis**
541 Protein samples were labeled utilizing the 16-plex TMT chemical labeling reagent (Thermo Fisher
542 Scientific), followed by MultiNotch MS3 LC–MS analysis employing an Orbitrap Fusion MS[77]. In
543 brief, 300 µg of proteins were reduced with 5 mM TCEP (tris(2-carboxyethyl) phosphine
544 hydrochloride), alkylated with 10 mM IAA (iodoacetamide), and quenched with 10 mM DTT
545 (dithiothreitol). The samples were subsequently diluted and precipitated using chloroform-
546 methanol[78]. The resulting pellets were resuspended in 50 µL of 200 mM EPPS buffer and
547 digested with Lys-C and trypsin at 37 °C overnight. Anhydrous acetonitrile was added to achieve
548 a final volume of 30%. TMTPro (16-plex) reagents were introduced to the peptides at a 2.8:1
549 (TMT reagent-to-peptide) ratio and incubated for 1 hour at room temperature. A label ratio check
550 was conducted to determine mixing ratios, labeling efficiency, and missed cleavages by pooling
551 1 µL from each sample. Based on these results, samples were mixed to achieve equal intensity
552 across TMT channels. The samples were dried to remove acetonitrile, desalted using C18 solid-
553 phase extraction (SPE) Sep-Pak (Waters), and vacuum-centrifuged. Five percent of the desalted
554 samples were reserved for global proteome analysis, while 95% were utilized for phosphopeptide
555 enrichment. Dried peptides intended for global proteomics were fractionated using the Pierce

556 High pH Reversed-Phase Peptide Fractionation Kit (ThermoFisher catalog # 84868) into nine
557 fractions per the modified manufacturer's protocol. The fractions were vacuum-centrifuged and
558 then reconstituted in 1% ACN/0.1% FA for LC–MS analysis. For optimal phosphopeptide
559 enrichment, phosphopeptides were collected using the High-Select TiO2 Phosphopeptide
560 Enrichment Kit (Thermo Fisher) and from the stored flow-through of Fe-NTA using the Pierce
561 High-Select Fe-NTA Phosphopeptide Enrichment Kit (Thermo Fisher), and both were combined.
562 Phosphopeptide samples were fractionated into nine fractions utilizing the Pierce High pH
563 Reversed-Phase Peptide Fractionation Kit. The fractions underwent analysis via LC-MS/MS
564 employing a nanoAQUITY UPLC (Waters) with a 50 cm (75 µm inner diameter) EASY-Spray
565 column (PepMap RSLC, C18, 2 µm, 100 Å) maintained at 60 °C, in conjunction with an Orbitrap
566 Fusion Lumos Tribrid Mass Spectrometer (Thermo Fisher). Peptide separation was conducted
567 at a flow rate of 300 nL/min using a linear gradient of 1-35% acetonitrile (0.1% FA) in water (0.1%
568 FA) over a duration of four hours, analyzed in SPS-MS3 (for global proteome) and neutral loss-
569 triggered MS3 (phosphopeptide) modes. For global proteome analysis, MS1 scans were
570 performed within a range of 375–1,500 m/z, with a resolution of 120 K, an AGC target of 4×1e5,
571 and a maximum IT of 50 ms. MS2 scans were conducted on MS1 scans of charges 2-7 using
572 0.7 m/z isolation, collision-induced dissociation at 32%, turbo scan, and a maximum IT of 50 ms.
573 MS3 scans employed specific precursor selection (SPS) of 10 isolation notches, an m/z range of
574 100–1,000, 45% CE, 50 K resolution, and an AGC target of 1e5. For phosphopeptide analysis,
575 MS1 scans were conducted within a range of 380-1,400 m/z, with a resolution of 120 K, an AGC
576 target of 4×1e5, and a maximum IT of 50 ms. MS2 scans were performed on MS1 scans of
577 charges 2-7 using 0.7 m/z isolation, collision-induced dissociation at 35%, turbo scan, and a
578 maximum IT of 50 ms. Neutral loss-triggered MS3 scans utilized a range of 100–1,000 m/z, 38%
579 CE, 60 K resolution, an AGC target of 1e5, and a maximum IT of 118 ms. Targeted loss masses
580 of 97.9763 and 79.9658 were employed.
581
582 **TMT data analysis**
583 Raw data files were processed utilizing Proteome Discoverer (PD) version 2.4.1.15 (Thermo
584 Scientific). In each of the TMT experiments, raw files from all fractions were consolidated and
585 analyzed using the SEQUEST HT search engine with a Homo sapiens UniProt protein database,
586 downloaded on 2019/12/13 (92,249 entries). Oxidation (M), phosphorylation (STY), and
587 deamidation (NQ) were designated as variable modifications, whereas cysteine
588 carbamidomethylation, TMT 16plex (K), and TMT 16plex (N-terminus) were specified as fixed
589 modifications. The precursor and fragment mass tolerances were set at 10 ppm and 0.6 Da,

590 respectively. A maximum of two missed trypsin cleavages were allowed. Searches employed a
591 reversed sequence decoy strategy to control the peptide false discovery rate (FDR), with a 1%
592 FDR established as the identification threshold. Phosphosite localization was determined using
593 the PD IMP-ptmRS node. The raw abundance of protein, phosphoprotein, and phosphopeptide
594 levels was normalized by sample loading normalization of each TMT channel after filtering out
595 protein/peptide assignments with valid values for less than 30% of samples and K-Nearest
596 Neighbors (KNN) imputation. Internal reference scaling based on pooled samples assigned for
597 each TMT plex was conducted across the entire cohort to eliminate batch effects, followed by
598 TMM normalization for median centralization between TMT batches[79,80]. Differential expression
599 analyses were performed using edgeR[81].
600
601 **Unsupervised clustering of proteomic data**
602 We used non-negative matrix factorization (NMF)-based clustering of global protein and
603 phosphoprotein expression data. In brief, we constructed a merged expression matrix of
604 normalized global protein and phosphoprotein data and selected the top 500 (phospho)proteins
605 with highest mean absolute deviations for unsupervised clustering to avoid noisy
606 (phospho)proteins in the data. The resulting matrix was then subjected to NMF analysis
607 leveraging the NMF R package[82]. To determine the optimal factorization rank *k* (number of
608 clusters) for the matrix, we tested a range of clusters between *k* = 2-7. For each *k*, we factorized
609 matrix *V* using 1,000 iterations with random initializations of W and H. To determine the optimal
610 factorization rank, we calculated two metrics for each k: (1) cophenetic correlation coefficient
611 (measuring how well the intrinsic structure of the data was recapitulated after clustering) and (2)
612 silhouette score (indicating how well each sample lies within its cluster compared to other
613 clusters[83]) (Figure S1CD). These metrics indicate the reproducibility of the clustering across
614 1,000 iterations. The cophenetic correlation coefficients for *k* = 2 and 3 were >0.95 and showed
615 a large drop at *k* = 4. The silhouette score at *k* = 2, 3 showed relatively higher scores than that
616 at *k* = 4. Based on these results and census heatmaps for each *k*, we chose *k* = 3 as the optimal
617 clustering number with sufficient samples in each class for downstream analysis.
618
619 **Gene set enrichment analysis (GSEA)**
620 Fold change values for gene expression between specified groups were calculated using edgeR.
621 We employed clusterProfiler (v2.1.2)[84]with a ranked gene list ordered by fold change value.
622 Enrichment tests were conducted for GOBP, Hallmark, KEGG, and Reactome gene sets
623 obtained from the Molecular Signatures Database (MSigDB v7.4). For the functional

624 characterization of each sample via single sample GSEA, we calculated normalized enrichment
625 scores (NES) of cancer-relevant gene sets by projecting the matrix of normalized expression
626 data onto GOBP, Hallmark, KEGG, and Reactome pathway gene sets. This was done using the
627 ssGSEA implementation available at https://github.com/broadinstitute/ssGSEA2.0 with the
628 following parameters: sample.norm.type = “log”, weight = 0.75, statistic = “area.under.RES”,
629 output.score.type = “NES”, nperm = 1000, min.overlap = 3, correl.type = “z.score”. To elucidate
630 gene set enrichment in each sample, we applied single sample GSEA on normalized expression
631 data[85]and calculated the normalized enrichment score (NES) for MSigDB terms.
632
633 **Over-representation analysis (ORA)**
634 ORA was performed to identify biological pathways and functional categories enriched in the
635 gene set of interest by using clusterProfiler (v2.1.2)[84]. Enrichment was calculated using a
636 hypergeometric test against the defined background, and p-values were adjusted for multiple
637 comparisons using the Benjamini-Hochberg method.
638
639 **Posttranslational modification signature enrichment analysis (PTM-SEA)**
640 To infer kinase activity based on phophosite-specific quantification data, we performed PTM-
641 SEA (https://github.com/broadinstitute/ssGSEA2.0), an adaptation of ssGSEA, that analyzes
642 site-specific signatures by scoring PTMsigDB bi-directional signature sets. PTMsigDB, sourced
643 from over 2,500 publications, contains modification site-specific signatures linked to
644 perturbations, kinase activities, and signaling pathways. Unlike other pathway databases,
645 PTMsigDB annotates each PTM site with its directional change upon specific perturbations or
646 signaling events, improving PTM-SEA's scoring. PTMsigDB’s main source is the
647 PhosphoSitePlus database[86]. The PTM-SEA requires two main inputs: a site-centric data matrix,
648 denoted as m and formatted in GCT v1.3, along with the PTM signatures database. In matrix m,
649 each row corresponds to a distinct phosphorylation site that is precisely localized to a particular
650 amino acid residue, with the columns indicating the measured abundances across different
651 samples. When multiple phosphorylation sites are identified on the same peptide, they are
652 transformed into individual site-specific entities for each site.
653
654 **Kinase enrichment analysis between subtypes**
655 To assess enrichment of activated kinases and phosphatases between groups based on
656 functional information such as known protein-protein interactions or kinase-substrate
657 annotations, we input phosphopeptide abundance change values (log2(fold change)) between

658 proteome subtypes into the RoKAI application (v0.8.3)[31] and obtained kinase activity scores at
659 proteome subtype level. Default parameters were used.
660
661 **Transcriptional regulatory network analysis**
662 To deduce the transcriptional regulatory activity of STLMS based on proteome subtype, we
663 utilized the RTN R package[87-89], which examines the link between specific transcription factors
664 (TFs) or gene regulators and all possible targets, as well as the enrichment status of a network
665 using proteome data. In summary, we employed normalized transcriptome data (n=27) and
666 Lambert's human TF list[90](1,612 TFs) to initially build a potential transcriptional regulatory
667 network (TRN). We then eliminated non-significant associations through permutation analysis
668 (permutation n=1000) and bootstrapping, followed by the ARACNe algorithm[91] to preserve direct
669 TF-target interactions, thereby removing redundant indirect interactions. The refined TRN
670 dataset includes the regulator gene name, its target gene name, and adjusted p-values as a
671 confidence measure. For enrichment analyses over a list of regulons in the refined TRN, we first
672 conducted a Master Regulator Analysis (MRA)[92]over a list of regulons, with adjustments for
673 multiple hypothesis testing. The MRA evaluates the overlap between each regulon and the
674 significantly dysregulated genes between subtypes, providing adjusted p-values, which resulted
675 in 600 confident regulons (adjusted p-values <0.05) for further analyses. We then determined
676 these regulons' enrichment scores between proteome subtypes using transcriptome
677 expressional fold change values and the GSEA-2T (two-tailed GSEA analysis) function from the
678 RTN R package. To identify similarities and differences in regulatory programs among samples
679 in our cohort, we selected 60 statistically robust regulons (adjusted p-values <1E-8), calculated
680 the activity score of each regulon in each sample, and visualized the data as a heatmap.
681 Additionally, we incorporated CRISPR perturbation scores (sourced from the Cancer
682 Dependency Map (DepMap) Project[93]) of 36 sarcoma cell lines and included this data in the
683 heatmap.
684
685 **Proteome subtype inference of TCGA STLMS samples**
686 To externally validate differences found in cohort 1 between proteome subtypes (P1, P2, and P3)
687 by using the TCGA STLMS transcriptome dataset, we constructed a random forest classifier of
688 our proteome subtypes based on transcriptome data of cohort 1 (n=27). To select informative
689 genes for model construction, we first conducted differential expression analyses between
690 subtypes and selected only differentially expressed genes (FDR <0.01, absolute value of fold
691 change >2) between subtypes (668 genes). Next, we divided normalized expression data of

692 selected genes into training (n=19) and test (n=8) datasets, followed by random forest model
693 training with 10-fold cross-validation using the tidymodel R package[94]. We applied the trained
694 classifier to the test dataset and got a 1.0 F-means value (which means all sample subtypes
695 were correctly predicted). We then applied the classifier to the TCGA STLMS transcriptome
696 dataset (n=53) and inferred proteomic subtypes in the dataset. TCGA sample IDs and inferred
697 proteomic subtypes are shown in Table S6.
698
699 **Assessment of immune status**
700 ImmuneScores were derived from the normalized protein expression matrix of 72 samples
701 utilizing the ESTIMATE (Estimation of STromal and Immune cells in MAlignant Tumor tissues
702 using Expression data) package in R[40]. To evaluate immune cell infiltration, we conducted IHC
703 with immune cell markers (CD3, CD4, CD8, CD68, CD163) and tallied the number of positive
704 cells for each marker across 3 TMA cores per sample. Subsequently, we computed the average
705 number of positive cells per core for each sample for further analysis.
706
707 **Tissue microarray (TMA) construction**
708 TMAs were created from FFPE blocks of 67 STLMS tissues in cohort 1 and 222 STLMS tissues
709 in cohort 2. Three 1-mm cores were extracted from each tissue paraffin block and placed into
710 tissue array blocks using a TMA arrayer (3DHistech). Tumor and normal regions were chosen
711 after a thorough examination of individual histologic slides and electronic image-based selection
712 of target areas for coring.
713
714 **Immunohistochemistry (IHC)**
715 Four-μm sections were cut from FFPE tissue blocks. Paraffin was removed with xylene, and
716 antigens were retrieval by heat-mediated epitope retrieval. Tissue sections were stained using a
717 Leica BOND-MAX IHC autostainer. The antibodies used were anti-CD3 (Leica Microsystems,
718 NCL-L-CD3-565, 1/200 dilution), anti-CD4 (Sigma, 104R-15, 1/100 dilution), anti-CD8 (Dako,
719 M7103, 1/1500 dilution), anti-CD68 (Dako, M0876, 1/200 dilution), anti-CD163 (Leica
720 Microsystems, NCL-CD163, 1/250 dilution), anti-MITF (Abcam, ab303530, 1/100 dilution), anti-
721 MYC (Abcam, ab32072, 1/200 dilution), anti-PAX6 (Abcam, ab195045, 1/100 dilution), anti-
722 AURKA (Cell Signaling Technology, #91590, 1/100 dilution), anti-AURKB (Abcam, ab45145,
723 1/100 dilution), anti-CDK1 (Abcam, ab133327, 1/250 dilution), anti-CDK2 (Cell Signaling
724 Technology, #18048, 1/200 dilution), anti-active YAP1 (Abcam, ab205270, 1/20,000, dilution),
725 anti-TEAD (Abcam, ab97460, 1/400 dilution), anti-NCOR1 (Invitrogen, PA1-844A, 1/1000

726 dilution), anti-RAD51 (Abcam, ab133534, 1/100 dilution), anti-BRCA2 (Sigma, HPA026815, 1/50
727 dilution), anti-RAD50 (Cell Signaling Technology, #61915, 1/4000 dilution), anti-phosphorylated
728 H2AX (Cell Signaling Technology, #9718, 1/200 dilution), anti-Ki67 (Biocare, PRM325, 1/100
729 dilution), anti-IRF1 (Abcam, ab243895, 1/100 dilution), anti-LGALS9 (Abcam, ab227046, 1/200
730 dilution). For MYC, MITF, CDK1, CDK2, NCOR1, YAP1, RAD50, RAD51, PAX6, IRF1, and
731 LGALS9 IHC, The staining intensity of each tumor cell was rated on a scale from 0 to 3+, and the
732 average was taken from three separate tissue cores for each case. To determine the total
733 weighted IHC score (IHC H-score, theoretically ranging from 0 to 300) for a sample, the
734 expression intensity of individual tumor regions (rated 0-3+) was multiplied by their respective
735 contributions (0-100%) to the overall tumor area, and these values were then summed. Two
736 pathologists independently evaluated all tissue samples. If there were any discrepancies, the
737 cases were re-evaluated until a consensus score was achieved. For Ki67, AURKA/B, TEAD,
738 BRCA2, and phosphorylated H2AX IHC, we assessed the fraction of positive tumor cells. BRCA2
739 loss was defined as less than 50% positivity.
740
741 **External datasets**
742 The TCGA STLMS dataset (53 samples) including clinical information, transcriptome data (count
743 data, TPM), and gene alterations with allele-specific copy number data (data for 1 sample was
744 not available) was obtained via the TCGAbiolink R package (version 2.30.0)[71]. Protein
745 expression data from reverse phase protein array (RPPA) experiments of the TCGA cohort was
746 downloaded from MD Anderson Cancer Center database
747 (https://tcpa.drbioright.org/rppa500/main.html).
748
749 **Data and code availability**
750 Our mass spectrometry proteomics data have been deposited in Harvard Dataverse and is
751 available under controlled access (DOI: 10.7910/DVN/BWOCM9). Raw and processed data of
752 MSK-IMPACT genomic sequencing and transcriptome data were deposited to the EGA
753 repository with the identifiers EGAS50000000594 and EGAS50000000595, respectively. Other
754 data derived from this study will be shared upon reasonable request. We did not generate original
755 code in this study. Any additional information required to reanalyze the data reported in this paper
756 is available from the lead contact upon request.
757
758 **Abbreviations (there are a lot more missing)**
759 STLMS - soft tissue leiomyosarcoma. IHC - immunohistochemistry. SNV - single nucleotide

variant. FFPE - formalin fixed paraffin embedded. GSEA - gene set enrichment analysis. PTM-SEA - posttranslational modification signature enrichment analysis. CIN - chromosomal instability. CNA - copy number alteration. DSB - DNA double-strand break. EMT - epithelial–mesenchymal transition. FDR - false discovery rate. APM - antigen-presenting machinery.

Acknowledgements

MHR expresses gratitude for the grants received from the NIH/NCI (R21 CA251992, R21 CA263262, and U01 CA263986), a Cycle for Survival Equinox Innovation Grant, an Investigator Grant from the Neuroendocrine Tumor Research Foundation (NETRF), and the support from the Farmer Family Foundation. Portions of the research were funded by the MSKCC NCI Cancer Center Support Grant (P30 CA008748). The funding organizations did not participate in the study's design, nor in the data collection, analysis, or interpretation.

Author Contributions

AT designed the study, carried out experiments, performed data analyses, and wrote a draft of the manuscript. MO performed sample processing and clinical data collection. YO carried out experiments and data analysis. RH and ZL supported mass spectrometric experiments and data analyses. NA provided expert advice on sarcoma pathology. DK consulted on the study and provided partial funding. JW wrote and edited the manuscript. MR designed and supervised the study, analyzed data, wrote and edited the manuscript, and provided funding.

Declaration of Interests

Atsushi Tanaka, Makiko Ogawa, Yusuke Otani, Ronald C. Hendrickson, Zhuoning Li, and Narasimhon Agaram have stated that they have no conflicts of interest concerning this research. David S. Klimstra served as a consultant for Paige.AI, while Michael H. Roehrl is part of the Scientific Advisory Board at Universal DX. J.Y.W. is the founder of Curandis. These companies did not participate in the study's design, implementation, data analysis, or any other related activities.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used Paperpal in order to check English grammar. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

794

795 **Supplemental information**

796 Document S1. Figures S1-S7

797 Table S1. Clinicopathological characteristics of patients in cohort 1, related to Figures 1-7

798 Table S2. Proteome and phosphoproteome expression data of cohort 1, related to Figures 1-7

799 Table S3. Differential protein expression analysis between subtypes, related to Figure 2

800 Table S4. Differential protein expression analysis between immune-cold and immune-hot
801 samples, related to Figure 7

802 Table S5. MSK-IMPACT summary with GISTIC 2.0 results of cohort 1, related to Figure 2

803 Table S6. Clinicopathological attributes and inferred proteome subtypes of the TCGA STLMS
804 cohort, related to Figures S1-S7

805

806 **Figure titles and legends**

807 **Figure 1. Integrative study design and clinicopathological features of proteome subtypes**

808 (A) Summary workflow and metrics of this study

809 (B) Summary plot of cohort 1 including clinicopathological information and proteome subtypes.
810 In the heatmap, proteins and phosphoproteins which show significant differences (FDR <0.05)
811 between proteome subtypes are shown.

812 (C) Kaplan-Meier curves of overall survival and recurrence-free survival between proteome
813 subtypes

814 (D) Differences of clinicopathological factors between proteome subtypes

815

816 **Figure 2. Landscape of genomic alterations and regulatory activity across proteome**
817 **subtypes**

818 (A) A boxplot showing ploidy between subtypes. P2/P3 show significantly higher ploidy than P1.

819 (B) Bar chart of whole genome duplication (WGD) frequency in each subtype on the left and
820 boxplot of chromosomal instability on the right. P2/P3 show significantly higher frequency of
821 WGD and a higher CIN score than P1.

822 (C) Oncoprint of integrated genomic alterations including SNV, Indel, and CNA. Genes with 8
823 events across cohort 1 (>12% samples) are shown. The bar chart on the right side of the
824 oncoprint shows the alteration frequency across cohort 1. The bar chart on the top shows the
825 number of genomic events of each sample.

826 (D) Bar plot showing arm-level amplifications and deletions. In general, P1 shows lower
827 frequency of arm-level events for each chromosome. P3 shows the highest frequency of arm-

828 level deletions for most chromosomes.
829 (E) Concordant with arm-level event frequencies for each chromosome, total event numbers of
830 amplifications are significantly higher in P2/P3 than in P1. For arm-level deletions, P3 shows the
831 highest event numbers compared with P1 or P2.
832 (F) Focal amplification/deletion patterns in each subtype are shown. P1 has a relatively lower
833 number of focal peak events. The 17p12 peak (NCOR1 and MAP2K4 gene loci) are shared
834 between the P2 and P3 subtypes.
835 (G) Venn diagram showing overlap of genes in in the recurrent focal peaks of each subtype.
836 (H) Balloon plot showing pathway enrichment analyses of Hallmark and KEGG terms between
837 subtypes. Only statistically significant results are shown.
838 (I) Regulatory activity profile of STLMS by subtype status based on transcriptome data.
839 Computed regulon activity scores are shown as a heatmap. Regulons with p<1E-8 from the
840 Master Regulator Analysis process are shown. The boxplot on the right side of the heatmap
841 shows DepMap CRISPER gene effect scores for 36 sarcoma cell lines.
842 (J) Boxplots of MYC, PAX6, MITF IHC scores by subtypes are shown. In addition, representative
843 IHC images are shown. Each transcription factor shows an expression profile of subtypes that is
844 concordant with regulon network analyses of transcriptome data. Bars, 50 μm.
845
846 **Figure 3. Impact of genomic alterations on the proteome**
847 (A) The correlations between CNA and mRNA (left) and protein expression (right) levels (y axes)
848 reveal both cis and trans effects. The presence of significant positive (red) and negative (green)
849 Spearman's correlations (FDR <0.1) between CNA and mRNA or protein is highlighted. The
850 numbers of significant cis and trans events corresponding to the indicated genomic loci specific
851 to mRNAs or proteins are represented by blue bars, while the overlap between CNA-mRNA and
852 CNA-protein events is depicted by black bars.
853 (B) Venn diagram showing cis correlation overlap between CNA-mRNA and CNA-protein.
854 (C) GOBP over-representatation analysis of cis correlation between CNA-protein is shown. Cell
855 cycle-related terms are significantly enriched.
856 (D) Genomic effect of cell cycle-related genes on mRNA and protein. Columns represent each
857 sample and are ordered by cell cycle score. The bar chart on the right side of heatmap shows
858 Spearman's correlation coefficients between targets and cell cycle score. "E2F targets" is based
859 on ssGSEA of the Hallmark E2F term. Kinase activity scores (CDK1,2 kinase activities) are from
860 PTM-GSE analysis. p_RB1 denotes phosphorylated protein of RB1. * denotes statistical
861 significance (p<0.05).

862 (E) Boxplots showing proteome subtype differences of GOBP cell cycle scores and Ki67-positive
863 tumor frequency.
864 (F) Boxplot showing AURKA protein expression (IHC score) by subtypes.
865 (G) Boxplot of E2F scores from ssGSEA by subtypes.
866 (H) Boxplot showing kinase activity scores of CDK1/2 by subtypes.
867 (I) Boxplot of Depmap CRISPR gene perturbation scores targeting CDKs and AURKA using 3
868 leiomyosarcoma cell lines. Score below -1 represents significant effect on cell survival when the
869 target gene is inhibited.
870
871 **Figure 4. Phosphoproteomic differences between subtypes**
872 (A) Oncoprint showing RTK-RAS pathways with 4 or more events. Amplification of MAP2K4 and
873 IFG1R are the top 2 most frequent alterations. * denotes statistically significant differences
874 (Fisher's exact test) in gene alteration frequency between subtypes.
875 (B) Over-representation analyses of GOBP terms. Phosphoproteins which are positively
876 upregulated in one subtype compared to other subtypes with FDR<0.05 are used for GOBP term
877 enrichment analyses.
878 (C) Kinase activity heatmap by subtypes. Z-scores of inferred kinase activities are plotted as a
879 heatmap. * denotes statistical significance. Kinase names are shown on the left side of the
880 heatmap.
881 (D) Kinase-substrate networks of the top 20 positively enriched kinases in each subtype are
882 shown. Note: all positively enriched kinases of P1 are shown due to its small number. Genes
883 shown as ellipses are kinases. Genes shown as rectangles are kinase substrates Colors of
884 kinase substrates denote the phosphorylation Z-score of the target's phosphorylation site. If
885 multiple phosphorylation sites exist in a substrate, phosphorylation averages are shown based
886 on color scale.
887 (E) Boxplot of PTM-GSEA AURKB activity score shows enrichment in P2/P3 compared to P1.
888 (F) Representative AURKB IHC image and boxplot of AURKB IHC score by subtype. Bar,100
889 μm.
890 (G) Networks of AURKB and inferred regulatory kinases are shown. Motif analysis of AURKB-
891 T232 identifies 194 kinases. Among them, kinases with significant enrichment (FDR<0.05) in
892 RoKAI app analysis between subtypes were selected to be shown here. Colors of kinases denote
893 Z-scores of kinase activity.
894
895 **Figure 5. NCOR1 expression corelates with muscular differentiation, stemness, and**

896 **immune status**
897 (A) Oncoprint of epigenetic regulators in cohort 1. NCOR1 gene is the top altered epigenetic
898 regulator. * denotes statistical difference of gene alteration frequency between subtypes.
899 (B) NCOR1 amplification shows higher frequency in P2/P3 compared to P1 in cohort 1 (butnot
900 statistically significant).
901 (C) NCOR1 copy number positively correlates with its protein level (Spearman's correlation
902 coefficient) in cohort 1.
903 (D) Boxplot of NCOR1 protein expression by subtypes in cohort 1.
904 (E) Top 5 positively and negatively enriched GOBP GSEA results are shown. A ranked gene list
905 based on fold change values between NCOR1-low and -high groups in cohort 1 was used for
906 GSEA.
907 (F) Boxplot of smooth muscle protein markers between NCOR1-low and NCOR1-high.
908 (G) Boxplot of GOBP_STEM_CELL_DIFFERENTIATION ssGSEA scores. In cohort 1, the
909 NCOR1-high group shows significantly higher scores compared to the NCOR1-low group.
910 (H) Boxplot of stemness markers between NCOR1-low/high groups. PROM1 (CD133) and
911 ALDH7A1 in the NCOR1-high group show significantly higher protein expression than the
912 NCOR1-low group in cohort 1.
913 (I) Boxplots of YAP1 IHC H-scores between NCOR1-low and -high groups in cohorts 1 and 2.
914 (J) Boxplots of TEAD IHC H-scores between NCOR1-low and -high groups in cohorts 1 and 2.
915
916 **Figure 6. Homologous recombination repair pathway**
917 (A) Oncoprint of HRR pathway components showing frequent LOH events in cohort 1.
918 (B) Boxplot of genomic alteration counts in the HRR pathway per sample shows high occurrence
919 in P2/P3, but not statistically significant.
920 (C) The HRD score is higher in P2/P3 compared to P1.
921 (D) The HRD score positively correlates with CIN.
922 (E) The RAD51 IHC scores are lower in P2/P3 compared to P1, which is concordant with the
923 HRD scores of these subtypes.
924 (F) Boxplot of HRD score comparison between paired primary and metastatic samples of P2 and
925 P3. Metastatic samples of P3 show a significant increase. Although overall not statistically
926 significant, 3 out of 4 samples of P2 show HRD score increases.
927 (G) UpSet plot of thetop 5 altered genes. BRCA2 is the top altered gene with multiple different
928 cooccurrence partners.
929 (H) Boxplot of HRD scores between BRCA2 LOH and non-LOH groups showing significant

930 increase of HRD scores in the LOH group.
931 (I) Boxplots of mRNA and protein expression between LOH and non-LOH groups for each HRR
932 gene. Among 37 HRR genes (*FANCD2* was excluded due to low LOH event number), 16 proteins
933 were quantified in cohort 1 and are shown here.
934 (J) Bar plot showing frequencies of BRCA2-retaining samples in each subtype. P2/P3 subtypes,
935 especially P2, show decrease of BRCA2 retention compared to P1.
936 (K) The RAD50 IHC score does not change significantly between HRD-low and HRD-high
937 groups, but it shows a downward trend in P2/P3 compared to P1.
938 (L) Boxplot of phosphorylated H2AX IHC scores by subtypes.
939 (M) Boxplot of XRCC5/6 protein expression by subtypes. P3 shows significantly higher
940 expression of XRCC5/6 compared to P1 or P2.
941 (N) Boxplot of PARP1 protein expression between HRD-low and HRD-high groups by subtypes.
942 (O) Kaplan-Meier analysis of overall and recurrence-free survival in combined cohorts 1 and 2,
943 stratified by median values of RAD51 H-score and BRCA2 tumor cell positivity.
944
945 **Figure 7. Tumor microenvironment profiling shows subtype differences and immune**
946 **evasion characteristics in both immune-cold and immune-hot tumors**
947 (A) A heatmap of immune signatures including copy number status, mRNA, and protein level. In
948 general, the ImmuneScore positively correlates with CD3 count, antigen presenting molecules,
949 and immune-related pathways. Bars on the right side of the heatmap denote Spearman's
950 correlation coefficients between ImmuneScores and each gene/signature.
951 (B) Boxplot showing the ImmuneScore by proteome subtypes for cohort 1.
952 (C) Boxplot showing the ImmuneScore between paired primary and metastatic tumors.
953 (D) Volcano plot showing differential protein expression results between immune-cold and
954 immune-hot. Proteins with statistical significance ($q<0.05$) are shown in red (log2(fold change)>1)
955 and blue (log2(fold change)<-1). The horizontal dashed line denotes $q=0.05$.
956 (E) GSE plot of antigen presenting pathway-related KEGG and REACTOME terms, showing
957 antigen presenting processes are significantly downregulated in immune-cold tumors compared
958 to immune-hot tumors.
959 (F) A bar plot of APM protein expression ratios between immune-cold and immune-hot in cohort
960 1. * denotes statistical significance in the differential expressional analysis shown in (D).
961 (G) Boxplot of IRF1 protein expression between immune-hot and immune-cold tumors
962 (H) A correlation plot (colored by Spearman correlation coefficient) including the ImmuneScore,
963 IRF1, and antigen processing machinery protein expression. *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$.

964 IFNgamma_response scores are ssGSEA scores of the
965 “Hallmark_Interferon_Gamma_Reponse” term.
966 (I) Boxplot of immune-suppressive molecules between immune-cold and immune-hot tumors
967 showing only TIM3 is significantly upregulated in immune-hot tumors vs. immune-cold tumors.
968 (J) Boxplot of LGALS9 protein expression by subtypes (left side) and representative IHC image
969 (right side). P2 shows a trend for higher expression of LGALS9 compared to P1 or P3. Bar, 50
970 μm.
971 (K) Boxplots of CD68-positive or CD163 -sitive cell counts by subtypes.
972 (L) Kaplan-Meier analyses of recurrence-free survival in combined cohorts 1 and 2, stratified by
973 median values of CD68-positive or CD163-positive cell counts.
974
975 **Figure 8. Integrated summary of proteome-defined STLMS subtypes**
976 Schematic overview of proteome-defined STLMS subtypes (P1-P3). Major genomic alterations,
977 pathway alterations, kinome alterations, DNA repair alterations, tumor microenvironment
978 features, prognostic trends, and hypothesis-generating therapeutic targets are summarized.
979 APM, antigen-processing machinery; EMT, epithelial-mesenchymal transition; HRD,
980 homologous recombination deficiency; LOH, loss of heterozygosity; NHEJ, non-homologous
981 end-joining; TME, tumor microenvironment.

References

1. Board, W.C.o.T.E. (2020). Soft Tissue and Bone Tumours.
2. Devaud, N., Vornicova, O., Abdul Razak, A.R., Khalili, K., Demicco, E.G., Mitric, C., Bernardini, M.Q., and Gladdy, R.A. (2022). Leiomyosarcoma: Current Clinical Management and Future Horizons. Surg Oncol Clin N Am *31*, 527-546. 10.1016/j.soc.2022.03.011.
3. Serrano, C., and George, S. (2013). Leiomyosarcoma. Hematol Oncol Clin North Am *27*, 957-974. 10.1016/j.hoc.2013.07.002.
4. Acem, I., van Houdt, W.J., Grünhagen, D.J., van der Graaf, W.T.A., Rueten-Budde, A.J., Gelderblom, H., Verhoef, C., and van de Sande, M.A.J. (2022). The role of perioperative chemotherapy in primary high-grade extremity soft tissue sarcoma: a risk-stratified analysis using PERSARC. European journal of cancer (Oxford, England : 1990) *165*, 71-80. 10.1016/j.ejca.2022.01.013.
5. Gronchi, A., Miah, A.B., Dei Tos, A.P., Abecassis, N., Bajpai, J., Bauer, S., Biagini, R., Bielack, S., Blay, J.Y., Bolle, S., et al. (2021). Soft tissue and visceral sarcomas: ESMO-EURACAN-GENTURIS Clinical Practice Guidelines for diagnosis, treatment and follow-up(☆). Ann Oncol *32*, 1348-1365. 10.1016/j.annonc.2021.07.006.
6. Pisters, P.W., Leung, D.H., Woodruff, J., Shi, W., and Brennan, M.F. (1996). Analysis of prognostic factors in 1,041 patients with localized soft tissue sarcomas of the extremities. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *14*, 1679-1689. 10.1200/jco.1996.14.5.1679.
7. Lewis, J.J., Leung, D., Heslin, M., Woodruff, J.M., and Brennan, M.F. (1997). Association of local recurrence with subsequent survival in extremity soft tissue sarcoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *15*, 646-652. 10.1200/jco.1997.15.2.646.
8. Trovik, C.S., Bauer, H.C., Alvegård, T.A., Anderson, H., Blomqvist, C., Berlin, O., Gustafson, P., Saeter, G., and Wallöe, A. (2000). Surgical margins, local recurrence and metastasis in soft tissue sarcomas: 559 surgically-treated patients from the Scandinavian Sarcoma Group Register. European journal of cancer (Oxford, England : 1990) *36*, 710-716. 10.1016/s0959-8049(99)00287-7.
9. Linch, M., Miah, A.B., Thway, K., Judson, I.R., and Benson, C. (2014). Systemic treatment of soft-tissue sarcoma-gold standard and novel therapies. Nat Rev Clin Oncol *11*, 187-202. 10.1038/nrclinonc.2014.26.
10. Savina, M., Le Cesne, A., Blay, J.Y., Ray-Coquard, I., Mir, O., Toulmonde, M., Cousin, S., Terrier, P., Ranchere-Vince, D., Meeus, P., et al. (2017). Patterns of care and outcomes of patients with METAstatic soft tissue SARComa in a real-life setting: the METASARC observational study. BMC Med *15*, 78. 10.1186/s12916-017-0831-7.
11. Chibon, F., Lagarde, P., Salas, S., Pérot, G., Brouste, V., Tirode, F., Lucchesi, C., de Reynies, A., Kauffmann, A., Bui, B., et al. (2010). Validated prediction of clinical outcome in sarcomas and multiple types of cancer on the basis of a gene expression signature related to genome complexity. Nature medicine *16*, 781-787. 10.1038/nm.2174.

1023 12. Network, T.C.G.A.R. (2017). Comprehensive and Integrated Genomic Characterization of
1024 Adult Soft Tissue Sarcomas. Cell *171*, 950-965.e928. 10.1016/j.cell.2017.10.014.
1025 13. Nacev, B.A., Sanchez-Vega, F., Smith, S.A., Antonescu, C.R., Rosenbaum, E., Shi, H., Tang,
1026 C., Socci, N.D., Rana, S., Gularte-Merida, R., et al. (2022). Clinical sequencing of soft tissue
1027 and bone sarcomas delineates diverse genomic landscapes and potential therapeutic
1028 targets. Nat Commun *13*, 3405. 10.1038/s41467-022-30453-x.
1029 14. Italiano, A., Kind, M., Stoeckle, E., Jones, N., Coindre, J.M., and Bui, B. (2011).
1030 Temsirolimus in advanced leiomyosarcomas: patterns of response and correlation with
1031 the activation of the mammalian target of rapamycin pathway. Anticancer Drugs *22*, 463-
1032 467. 10.1097/CAD.0b013e3283442074.
1033 15. Schwartz, G.K., Tap, W.D., Qin, L.X., Livingston, M.B., Undevia, S.D., Chmielowski, B.,
1034 Agulnik, M., Schuetze, S.M., Reed, D.R., Okuno, S.H., et al. (2013). Cixutumumab and
1035 temsirolimus for patients with bone and soft-tissue sarcoma: a multicentre, open-label,
1036 phase 2 trial. The Lancet. Oncology *14*, 371-382. 10.1016/s1470-2045(13)70049-4.
1037 16. Burns, J., Wilding, C.P., R, L.J., and P, H.H. (2020). Proteomic research in sarcomas -
1038 current status and future opportunities. Semin Cancer Biol *61*, 56-70.
1039 10.1016/j.semcancer.2019.11.003.
1040 17. Gillette, M.A., Satpathy, S., Cao, S., Dhanasekaran, S.M., Vasaikar, S.V., Krug, K., Petralia,
1041 F., Li, Y., Liang, W.W., Reva, B., et al. (2020). Proteogenomic Characterization Reveals
1042 Therapeutic Vulnerabilities in Lung Adenocarcinoma. Cell *182*, 200-225.e235.
1043 10.1016/j.cell.2020.06.013.
1044 18. Zhang, B., Wang, J., Wang, X., Zhu, J., Liu, Q., Shi, Z., Chambers, M.C., Zimmerman, L.J.,
1045 Shaddox, K.F., Kim, S., et al. (2014). Proteogenomic characterization of human colon and
1046 rectal cancer. Nature *513*, 382-387. 10.1038/nature13438.
1047 19. Burns, J., Wilding, C.P., Krasny, L., Zhu, X., Chadha, M., Tam, Y.B., Ps, H., Mahalingam,
1048 A.H., Lee, A.T.J., Arthur, A., et al. (2023). The proteomic landscape of soft tissue
1049 sarcomas. Nat Commun *14*, 3834. 10.1038/s41467-023-39486-2.
1050 20. Tang, S., Wang, Y., Luo, R., Fang, R., Liu, Y., Xiang, H., Ran, P., Tong, Y., Sun, M., Tan, S., et
1051 al. (2024). Proteomic characterization identifies clinically relevant subgroups of soft
1052 tissue sarcoma. Nat Commun *15*, 1381. 10.1038/s41467-024-45306-y.
1053 21. AACR Project GENIE: Powering Precision Medicine through an International Consortium.
1054 (2017). Cancer Discov *7*, 818-831. 10.1158/2159-8290.Cd-17-0151.
1055 22. Ucar, D., and Lin, D.I. (2015). Amplification of the bromodomain-containing protein 4
1056 gene in ovarian high-grade serous carcinoma is associated with worse prognosis and
1057 survival. Mol Clin Oncol *3*, 1291-1294. 10.3892/mco.2015.622.
1058 23. Brahmi, M., Lesluyes, T., Dufresne, A., Toulmonde, M., Italiano, A., Mir, O., Le Cesne, A.,
1059 Valentin, T., Chevreau, C., Bonvalot, S., et al. (2021). Expression and prognostic
1060 significance of PDGF ligands and receptors across soft tissue sarcomas. ESMO Open *6*,
1061 100037. 10.1016/j.esmoop.2020.100037.
1062 24. Peng, L., Li, J., Wu, J., Xu, B., Wang, Z., Giamas, G., Stebbing, J., and Yu, Z. (2021). A Pan-
1063 Cancer Analysis of SMARCA4 Alterations in Human Cancers. Front Immunol *12*, 762598.
1064 10.3389/fimmu.2021.762598.

1065 25. Liang, J., Li, B., Yuan, L., and Ye, Z. (2015). Prognostic value of IGF-1R expression in bone
1066 and soft tissue sarcomas: a meta-analysis. Onco Targets Ther *8*, 1949-1955.
1067 10.2147/ott.S88293.
1068 26. Li, Y., Li, Y., Liu, Y., Xie, P., Li, F., and Li, G. (2014). PAX6, a novel target of microRNA-7,
1069 promotes cellular proliferation and invasion in human colorectal cancer cells. Dig Dis Sci
1070 *59*, 598-606. 10.1007/s10620-013-2929-x.
1071 27. Ooki, A., Dinalankara, W., Marchionni, L., Tsay, J.J., Goparaju, C., Maleki, Z., Rom, W.N.,
1072 Pass, H.I., and Hoque, M.O. (2018). Epigenetically regulated PAX6 drives cancer cells
1073 toward a stem-like state via GLI-SOX2 signaling axis in lung adenocarcinoma. Oncogene
1074 *37*, 5967-5981. 10.1038/s41388-018-0373-2.
1075 28. Jing, N., Du, X., Liang, Y., Tao, Z., Bao, S., Xiao, H., Dong, B., Gao, W.Q., and Fang, Y.X.
1076 (2024). PAX6 promotes neuroendocrine phenotypes of prostate cancer via enhancing
1077 MET/STAT5A-mediated chromatin accessibility. J Exp Clin Cancer Res *43*, 144.
1078 10.1186/s13046-024-03064-1.
1079 29. Ballotti, R., Cheli, Y., and Bertolotto, C. (2020). The complex relationship between MITF
1080 and the immune system: a Melanoma ImmunoTherapy (response) Factor? Molecular
1081 cancer *19*, 170. 10.1186/s12943-020-01290-7.
1082 30. Du, R., Huang, C., Liu, K., Li, X., and Dong, Z. (2021). Targeting AURKA in Cancer: molecular
1083 mechanisms and opportunities for Cancer therapy. Molecular cancer *20*, 15.
1084 10.1186/s12943-020-01305-3.
1085 31. Yilmaz, S., Ayati, M., Schlatzer, D., Cicek, A.E., Chance, M.R., and Koyuturk, M. (2021).
1086 Robust inference of kinase activity using functional networks. Nat Commun *12*, 1177.
1087 10.1038/s41467-021-21211-6.
1088 32. Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the
1089 genome integrates intrinsic and environmental signals. Nature genetics *33 Suppl*, 245-
1090 254. 10.1038/ng1089.
1091 33. Li, X., and Heyer, W.D. (2008). Homologous recombination in DNA repair and DNA
1092 damage tolerance. Cell Res *18*, 99-113. 10.1038/cr.2008.1.
1093 34. Lord, C.J., and Ashworth, A. (2016). BRCAness revisited. Nature reviews. Cancer *16*, 110-
1094 120. 10.1038/nrc.2015.21.
1095 35. Turajlic, S., and Swanton, C. (2016). Metastasis as an evolutionary process. Science (New
1096 York, N.Y.) *352*, 169-175. 10.1126/science.aaf2784.
1097 36. Bakhoum, S.F., Ngo, B., Laughney, A.M., Cavallo, J.A., Murphy, C.J., Ly, P., Shah, P., Sriram,
1098 R.K., Watkins, T.B.K., Taunk, N.K., et al. (2018). Chromosomal instability drives metastasis
1099 through a cytosolic DNA response. Nature *553*, 467-472. 10.1038/nature25432.
1100 37. Planas-Paz, L., Pliego-Mendieta, A., Hagedorn, C., Aguilera-Garcia, D., Haberecker, M.,
1101 Arnold, F., Herzog, M., Bankel, L., Guggenberger, R., Steiner, S., et al. (2023). Unravelling
1102 homologous recombination repair deficiency and therapeutic opportunities in soft tissue
1103 and bone sarcoma. EMBO Mol Med *15*, e16863. 10.15252/emmm.202216863.
1104 38. Baumann, P., and West, S.C. (1998). Role of the human RAD51 protein in homologous
1105 recombination and double-stranded-break repair. Trends Biochem Sci *23*, 247-251.
1106 10.1016/s0968-0004(98)01232-8.

1107 39. Mansour, W.Y., Rhein, T., and Dahm-Daphi, J. (2010). The alternative end-joining
1108 pathway for repair of DNA double-strand breaks requires PARP1 but is not dependent
1109 upon microhomologies. Nucleic Acids Res *38*, 6065-6077. 10.1093/nar/gkq387.
1110 40. Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W.,
1111 Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and
1112 stromal and immune cell admixture from expression data. Nat Commun *4*, 2612.
1113 10.1038/ncomms3612.
1114 41. Zhou, F. (2009). Molecular mechanisms of IFN-gamma to up-regulate MHC class I antigen
1115 processing and presentation. Int Rev Immunol *28*, 239-260.
1116 10.1080/08830180902978120.
1117 42. Der, S.D., Zhou, A., Williams, B.R., and Silverman, R.H. (1998). Identification of genes
1118 differentially regulated by interferon alpha, beta, or gamma using oligonucleotide arrays.
1119 Proceedings of the National Academy of Sciences of the United States of America *95*,
1120 15623-15628. 10.1073/pnas.95.26.15623.
1121 43. van den Elsen, P.J., Holling, T.M., van der Stoep, N., and Boss, J.M. (2003). DNA
1122 methylation and expression of major histocompatibility complex class I and class II
1123 transactivator genes in human developmental tumor cells and in T cell malignancies.
1124 Clinical Immunology *109*, 46-52. 10.1016/s1521-6616(03)00200-6.
1125 44. Mora-Garcia Mde, L., Duenas-Gonzalez, A., Hernandez-Montes, J., De la Cruz-Hernandez,
1126 E., Perez-Cardenas, E., Weiss-Steider, B., Santiago-Osorio, E., Ortiz-Navarrete, V.F.,
1127 Rosales, V.H., Cantu, D., et al. (2006). Up-regulation of HLA class-I antigen expression and
1128 antigen-specific CTL response in cervical cancer cells by the demethylating agent
1129 hydralazine and the histone deacetylase inhibitor valproic acid. J Transl Med *4*, 55.
1130 10.1186/1479-5876-4-55.
1131 45. Angell, T.E., Lechner, M.G., Jang, J.K., LoPresti, J.S., and Epstein, A.L. (2014). MHC class I
1132 loss is a frequent mechanism of immune escape in papillary thyroid cancer that is
1133 reversed by interferon and selumetinib treatment in vitro. Clin Cancer Res *20*, 6034-
1134 6044. 10.1158/1078-0432.CCR-14-0879.
1135 46. Hu, J., Li, X., Yang, L., and Li, H. (2022). Hypoxia, a key factor in the immune
1136 microenvironment. Biomed Pharmacother *151*, 113068. 10.1016/j.biopha.2022.113068.
1137 47. Abou Khouzam, R., Brodaczewska, K., Filipiak, A., Zeinelabdin, N.A., Buart, S., Szczylik, C.,
1138 Kieda, C., and Chouaib, S. (2020). Tumor Hypoxia Regulates Immune Escape/Invasion:
1139 Influence on Angiogenesis and Potential Impact of Hypoxic Biomarkers on Cancer
1140 Therapies. Front Immunol *11*, 613114. 10.3389/fimmu.2020.613114.
1141 48. Pietrobon, V., and Marincola, F.M. (2021). Hypoxia and the phenomenon of immune
1142 exclusion. J Transl Med *19*, 9. 10.1186/s12967-020-02667-4.
1143 49. Kikutake, C., and Suyama, M. (2023). Pan-cancer analysis of whole-genome doubling and
1144 its association with patient prognosis. BMC cancer *23*, 619. 10.1186/s12885-023-11132-
1145 6.
1146 50. Baker, T.M., Lai, S., Lynch, A.R., Lesluyes, T., Yan, H., Ogilvie, H.A., Verfaillie, A., Dentro,
1147 S., Bowes, A.L., Pillay, N., et al. (2024). The History of Chromosomal Instability in Genome-
1148 Doubled Tumors. Cancer Discov *14*, 1810-1822. 10.1158/2159-8290.Cd-23-1249.

1149 51. Kaufmann, T.L., Petkovic, M., Watkins, T.B.K., Colliver, E.C., Laskina, S., Thapa, N.,
1150 Minussi, D.C., Navin, N., Swanton, C., Van Loo, P., et al. (2022). MEDICC2: whole-genome
1151 doubling aware copy-number phylogenies for cancer evolution. Genome biology *23*, 241.
1152 10.1186/s13059-022-02794-9.
1153 52. Nikhil, K., and Shah, K. (2024). The significant others of aurora kinase a in cancer:
1154 combination is the key. Biomark Res *12*, 109. 10.1186/s40364-024-00651-4.
1155 53. Jamalzadeh, S., Dai, J., Lavikka, K., Li, Y., Jiang, J., Huhtinen, K., Virtanen, A., Oikkonen, J.,
1156 Hietanen, S., Hynninen, J., et al. (2024). Genome-wide quantification of copy-number
1157 aberration impact on gene expression in ovarian high-grade serous carcinoma. BMC
1158 cancer *24*, 173. 10.1186/s12885-024-11895-6.
1159 54. Portella, G., Passaro, C., and Chieffi, P. (2011). Aurora B: a new prognostic marker and
1160 therapeutic target in cancer. Curr Med Chem *18*, 482-496.
1161 10.2174/092986711794480203.
1162 55. Gautschi, O., Heighway, J., Mack, P.C., Purnell, P.R., Lara, P.N., Jr., and Gandara, D.R.
1163 (2008). Aurora kinases as anticancer drug targets. Clin Cancer Res *14*, 1639-1648.
1164 10.1158/1078-0432.Ccr-07-2179.
1165 56. Dang, C.V. (2012). MYC on the path to cancer. Cell *149*, 22-35.
1166 10.1016/j.cell.2012.03.003.
1167 57. Yamamoto, H., Williams, E.G., Mouchiroud, L., Cantó, C., Fan, W., Downes, M., Héligon,
1168 C., Barish, G.D., Desvergne, B., Evans, R.M., et al. (2011). NCoR1 is a conserved
1169 physiological modulator of muscle mass and oxidative function. Cell *147*, 827-839.
1170 10.1016/j.cell.2011.10.017.
1171 58. Kohn, E.C., Lee, J.M., and Ivy, S.P. (2017). The HRD Decision-Which PARP Inhibitor to Use
1172 for Whom and When. Clin Cancer Res *23*, 7155-7157. 10.1158/1078-0432.Ccr-17-2186.
1173 59. Slade, D., and Loizou, J.I. (2023). Leveraging homologous recombination repair deficiency
1174 in sarcoma. EMBO Mol Med *15*, e17453. 10.15252/emmm.202317453.
1175 60. Dall, G., Vandenberg, C.J., Nesic, K., Ratnayake, G., Zhu, W., Vissers, J.H.A., Bedő, J.,
1176 Penington, J., Wakefield, M.J., Kee, D., et al. (2023). Targeting homologous
1177 recombination deficiency in uterine leiomyosarcoma. J Exp Clin Cancer Res *42*, 112.
1178 10.1186/s13046-023-02687-0.
1179 61. Mimura, K., Kua, L.F., Xiao, J.F., Asuncion, B.R., Nakayama, Y., Syn, N., Fazreen, Z., Soong,
1180 R., Kono, K., and Yong, W.P. (2021). Combined inhibition of PD-1/PD-L1, Lag-3, and Tim-
1181 3 axes augments antitumor immunity in gastric cancer-T cell coculture models. Gastric
1182 cancer : official journal of the International Gastric Cancer Association and the Japanese
1183 Gastric Cancer Association *24*, 611-623. 10.1007/s10120-020-01151-8.
1184 62. Cheng, D.T., Mitchell, T.N., Zehir, A., Shah, R.H., Benayed, R., Syed, A., Chandramohan,
1185 R., Liu, Z.Y., Won, H.H., Scott, S.N., et al. (2015). Memorial Sloan Kettering-Integrated
1186 Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-
1187 Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology.
1188 The Journal of molecular diagnostics : JMD *17*, 251-264. 10.1016/j.jmoldx.2014.12.006.
1189 63. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-
1190 Wheeler transform. Bioinformatics (Oxford, England) *25*, 1754-1760.
1191 10.1093/bioinformatics/btp324.

1192 64. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis,
1193 A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation
1194 discovery and genotyping using next-generation DNA sequencing data. Nature genetics
1195 *43*, 491-498. 10.1038/ng.806.
1196 65. Mose, L.E., Wilkerson, M.D., Hayes, D.N., Perou, C.M., and Parker, J.S. (2014). ABRA:
1197 improved coding indel detection via assembly-based realignment. Bioinformatics
1198 (Oxford, England) *30*, 2813-2815. 10.1093/bioinformatics/btu376.
1199 66. Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel,
1200 S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point
1201 mutations in impure and heterogeneous cancer samples. Nature biotechnology *31*, 213-
1202 219. 10.1038/nbt.2514.
1203 67. Ye, K., Schulz, M.H., Long, Q., Apweiler, R., and Ning, Z. (2009). Pindel: a pattern growth
1204 approach to detect break points of large deletions and medium sized insertions from
1205 paired-end short reads. Bioinformatics (Oxford, England) *25*, 2865-2871.
1206 10.1093/bioinformatics/btp394.
1207 68. Lai, Z., Markovets, A., Ahdesmaki, M., Chapman, B., Hofmann, O., McEwen, R., Johnson,
1208 J., Dougherty, B., Barrett, J.C., and Dry, J.R. (2016). VarDict: a novel and versatile variant
1209 caller for next-generation sequencing in cancer research. Nucleic Acids Res *44*, e108.
1210 10.1093/nar/gkw227.
1211 69. Shen, R., and Seshan, V.E. (2016). FACETS: allele-specific copy number and clonal
1212 heterogeneity analysis tool for high-throughput DNA sequencing. Nucleic Acids Res *44*,
1213 e131. 10.1093/nar/gkw520.
1214 70. Bielski, C.M., Zehir, A., Penson, A.V., Donoghue, M.T.A., Chatila, W., Armenia, J., Chang,
1215 M.T., Schram, A.M., Jonsson, P., Bandlamudi, C., et al. (2018). Genome doubling shapes
1216 the evolution and prognosis of advanced cancers. Nature genetics *50*, 1189-1195.
1217 10.1038/s41588-018-0165-1.
1218 71. Colaprico, A., Silva, T.C., Olsen, C., Garofano, L., Cava, C., Garolini, D., Sabedot, T.S.,
1219 Malta, T.M., Pagnotta, S.M., Castiglioni, I., et al. (2016). TCGAbiolinks: an R/Bioconductor
1220 package for integrative analysis of TCGA data. Nucleic Acids Res *44*, e71.
1221 10.1093/nar/gkv1507.
1222 72. Sztupinszki, Z., Diossy, M., Krzystanek, M., Reiniger, L., Csabai, I., Favero, F., Birkbak, N.J.,
1223 Eklund, A.C., Syed, A., and Szallasi, Z. (2018). Migrating the SNP array-based homologous
1224 recombination deficiency measures to next generation sequencing data of breast cancer.
1225 NPJ Breast Cancer *4*, 16. 10.1038/s41523-018-0066-6.
1226 73. Shi, Z., Chen, B., Han, X., Gu, W., Liang, S., and Wu, L. (2023). Genomic and molecular
1227 landscape of homologous recombination deficiency across multiple cancer types.
1228 Scientific reports *13*, 8899. 10.1038/s41598-023-35092-w.
1229 74. Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., and Getz, G.
1230 (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal
1231 somatic copy-number alteration in human cancers. Genome Biol *12*, R41. 10.1186/gb-
1232 2011-12-4-r41.

1233 75. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson,
1234 M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics
1235 (Oxford, England) *29*, 15-21. 10.1093/bioinformatics/bts635.
1236 76. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq
1237 data with or without a reference genome. BMC bioinformatics *12*, 323. 10.1186/1471-
1238 2105-12-323.
1239 77. McAlister, G.C., Nusinow, D.P., Jedrychowski, M.P., Wuhr, M., Huttlin, E.L., Erickson, B.K.,
1240 Rad, R., Haas, W., and Gygi, S.P. (2014). MultiNotch MS3 Enables Accurate, Sensitive, and
1241 Multiplexed Detection of Differential Expression across Cancer Cell Line Proteomes.
1242 Analytical Chemistry *86*, 7150-7158. 10.1021/ac502040v.
1243 78. Navarrete-Perea, J., Yu, Q., Gygi, S.P., and Paulo, J.A. (2018). Streamlined Tandem Mass
1244 Tag (SL-TMT) Protocol: An Efficient Strategy for Quantitative (Phospho)proteome
1245 Profiling Using Tandem Mass Tag-Synchronous Precursor Selection-MS3. J Proteome Res
1246 *17*, 2226-2236. 10.1021/acs.jproteome.8b00217.
1247 79. Liu, H., Sadygov, R.G., and Yates, J.R., 3rd (2004). A model for random sampling and
1248 estimation of relative protein abundance in shotgun proteomics. Analytical chemistry *76*,
1249 4193-4201. 10.1021/ac0498563.
1250 80. Plubell, D.L., Wilmarth, P.A., Zhao, Y., Fenton, A.M., Minnier, J., Reddy, A.P., Klimek, J.,
1251 Yang, X., David, L.L., and Pamir, N. (2017). Extended Multiplexing of Tandem Mass Tags
1252 (TMT) Labeling Reveals Age and High Fat Diet Specific Proteome Changes in Mouse
1253 Epididymal Adipose Tissue. Mol Cell Proteomics *16*, 873-890.
1254 10.1074/mcp.M116.065524.
1255 81. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package
1256 for differential expression analysis of digital gene expression data. Bioinformatics
1257 (Oxford, England) *26*, 139-140. 10.1093/bioinformatics/btp616.
1258 82. Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix
1259 factorization. BMC bioinformatics *11*, 367. 10.1186/1471-2105-11-367.
1260 83. Kim, H., and Park, H. (2007). Sparse non-negative matrix factorizations via alternating
1261 non-negativity-constrained least squares for microarray data analysis. Bioinformatics
1262 (Oxford, England) *23*, 1495-1502. 10.1093/bioinformatics/btm134.
1263 84. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for
1264 comparing biological themes among gene clusters. Omics *16*, 284-287.
1265 10.1089/omi.2011.0118.
1266 85. Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C.,
1267 Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that
1268 oncogenic KRAS-driven cancers require TBK1. Nature *462*, 108-112.
1269 10.1038/nature08460.
1270 86. Hornbeck, P.V., Zhang, B., Murray, B., Kornhauser, J.M., Latham, V., and Skrzypek, E.
1271 (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res *43*,
1272 D512-520. 10.1093/nar/gku1267.
1273 87. Castro, M.A., de Santiago, I., Campbell, T.M., Vaughn, C., Hickey, T.E., Ross, E., Tilley,
1274 W.D., Markowetz, F., Ponder, B.A., and Meyer, K.B. (2016). Regulators of genetic risk of

1275 breast cancer identified by integrative network analysis. Nature genetics *48*, 12-21.
1276 10.1038/ng.3458.
1277 88. Fletcher, M.N., Castro, M.A., Wang, X., de Santiago, I., O'Reilly, M., Chin, S.F., Rueda,
1278 O.M., Caldas, C., Ponder, B.A., Markowetz, F., and Meyer, K.B. (2013). Master regulators
1279 of FGFR2 signalling and breast cancer risk. Nat Commun *4*, 2464. 10.1038/ncomms3464.
1280 89. Robertson, A.G., Kim, J., Al-Ahmadie, H., Bellmunt, J., Guo, G., Cherniack, A.D., Hinoue,
1281 T., Laird, P.W., Hoadley, K.A., Akbani, R., et al. (2017). Comprehensive Molecular
1282 Characterization of Muscle-Invasive Bladder Cancer. Cell *171*, 540-556.e525.
1283 10.1016/j.cell.2017.09.007.
1284 90. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J.,
1285 Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell *172*, 650-
1286 665. 10.1016/j.cell.2018.01.029.
1287 91. Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., and
1288 Califano, A. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory
1289 networks in a mammalian cellular context. BMC bioinformatics *7 Suppl 1*, S7.
1290 10.1186/1471-2105-7-s1-s7.
1291 92. Carro, M.S., Lim, W.K., Alvarez, M.J., Bollo, R.J., Zhao, X., Snyder, E.Y., Sulman, E.P., Anne,
1292 S.L., Doetsch, F., Colman, H., et al. (2010). The transcriptional network for mesenchymal
1293 transformation of brain tumours. Nature *463*, 318-325. 10.1038/nature08712.
1294 93. Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill,
1295 S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a Cancer
1296 Dependency Map. Cell *170*, 564-576.e516. 10.1016/j.cell.2017.06.010.
1297 94. Kuhn M, W.H. (2020). Tidymodels: a collection of packages for modeling and machine
1298 learning using tidyverse principles.
1299

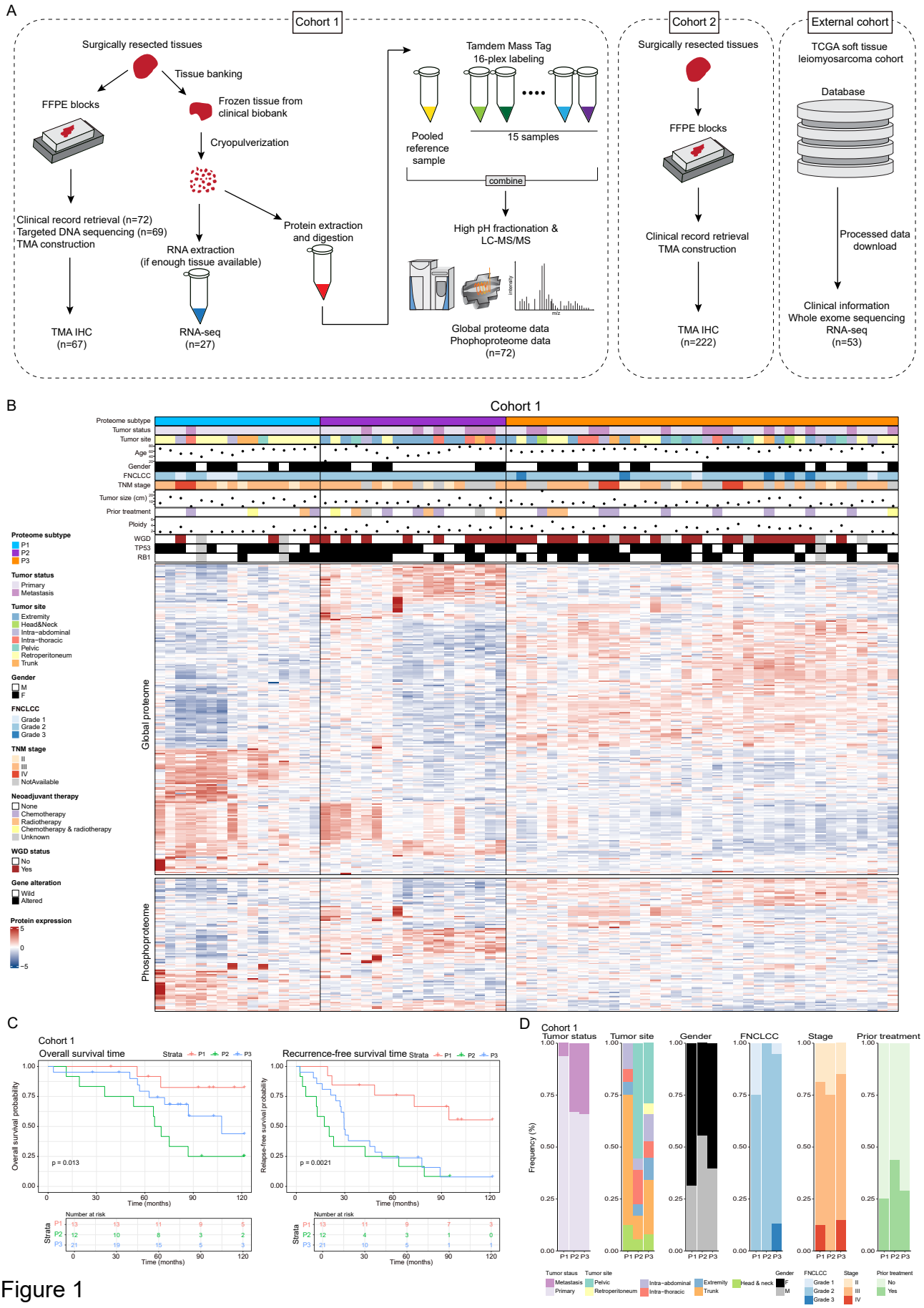

A
Cohort 1
Surgically resected tissues
Tissue banking
FFPE blocks
Frozen tissue from clinical biobank
Cryopulverization
Clinical record retrieval (n=72)
Targeted DNA sequencing (n=69)
TMA construction
RNA extraction (if enough tissue available)
Protein extraction and digestion
TMA IHC (n=67)
RNA-seq (n=27)
Tamdem Mass Tag 16-plex labeling
Pooled reference sample
15 samples
combine
High pH fractionation & LC-MS/MS
Global proteome data
Phophoproteome data
(n=72)
Cohort 2
Surgically resected tissues
FFPE blocks
Clinical record retrieval
TMA construction
TMA IHC
(n=222)
External cohort
TCGA soft tissue leiomyosarcoma cohort
Database
Processed data download
Clinical information
Whole exome sequencing
RNA-seq
(n=53)
B
Cohort 1
Proteome subtype
Tumor status
Tumor site
Age
Gender
FNCLCC
TNM stage
Tumor size (cm)
Prior treatment
Ploidy
WGD
TP53
RB1
Global proteome
Phosphoproteome
Proteome subtype
P1
P2
P3
Tumor status
Primary
Metastasis
Tumor site
Extremity
Head&Neck
Intra−abdominal
Intra−thoracic
Pelvic
Retroperitoneum
Trunk
Gender
M
F
FNCLCC
Grade 1
Grade 2
Grade 3
TNM stage
II
III
IV
NotAvailable
Neoadjuvant therapy
None
Chemotherapy
Radiotherapy
Chemotherapy & radiotherapy
Unknown
WGD status
No
Yes
Gene alteration
Wild
Altered
Protein expression
5
0
−5
C
Cohort 1
Overall survival time
Strata P1 P2 P3
Overall survival probability
p = 0.013
Time (months)
Number at risk
Recurrence-free survival time
Relapse-free survival probability
p = 0.0021
D
Cohort 1
Tumor status
Tumor site
Gender
FNCLCC
Stage
Prior treatment
Frequency (%)
P1 P2 P3
Metastasis
Primary
Pelvic
Retroperitoneum
Intra−abdominal
Intra−thoracic
Extremity
Trunk
Head & neck
F
M
Grade 1
Grade 2
Grade 3
II
III
IV
No
Yes

Figure 1

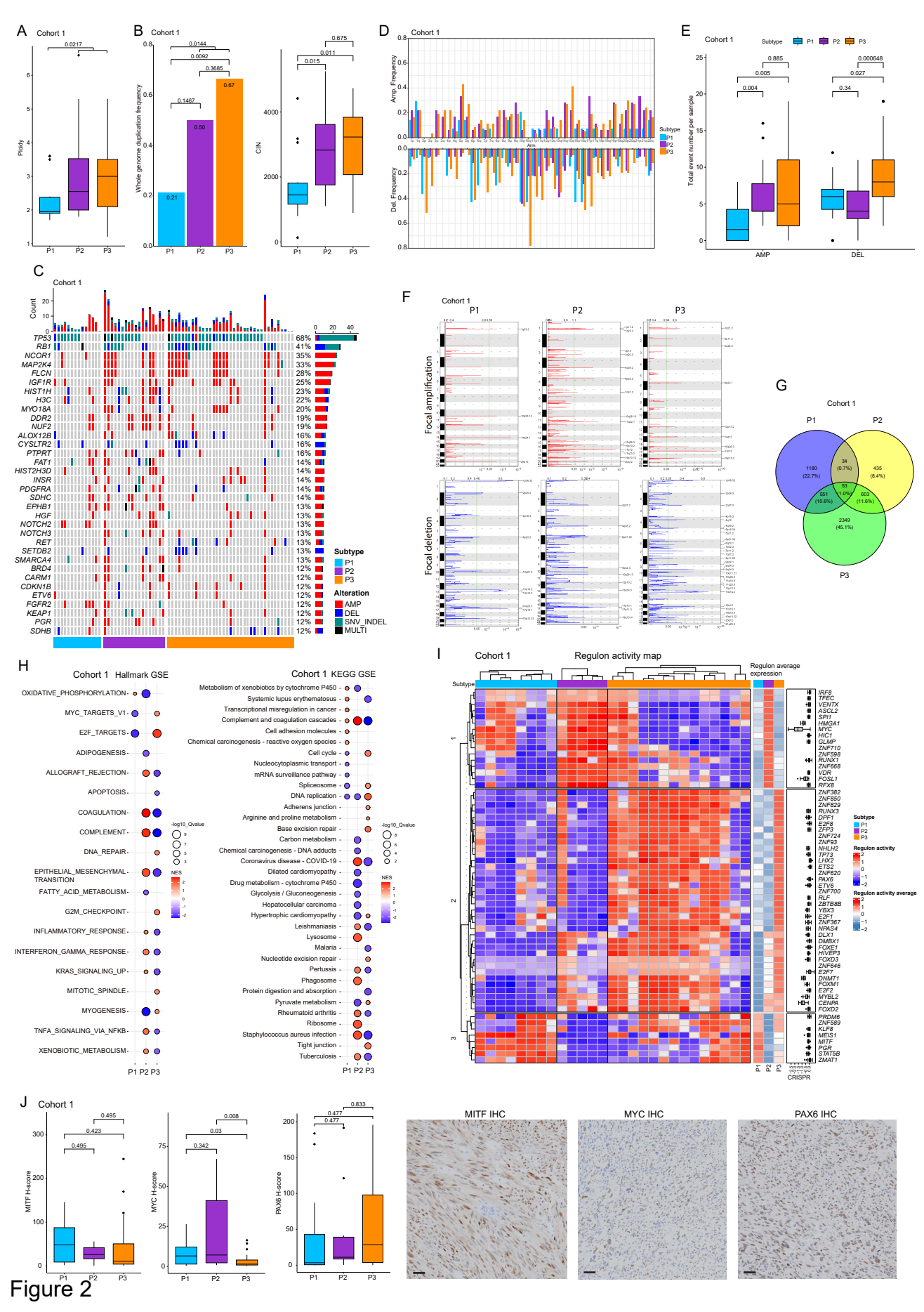

A
Cohort 1
Ploidy
0.0217
P1
P2
P3
B
Cohort 1
Whole genome duplication frequency
0.0144
0.0092
0.3685
0.1467
0.21
0.50
0.67
CIN
0.675
0.011
0.015
D
Cohort 1
Amp. Frequency
Del. Frequency
Arm
Subtype
P1
P2
P3
E
Cohort 1
Subtype
Total event number per sample
0.885
0.005
0.004
0.000648
0.027
0.34
AMP
DEL
C
Cohort 1
Count
TP53
RB1
NCOR1
MAP2K4
FLCN
IGF1R
HIST1H
H3C
MYO18A
DDR2
NUF2
ALOX12B
CYSLTR2
PTPRT
FAT1
HIST2H3D
INSR
PDGFRA
SDHC
EPHB1
HGF
NOTCH2
NOTCH3
RET
SETDB2
SMARCA4
BRD4
CARM1
CDKN1B
ETV6
FGFR2
KEAP1
PGR
SDHB
68%
41%
35%
33%
28%
25%
23%
22%
20%
19%
19%
16%
16%
16%
14%
14%
14%
14%
14%
13%
13%
13%
13%
13%
13%
12%
12%
12%
12%
12%
12%
12%
Subtype
P1
P2
P3
Alteration
AMP
DEL
SNV_INDEL
MULTI
F
Cohort 1
P1
P2
P3
Focal amplification
Focal deletion
G
Cohort 1
P1
P2
P3
1180 (22.7%)
34 (0.7%)
435 (8.4%)
551 (10.6%)
53 (1.0%)
603 (11.6%)
2349 (45.1%)
H
Cohort 1 Hallmark GSE
OXIDATIVE_PHOSPHORYLATION
MYC_TARGETS_V1
E2F_TARGETS
ADIPOGENESIS
ALLOGRAFT_REJECTION
APOPTOSIS
COAGULATION
COMPLEMENT
DNA_REPAIR
EPITHELIAL_MESENCHYMAL_TRANSITION
FATTY_ACID_METABOLISM
G2M_CHECKPOINT
INFLAMMATORY_RESPONSE
INTERFERON_GAMMA_RESPONSE
KRAS_SIGNALING_UP
MITOTIC_SPINDLE
MYOGENESIS
TNFA_SIGNALING_VIA_NFKB
XENOBIOTIC_METABOLISM
P1 P2 P3
-log10_Qvalue
NES
Cohort 1 KEGG GSE
Metabolism of xenobiotics by cytochrome P450
Systemic lupus erythematosus
Transcriptional misregulation in cancer
Complement and coagulation cascades
Cell adhesion molecules
Chemical carcinogenesis - reactive oxygen species
Cell cycle
Nucleocytoplasmic transport
mRNA surveillance pathway
Spliceosome
DNA replication
Adherens junction
Arginine and proline metabolism
Base excision repair
Carbon metabolism
Chemical carcinogenesis - DNA adducts
Coronavirus disease - COVID-19
Dilated cardiomyopathy
Drug metabolism - cytochrome P450
Glycolysis / Gluconeogenesis
Hepatocellular carcinoma
Hypertrophic cardiomyopathy
Leishmaniasis
Lysosome
Malaria
Nucleotide excision repair
Pertussis
Phagosome
Protein digestion and absorption
Pyruvate metabolism
Rheumatoid arthritis
Ribosome
Staphylococcus aureus infection
Tight junction
Tuberculosis
P1 P2 P3
I
Cohort 1
Regulon activity map
Subtype
Regulon average expression
IRF8
TFEC
VENTX
ASCL2
SPI1
HMGA1
MYC
HIC1
GLMP
ZNF710
ZNF598
RUNX1
ZNF668
VDR
FOSL1
RFX8
ZNF382
ZNF850
ZNF829
RUNX3
DPF1
E2F8
ZFP3
ZNF724
ZNF93
NHLH2
TFF3
LHX2
ETS2
NFE2L2
PAX6
ETV6
ZNF700
RLF
ZBTB8B
YBX3
E2F1
ZNF367
NPAS4
DLX1
DMRT1
FOXE1
HIVEP3
FOXD3
ZNF646
E2F7
DNMT1
FOXM1
E2F2
MYBL2
CENPA
FOXD2
PRDM6
ZNF589
KLF8
MEIS1
MITF
PGR
STAT5B
ZMAT1
Subtype
P1
P2
P3
Regulon activity
Regulon activity average
CRISPR
J
Cohort 1
MITF H-score
0.495
0.423
0.495
MYC H-score
0.008
0.03
0.342
PAX6 H-score
0.833
0.477
0.477
P1
P2
P3
MITF IHC
MYC IHC
PAX6 IHC

Figure 2

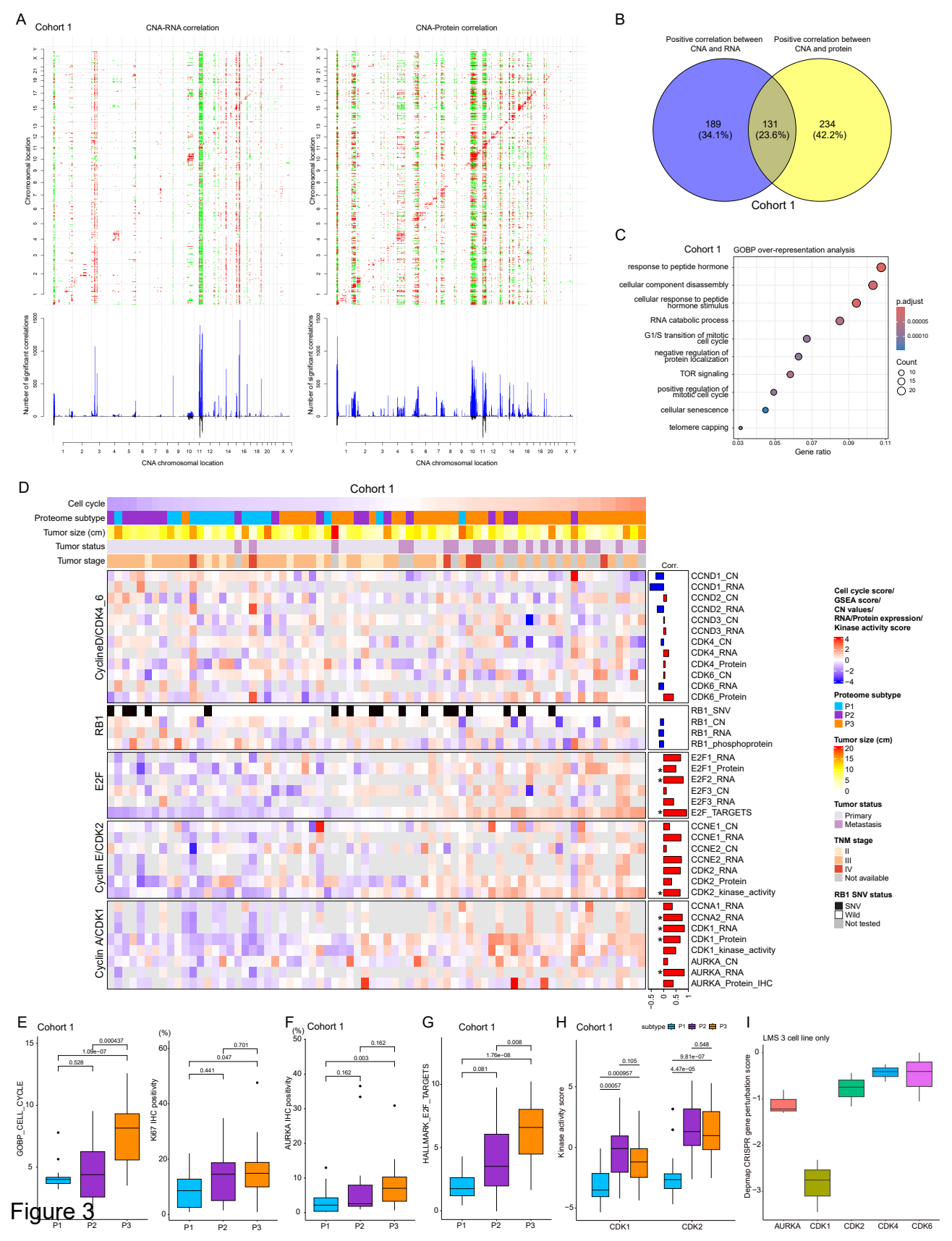

A
Cohort 1
CNA-RNA correlation
CNA-Protein correlation
Chromosomal location
Number of significant correlations
CNA chromosomal location
B
Positive correlation between CNA and RNA
Positive correlation between CNA and protein
189 (34.1%)
131 (23.6%)
234 (42.2%)
Cohort 1
C
Cohort 1
GOBP over-representation analysis
response to peptide hormone
cellular component disassembly
cellular response to peptide hormone stimulus
RNA catabolic process
G1/S transition of mitotic cell cycle
negative regulation of protein localization
TOR signaling
positive regulation of mitotic cell cycle
cellular senescence
telomere capping
Gene ratio
p.adjust
Count
D
Cohort 1
Cell cycle
Proteome subtype
Tumor size (cm)
Tumor status
Tumor stage
Corr.
CyclineD/CDK4_6
RB1
E2F
Cyclin E/CDK2
Cyclin A/CDK1
CCND1_CN
CCND1_RNA
CCND2_CN
CCND2_RNA
CCND3_CN
CCND3_RNA
CDK4_CN
CDK4_RNA
CDK4_Protein
CDK6_CN
CDK6_RNA
CDK6_Protein
RB1_SNV
RB1_CN
RB1_RNA
RB1_phosphoprotein
E2F1_RNA
E2F1_Protein
E2F2_RNA
E2F3_CN
E2F3_RNA
E2F_TARGETS
CCNE1_CN
CCNE1_RNA
CCNE2_CN
CCNE2_RNA
CDK2_RNA
CDK2_Protein
CDK2_kinase_activity
CCNA1_RNA
CCNA2_RNA
CDK1_RNA
CDK1_Protein
CDK1_kinase_activity
AURKA_CN
AURKA_RNA
AURKA_Protein_IHC
Cell cycle score/ GSEA score/ CN values/ RNA/Protein expression/ Kinase activity score
Proteome subtype
P1
P2
P3
Tumor size (cm)
Tumor status
Primary
Metastasis
TNM stage
II
III
IV
Not available
RB1 SNV status
SNV
Wild
Not tested
E
Cohort 1
GOBP_CELL_CYCLE
P1
P2
P3
F
Cohort 1
AURKA IHC positivity
(%)
G
Cohort 1
HALLMARK_E2F_TARGETS
H
Cohort 1
subtype P1 P2 P3
Kinase activity score
CDK1
CDK2
I
LMS 3 cell line only
Depmap CRISPR gene perturbation score
AURKA
CDK1
CDK2
CDK4
CDK6

Figure 3

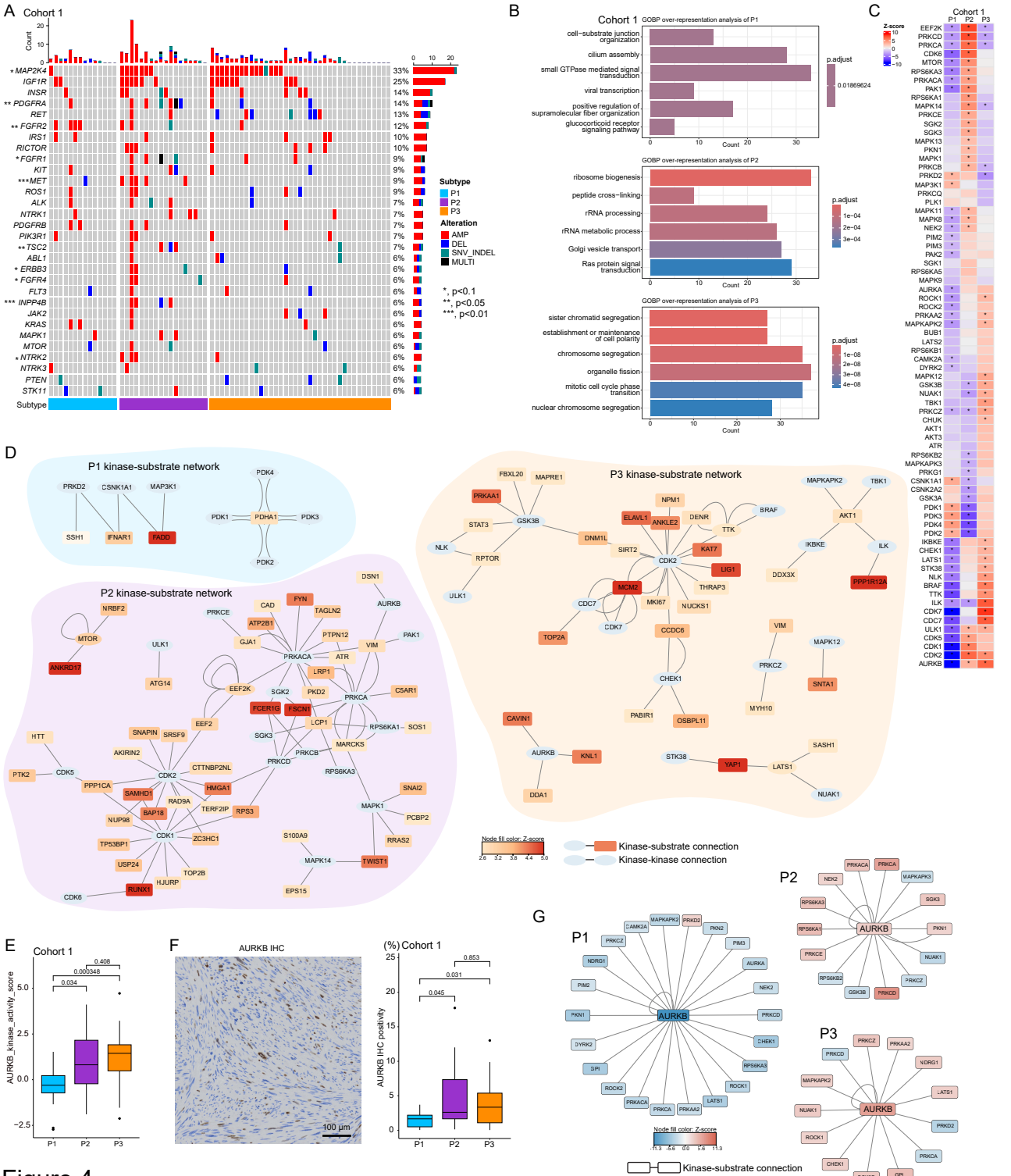

A
Cohort 1
Count
MAP2K4 33%
IGF1R 25%
INSR 14%
PDGFRA 14%
RET 13%
FGFR2 12%
IRS1 10%
RICTOR 10%
FGFR1 9%
KIT 9%
MET 9%
ROS1 9%
ALK 7%
NTRK1 7%
PDGFRB 7%
PIK3R1 7%
TSC2 7%
ABL1 6%
ERBB3 6%
FGFR4 6%
FLT3 6%
INPP4B 6%
JAK2 6%
KRAS 6%
MAPK1 6%
MTOR 6%
NTRK2 6%
NTRK3 6%
PTEN 6%
STK11 6%
Subtype
Subtype
P1
P2
P3
Alteration
AMP
DEL
SNV_INDEL
MULTI
*, p<0.1
**, p<0.05
***, p<0.01
B
Cohort 1
GOBP over-representation analysis of P1
cell−substrate junction organization
cilium assembly
small GTPase mediated signal transduction
viral transcription
positive regulation of supramolecular fiber organization
glucocorticoid receptor signaling pathway
p.adjust
0.01869624
Count
GOBP over-representation analysis of P2
ribosome biogenesis
peptide cross−linking
rRNA processing
rRNA metabolic process
Golgi vesicle transport
Ras protein signal transduction
p.adjust
1e−04
2e−04
3e−04
Count
GOBP over-representation analysis of P3
sister chromatid segregation
establishment of maintenance of cell polarity
chromosome segregation
organelle fission
mitotic cell cycle phase transition
nuclear chromosome segregation
p.adjust
1e−08
2e−08
3e−08
4e−08
Count
C
Cohort 1
P1 P2 P3
Z-score
EEF2K
PRKCD
PRKCA
CDK6
MTOR
RPS6KA3
PRKACA
PAK1
RPS6KA1
MAPK14
PRKCE
SGK2
SGK3
MAPK13
PKN1
MAPK1
PRKCB
PRKD2
MAP3K1
PRKCQ
PLK1
MAPK11
MAPK8
NEK2
PIM2
PIM3
PAK2
SGK1
RPS6KA5
MAPK9
AURKA
ROCK1
ROCK2
PRKAA2
MAPKAPK2
BUB1
LATS2
RPS6KB1
CAMKIA
DYRK2
MAPK12
GSK3B
NUAK1
TBK1
PRKCZ
CHUK
AKT1
AKT3
ATR
RPS6KB2
MAPKAPK3
PRKG1
CSNK1A1
CSNK2A2
GSK3A
PDK1
PDK3
PDK4
PDK2
IKBKE
CHEK1
LATS1
STK38
NLK
BRAF
TTK
ILK
CDK7
CDC7
ULK1
CDK5
CDK1
CDK2
AURKB
D
P1 kinase-substrate network
PRKD2
CSNK1A1
MAP3K1
SSH1
IFNAR1
FADD
PDK4
PDK1
PDHA1
PDK3
PDK2
P2 kinase-substrate network
NRBF2
MTOR
ANKRD17
ULK1
ATG14
PRKCE
CAD
FYN
TAGLN2
ATP2B1
GJA1
PTPN12
ATR
VIM
PAK1
DSN1
AURKB
PRKACA
LRP1
EEF2K
SGK2
PKD2
C5AR1
PRKCA
FCER1G
FSCN1
EEF2
LCP1
RPS6KA1
SOS1
SGK3
MARCKS
PRKCB
PRKCD
RPS6KA3
HTT
SNAPIN
SRSF9
AKIRIN2
PTK2
CDK5
CDK2
CTTNBP2NL
PPP1CA
SAMHD1
HMGA1
RAD9A
BAP18
TERF2IP
RPS3
NUP98
CDK1
ZC3HC1
TP53BP1
USP24
TOP2B
HJURP
RUNX1
CDK6
S100A9
MAPK14
TWIST1
EPS15
SNAI2
MAPK1
PCBP2
RRAS2
P3 kinase-substrate network
FBXL20
MAPRE1
PRKAA1
GSK3B
STAT3
NLK
RPTOR
ULK1
DNM1L
SIRT2
CDK2
NPM1
ELAVL1
ANKLE2
DENR
TTK
BRAF
KAT7
LIG1
THRAP3
MCM2
CDC7
MKI67
NUCKS1
CDK7
TOP2A
CCDC6
CHEK1
PABIR1
OSBPL11
CAVIN1
AURKB
KNL1
DDA1
STK38
YAP1
LATS1
SASH1
NJAK1
MAPKAPK2
TBK1
AKT1
IKBKE
ILK
DDX3X
PPP1R12A
VIM
MAPK12
PRKCZ
SNTA1
MYH10
Node fill color: Z-score
Kinase-substrate connection
Kinase-kinase connection
E
Cohort 1
AURKB_kinase_activity_score
0.408
0.000348
0.034
P1
P2
P3
F
AURKB IHC
100 μm
(%) Cohort 1
AURKB IHC positivity
0.853
0.031
0.045
P1
P2
P3
G
P1
AURKB
P2
AURKB
P3
AURKB
Node fill color: Z-score
Kinase-substrate connection

Figure 4

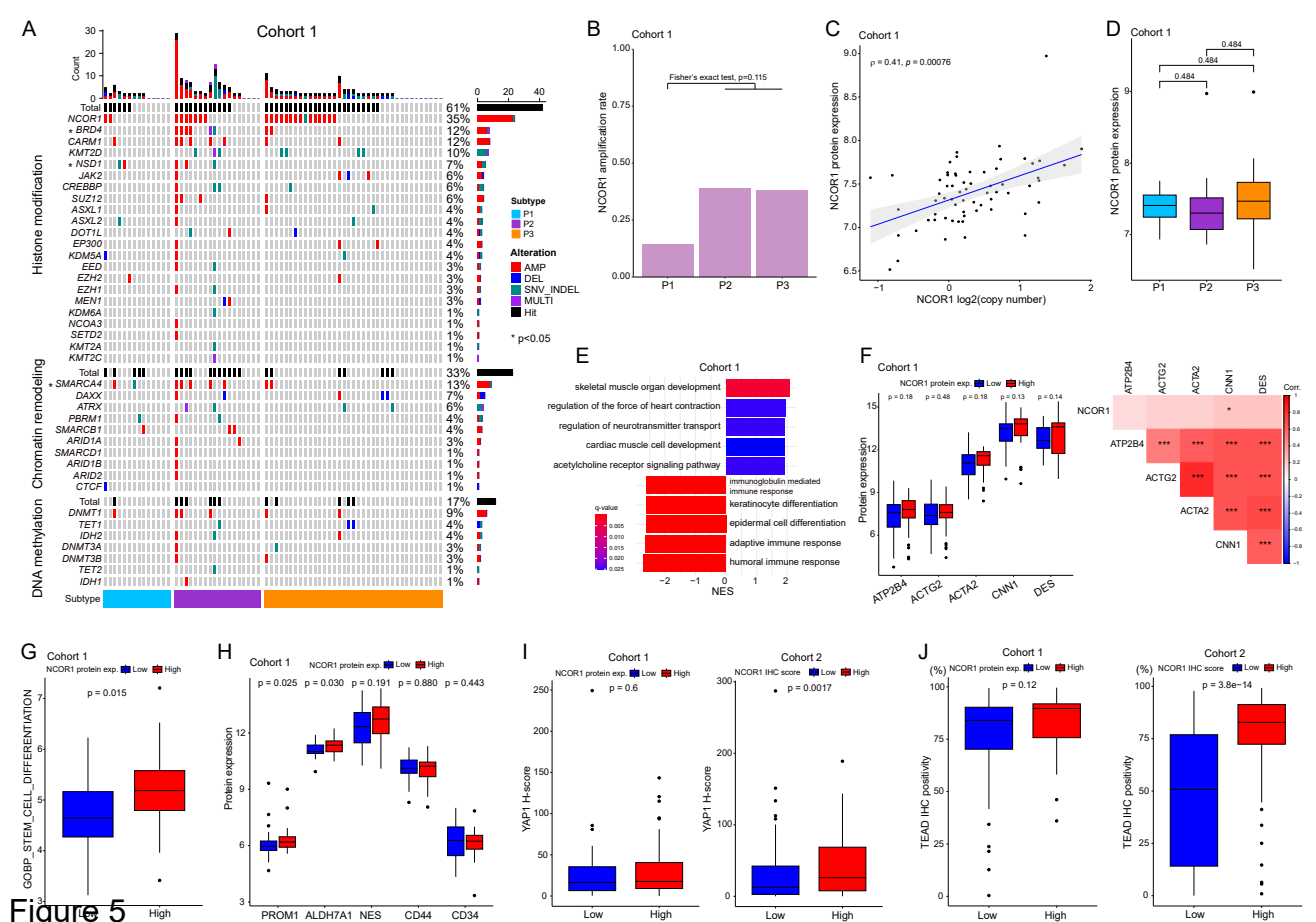

A
Cohort 1
Count
Histone modification
Chromatin remodeling
DNA methylation
Total
NCOR1
* BRD4
CARM1
KMT2D
* NSD1
JAK2
CREBBP
SUZ12
ASXL1
ASXL2
DOT1L
EP300
KDM5A
EED
EZH2
EZH1
MEN1
KDM6A
NCOA3
SETD2
KMT2A
KMT2C
* SMARCA4
DAXX
ATRX
PBRM1
SMARCB1
ARID1A
SMARCD1
ARID1B
ARID2
CTCF
DNMT1
TET1
IDH2
DNMT3A
DNMT3B
TET2
IDH1
Subtype
Alteration
AMP
DEL
SNV_INDEL
MULTI
Hit
* p<0.05
B
Cohort 1
Fisher's exact test, p=0.115
NCOR1 amplification rate
P1
P2
P3
C
Cohort 1
ρ = 0.41, p = 0.00076
NCOR1 protein expression
NCOR1 log2(copy number)
D
Cohort 1
NCOR1 protein expression
E
Cohort 1
skeletal muscle organ development
regulation of the force of heart contraction
regulation of neurotransmitter transport
cardiac muscle cell development
acetylcholine receptor signaling pathway
immunoglobulin mediated immune response
keratinocyte differentiation
epidermal cell differentiation
adaptive immune response
humoral immune response
q-value
NES
F
Cohort 1
NCOR1 protein exp. Low High
Protein expression
ATP2B4
ACTG2
ACTA2
CNN1
DES
Corr.
G
Cohort 1
NCOR1 protein exp. Low High
p = 0.015
GOBP_STEM_CELL_DIFFERENTIATION
Low
High
H
Cohort 1
NCOR1 protein exp. Low High
p = 0.025
p = 0.030
p = 0.191
p = 0.880
p = 0.443
Protein expression
PROM1
ALDH7A1
NES
CD44
CD34
I
Cohort 1
NCOR1 protein exp. Low High
p = 0.6
YAP1 H-score
Cohort 2
NCOR1 IHC score Low High
p = 0.0017
YAP1 H-score
J
Cohort 1
(%) NCOR1 protein exp. Low High
p = 0.12
TEAD IHC positivity
Cohort 2
(%) NCOR1 IHC score Low High
p = 3.8e−14
TEAD IHC positivity

Figure 5

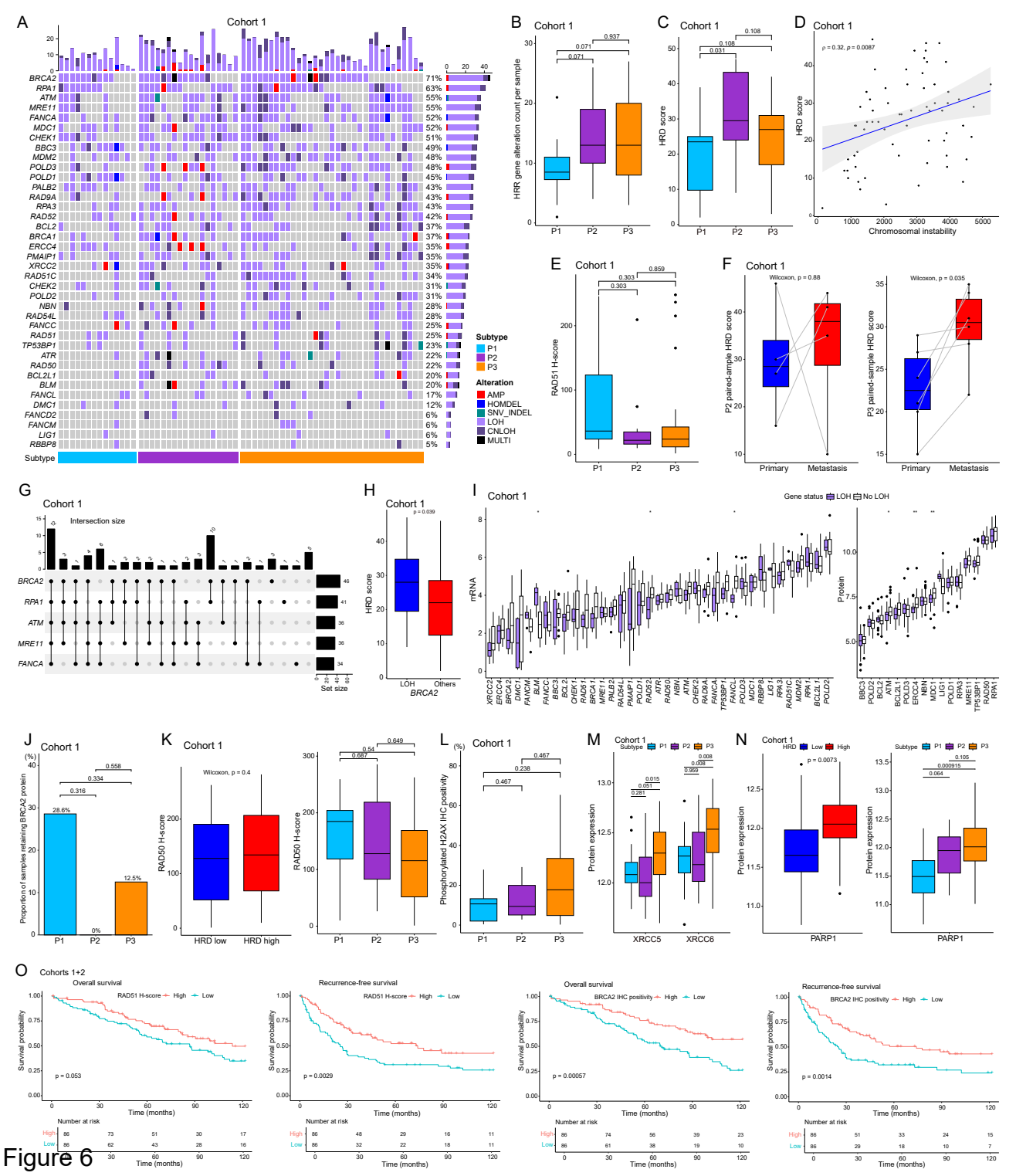

A
Cohort 1
BRCA2 71%
RPA1 63%
ATM 55%
MRE11 55%
FANCA 52%
MDC1 52%
CHEK1 51%
BBC3 49%
MDM2 48%
POLD3 48%
POLD1 45%
PALB2 43%
RAD9A 43%
RPA3 43%
RAD52 42%
BCL2 37%
BRCA1 37%
ERCC4 35%
PMAIP1 35%
XRCC2 35%
RAD51C 34%
CHEK2 31%
POLD2 31%
NBN 28%
RAD54L 28%
FANCC 25%
RAD51 25%
TP53BP1 23%
ATR 22%
RAD50 22%
BCL2L1 20%
BLM 20%
FANCL 17%
DMC1 12%
FANCD2 6%
FANCM 6%
LIG1 6%
RBBP8 5%
Subtype
Subtype
P1
P2
P3
Alteration
AMP
HOMDEL
SNV_INDEL
LOH
CNLOH
MULTI
B
Cohort 1
HRR gene alteration count per sample
0.071
0.071
0.937
P1
P2
P3
C
Cohort 1
HRD score
0.031
0.108
0.108
P1
P2
P3
D
Cohort 1
ρ = 0.32, p = 0.0087
HRD score
Chromosomal instability
E
Cohort 1
RAD51 H-score
0.303
0.303
0.859
P1
P2
P3
F
Cohort 1
Wilcoxon, p = 0.88
P2 paired-ample HRD score
Primary
Metastasis
Wilcoxon, p = 0.035
P3 paired-sample HRD score
Primary
Metastasis
G
Cohort 1
Intersection size
BRCA2
RPA1
ATM
MRE11
FANCA
46
41
36
36
34
Set size
H
Cohort 1
p = 0.039
HRD score
LOH
Others
BRCA2
I
Cohort 1
Gene status LOH No LOH
mRNA
XRCC2
ERCC4
BRCA2
DMC1
FANCM
BLM
FANCC
BBC3
BCL2
CHEK1
CHEK2
BRCA1
MRE11
PALB2
RAD54L
PMAIP1
POLD1
RAD52
ATR
RAD50
NBN
ATM
CHEK2
RAD9A
FANCA
TP53BP1
FANCL
POLD3
MDC1
RBBP8
LIG1
RPA3
RAD51C
MDM2
RPA1
BCL2L1
POLD2
Protein
BBC3
POLD2
BCL2
ATM
BCL2L1
POLD3
ERCC4
NBN
MDC1
LIG1
POLD1
RPA3
MRE11
TP53BP1
RAD50
RPA1
J
Cohort 1
(%)
Proportion of samples retaining BRCA2 protein
0.316
0.334
0.558
28.6%
0%
12.5%
P1
P2
P3
K
Cohort 1
Wilcoxon, p = 0.4
RAD50 H-score
HRD low
HRD high
0.687
0.54
0.649
RAD50 H-score
P1
P2
P3
L
Cohort 1
(%)
Phosphorylated H2AX IHC positivity
0.467
0.238
0.467
P1
P2
P3
M
Cohort 1
Subtype P1 P2 P3
Protein expression
0.281
0.015
0.051
0.959
0.008
0.008
XRCC5
XRCC6
N
Cohort 1
HRD Low High
p = 0.0073
Protein expression
PARP1
Subtype P1 P2 P3
0.064
0.000915
0.105
Protein expression
PARP1
O
Cohorts 1+2
Overall survival
RAD51 H-score High Low
p = 0.053
Survival probability
Time (months)
Number at risk
High 86 73 51 30 17
Low 86 62 43 28 16
Recurrence-free survival
RAD51 H-score High Low
p = 0.0029
Survival probability
Time (months)
Number at risk
High 86 48 29 16 11
Low 86 32 22 18 11
Overall survival
BRCA2 IHC positivity High Low
p = 0.00057
Survival probability
Time (months)
Number at risk
High 86 74 56 39 23
Low 86 61 38 19 10
Recurrence-free survival
BRCA2 IHC positivity High Low
p = 0.0014
Survival probability
Time (months)
Number at risk
High 86 51 33 24 15
Low 86 29 18 10 7

Figure 6

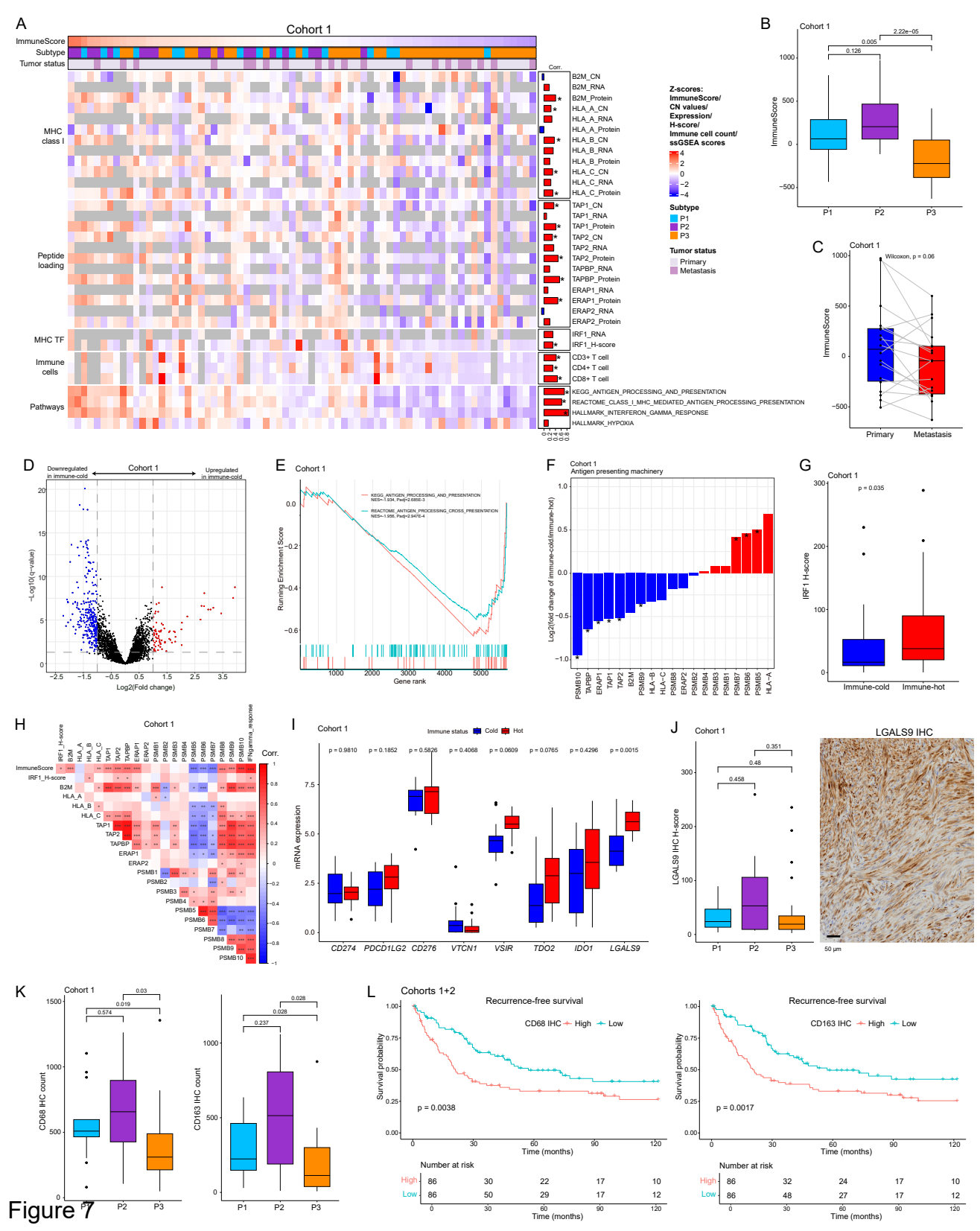
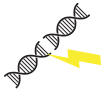

A
Cohort 1
ImmuneScore
Subtype
Tumor status
MHC class I
Peptide loading
MHC TF
Immune cells
Pathways
Corr.
B2M_CN
B2M_RNA
B2M_Protein
HLA_A_CN
HLA_A_RNA
HLA_A_Protein
HLA_B_CN
HLA_B_RNA
HLA_B_Protein
HLA_C_CN
HLA_C_RNA
HLA_C_Protein
TAP1_CN
TAP1_RNA
TAP1_Protein
TAP2_CN
TAP2_RNA
TAP2_Protein
TAPBP_RNA
TAPBP_Protein
ERAP1_RNA
ERAP1_Protein
ERAP2_RNA
ERAP2_Protein
IRF1_RNA
IRF1_H-score
CD3+ T cell
CD4+ T cell
CD8+ T cell
KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION
REACTOME_CLASS_I_MHC_MEDIATED_ANTIGEN_PROCESSING_PRESENTATION
HALLMARK_INTERFERON_GAMMA_RESPONSE
HALLMARK_HYPOXIA
Z-scores:
ImmuneScore/
CN values/
Expression/
H-score/
Immune cell count/
ssGSEA scores
Subtype
P1
P2
P3
Tumor status
Primary
Metastasis
B
Cohort 1
ImmuneScore
0.126
0.005
2.22e−05
P1
P2
P3
C
Cohort 1
Wilcoxon, p = 0.06
ImmuneScore
Primary
Metastasis
D
Cohort 1
Downregulated in immune-cold
Upregulated in immune-cold
−Log10(q-value)
Log2(Fold change)
E
Cohort 1
KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION
NES=-1.934, Padj=2.685E-3
REACTOME_ANTIGEN_PROCESSING_CROSS_PRESENTATION
NES=-1.956, Padj=2.947E-4
Running Enrichment Score
Gene rank
F
Cohort 1
Antigen presenting machinery
Log2(fold change of immune-cold/immune-hot)
PSMB10
TAPBP
ERAP1
TAP1
TAP2
B2M
PSMB9
HLA_B
HLA_C
PSMB8
ERAP2
PSMB2
PSMB4
PSMB3
PSMB1
PSMB7
PSMB6
PSMB5
HLA_A
G
Cohort 1
p = 0.035
IRF1 H-score
Immune-cold
Immune-hot
H
Cohort 1
IRF1_H-score
B2M
HLA_A
HLA_B
HLA_C
TAP1
TAP2
TAPBP
ERAP1
ERAP2
PSMB1
PSMB2
PSMB3
PSMB4
PSMB5
PSMB6
PSMB7
PSMB8
PSMB9
PSMB10
IFN.gamma_response
ImmuneScore
Corr.
I
Cohort 1
Immune status
Cold
Hot
p = 0.9810
p = 0.1852
p = 0.5826
p = 0.4068
p = 0.0609
p = 0.0765
p = 0.4296
p = 0.0015
mRNA expression
CD274
PDCD1LG2
CD276
VTCN1
VSIR
TDO2
IDO1
LGALS9
J
Cohort 1
0.351
0.48
0.458
LGALS9 IHC H-score
P1
P2
P3
LGALS9 IHC
50 μm
K
Cohort 1
0.03
0.019
0.574
CD68 IHC count
0.028
0.028
0.237
CD163 IHC count
P1
P2
P3
L
Cohorts 1+2
Recurrence-free survival
CD68 IHC
High
Low
p = 0.0038
Survival probability
Time (months)
Number at risk
High 86 30 22 17 10
Low 86 50 29 17 12
Recurrence-free survival
CD163 IHC
High
Low
p = 0.0017
Survival probability
Time (months)
Number at risk
High 86 32 24 17 10
Low 86 48 27 17 12

Figure 7

Proteome subtypes of STLMS

	P1	P2	P3	
Genomic alterations	Genomically stable *FGFR2* amp	Genomically unstable RTK drivers (*IGF1R*, *PDGFRA*) SWI/SNF (*SMARCA4*, *BRD4*) enriched	Genomically unstable Highest arm-level deletions	Molecular characteristics
Pathway alterations	Metabolic/OXPHOS programs Low proliferation	EMT, NF-κB/inflammation High proliferation	DNA repair programs Smooth-muscle differentiation High proliferation	Molecular characteristics
Kinome alterations	CSNK, PDK high AURKA/B low	CDK1/2/5/6 high AURKA/B high mTOR, MAPK1(ERK2) high	CDK1/2/7 high AURKA/B high BRAF, AKT1/3 high	Molecular characteristics
DNA repair	HRD low	HRD high *BRCA2* LOH/protein loss frequent	HRD high Shift toward NHEJ (XRCC5/6, PARP1)	Molecular characteristics
TME	Intact APM/IRF1 axis	Immunosuppresive LGALS9 high M2 macrophages enriched	Immune-cold tendency APM down regulated	Molecular characteristics
Prognostic significance	Favorable outcome	Poor outcome	Poor outcome	Clinical implications
Possible drug targets	FGFR2 PDK1-4/metabolic axis	IGF1R/PDGFRA AURKA/B, CDKs, mTOR/ERK Immuno-modulation via LGALS9	CDKs, AURKA/B, PARP	Clinical implications



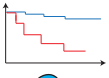
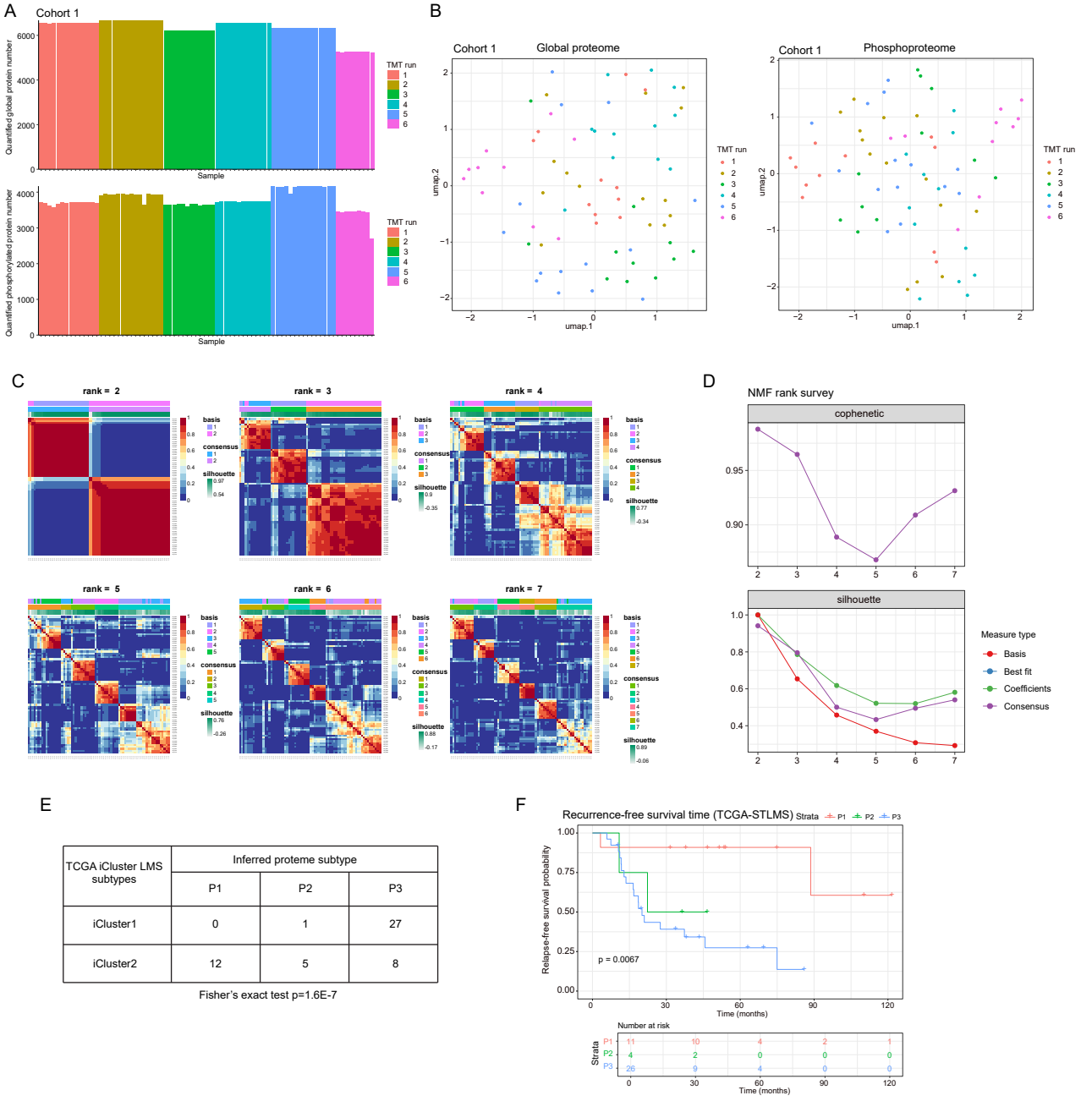


Figure 8



TCGA iCluster LMS subtypes	Inferred proteme subtype		
	P1	P2	P3
iCluster1	0	1	27
iCluster2	12	5	8

Fisher's exact test p=1.6E-7

Figure S1. Evaluation of proteome data quality, batch effects, and robust subtype inference, related to Figure 1

(A) Bar plots showing the number of quantified proteins per sample for (top) global proteome and (bottom) phosphoproteome datasets. Samples are grouped and color-coded by TMT plex MS runs (1-6). The relatively consistent number of quantified proteins across samples and batches indicates robust and comparable proteome/phosphoproteome coverage across the TMT experiments.

(B) UMAP plots of (left) global proteome and (right) phosphoproteome profiles, colored by TMT plex MS runs (n=6). Each point represents a sample. Samples are well intermixed across TMT batches, indicating no detectable batch effect in both global proteome and phosphoproteome datasets.

(C) Consensus heatmaps for non-negative matrix factorization (NMF) clustering at ranks k=2-7, showing clustering stability across different cluster numbers. Each heatmap illustrates consensus values between sample pairs, with color bars indicating sample assignments in basis and consensus groupings. Silhouette width values are shown for each rank, reflecting clustering quality.

(D) Summary plots of NMF rank survey metrics: cophenetic coefficient (top) and silhouette width (bottom) across ranks. Rank 3 shows a relatively high silhouette score and a drop in cophenetic coefficient, suggesting it as the optimal cluster number balancing stability and separation.

(E) Association between proteome-defined subtypes (P1-P3) and TCGA iCluster LMS subtypes (iCluster1 and iCluster2) in the TCGA STLMS cohort. The distribution shows significant concordance between iCluster1 and subtype P3, and iCluster2 and subtypes P1/P2 (Fisher's exact test, p=1.6×10E-7).

(F) Kaplan-Meier analysis of recurrence-free survival in the TCGA STLMS cohort, stratified by proteome-based subtypes. Stratification by proteome-based subtypes shows significant survival differences between these subtypes. The P2/P3 subtypes show poorer recurrence-free survival compared to P1.

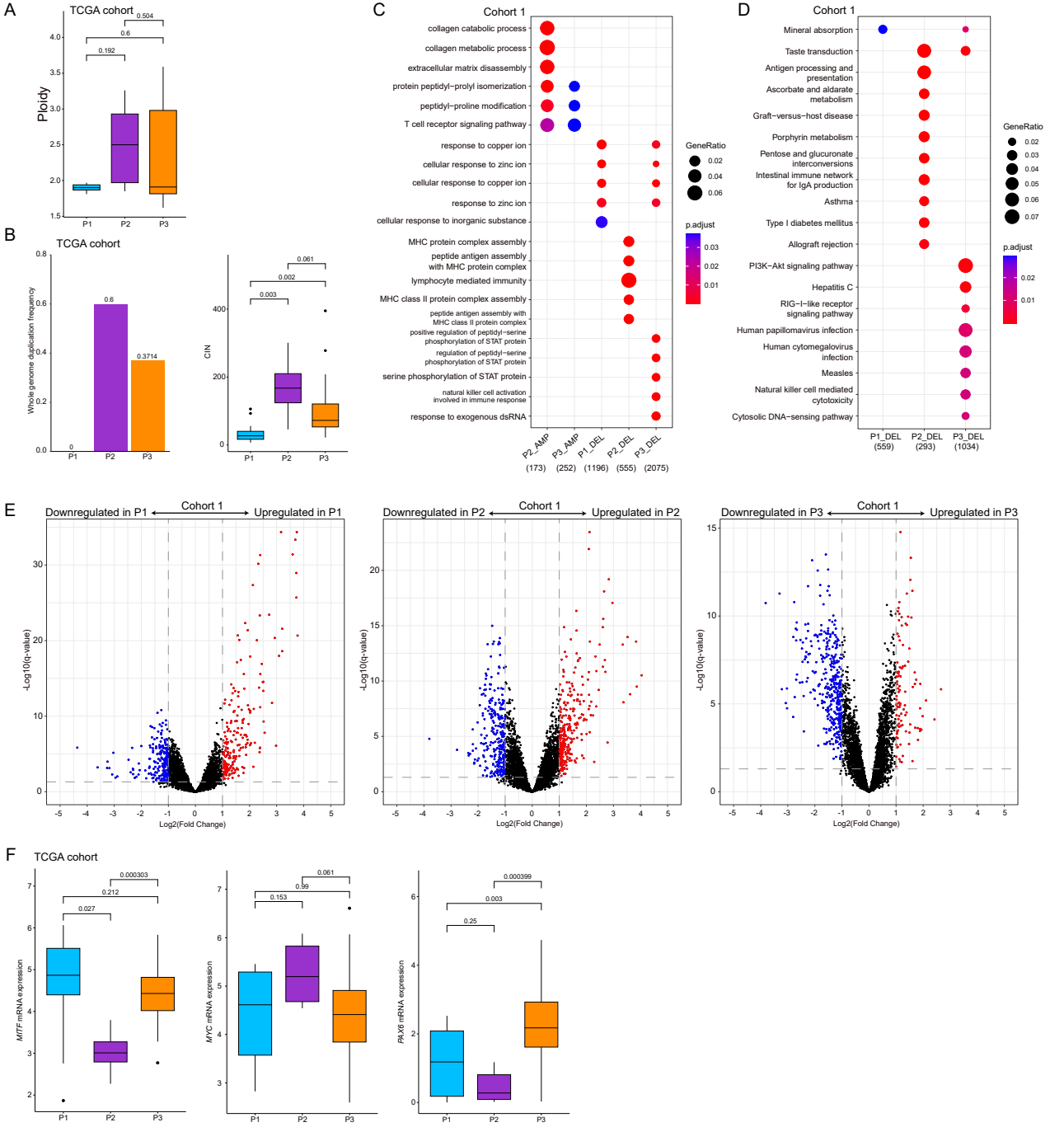


Figure S2. Genomic instability features and regulator profiles in the TCGA STLMS cohort and differential expression analyses of cohort 1, related to Figure 2

(A) Boxplot showing ploidy distributions by inferred proteome subtypes in the TCGA STLMS cohort (n=53).

(B) Bar chart of whole genome duplication frequency by inferred proteome subtypes in the TCGA STLMS cohort (n=53) on the left and boxplot of CIN in the TCGA STLMS cohort (n=52) on the right.

(C) Balloon plot showing GOBP over-representation analysis of genes involved in focal peak amplifications or deletions by subtype in cohort 1.

(D) Balloon plot showing KEGG over-representation analysis of genes involved in focal peak amplifications or deletions by subtype in cohort 1.

(E) Volcano plots showing differentially expressed proteins between subtypes in our cohort 1.

(F) mRNA expression profile of representative regulons by subtypes in the TCGA STLMS dataset.

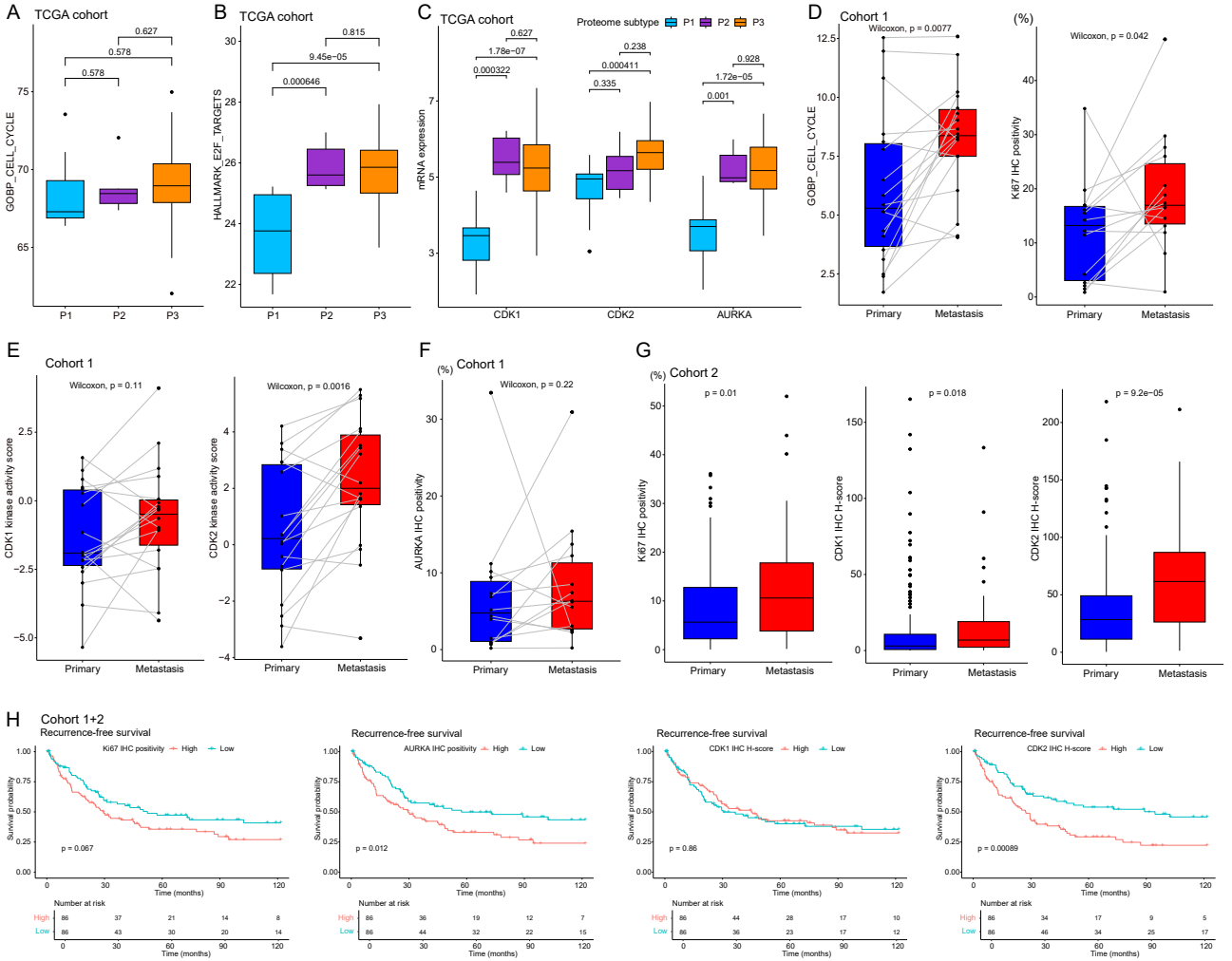


Figure S3. Cell cycle feature differences in the TCGA STLMS transcriptome data and comparison of metastatic samples in cohort 1 against matched primary samples, related to Figure 3

(A) Boxplot showing cell cycle score differences by proteome subtypes in the TCGA STLMS cohort.
(B) Boxplot of "E2F targets" scores of ssGSEA by subtypes in the TCGA STLMS cohort. P2/P3 show significantly higher scores than P1.
(C) Boxplot of mRNA expression of CDK1, CDK2, and AURKA in the TCGA cohort. P2/P3 show significantly higher expressions of CDK1 and AURKA than P1. CDK2 mRNA expression in P3 is higher than in P1.
(D) Boxplot on the left side showing cell cycle score differences between 18 paired primary and metastatic samples in cohort 1. Boxplot on the right side showing Ki67-positive tumor cell frequency based on IHC by subtypes (14 paired samples) in cohort 1.
(E) Boxplot showing kinase activity scores of CDK1/2 between primary and metastatic samples. 18 paired samples were analyzed from cohort 1.
(F) Boxplot of AURKA IHC scores. 14 paired samples were available for assessment.
(G) Boxplot of Ki67, CDK1, and CDK2 IHC scores between primary and metastasis samples in cohort 2, confirming metastatic samples show significant increases of these markers.
(H) Kaplan-Meier analysis of recurrence-free survival in combined cohorts 1 and 2, stratified by Ki67, AURKA median positivities, CDK1/2 IHC H-score median values. Patients with high AURKA positivity or CDK2-high H-scores show poorer outcomes, concordant with poor survival outcome of the P2/P3 subtypes (cell cycle-enriched subtypes).

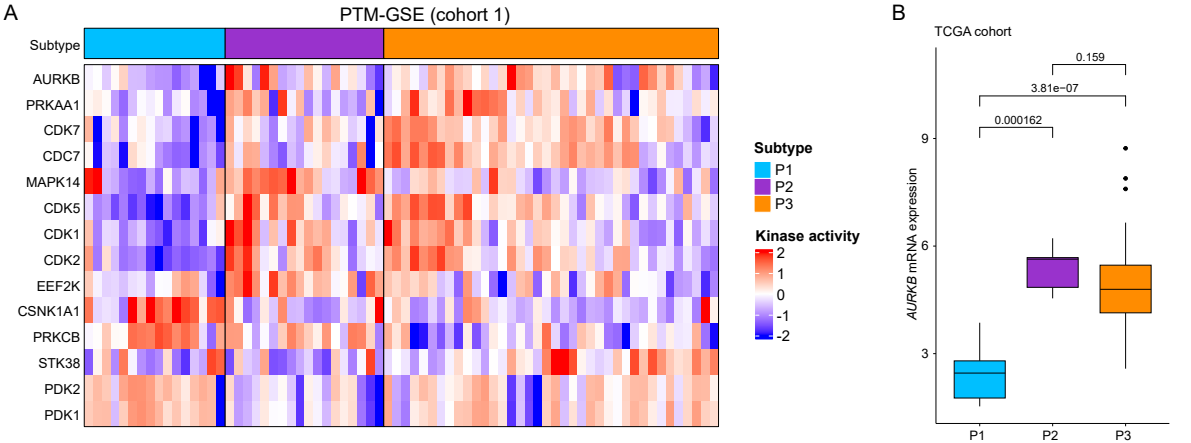


Figure S4. Phosphorylation landscape of cohort 1 and validation of findings in the independent cohort 2 and the external TCGA STLMS cohort, related to Figure 4
(A) A heatmap of PTM-GSE scores. Terms with significant differences between subtypes (ANOVA q<0.01) are shown. AURKB activity is high in P2/P3, consistent with the RoKAI results.
(B) Boxplot of AURKB mRNA expression by subtypes in the TCGA cohort. AURKB mRNA expression is significantly higher in P2/P3 compared to P1.

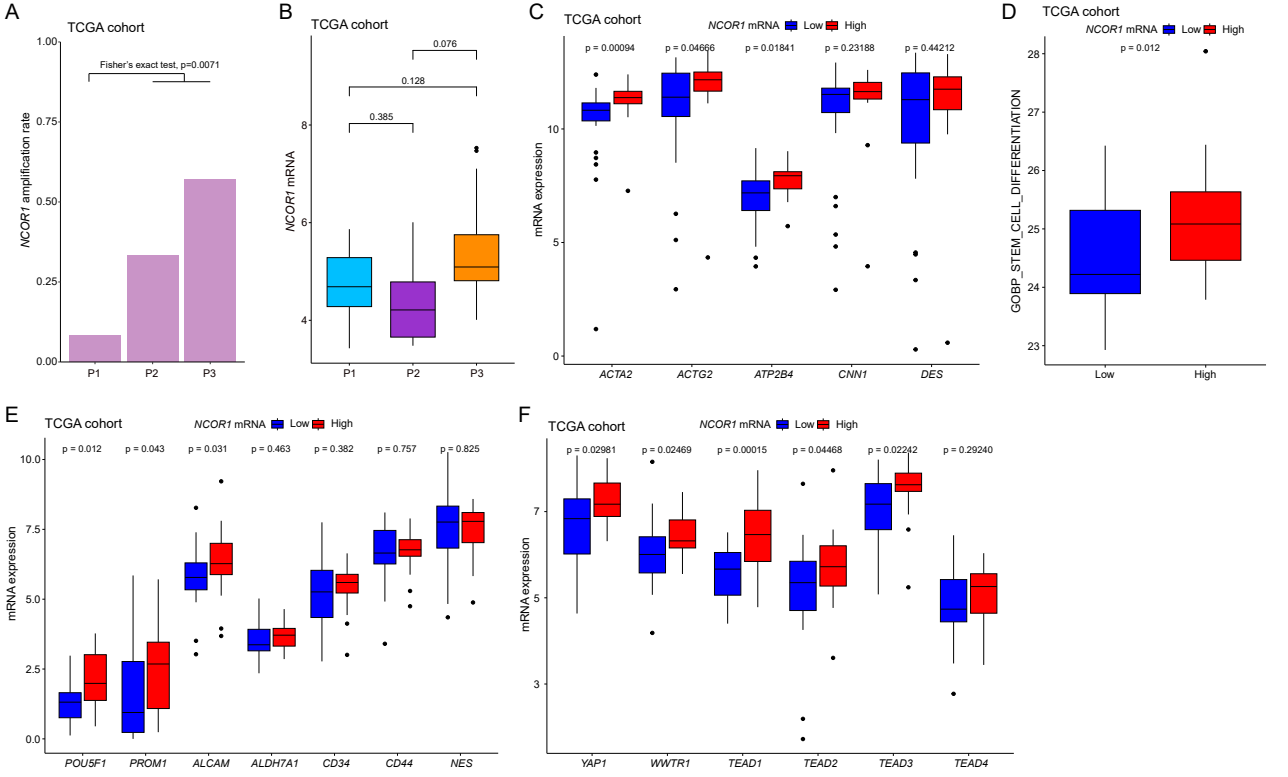


Figure S5. NCOR1 characteristics in the external TCGA STLMS dataset, related to Figure 5

(A) NCOR1 amplification shows higher frequency in P2/P3 compared to P1.

(B) NCOR1 mRNA expression in P3 trends higher than in other subtypes, but not statistically significant.

(C) mRNA expression of smooth muscle markers shows higher levels in the NCOR1-high group.

(D) Stem cell differentiation scores (GOBP ssGSEA scores) are significantly higher in the NCOR1-high group.

(E) mRNA expression of POU5F1, PROM1 (CD133), and ALCAM are significantly higher in the NCOR1-high group.

(F) Boxplots of YAP1/TAZ (encoded by WWTR1) related molecules. mRNA expression of YAP1, WWTR1, and TEAD1/2/3 is significantly higher in the NCOR1-high group.

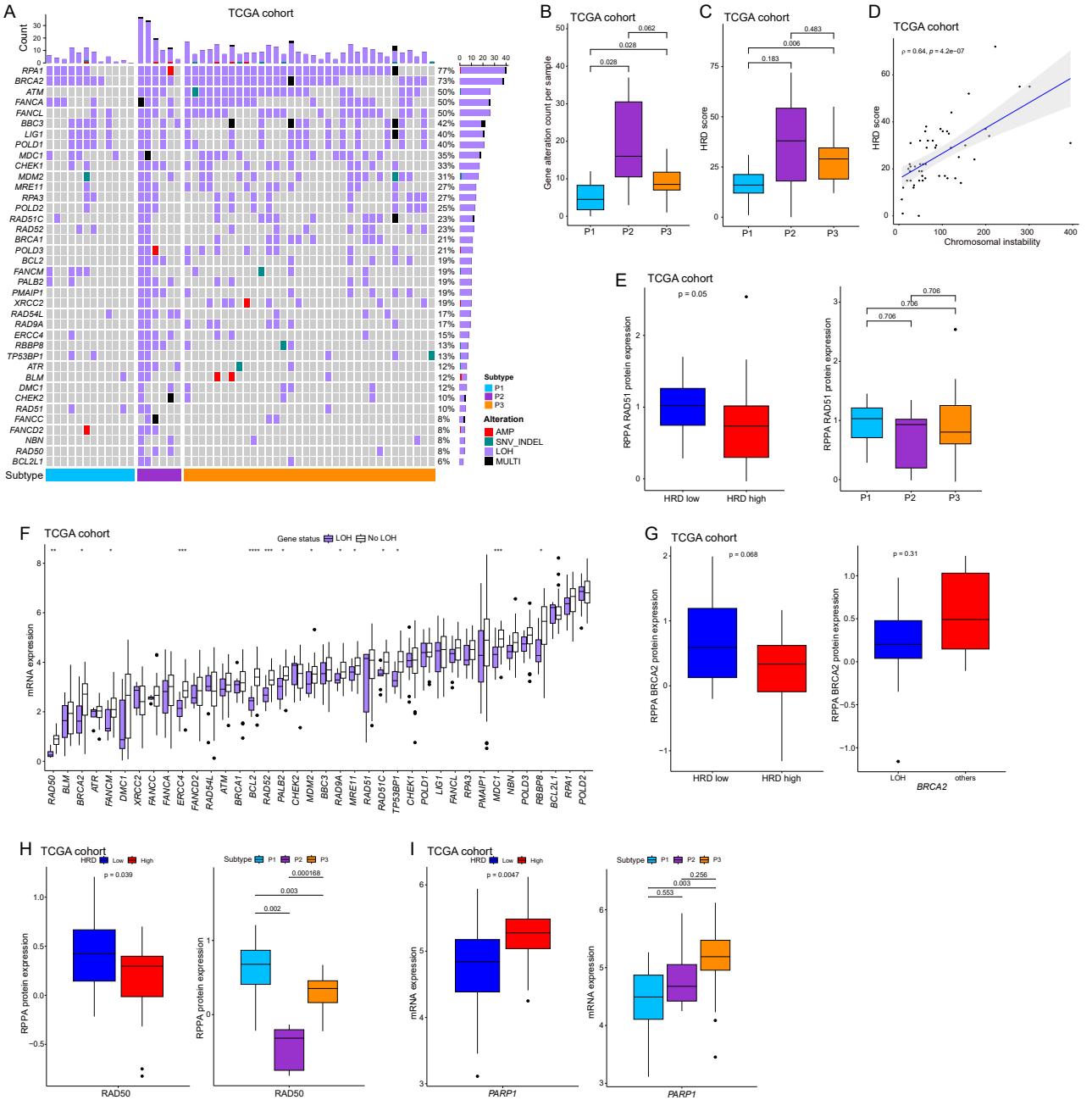


Figure S6. HRR pathway in validation cohort 2 and the external TCGA cohort, related to Figure 6

(A) Oncoprint of HRR pathway coponents showing frequent LOH events in the TCGA STLMS cohort.

(B) Boxplot of genomic alteration counts in the HRR pathway per sample showing significantly higher occurrence in P2/P3 in the TCGA cohort.

(C) The HRD score is higher in P2/P3 compared to P1 in the TCGA cohort.

(D) The HRD score positively correlates with CIN in the TCGA cohort.

(E) The RAD51 protein expression is lower in the HRD-low group and P2/P3 vs. P1, but not significant in the TCGA cohort.

(F) Boxplots of mRNA expression between LOH and non-LOH groups for each HRR gene. All 38 HRR genes are shown. Most genes, including BRCA2, show lower expression in the LOH group compared to the non-LOH group. Note: all genes with statistical significance show expressional decrease in the LOH group in the TCGA cohort.

(G) Boxplots of BRCA2 protein expression between HRD-low and HRD-high groups and BRCA2 LOH and non-LOH groups. BRCA2 protein expression shows the same trend as in cohort 1.

(H) Boxplots of RAD50 protein expression between HRD-low and HRD-high groups and by proteome subtypes. RAD50 protein expression is significantly downregulated in the HRD-low group and in P2/P3 subtypes.

(I) Boxplot of PARP1 mRNA expression between HRD-low and HRD-high groups and by proteome subtypes. PARP1 mRNA expression is significantly higher in the HRD-high group and the P3 subtype.

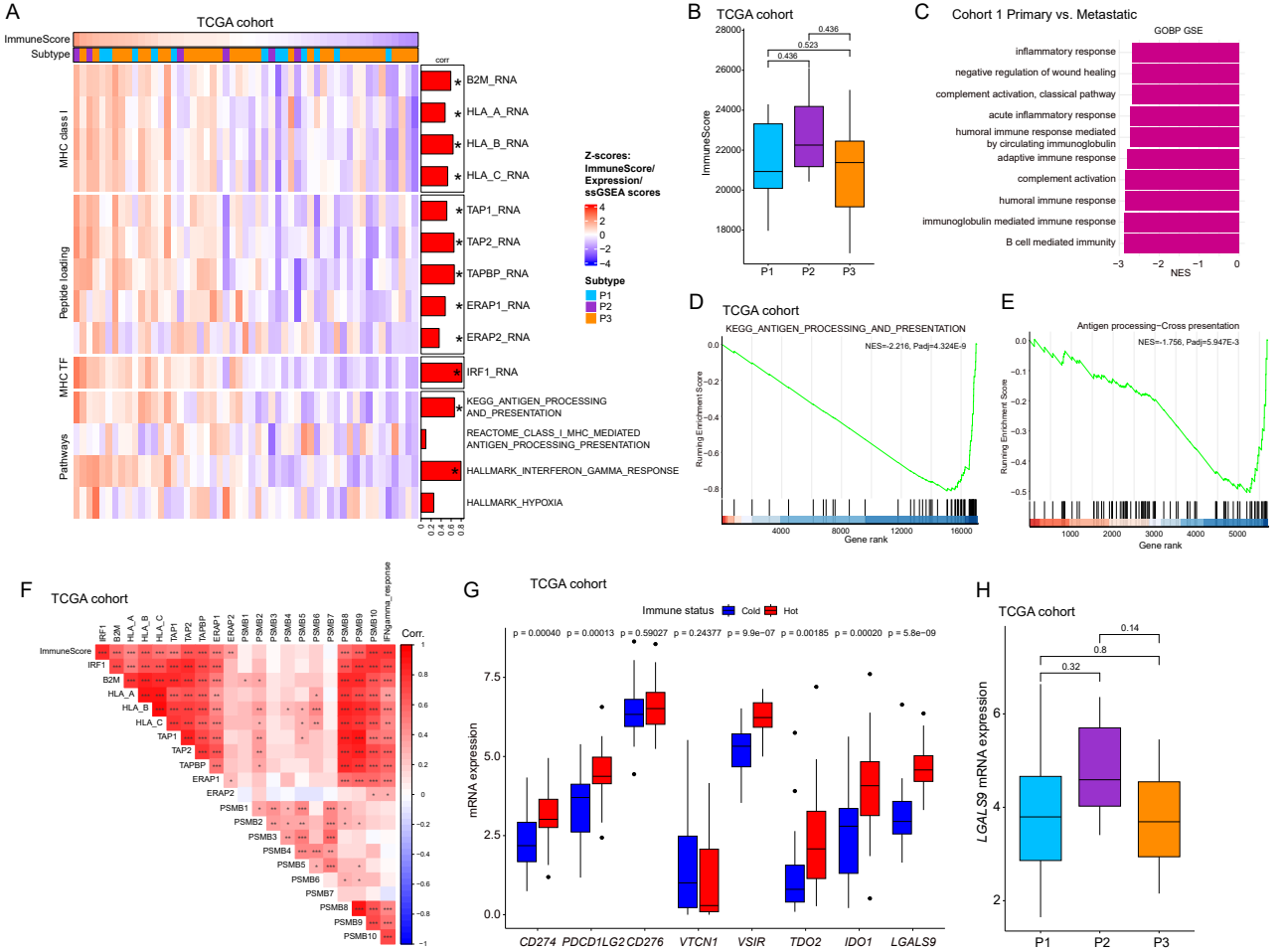


Figure S7. Tumor microenvironment analysis of the external TCGA STLMS dataset and comparison of primary vs. metastatic lesions, related to Figure 7

(A) A heatmap with immune signatures of the TCGA dataset. Immune-cold tumors show low expression of MHC class I molecules and peptide loading molecules, confirming findings in cohort 1. Bars on the right side of the heatmap denote Spearman's correlation coefficients between ImmuneScores and each gene/signature.

(B) Boxplot showing ImmuneScores by inferred proteome subtypes. Although not statistically significant, P3 shows lower scores compared to other subtypes.

(C) Top 10 significantly downregulated GOBP terms. GSEA of differential expression between primary and metastatic tumors in cohort 1 shows significant downregulation of immune-related pathways in metastatic tumors compared to primary tumors.

(D) GSE plot of KEGG pathway analyses showing significant downregulation of antigen processing pathway coponents in immune-cold tumors compared to immune-hot tumors in the TCGA dataset.

(E) GSE plot of Reactome pathway analyses showing significant downregulation of antigen processing pathway components in metastatic tumors compared to primary tumors in cohort 1.

(F) A correlation plot (Spearman's correlation coefficients) including the ImmuneScore, IRF1, and antigen processing machinery protein expression in the TCGA cohort.
*, p<0.05; **, p<0.01; ***, p<0.001. Most genes and ImmuneScore show positive correlation with each other. IFNgamma_response scores are ssGSEA scores of the "Hallmark_Interferon_Gamma_Reponse" term.

(G) mRNA expressions of immunosuppressive genes in the TCGA cohort. All genes except CD276 and VTCN1 show significantly higher expression in immune-hot tumors compared to immune-cold tumors.

(H) Boxplot of LGALS9 mRNA expression by subtypes in the TCGA cohort. P2 shows higher expression of LGALS9 compared to P1/P3, which is consistent with the findings in cohort 1.