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

Atsushi Tanaka[1,2,3,12]*, 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], Wenyi Wei[1] and Michael H. A. Roehrl[1,3,11]*

[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 leiomyosarcoma

Abstract

Soft tissue leiomyosarcoma (STLMS) is an aggressive malignancy for which robust molecular subclassification and mechanism-based therapeutic strategies still remain limited. We performed integrative proteogenomic analyses of primary and metastatic STLMS to define subtype-associated molecular programs. Joint analysis of the proteome and phosphoproteome identified 3 biologically distinct subtypes. P1 was characterized by relative genomic stability, low proliferative activity, and enrichment of FGFR2- and PDK-associated signaling. In contrast, P2 and P3 showed greater chromosomal instability and more aggressive clinical behavior, but with distinct molecular features. Notably, P2 was associated with inflammatory and RTK-RAS pathway programs, activation of CDK-AURKA/B-mTOR-ERK kinase networks, IGF1R/PDGFRA alterations, and the poorest outcomes. On the other hand, P3 showed strong cell cycle and DNA repair programs, elevated NCOR1 expression, and increased expression of nonhomologous end joining components, including PARP1. Homologous recombination deficiency analyses distinguished HRD-low P1 from HRD-high P2/P3, and paired analyses suggested increased HRD-related features in metastatic lesions within P3. Immune profiling identified an immune-hot yet potentially suppressive state in P2, marked by higher LGALS9 expression and M2-like macrophage infiltration. To support clinical translation, we developed a tissue microarray-based immunohistochemical classifier that enabled surrogate assignment of proteome-defined subtypes in an independent cohort and showed recurrence-free survival differences across inferred subtypes. These findings together establish a proteogenomic framework for STLMS heterogeneity and nominate subtype-associated biological vulnerabilities for future translational and clinical investigation.

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 largely 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, recurrence and distant metastasis remain major causes of mortality in STLMS. 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]. Hence, improved biologic stratification may help clarify disease heterogeneity and identify clinically relevant subsets of STLMS.

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]. Multi-omic 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 signaling 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 execute complex cellular and biochemical functions and are therefore important therapeutic targets in cancer[16]. Because mRNA abundance does not consistently predict protein levels[17,18], proteomics provides complementary insight into tumor biology beyond nucleic acid-based approaches. Two recent pan-soft tissue sarcoma proteomic studies by Burns et al.[19] and Tang et al.[20] included leiomyosarcoma cases, but both were based on formalin-fixed, paraffin-embedded specimens and included LMS, including uterine leiomyosarcoma, rather than a cohort restricted to soft tissue LMS. Burns et al. performed global proteomic profiling and identified three LMS proteomic subtypes, whereas Tang et al. incorporated proteomic and phosphoproteomic analyses in a pan-sarcoma setting. However, a clearly matched soft tissue LMS-focused multi-omic cohort integrating DNA sequencing, transcriptome, proteome, and phosphoproteome has remained lacking. In addition, prior studies did not systematically analyze metastatic lesions,

86 particularly paired primary-metastatic samples, which limited insight into proteogenomic changes
87 associated with metastatic progression. Importantly, the use of FFPE material is a major
88 constraint for deep proteomic profiling and an even greater limitation for phosphoproteomic
89 analysis, because phosphorylation signals are particularly vulnerable to pre-analytical and
90 fixation-related effects.

91 Here, we performed an integrative proteogenomic analysis of freshly frozen STLMS
92 specimens with matched phosphoproteomic profiling. We analyzed 72 primary and metastatic
93 STLMS samples, including 18 paired primary-metastatic lesions, together with matched DNA
94 sequencing and transcriptomic data in a subset of cases, and incorporated an independent
95 validation cohort of 219 STLMS FFPE samples as tissue microarrays. By incorporating
96 phosphoproteomic information into subtype discovery, our study was designed to capture
97 signaling pathway activity more directly, since phosphorylation is a key regulatory layer in tumor
98 progression that is not fully reflected by DNA, RNA, or total protein abundance alone.
99

100 Results

101 Overview of study cohorts

102 We analyzed two independent STLMS cohorts with complementary purposes. Cohort 1
103 comprised 72 frozen primary and metastatic STLMS samples and served as the discovery cohort
104 for integrative proteogenomic analyses, including global proteomics, phosphoproteomics,
105 targeted DNA sequencing, and transcriptomic profiling in a subset of samples. Cohort 2
106 comprised 219 formalin-fixed, paraffin-embedded STLMS cases assembled into tissue
107 microarrays and served as an orthogonal validation cohort for immunohistochemical evaluation
108 of selected markers. We also used the TCGA soft tissue leiomyosarcoma cohort as an external
109 multi-omic comparison dataset, as it provides broadly accessible molecular and clinical data.
110 Although two additional sarcoma proteomic datasets have recently been reported[19,20], limited
111 access to accompanying genomic information and detailed sarcoma subtype annotation
112 restricted their use for external comparison in this study. Unless otherwise specified, subtype
113 discovery was performed in cohort 1, immunohistochemical validation was performed in cohort
114 2, and external comparison was performed using TCGA.
115

116 Unsupervised clustering of proteome identifies 3 proteomic subtypes with clinical
117 relevance

118 We obtained 72 frozen tissues of primary and metastatic STLMS (cohort 1) for liquid
119 chromatography-mass spectrometry (LC-MS) analysis (Figure 1A). For independent cohort 2,

120 we obtained FFPE blocks of 219 STLMS cases and constructed tissue microarray blocks (Figure
121 1A). Patient demographics are summarized in Table S1. LC-MS analysis of global proteomes
122 quantified 10,249 proteins in total and 7,325 proteins in at least 30% of samples (Table S2A).
123 For phosphoproteomics, the analysis quantified 6,542 phosphorylated proteins in total and 4,584
124 in at least 30% of samples (Figure S1A, Table S2B). Batch correction removed TMT multiplexing
125 effects from the data (Figure S1B). Using non-negative matrix factorization (NMF) on filtered
126 global proteome and phosphoproteome data, we identified three distinct proteome-based
127 subtypes (P1, P2, and P3) of STLMS (Figures 1BC, S1CD). Survival analyses of patients who
128 underwent complete surgical resection showed significant differences in overall survival (OS)
129 and recurrence-free survival (RFS) among subtypes, with P2 showing the poorest survival. We
130 next performed transcriptome-based subtype inference in the TCGA STLMS dataset as an
131 orthogonal external comparison, in which inferred P2/P3-like tumors showed shorter recurrence-
132 free survival and overlapped with TCGA iCluster1 (Fig. S1EF). Analysis of clinical attributes
133 (Figure 1D) did not show significant differences in tumor size, age, tumor status (primary lesion
134 or metastatic lesion), gender, TNM stage, or prior treatment between proteomic subtypes.
135 However, metastatic samples were enriched in P2/P3 subtypes compared to P1 (p=0.0308).
136 Primary tumor site distribution varied significantly among subtypes (p=0.01735), with trunk
137 samples predominantly in P1 and pelvis samples in P2. FNCLCC Grade 1 tumors were only seen
138 in P1, while FNCLCC Grade 3 tumors were only seen in P3.
139
140 **Differences of proteogenomic landscape and transcription regulatory networks between**
141 **proteome-based subtypes**
142 We evaluated genomic alterations in 69 samples by MSK-IMPACT targeted sequencing. Notably,
143 P2/3 subtypes showed higher ploidy, whole genome duplication, and chromosomal instability
144 versus P1 (Figure 2AB, findings supported by analysis of the TCGA STLMS cohort (Figure
145 S2AB). Genomically, *TP53* (68%) and *RB1* (41%) were most frequently altered, followed by
146 *NCOR1* (35%), *MAP2K4* (33%), and *FLCN* (28%) (Figure 2C). The GENIE STLMS cohort (v15.0,
147 n=774)[21] showed similar frequencies. On the other hand, P2/P3 showed frequent genomic
148 alterations in *NCOR1*, *MAP2K4*, *FLCN*, and *IGF1R*. P2 showed commonly alterations in
149 oncogenic drivers like *IGF1R* or *PDGFRA* and SWI/SNIF complex genes, associated with poor
150 outcomes[22-25]. Somatic copy number alteration (SCNA) analysis revealed more arm-level
151 amplifications in P2/P3 than in P1, while P3 showed higher deletion frequencies (Figure 2DE).
152 Each subtype showed unique focal amplification/deletion patterns, with P1 showing fewer focal
153 peaks (Figure 2F). The 17p12 peak, containing *NCOR1* and *MAP2K4*, was shared between P2

154 and P3. *NCOR1* and *MAP2K4* alterations occurred frequently together (p=1.03E-14, Fisher's
155 exact test). We analyzed gene overlap in recurrent focal peaks, finding 1,818, 1,125, and 3,556
156 involved genes, respectively (Figure 2G). Notably, Gene ontology analyses showed P1 focal
157 deletion peaks enriched in copper/zinc ion processes, P2 focal peaks in extracellular matrix
158 processes, and P3 focal peaks in T-cell receptor signaling (Figure S2C). KEGG analysis showed
159 P2/P3 deletion peaks enriched in immune-related pathways, while only P3 showed deletions
160 involving in PI3K-Akt oncogenic signaling (Figure S2D).
161 Pairwise differential expression analyses between subtypes (P1 vs. P2/P3, P2 vs. P1/P3,
162 P3 vs. P1/P2) revealed 2,277 dysregulated proteins in P1 vs. others (207 upregulated >2-fold,
163 211 downregulated >2-fold), 2,503 in P2 vs. others (282 upregulated >2-fold, 286 downregulated
164 >2-fold), and 2,652 in P3 vs. others (106 upregulated >2-fold, 377 downregulated >2-fold) with
165 q-values <0.05 (Figure S2E) (Table S3ABC). KEGG and Hallmark analyses identified cancer-
166 related pathways in each subtype (Figure 2H). P1 showed low proliferation and enriched
167 metabolic processes. P2 showed enrichment of epithelial-mesenchymal transition (EMT), NF-κB
168 pathway, inflammatory responses, and KRAS signaling, with downregulated oxidative
169 phosphorylation. P3 exhibited cell cycle activation, DNA repair, low EMT, and higher muscle
170 differentiation. Transcriptome regulatory network analysis identified 60 dysregulated regulons
171 across proteome subtypes (Figure 2I). The regulons formed 3 clusters. Cluster 1 showed high
172 activity in the P2 subtype with poor outcome (Figure 1B), with MYC enrichment confirmed by IHC
173 (Figure 2J). Cluster 2 showed high activity in P3, with enriched cell cycle-related transcription
174 factors. DNMT1 enrichment in P3 suggests methylation differences. PAX6, a tumor progression
175 factor[26-28], was upregulated in P3 and confirmed by IHC (Figure 2J). Cluster 3 showed high
176 activity in P1 and some P3 samples. MITF, regulating immune response[29], was enriched in P1
177 (Figure 2J), which is also supported by TCGA STLMS dataset (Figures 2J, S2F).
178
179 **Impact of genomic alterations on proteomic subtypes and cell cycle-related components**
180 We studied *cis* effects of copy number alteration (CNA) on mRNA and protein. Analysis found
181 320 and 365 *cis* correlations of CNA-mRNA and CNA-protein respectively (FDR <0.1) (Figure
182 3A), with 131 genes overlapping (Figure 3B). GOBP over-representation analysis of cis regulated
183 genes that showed positive CNA-protein abundance correlations revealed significant enrichment
184 of cell cycle-related terms (Figure 3C). Most cell cycle components showed positive correlation
185 with the cell cycle score at CNA, mRNA, and (phosho)protein levels (Figure 3D). The P3 subtype
186 showed highest cell cycle activity, Ki67 immunolabeling (Figure 3E), and AURKA protein
187 abundance (Figure 3F), an emerging cancer treatment target[30]. E2F targets were enriched in

188 P2/P3 (Figure 3G), with CDK1/2 kinase activity significantly higher than in P1 (Figure 3H).
189 Independent TCGA STLMS data provided orthogonal support for these findings (Figure S3A-C).
190 DepMap CRISPR data showed that CDK1 and AURKA perturbation reduced cell viability in
191 STLMS sarcoma cell lines, therefore supporting these molecules as candidate subtype-
192 associated dependencies that warrant functional testing in LMS-specific models (Figure 3I).
193 Metastatic STLMS samples showed increased cell cycle score, proliferation rate, and CDK2
194 kinase activity relative to primary site STLMS (Figure S3DE), with a trend for elevated AURKA
195 protein expression (Figure S3F). IHC scores in the independent TMA-based cohort 2 supported
196 these findings (Figure S3G). Combined cohort analyses of recurrence-free survival showed poor
197 outcomes associated with increased expression of cell cycle-related components (Figure S3H),
198 supporting further evaluation of AURKA- and CDK-associated programs in aggressive STLMS.
199

200 **Phosphoproteomic differences across proteomic subtypes and receptor tyrosine kinase**
201 **signaling**

202 We investigated phosphoproteomic differences across STLMS subtypes focusing on receptor
203 tyrosine kinase molecules (RTK-RAS pathways). Among 69 DNA-sequenced samples, 56
204 samples (81.2%) had at least one genomic alteration in RTK-RAS pathway genes. *MAP2K4* and
205 *IGF1R* were the most frequent events. Though alteration frequencies were higher in P2/P3 vs.
206 P1, differences were not statistically significant (Figure 4A). Genomic alteration frequencies of
207 *PDGFRA* (overall frequency,14%), *FGFR2* (12%), *MET* (9%), *TSC2* (7%), *ERBB3* (6%), *FGFR4*
208 (6%), *INPP4B* (6%) differed significantly among proteomic subtypes. All these genes except
209 *FGFR2* showed enrichment in P2, while *FGFR2* was enriched in P1. Differential
210 phosphoproteomic analysis across subtypes showed enrichment of Ras protein signal
211 transduction in P2 and cell cycle-related terms in P3 (Figure 4B), consistent with the RTK–RAS
212 genomic alterations observed in P2 and the cell cycle-related features shown in Figures 2 and 3.
213 We next inferred kinase activity using phosphopeptide expression data and the RoKAI app[31] to
214 generate kinase-substrate networks (Figure 4CD). CSNK1A1 and PDK family kinases (PDK1,
215 PDK2, PDK3, PDK4) were enriched in P1. AURKA activity was downregulated in P1, consistent
216 with AURKA IHC assessment (Figure 3F). CDKs (CDK1, 2, 5, 6), AURKB, mTOR, and MAPK1
217 (ERK2) were enriched in P2. CDKs (CDK2, 7), AURKB, ROCK1, and BRAF were enriched in
218 P3. AKT1/3 also showed enrichment in P3 versus P1/P2, although this difference did not reach
219 statistically significance. Overall, P1 showed a distinct kinase activity profile from P2/P3, whereas
220 P2 and P3 shared kinase activities such as AURKB or CDK2. This difference may reflect distinct
221 genomic alteration patterns across subtypes (Figure 4A). PTM-GSEA showed similar subtype-

222 associated differences in kinase activity (Figure S4A). AURKB kinase activity scores were higher
223 in P2/P3 than P1 (Figure 4E). Consistently, AURKB IHC positivity was higher in P2/P3 than in
224 P1 (Figure 4F), and AURKB mRNA expression in the TCGA cohort showed the same pattern
225 (Figure S4B). According to PhosphoSitePlus database
226 (https://www.phosphosite.org/homeAction), T232 is a well-characterized activating
227 phosphorylation site of AURKB kinase. Motif analysis centered on AURKB-T232 identified
228 candidate upstream kinases associated with this phosphosite pattern (Figure 4G). Most
229 candidate upstream kinases were more active in P2/P3, whereas several AURKB-associated
230 kinases showed lower activity in P1.
231
232 **NCOR1 expression is associated with progenitor-like and differentiation-related features**
233 **in STLMS**
234 Epigenetic regulators influence gene expression through multiple chromatin-related
235 mechanisms, including histone modification, DNA methylation, transcription factor binding, and
236 chromatin accessibility[32]. Given the frequent NCOR1 amplification observed in our cohort (Figure
237 2C), we next examined genomic alterations in epigenetic regulator genes in cohort 1 (Figure 5A).
238 *NCOR1* showed the highest alteration frequency (35%), followed by *SMARCA4* (13%), *BRD4*
239 (12%), and *CARM1* (12%). Alteration frequencies of *BRD4*, *NSD1*, and *SMARCA4* differed
240 significantly among proteomic subtypes. *NCOR1* amplification was observed in more than 30%
241 of P2/P3 tumors, compared with 14.2% of P1 tumors (Figure 5B). *NCOR1* copy number
242 correlated with NCOR1 protein abundance, with P3 showing a trend toward the highest
243 expression. Similarly, P2/P3 tumors in the TCGA STLMS cohort showed higher frequencies of
244 *NCOR1* amplification, and *NCOR1* mRNA expression was highest in P3 (Figure S5A, B).
245 Because *NCOR1* was one of the most frequently altered epigenetic regulators and was enriched
246 in the poor-outcome P3 subtype, we further examined molecular features associated with high
247 versus low NCOR1 protein expression. Tumor samples were divided into 2 groups by median
248 NCOR1 LC-MS based protein level, followed by differential expression analysis and gene set
249 enrichment analysis (GSEA). GSEA showed enrichment of muscle organ development-related
250 terms in the NCOR1-high group, whereas immune-related and epithelial process-related terms
251 were relatively depleted (Figure 5E). Smooth muscle differentiation markers also showed higher
252 expression in the NCOR1-high group, although some differences did not reach statistical
253 significance (Figure 5F), and similar patterns were observed in the TCGA cohort (Figure S5C).
254 In addition, the NCOR1-high group showed higher stemness scores (Figures 5G and S5D),
255 together with increased expression of PROM1 (CD133) and ALDH7A1 at the protein level (Figure

256 5H). In the TCGA cohort, the NCOR1-high group also showed higher mRNA expression of
257 *POU5F1* (*OCT3/4*), *PROM1* (*CD133*), and *ALCAM* (Figure S5E). CD133 immunohistochemistry
258 in TMA cohort 2 showed a similar trend, with higher CD133 expression in the NCOR1 IHC-high
259 group than in the NCOR1 IHC-low group. YAP1/TEAD-related signals were also higher in the
260 NCOR1-high group (Figure 5I, J), and similar enrichment of YAP1/TAZ (encoded by
261 *WWTR1*)/TEAD-related features was observed in the TCGA dataset (Figure S5F). Together,
262 these findings indicate that high NCOR1 expression is associated with progenitor-like and
263 differentiation-related molecular features in STLMS; however, functional studies will be required
264 to determine whether NCOR1 directly regulates these programs.
265

266 **Enrichment of homologous recombination deficiency and nonhomologous end joining-**
267 **related features in poor-outcome subtypes**

268 We observed increased chromosomal instability (CIN) in P2/P3 (Figures 2B and S2B), together
269 with whole-genome duplication and enrichment of DNA repair-related programs in P3 (Figure
270 2H). Because homologous recombination repair (HRR) is a major pathway for the repair of DNA
271 double-strand breaks (DSBs)[33], and its impairment has been associated with genomic instability
272 and adverse clinical behavior in multiple tumor types[34-36], we next investigated HRR-related
273 alterations in STLMS. Using a set of 64 HRR pathway genes defined from the TCGA sarcoma
274 cohort[37], we identified alterations in 38 genes in cohort 1. Integration of allele-specific copy
275 number and variant data showed frequent loss of heterozygosity (LOH) affecting HRR pathway
276 genes (Figure 6A), with HRR-related events more commonly observed in P2/P3 than in P1
277 (Figure 6B). The homologous recombination deficiency (HRD) score was highest in P2 (Figure
278 6C) and positively correlated with CIN (Figure 6D). Similar patterns were observed in the TCGA
279 dataset (Figure S6A–D).

280 To further assess HRR-related features, we examined RAD51 protein abundance by IHC,
281 as RAD51 is a key recombinase in homologous recombination[38]. RAD51 staining showed a trend
282 toward lower expression in P2/P3 than in P1 (Figure 6E), and this pattern was also supported by
283 TCGA STLMS RPPA data (Figure S6E). In paired samples from the same patients, metastatic
284 lesions in P3 showed higher HRD scores than matched primary lesions (Figure 6F), consistent
285 with the association between genomic instability and metastatic progression reported in prior
286 studies[35,36].

287 We next examined genomic alteration patterns in the five most frequently altered HRR
288 genes in STLMS, namely BRCA2, RPA1, ATM, MRE11, and FANCA, in cohort 1. The most
289 frequent pattern involved concurrent alteration of all five genes (Figure 6G). Among single-gene

290 and two-gene events, BRCA2 alterations were the most common (total events = 17). Tumors
291 with BRCA2 LOH showed higher HRD scores than BRCA2-intact tumors (Figure 6H). More
292 broadly, assessment of the remaining 37 HRR genes showed that LOH-positive tumors tended
293 to have lower mRNA and protein expression of the affected genes (Figure 6I), a pattern that was
294 also supported by the TCGA STLMS mRNA dataset (Figure S6F). TCGA STLMS RPPA data
295 further showed reduced BRCA2 protein levels in HRD-high tumors and in BRCA2-LOH tumors
296 (Figure S6G). In our cohort, BRCA2 loss by IHC was frequent in P2 (Figure 6J), which was also
297 the subtype with the highest HRD score (Figure 6C).
298 We also evaluated RAD50 protein abundance by IHC because RAD50 is a core
299 component of the MRN complex involved in early DSB recognition. Although RAD50 expression
300 did not differ significantly according to HRD status alone in cohort 1, P2/P3 showed lower RAD50
301 expression than P1 (Figure 6K), and a similar pattern was observed in the TCGA dataset (Figure
302 S6H). Assessment of phosphorylated H2AX, a marker of DNA damage signaling, showed
303 increased positivity in P2/P3 relative to P1 (Figure 6L), consistent with altered DSB-associated
304 signaling across subtypes.
305 We next examined non-homologous end joining (NHEJ), an alternative pathway for DSB
306 repair. Specifically, we assessed classical NHEJ components KU70/KU80 (encoded by
307 XRCC6/XRCC5) and the alternative NHEJ-associated factor PARP1[39]. XRCC5/6 expression
308 was increased in P3 relative to P1, whereas P2 did not differ significantly from P1 (Figure 6M).
309 PARP1 expression was higher in the HRD-high group and in P3 (Figure 6N), with P2 also
310 showing a near-significant increase, consistent with increased expression of selected NHEJ-
311 related components, particularly in P3. Similar findings were observed in the TCGA dataset
312 (Figure S6I). Finally, using RAD51 and BRCA2 IHC data from cohorts 1 and 2, we found that
313 lower expression of both proteins was significantly associated with poorer patient outcome
314 (Figure 6O).
315
316 **Tumor immune microenvironment profiling identifies subtype-associated immune states**
317 **in STLMS**
318 To characterize immune microenvironment differences across STLMS, we performed integrative
319 proteogenomic analyses together with IHC-based assessment of immune cell infiltration.
320 Immune microenvironment status was evaluated using the ESTIMATE ImmuneScore[40], which
321 correlated with CD3+ T-cell counts by IHC (Figure 7A). The ImmuneScore also correlated with
322 antigen-presenting machinery (APM) components, including MHC class I pathway members and
323 IRF1 (Figure 7A). In addition, APM-related molecules showed concordance across genomic copy

number, mRNA, and protein expression (Figure 7A), suggesting that genomic alterations may contribute to immune microenvironment differences in STLMS. The P3 subtype showed a significantly lower ImmuneScore (Figure 7B), a finding that was also supported by the TCGA dataset (Figure S7A, B). Paired-sample analyses further showed that metastatic tumors had lower ImmuneScores (Figure 7C) and reduced expression of immune-related processes (Figure S7C).

Comparison of immune-cold and immune-hot tumors identified 356 differentially expressed proteins, including 75 upregulated and 281 downregulated in immune-cold tumors (Figure 7D and Table S4). GSEA showed downregulation of antigen-presentation-related pathways in immune-cold tumors (Figure 7E), which was also supported by the TCGA STLMS dataset (Figure S7D). Metastatic tumors showed similar downregulation of antigen-processing pathways (Figure S7E). ERAP1, TAP1/2, MHC class I complex molecules (HLA-B/C and B2M), and proteasome components were downregulated in immune-cold tumors (Figure 7F). IRF1, a key transcriptional regulator of MHC class I and APM molecules[41], also showed lower expression in immune-cold tumors (Figure 7G). We observed positive correlations between IRF1 and MHC class I molecules (HLA-B/C and B2M), although these correlations did not reach statistical significance. IRF1 also correlated positively with proteins involved in antigen processing, including TAP1, TAP2, TAPBP, ERAP1, PSMB8, and PSMB9, whereas the catalytic proteasome subunits PSMB5 and PSMB7 showed negative correlations (Figure 7H). Similar findings were observed in the TCGA transcriptomic dataset (Figure S7F). In addition, IRF1 expression correlated with the IFNγ-JAK-STAT pathway, consistent with its known upstream regulation[42]. By contrast, hypoxia score did not show a significant correlation with the STLMS ImmuneScore (Figure 7A).

We next examined the expression of immunoregulatory molecules in immune-hot tumors, which may attenuate antitumor immune responses. Transcriptomic analysis of cohort 1 showed that LGALS9 was upregulated in immune-hot tumors (Figure 7I). Consistently, the TCGA STLMS dataset showed increased expression of several immunoregulatory molecules, including LGALS9 (Figure S7G). Because LGALS9 was identified in both cohorts, we further evaluated its protein expression by IHC in cohorts 1 and 2. IHC showed LGALS9 expression in tumor cells, with a trend toward higher expression in the P2 subtype, which was relatively immune-hot (Figure 7J). Notably, LGALS9 protein expression was not detected by IHC in benign smooth muscle tissues, including vascular and intestinal smooth muscle. A similar pattern was observed in the TCGA STLMS cohort (Figure S7H).

To further characterize the immune microenvironment of P2, we assessed tumor-

358 associated macrophages by IHC using CD68 as a pan-macrophage marker and CD163 as an
359 M2-like macrophage marker. CD163-positive macrophages were enriched in the P2 subtype
360 (Figure 7K). In survival analyses, higher CD68-positive or CD163-positive cell counts were
361 associated with shorter recurrence-free survival (Figure 7L). Together, these findings indicate
362 that the P2 subtype is associated with an immune-hot yet potentially suppressive
363 microenvironment characterized by higher LGALS9 expression and increased M2-like
364 macrophage infiltration. However, these data do not establish that LGALS9 directly mediates
365 macrophage recruitment or immune evasion, which requires future in-depth investigation.
366

367 **Development and validation of an IHC-based classifier for proteome-defined STLMS**
368 **subtypes**

369 To enable practical subtype assignment, we developed a TMA-based IHC classifier for the three
370 proteome-defined STLMS subtypes P1, P2, and P3 (Figure 8A). The candidate pool included 23
371 IHC markers, comprising 18 markers assessed in this study and 5 additional subtype-informative
372 markers identified through subtype-specific differential expression analyses comparing each
373 subtype against the other two subtypes combined. Using cohort 1 as the training set, we
374 systematically tested all 3-, 4-, 5-, and 6-marker combinations by multinomial logistic regression
375 with 10-fold cross-validation and selected the optimal model on the basis of subtype-
376 discriminative performance. This approach identified a 6-marker classifier consisting of CD133,
377 MYC, CDK2, AURKA, AURKB, and HLA-ABC (Figure 8A). The final model used a multinomial
378 logistic regression framework with a predefined probability threshold for subtype assignment;
379 cases not meeting the threshold were designated unclassified. The 6 markers showed subtype-
380 associated expression patterns in both cohort 1 and cohort 2, supporting reproducibility across
381 independent TMA cohorts (Figure 8B). When the locked classifier was applied to cohort 2, the
382 inferred proteomic subtypes showed significant differences again in recurrence-free survival
383 (Figure 8C, log-rank $p = 0.039$). P1 was associated with the most favorable outcome, whereas
384 P2 and P3 showed less favorable recurrence-free survival.

385 To summarize the major biological and clinical features of each subtype, we generated
386 an integrative schematic of the proteome-defined STLMS subtypes (Figure 8D). This overview
387 highlights subtype-associated genomic alterations, pathway and kinome features, DNA repair
388 states, tumor microenvironment characteristics, prognostic associations, and potential
389 therapeutic vulnerabilities. Collectively, these findings support the clinical feasibility of IHC-based
390 subtype assignment and provide a concise framework for biological interpretation of the three
391 proteome-defined STLMS subtypes.

392 **Discussion**

393 In this integrative proteogenomic analysis of STLMS, we identified three proteome-defined
394 subtypes (P1, P2, and P3) with distinct molecular and clinicopathologic features (Figure 8).
395 These subtypes were associated with differences in patient outcome, genomic alterations,
396 signaling pathway activity, and tumor immune microenvironment, providing a framework for
397 understanding biological heterogeneity in STLMS.

398 Unsupervised clustering identified P2 as the subtype with the poorest outcome, showing
399 the shortest overall survival (OS) and recurrence-free survival (RFS). Importantly, survival
400 analyses were restricted to primary tumors with complete (R0) resection, supporting that the
401 adverse outcomes observed in P2/P3 are consistent with intrinsically aggressive tumor biology
402 that is evident before metastasis. In the overall cohort, both P2 and P3 were enriched for
403 metastatic lesions and higher-grade tumors, further supporting their association with clinically
404 aggressive disease biology. External comparison using the TCGA STLMS dataset provided
405 orthogonal support for this subtype framework. Although the TCGA study identified two major
406 subtypes (iClusters 1 and 2)[12], it lacked comprehensive phosphoproteomic profiling. Our
407 integrated analyses identified three proteome-defined subtypes, providing an additional layer of
408 molecular characterization of STLMS heterogeneity.

409 Genomic analyses revealed differences in genomic instability across subtypes. High
410 ploidy and whole-genome duplication in P2 and P3 were consistent with a more genomically
411 unstable phenotype associated with aggressive clinical behavior[43]. Increased chromosomal
412 instability (CIN) in these subtypes may contribute to tumor heterogeneity and therapy
413 resistance[44]. Alterations in oncogenic drivers such as IGF1R and PDGFRA were more frequent
414 in P2/P3, while recurrent alterations in chromatin-regulatory genes, including SMARCA4, BRD4,
415 NCOR1, and MAP2K4, pointed to subtype-associated differences in chromatin regulation and
416 signaling programs. These findings support biologically distinct routes of tumor progression,
417 although functional studies are needed to define their causal roles.

418 Copy number alterations correlated with mRNA and protein abundance, particularly for
419 genes related to cell-cycle regulation (Figure 3). P3 showed the strongest cell-cycle activity
420 together with elevated AURKA expression, nominating CDK- and Aurora-related programs as
421 candidate dependencies in aggressive STLMS subtypes. DepMap CRISPR-based perturbation
422 data support the biological relevance of some of these molecules[45,46], although disease-specific
423 validation in LMS models remains necessary.

424 The three subtypes also showed distinct pathway and kinome features. P1 was
425 characterized by lower proliferation and metabolic programs. P2 showed enrichment of EMT,

426 NF-κB signaling, immune-related programs, and RTK-RAS pathway features, whereas P3 was
427 characterized by cell-cycle activation and DNA repair-related programs (Figures 2, 3, 6, and 7).
428 Phosphoproteomic analyses identified subtype-associated kinase activity patterns consistent
429 with these biological states. P1 showed relative enrichment of CSNK1A1 and PDK family kinases
430 with lower AURKA/B activity. P2 was characterized by RTK-RAS/ERK-mTOR and cell-cycle-
431 related kinase signals, including CDK1/2/5/6, AURKB, mTOR, and MAPK1 (ERK2). P3 showed
432 CDK2/7, AURKB, ROCK1, and BRAF activity together with relatively increased AKT1/3 kinase
433 activity. Across subtypes, AURKB kinase activity was higher in P2/P3 than in P1 and this pattern
434 was supported by IHC (Figure 4E, F). Together, these kinome analyses nominate subtype-
435 associated signaling dependencies for future functional testing.

436 Transcriptional regulatory network analysis identified distinct transcription factor
437 programs across subtypes (Figure 2I). P2 showed higher activity of regulators including MYC,
438 consistent with its aggressive phenotype. P3 showed enrichment of cell-cycle-associated
439 regulators, including E2F family members, DNMT1, and PAX6. These observations provide a
440 framework for future studies of subtype-specific regulatory circuitry in STLMS.

441 High NCOR1 expression was associated with increased smooth muscle marker
442 expression together with higher expression of progenitor-like markers, including PROM1
443 (CD133) and ALDH7A1 (Figure 5). Increased YAP1/TAZ-TEAD-related signals in NCOR1-high
444 tumors further support the possibility that differentiation-associated and progenitor-like programs
445 coexist in a lineage-related tumor state. However, these observations are associative and do not
446 establish a direct causal role for NCOR1 in stemness maintenance or smooth muscle
447 differentiation[47].

448 Our data also showed that HRD-related features were enriched in both P2 and P3,
449 together with increased expression of selected NHEJ-related components (Figure 6)[48]. Elevated
450 HRD scores were associated with increased chromosomal instability, and loss of heterozygosity
451 in HRR genes, particularly BRCA2, was associated with reduced HRR-related expression.
452 Increased phosphorylated H2AX positivity in P2/P3 was consistent with altered DNA damage
453 signaling across subtypes (Figure 6L). PARP1 expression was also increased in HRD-high
454 tumors and in P3, supporting the possibility that DNA damage response-targeted strategies may
455 merit future evaluation in selected STLMS subsets. These implications remain hypothesis-
456 generating until validated experimentally[49,50].

457 Immune profiling showed that P3 and metastatic tumors had lower ImmuneScores
458 together with downregulation of antigen-presenting machinery components, supporting an
459 immune-cold phenotype in these contexts. By contrast, immune-hot tumors showed increased

expression of immunoregulatory molecules, including LGALS9, together with enrichment of M2-like macrophages. These findings suggest that STLMS contains biologically distinct immune states, including an immune-hot yet potentially suppressive microenvironment in P2. However, our data do not establish that LGALS9 directly drives macrophage recruitment or immune evasion.

To advance the clinical translation of the proteomic subtype framework, we developed a tissue microarray (TMA)-based immunohistochemistry (IHC) classifier utilizing a limited marker panel and applied the established model to an independent validation cohort. The inferred subtypes demonstrated significant differences in recurrence-free survival, thereby supporting the feasibility of approximating proteome-defined STLMS subtypes through a pathology-based surrogate assay. Although further refinement and external validation are necessary prior to clinical application, this approach offers a practical bridge between proteogenomic subtype discovery and clinically deployable patient stratification.

In conclusion, this study offers a comprehensive proteogenomic characterization of STLMS, integrating phosphoproteomic analysis of subtype-associated kinase programs. By delineating proteome-based subtypes with distinct genomic, signaling, DNA repair, and immune characteristics, this research serves as a valuable resource for future biological and translational investigations of STLMS. The findings highlight CDK/Aurora-related signaling, HRD-associated features, and subtype-specific immune states as potential areas for further exploration. However, additional orthogonal validation, mechanistic studies, and further refinement and external validation of practical surrogate assays are required before these findings can be applied to clinical stratification or treatment selection in routine practice.

484 ## Materials and Methods

485 ### Clinical specimens and pathological data

486 We examined tissues from two cohorts of STLMS patients. Cohort 1 consists of 72 primary and
487 metastatic STLMS tissues from 54 patients, including 18 primary-metastasis-paired samples.
488 Cohort 2 consists of 219 surgically resected STLMS samples. The study was approved by the
489 Institutional Review Board (MSKCC #16-1683, BIDMC #2023P000933). Clinical data, such as
490 patient demographics, treatment history, recurrence status, clinical targeted MSK-IMPACT
491 sequencing results, or histologic grade were retrieved from medical records in an anonymized
492 fashion. Tumor content ratios for all samples were verified by board-certified pathologists.
493 Clinicopathologic features and prognosis information, including gender, age, pathological TNM
494 stage, anatomic site, are summarized in Table S1.
495

496 ### Targeted cancer gene sequencing

497 MSK-IMPACT targeted sequencing was conducted as previously outlined[51]. In summary,
498 genomic DNA was extracted from both tumor and adjacent normal tissues using the DNeasy
499 Tissue kit and the EZ1 Advanced XL system (Qiagen). The extracted DNA was fragmented with
500 the Covaris E200 device. Custom DNA probes were crafted for targeted sequencing of all exons
501 and selected introns across 505 genes. These probes were synthesized using the NimbleGen
502 SeqCap EZ library custom oligo system and were biotinylated. Sequencing libraries were
503 prepared following the KAPA HTP protocol (Kapa Biosystems) and the Biomek FX system
504 (Beckman Coulter), and sequencing was performed on the Illumina HiSeq 2500 to achieve high,
505 uniform coverage (>500x median coverage). Demultiplexed fastq files underwent trimming with
506 Trim Galore v0.2.5mod (https://github.com/FelixKrueger/TrimGalore) to remove adapters and
507 short reads. These files were then aligned to the human reference genome (GRCh37) using
508 BWA-MEM v0.7.5a (with arguments -M -t 6)[52], and Picard Tools v2.9
509 (https://github.com/broadinstitute/picard) was used. AddOrReplaceReadGroups was applied to
510 annotate read groups, and MarkDuplicates was used to identify PCR duplicates. Genomic
511 regions were pinpointed using FindCoveredIntervals from the GATK Tool Kit v3.3-0[53]and
512 underwent indel realignment with Assembly Based ReAligner (ABRA) v2.12[54]. GATK
513 BaseRecalibrator was employed to identify systematic errors in base quality scores. Variant
514 calling was executed in paired tumor/normal mode using MuTect v1.1.4[55]for single nucleotide
515 variants (SNV). Pindel v0.2.5a7[56]was utilized for small insertions and deletions (indels). Vardict
516 v1.5.1[57]served as the variant caller to report both SNVs and indels. The vcfs generated by
517 MuTect, Vardict, and Pindel were subsequently combined. Additionally, copy-number variants,

518 including chromosomal instability (CIS) and whole-genome doubling (WGD), were identified
519 using FACETS[58]. The resulting variants were annotated using vcf2maf v1.6.14
520 (https://github.com/mskcc/vcf2maf), which employs Ensembl's Variant Effect Predictor v86. In
521 conclusion, we successfully conducted MSK-IMPACT sequencing on 69 out of 72 samples in
522 cohort 1 (LM_010P, LM_020P, LM_043P lacked sufficient DNA for sequencing). Amplifications
523 and deep deletions were identified from the GISTIC result table named
524 "allthresholded.bygenes.txt" (part of the GISTIC output used to determine the copy number status
525 of each gene in each sample). Genes with a value of +2 were classified as amplifications, while
526 those with a value of -2 were considered deep deletions. Sequencing results are detailed in Table
527 S5.
528
529 **Allele-specific copynumber analysis**
530 FACETS analyses were conducted to ascertain genome-wide allele-specific and absolute DNA
531 copy numbers for all target sequencing samples except 4 samples (FACETS version 0.5.6)[58].
532 Allele-specific copy number data were utilized to estimate tumor purity and ploidy. Prior to further
533 analyses, total copy number log ratios were adjusted for ploidy and purity. Tumors exhibiting
534 whole-genome doubling (WGD) were identified as those in which more than 50% of the
535 autosomal genome possessed a major copy number ≥2, where the major copy number is defined
536 as the number of copies of the most prevalent allele present in the sample[59]. Loss of
537 heterozygosity (LOH) was defined as the loss of one allele in cohort 1. For the TCGA STLMS
538 cohort, allele-specific copy number data were obtained via the R TCGAbiolink package[60], and
539 LOH was defined using the same criteria.
540
541 **HRD score calculation**
542 The scarHRD R package was utilized to calculate the HRD score of our cohort 1 samples[61]. This
543 metric was derived from the simple addition of three types of genetic alterations: loss of
544 heterozygosity (LOH), large-scale transitions (LST), and telomeric allelic imbalance (TAI) based
545 on allele-specific copy number data. LOH refers to the elimination of genomic segments
546 exceeding 15 Mb in size, excluding entire chromosomes. LST encompasses chromosomal
547 breakages between adjacent regions of at least 10 Mb, with breaks occurring no more than 3 Mb
548 apart. TAI quantifies the disparities in allele sequence contributions within the telomeric regions
549 of chromosomes. For the TCGA cohort, we downloaded precalculated HRD score from the prior
550 study[62].
551

552 **Chromosomal instability inference**

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

556

557 **Recurrent somatic copy number alteration detection**

558 Recurrent somatic copy number alterations (SCNAs) were identified using the Genomic
559 Identification of Significant Targets in Cancer (GISTIC, version 2.0.23)[63]to ascertain which SCNA
560 regions exhibited amplification or deletion at a frequency greater than expected by chance, with
561 a q value threshold of 0.05. The analysis employed the following parameters: Amplification
562 Threshold = 0.3, Deletion Threshold = 0.3, Cap Values = 1.5, Confidence Level = 0.99, Join
563 Segment Size = 4, Arm Level Peel Off = 1, and Sample Normalization Method = mean. Default
564 values were applied for all other parameters.

565

566 **RNA sequencing**

567 For each sample, 20-30 mg of frozen tissue was homogenized in 1 mL of TRIzol Reagent
568 (ThermoFisher catalog # 15596018), followed by phase separation induced with 200 µL of
569 chloroform. RNA was extracted from 350 µL of the aqueous phase utilizing the miRNeasy Mini
570 Kit (Qiagen catalog # 217004) on the QIAcube Connect (Qiagen) in accordance with the
571 manufacturer's protocol. Samples were eluted in 34 µL of RNase-free water. Subsequent to
572 RiboGreen quantification and quality control via the Agilent BioAnalyzer, 500 ng of total RNA with
573 RIN values ranging from 5.9 to 9.9 underwent polyA selection and TruSeq library preparation as
574 per the instructions provided by Illumina (TruSeq Stranded mRNA LT Kit, catalog # RS-122-
575 2102) with 8 cycles of PCR. Samples were barcoded and sequenced on a NovaSeq 6000 in a
576 PE100 run, employing the NovaSeq 6000 S1 Reagent Kit (200 Cycles) (Illumina). Read quality
577 was evaluated using FastQC. Trimmed raw reads were aligned to the human genome version
578 hg19 using STAR (v2.7.5a)[64]with default parameters, and gene annotation was incorporated
579 using Gencode v36, followed by the calculation of gene-level count values and TPM values via
580 the RSEM tool (version 1.3.1)[65].

581

582 **Tissue proteome extraction**

583 Aliquots of 5 mg of frozen tissue were lysed using a lysis buffer composed of 8 M urea and
584 200 mM EPPS (4-(2-Hydroxyethyl)-1-piperazinepropanesulfonic acid), at pH 8.5, supplemented
585 with protease inhibitor (complete mini EDTA-free, Roche) and phosphatase inhibitors (cocktails

586 2 and 3, Sigma). This was followed by homogenization using a Bead Ruptor. Subsequently, the
587 tissue mixture underwent sonication through 12 cycles of 1-minute sonication at 120 W power
588 (FB120, Fisher Scientific), with intermittent cooling. Post-centrifugation at 18,000 g for 10 minutes
589 at 4 °C, the supernatant containing all soluble proteins was collected. Protein concentrations
590 were quantified using BCA assays (Pierce). Pooled protein samples were prepared and stored,
591 with equal amounts of all samples combined to assess MS run quality.
592
593 **Liquid chromatography- mass spectrometry (LC-MS) proteomic analysis**
594 Protein samples were labeled utilizing the 16-plex TMT chemical labeling reagent (Thermo Fisher
595 Scientific), followed by MultiNotch MS3 LC–MS analysis employing an Orbitrap Fusion MS[66]. In
596 brief, 300 µg of proteins were reduced with 5 mM TCEP (tris(2-carboxyethyl) phosphine
597 hydrochloride), alkylated with 10 mM IAA (iodoacetamide), and quenched with 10 mM DTT
598 (dithiothreitol). The samples were subsequently diluted and precipitated using chloroform-
599 methanol[67]. The resulting pellets were resuspended in 50 µL of 200 mM EPPS buffer and
600 digested with Lys-C and trypsin at 37 °C overnight. Anhydrous acetonitrile was added to achieve
601 a final volume of 30%. TMTPro (16-plex) reagents were introduced to the peptides at a 2.8:1
602 (TMT reagent-to-peptide) ratio and incubated for 1 hour at room temperature. A label ratio check
603 was conducted to determine mixing ratios, labeling efficiency, and missed cleavages by pooling
604 1 µL from each sample. Based on these results, samples were mixed to achieve equal intensity
605 across TMT channels. The samples were dried to remove acetonitrile, desalted using C18 solid-
606 phase extraction (SPE) Sep-Pak (Waters), and vacuum-centrifuged. Five percent of the desalted
607 samples were reserved for global proteome analysis, while 95% were utilized for phosphopeptide
608 enrichment. Dried peptides intended for global proteomics were fractionated using the Pierce
609 High pH Reversed-Phase Peptide Fractionation Kit (ThermoFisher catalog # 84868) into nine
610 fractions per the modified manufacturer's protocol. The fractions were vacuum-centrifuged and
611 then reconstituted in 1% ACN/0.1% FA for LC–MS analysis. For optimal phosphopeptide
612 enrichment, phosphopeptides were collected using the High-Select TiO2 Phosphopeptide
613 Enrichment Kit (Thermo Fisher) and from the stored flow-through of Fe-NTA using the Pierce
614 High-Select Fe-NTA Phosphopeptide Enrichment Kit (Thermo Fisher), and both were combined.
615 Phosphopeptide samples were fractionated into nine fractions utilizing the Pierce High pH
616 Reversed-Phase Peptide Fractionation Kit. The fractions underwent analysis via LC-MS/MS
617 employing a nanoAQUITY UPLC (Waters) with a 50 cm (75 µm inner diameter) EASY-Spray
618 column (PepMap RSLC, C18, 2 µm, 100 Å) maintained at 60 °C, in conjunction with an Orbitrap
619 Fusion Lumos Tribrid Mass Spectrometer (Thermo Fisher). Peptide separation was conducted

620 at a flow rate of 300 nL/min using a linear gradient of 1-35% acetonitrile (0.1% FA) in water (0.1%
621 FA) over a duration of four hours, analyzed in SPS-MS3 (for global proteome) and neutral loss-
622 triggered MS3 (phosphopeptide) modes. For global proteome analysis, MS1 scans were
623 performed within a range of 375–1,500 m/z, with a resolution of 120 K, an AGC target of 4×1e5,
624 and a maximum IT of 50 ms. MS2 scans were conducted on MS1 scans of charges 2-7 using
625 0.7 m/z isolation, collision-induced dissociation at 32%, turbo scan, and a maximum IT of 50 ms.
626 MS3 scans employed specific precursor selection (SPS) of 10 isolation notches, an m/z range of
627 100–1,000, 45% CE, 50 K resolution, and an AGC target of 1e5. For phosphopeptide analysis,
628 MS1 scans were conducted within a range of 380-1,400 m/z, with a resolution of 120 K, an AGC
629 target of 4×1e5, and a maximum IT of 50 ms. MS2 scans were performed on MS1 scans of
630 charges 2-7 using 0.7 m/z isolation, collision-induced dissociation at 35%, turbo scan, and a
631 maximum IT of 50 ms. Neutral loss-triggered MS3 scans utilized a range of 100–1,000 m/z, 38%
632 CE, 60 K resolution, an AGC target of 1e5, and a maximum IT of 118 ms. Targeted loss masses
633 of 97.9763 and 79.9658 were employed.
634
635 **TMT data analysis**
636 Raw data files were processed utilizing Proteome Discoverer (PD) version 2.4.1.15 (Thermo
637 Scientific). In each of the TMT experiments, raw files from all fractions were consolidated and
638 analyzed using the SEQUEST HT search engine with a Homo sapiens UniProt protein database,
639 downloaded on 2019/12/13 (92,249 entries). Oxidation (M), phosphorylation (STY), and
640 deamidation (NQ) were designated as variable modifications, whereas cysteine
641 carbamidomethylation, TMT 16plex (K), and TMT 16plex (N-terminus) were specified as fixed
642 modifications. The precursor and fragment mass tolerances were set at 10 ppm and 0.6 Da,
643 respectively. A maximum of two missed trypsin cleavages were allowed. Searches employed a
644 reversed sequence decoy strategy to control the peptide false discovery rate (FDR), with a 1%
645 FDR established as the identification threshold. Phosphosite localization was determined using
646 the PD IMP-ptmRS node. The raw abundance of protein, phosphoprotein, and phosphopeptide
647 levels was normalized by sample loading normalization of each TMT channel after filtering out
648 protein/peptide assignments with valid values for less than 30% of samples and K-Nearest
649 Neighbors (KNN) imputation. Internal reference scaling based on pooled samples assigned for
650 each TMT plex was conducted across the entire cohort to eliminate batch effects, followed by
651 TMM normalization for median centralization between TMT batches[68,69]. Differential expression
652 analyses were performed using edgeR[70].
653

654 **Unsupervised clustering of proteomic data**

655 We used non-negative matrix factorization (NMF)-based clustering of global protein and
656 phosphoprotein expression data. In brief, we constructed a merged expression matrix of
657 normalized global protein and phosphoprotein data and selected the top 500 (phospho)proteins
658 with highest mean absolute deviations for unsupervised clustering to avoid noisy
659 (phospho)proteins in the data. The resulting matrix was then subjected to NMF analysis
660 leveraging the NMF R package[71]. To determine the optimal factorization rank *k* (number of
661 clusters) for the matrix, we tested a range of clusters between *k* = 2-7. For each *k*, we factorized
662 matrix *V* using 1,000 iterations with random initializations of W and H. To determine the optimal
663 factorization rank, we calculated two metrics for each k: (1) cophenetic correlation coefficient
664 (measuring how well the intrinsic structure of the data was recapitulated after clustering) and (2)
665 silhouette score (indicating how well each sample lies within its cluster compared to other
666 clusters[72]) (Figure S1CD). These metrics indicate the reproducibility of the clustering across
667 1,000 iterations. The cophenetic correlation coefficients for *k* = 2 and 3 were >0.95 and showed
668 a large drop at *k* = 4. The silhouette score at *k* = 2, 3 showed relatively higher scores than that
669 at *k* = 4. Based on these results and census heatmaps for each *k*, we chose *k* = 3 as the optimal
670 clustering number with sufficient samples in each class for downstream analysis.

671

672 **Gene set enrichment analysis (GSEA)**

673 Fold change values for gene expression between specified groups were calculated using edgeR.
674 We employed clusterProfiler (v2.1.2)[73]with a ranked gene list ordered by fold change value.
675 Enrichment tests were conducted for GOBP, Hallmark, KEGG, and Reactome gene sets
676 obtained from the Molecular Signatures Database (MSigDB v7.4). For the functional
677 characterization of each sample via single sample GSEA, we calculated normalized enrichment
678 scores (NES) of cancer-relevant gene sets by projecting the matrix of normalized expression
679 data onto GOBP, Hallmark, KEGG, and Reactome pathway gene sets. This was done using the
680 ssGSEA implementation available at https://github.com/broadinstitute/ssGSEA2.0 with the
681 following parameters: sample.norm.type = “log”, weight = 0.75, statistic = “area.under.RES”,
682 output.score.type = “NES”, nperm = 1000, min.overlap = 3, correl.type = “z.score”. To elucidate
683 gene set enrichment in each sample, we applied single sample GSEA on normalized expression
684 data[74]and calculated the normalized enrichment score (NES) for MSigDB terms.

685

686 **Over-representation analysis (ORA)**

687 ORA was performed to identify biological pathways and functional categories enriched in the

688 gene set of interest by using clusterProfiler (v2.1.2)[73]. Enrichment was calculated using a
689 hypergeometric test against the defined background, and p-values were adjusted for multiple
690 comparisons using the Benjamini-Hochberg method.
691
692 **Posttranslational modification signature enrichment analysis (PTM-SEA)**
693 To infer kinase activity based on phophosite-specific quantification data, we performed PTM-
694 SEA (https://github.com/broadinstitute/ssGSEA2.0), an adaptation of ssGSEA, that analyzes
695 site-specific signatures by scoring PTMsigDB bi-directional signature sets. PTMsigDB, sourced
696 from over 2,500 publications, contains modification site-specific signatures linked to
697 perturbations, kinase activities, and signaling pathways. Unlike other pathway databases,
698 PTMsigDB annotates each PTM site with its directional change upon specific perturbations or
699 signaling events, improving PTM-SEA's scoring. PTMsigDB's main source is the
700 PhosphoSitePlus database[75]. The PTM-SEA requires two main inputs: a site-centric data matrix,
701 denoted as m and formatted in GCT v1.3, along with the PTM signatures database. In matrix m,
702 each row corresponds to a distinct phosphorylation site that is precisely localized to a particular
703 amino acid residue, with the columns indicating the measured abundances across different
704 samples. When multiple phosphorylation sites are identified on the same peptide, they are
705 transformed into individual site-specific entities for each site.
706
707 **Kinase enrichment analysis between subtypes**
708 To assess enrichment of activated kinases and phosphatases between groups based on
709 functional information such as known protein-protein interactions or kinase-substrate
710 annotations, we input phosphopeptide abundance change values (log2(fold change)) between
711 proteome subtypes into the RoKAI application (v0.8.3)[31] and obtained kinase activity scores at
712 proteome subtype level. Default parameters were used.
713
714 **Transcriptional regulatory network analysis**
715 To deduce the transcriptional regulatory activity of STLMS based on proteome subtype, we
716 utilized the RTN R package[76-78], which examines the link between specific transcription factors
717 (TFs) or gene regulators and all possible targets, as well as the enrichment status of a network
718 using proteome data. In summary, we employed normalized transcriptome data (n=27) and
719 Lambert's human TF list[79](1,612 TFs) to initially build a potential transcriptional regulatory
720 network (TRN). We then eliminated non-significant associations through permutation analysis
721 (permutation n=1000) and bootstrapping, followed by the ARACNe algorithm[80] to preserve direct

722 TF-target interactions, thereby removing redundant indirect interactions. The refined TRN
723 dataset includes the regulator gene name, its target gene name, and adjusted p-values as a
724 confidence measure. For enrichment analyses over a list of regulons in the refined TRN, we first
725 conducted a Master Regulator Analysis (MRA)[81]over a list of regulons, with adjustments for
726 multiple hypothesis testing. The MRA evaluates the overlap between each regulon and the
727 significantly dysregulated genes between subtypes, providing adjusted p-values, which resulted
728 in 600 confident regulons (adjusted p-values <0.05) for further analyses. We then determined
729 these regulons' enrichment scores between proteome subtypes using transcriptome
730 expressional fold change values and the GSEA-2T (two-tailed GSEA analysis) function from the
731 RTN R package. To identify similarities and differences in regulatory programs among samples
732 in our cohort, we selected 60 statistically robust regulons (adjusted p-values <1E-8), calculated
733 the activity score of each regulon in each sample, and visualized the data as a heatmap.
734 Additionally, we incorporated CRISPR perturbation scores (sourced from the Cancer
735 Dependency Map (DepMap) Project[82]) of 36 sarcoma cell lines and included this data in the
736 heatmap.
737
738 **Transcriptome-based inference of proteome-defined subtypes in the TCGA STLMS cohort**
739 To perform an external comparison of the proteome-defined subtypes identified in cohort 1, we
740 inferred subtype labels in the TCGA STLMS transcriptome dataset using a classifier trained on
741 the subset of cohort 1 samples with matched transcriptome data and known proteome subtype
742 assignments. Specifically, among cohort 1 samples, 27 cases had both transcriptome data and
743 proteome-defined subtype labels (P1, P2, and P3) derived from proteomic and
744 phosphoproteomic analyses. These 27 samples were used as the reference set for classifier
745 construction. To select informative features, we first performed differential expression analyses
746 across the proteome-defined subtypes within this 27-sample transcriptome subset and selected
747 genes that were differentially expressed between subtypes at FDR < 0.01 and absolute fold
748 change > 2. This yielded 668 genes for model construction. Normalized expression data for these
749 genes were then divided into training (n = 19) and test (n = 8) sets, and a random forest classifier
750 was trained using the tidymodels R package[83] with 10-fold cross-validation. The classifier
751 achieved an F-measure of 1.0 in the held-out test set. We then applied the trained model to the
752 TCGA STLMS transcriptome dataset (n = 53) to assign transcriptome-inferred labels
753 corresponding to the proteome-defined subtypes. TCGA sample IDs and inferred subtype
754 assignments are listed in Table S6.
755

756 **Assessment of immune status**

757 ImmuneScores were derived from the normalized protein expression matrix of 72 samples
758 utilizing the ESTIMATE (Estimation of STromal and Immune cells in MAlignant Tumor tissues
759 using Expression data) package in R[40]. To evaluate immune cell infiltration, we conducted IHC
760 with immune cell markers (CD3, CD4, CD8, CD68, CD163) and tallied the number of positive
761 cells for each marker across 3 TMA cores per sample. Subsequently, we computed the average
762 number of positive cells per core for each sample for further analysis.

763

764 **Tissue microarray (TMA) construction**

765 TMAs were created from FFPE blocks of 63 STLMS tissues in cohort 1 and 219 STLMS tissues
766 in cohort 2. Three 1-mm cores were extracted from each tissue paraffin block and placed into
767 tissue array blocks using a TMA arrayer (3DHistech). Tumor and normal regions were chosen
768 after a thorough examination of individual histologic slides and electronic image-based selection
769 of target areas for coring.

770

771 **Immunohistochemistry (IHC)**

772 Four-μm sections were cut from FFPE tissue blocks. Paraffin was removed with xylene, and
773 antigens were retrieval by heat-mediated epitope retrieval. Tissue sections were stained using a
774 Leica BOND-MAX IHC autostainer. The antibodies used were anti-CD3 (Leica Microsystems,
775 NCL-L-CD3-565, 1/200 dilution), anti-CD4 (Sigma, 104R-15, 1/100 dilution), anti-CD8 (Dako,
776 M7103, 1/1500 dilution), anti-CD68 (Dako, M0876, 1/200 dilution), anti-CD163 (Leica
777 Microsystems, NCL-CD163, 1/250 dilution), anti-MITF (Abcam, ab303530, 1/100 dilution), anti-
778 MYC (Abcam, ab32072, 1/200 dilution), anti-PAX6 (Abcam, ab195045, 1/100 dilution), anti-
779 AURKA (Cell Signaling Technology, #91590, 1/100 dilution), anti-AURKB (Abcam, ab45145,
780 1/100 dilution), anti-CDK1 (Abcam, ab133327, 1/250 dilution), anti-CDK2 (Cell Signaling
781 Technology, #18048, 1/200 dilution), anti-active YAP1 (Abcam, ab205270, 1/20,000, dilution),
782 anti-TEAD (Abcam, ab97460, 1/400 dilution), anti-NCOR1 (Invitrogen, PA1-844A, 1/1000
783 dilution), anti-RAD51 (Abcam, ab133534, 1/100 dilution), anti-BRCA2 (Sigma, HPA026815, 1/50
784 dilution), anti-RAD50 (Cell Signaling Technology, #61915, 1/4000 dilution), anti-phosphorylated
785 H2AX (Cell Signaling Technology, #9718, 1/200 dilution), anti-Ki67 (Biocare, PRM325, 1/100
786 dilution), anti-IRF1 (Abcam, ab243895, 1/100 dilution), anti-LGALS9 (Abcam, ab227046, 1/200
787 dilution). For MYC, MITF, CDK1, CDK2, NCOR1, YAP1, RAD50, RAD51, PAX6, IRF1, and
788 LGALS9 IHC. The staining intensity of each tumor cell was rated on a scale from 0 to 3+, and the
789 average was taken from three separate tissue cores for each case. To determine the total

790 weighted IHC score (IHC H-score, theoretically ranging from 0 to 300) for a sample, the
791 expression intensity of individual tumor regions (rated 0-3+) was multiplied by their respective
792 contributions (0-100%) to the overall tumor area, and these values were then summed. Two
793 pathologists independently evaluated all tissue samples. If there were any discrepancies, the
794 cases were re-evaluated until a consensus score was achieved. For Ki67, AURKA/B, TEAD,
795 BRCA2, and phosphorylated H2AX IHC, we assessed the fraction of positive tumor cells. BRCA2
796 loss was defined as less than 50% positivity.
797
798 **Development of an IHC-based classifier for LMS subtype prediction**
799 To develop an IHC-based classifier for LMS subtype assignment, we used the discovery cohort
800 with matched proteomic subtype labels as the training set. A total of 23 candidate IHC markers
801 were evaluated, including 18 markers assessed in this study and 5 additional subtype-informative
802 markers identified through differential expression analyses comparing each subtype with the other
803 two subtypes combined. Only cases with complete IHC data were included in model development.
804 All possible combinations of 3, 4, 5, and 6 markers were systematically evaluated using
805 multinomial logistic regression models to predict the three LMS subtypes P1, P2, and P3. Model
806 performance was assessed by 10-fold cross-validation, and candidate marker sets were ranked
807 according to subtype-discriminative performance. To improve assignment confidence, a rejection
808 rule was applied such that a subtype was assigned only when the highest predicted probability
809 exceeded a predefined subtype-specific threshold; otherwise, the case was labeled unclassified.
810 Subtype-specific probability thresholds and a margin threshold were optimized within the training
811 cohort. Final model selection was based on 10-fold cross-validation metrics, including macro-F1
812 score, classification coverage, accuracy among classified cases, and the number of unclassified
813 cases. The final selected model included 6 markers, CD133, MYC, CDK2, AURKA, AURKB, and
814 HLA-ABC, with subtype-specific probability thresholds of 0.73 for P1, 0.65 for P2, and 0.57 for
815 P3. In the training cohort, this model classified 43 of 62 cases, corresponding to 69.4% coverage,
816 with 79.1% accuracy and a macro-F1 score of 0.613 among classified cases. The locked 6-marker
817 classifier was then applied to the independent validation cohort for downstream analyses. The
818 final classifier model, preprocessing files, and prediction script are available in a GitHub repository
819 (https://github.com/atanakapth/lms-ihc-classifier).
820
821 **Survival time analysis**
822 Prognostic associations of genes and proteins of interest with overall survival (OS) and
823 recurrence-free survival (RFS) were assessed using Kaplan–Meier survival analysis. For each

target gene or protein, patients were classified into groups based on the predefined expression cutoff or genomic status. The analysis was restricted to patients with primary lesions who underwent R0 surgical resection. Differences in OS and RFS between groups were compared using the log-rank test. Owing to the limited sample size, TNM stage-specific and tumor site-specific analyses were not performed. Statistical significance was defined as a two-sided $p < 0.05$.

External datasets

The TCGA STLMS dataset (53 samples) including clinical information, transcriptome data (count data, TPM), and gene alterations with allele-specific copy number data (data for 1 sample was not available) was obtained via the TCGAbiolink R package (version 2.30.0)[60]. Protein expression data from reverse phase protein array (RPPA) experiments of the TCGA cohort was downloaded from MD Anderson Cancer Center database (https://tcpa.drbioright.org/rppa500/main.html).

Data and code availability

Our mass spectrometry proteomics data have been deposited in Harvard Dataverse and is available under controlled access (DOI: 10.7910/DVN/BWOCM9). Raw and processed data of MSK-IMPACT genomic sequencing and transcriptome data were deposited to the EGA repository with the identifiers EGAS50000000594 and EGAS50000000595, respectively. Other data derived from this study will be shared upon reasonable request. We did not generate original code in this study. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Abbreviations (there are a lot more missing)

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 and WW wrote and edited the manuscript. MR 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.

884 **Supplemental information**

885 Document S1. Figures S1-S7

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

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

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

889 Table S4. Differential protein expression analysis between immune-cold and immune-hot

890 samples, related to Figure 7

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

892 Table S6. Clinicopathological attributes and inferred proteome subtypes of the TCGA STLMS

893 cohort, related to Figures S1-S7

894 **Figure titles and legends**

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

896 (A) Summary of the workflow and metrics of this study.

897 (B) Summary plot of cohort 1 including clinicopathological information and proteome subtypes.
898 In the heatmap, proteins and phosphoproteins that show significant differences (FDR <0.05)
899 between proteome subtypes are shown.

900 (C) Kaplan-Meier curves of overall survival and recurrence-free survival between proteome
901 subtypes.

902 (D) Differences of clinicopathological factors between proteome subtypes.

903

904 **Figure 2. Landscape of genomic alterations and regulatory activity across proteome**
905 **subtypes.**

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

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

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

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

917 (E) Concordant with arm-level event frequencies for each chromosome, total event numbers of
918 amplifications are significantly higher in P2/P3 than in P1. For arm-level deletions, P3 shows the
919 highest event numbers compared with P1 or P2.

920 (F) Focal amplification/deletion patterns in each subtype are shown. P1 has a relatively lower
921 number of focal peak events. The 17p12 peak (NCOR1 and MAP2K4 gene loci) is shared
922 between the P2 and P3 subtypes.

923 (G) Venn diagram showing overlap of genes in in the recurrent focal peaks of each subtype.

924 (H) Balloon plot showing pathway enrichment analyses of Hallmark and KEGG terms between
925 subtypes. Only statistically significant results are shown.

926 (I) Regulatory activity profile of STLMS by subtype status based on transcriptome data.
927 Computed regulon activity scores are shown as a heatmap. Regulons with p<1E-8 from the

928 Master Regulator Analysis process are shown. The boxplot on the right side of the heatmap
929 shows DepMap CRISPER gene effect scores for 36 sarcoma cell lines.
930 (J) Boxplots of MYC, PAX6, MITF IHC scores by subtypes are shown. In addition, representative
931 IHC images are shown. Each transcription factor shows an expression profile of subtypes that is
932 concordant with regulon network analyses of transcriptome data. Bars, 50 μm.
933
934 **Figure 3. Impact of genomic alterations on the proteome.**
935 (A) The correlations between CNA and mRNA (left) and protein expression (right) levels (y axes)
936 reveal both cis and trans effects. The presence of significant positive (red) and negative (green)
937 Spearman's correlations (FDR <0.1) between CNA and mRNA or protein is highlighted. The
938 numbers of significant cis and trans events corresponding to the indicated genomic loci specific
939 to mRNAs or proteins are represented by blue bars, while the overlap between CNA-mRNA and
940 CNA-protein events is depicted by black bars.
941 (B) Venn diagram showing cis correlation overlap between CNA-mRNA and CNA-protein.
942 (C) GOBP over-representation analysis of cis correlation between CNA-protein is shown. Cell
943 cycle-related terms are significantly enriched.
944 (D) Genomic effect of cell cycle-related genes on mRNA and protein. Columns represent each
945 sample and are ordered by cell cycle score. The bar chart on the right side of heatmap shows
946 Spearman's correlation coefficients between targets and cell cycle score. "E2F targets" is based
947 on ssGSEA of the Hallmark E2F term. Kinase activity scores (CDK1,2 kinase activities) are from
948 PTM-GSE analysis. p_RB1 denotes phosphorylated protein of RB1. * denotes statistical
949 significance (p<0.05).
950 (E) Boxplots showing proteome subtype differences of GOBP cell cycle scores and Ki67-positive
951 tumor frequency.
952 (F) Boxplot showing AURKA protein expression (IHC score) by subtypes.
953 (G) Boxplot of E2F scores from ssGSEA by subtypes.
954 (H) Boxplot showing kinase activity scores of CDK1/2 by subtypes.
955 (I) Boxplot of Depmap CRISPR gene perturbation scores targeting CDKs and AURKA using 3
956 leiomyosarcoma cell lines. Score below -1 represents significant effect on cell survival when the
957 target gene is inhibited.
958
959 **Figure 4. Phosphoproteomic differences between subtypes.**
960 (A) Oncoprint showing RTK-RAS pathways with 4 or more events. Amplification of MAP2K4 and
961 IFG1R are the top 2 most frequent alterations. * denotes statistically significant differences

962 (Fisher's exact test) in gene alteration frequency between subtypes.
963 (B) Over-representation analyses of GOBP terms. Phosphoproteins which are positively
964 upregulated in one subtype compared to other subtypes with FDR<0.05 are used for GOBP term
965 enrichment analyses.
966 (C) Kinase activity heatmap by subtypes. Z-scores of inferred kinase activities are plotted as a
967 heatmap. * denotes statistical significance. Kinase names are shown on the left side of the
968 heatmap.
969 (D) Kinase-substrate networks of the top 20 positively enriched kinases in each subtype are
970 shown. Note: all positively enriched kinases of P1 are shown due to its small number. Genes
971 shown as ellipses are kinases. Genes shown as rectangles are kinase substrates Colors of
972 kinase substrates denote the phosphorylation Z-score of the target's phosphorylation site. If
973 multiple phosphorylation sites exist in a substrate, phosphorylation averages are shown based
974 on color scale.
975 (E) Boxplot of PTM-GSEA AURKB activity score shows enrichment in P2/P3 compared to P1.
976 (F) Representative AURKB IHC image and boxplot of AURKB IHC score by subtype. Bar,100
977 μm.
978 (G) Networks of AURKB and inferred regulatory kinases are shown. Motif analysis of AURKB-
979 T232 identifies 194 kinases. Among them, kinases with significant enrichment (FDR<0.05) in
980 RoKAI app analysis between subtypes were selected to be shown here. Colors of kinases denote
981 Z-scores of kinase activity.
982
983 **Figure 5. NCOR1 expression corelates with muscular differentiation, stemness, and**
984 **immune status.**
985 (A) Oncoprint of epigenetic regulators in cohort 1. NCOR1 gene is the top altered epigenetic
986 regulator. * denotes statistical difference of gene alteration frequency between subtypes.
987 (B) NCOR1 amplification shows higher frequency in P2/P3 compared to P1 in cohort 1 (but not
988 statistically significant).
989 (C) NCOR1 copy number positively correlates with its protein level (Spearman's correlation
990 coefficient) in cohort 1.
991 (D) Boxplot of NCOR1 protein expression by subtypes in cohort 1.
992 (E) Top 5 positively and negatively enriched GOBP GSEA results are shown. A ranked gene list
993 based on fold change values between NCOR1-low and -high groups in cohort 1 was used for
994 GSEA.
995 (F) Boxplot of smooth muscle protein markers between NCOR1-low and NCOR1-high.

996 (G) Boxplot of GOBP_STEM_CELL_DIFFERENTIATION ssGSEA scores. In cohort 1, the
997 NCOR1-high group shows significantly higher scores compared to the NCOR1-low group.
998 (H) Boxplot of stemness markers between NCOR1-low/high groups. PROM1 (CD133) and
999 ALDH7A1 in the NCOR1-high group show significantly higher protein expression than the
1000 NCOR1-low group in cohort 1.
1001 (I) Boxplots of YAP1 IHC H-scores between NCOR1-low and -high groups in cohorts 1 and 2.
1002 (J) Boxplots of TEAD IHC H-scores between NCOR1-low and -high groups in cohorts 1 and 2.
1003
1004 **Figure 6. Homologous recombination repair pathway.**
1005 (A) Oncoprint of HRR pathway components showing frequent LOH events in cohort 1.
1006 (B) Boxplot of genomic alteration counts in the HRR pathway per sample shows high occurrence
1007 in P2/P3, but not statistically significant.
1008 (C) The HRD score is higher in P2/P3 compared to P1.
1009 (D) The HRD score positively correlates with CIN.
1010 (E) The RAD51 IHC scores are lower in P2/P3 compared to P1, which is concordant with the
1011 HRD scores of these subtypes.
1012 (F) Boxplot of HRD score comparison between paired primary and metastatic samples of P2 and
1013 P3. Metastatic samples of P3 show a significant increase. Although overall not statistically
1014 significant, 3 out of 4 samples of P2 show HRD score increases.
1015 (G) UpSet plot of thetop 5 altered genes. *BRCA2* is the top altered gene with multiple different
1016 cooccurrence partners.
1017 (H) Boxplot of HRD scores between *BRCA2* LOH and non-LOH groups showing significant
1018 increase of HRD scores in the LOH group.
1019 (I) Boxplots of mRNA and protein expression between LOH and non-LOH groups for each HRR
1020 gene. Among 37 HRR genes (*FANCD2* was excluded due to low LOH event number), 16 proteins
1021 were quantified in cohort 1 and are shown here.
1022 (J) Bar plot showing frequencies of BRCA2-retaining samples in each subtype. P2/P3 subtypes,
1023 especially P2, show decrease of BRCA2 retention compared to P1.
1024 (K) The RAD50 IHC score does not change significantly between HRD-low and HRD-high
1025 groups, but it shows a downward trend in P2/P3 compared to P1.
1026 (L) Boxplot of phosphorylated H2AX IHC scores by subtypes.
1027 (M) Boxplot of XRCC5/6 protein expression by subtypes. P3 shows significantly higher
1028 expression of XRCC5/6 compared to P1 or P2.
1029 (N) Boxplot of PARP1 protein expression between HRD-low and HRD-high groups by subtypes.

1030 (O) Kaplan-Meier analysis of overall and recurrence-free survival in combined cohorts 1 and 2,
1031 stratified by median values of RAD51 H-score and BRCA2 tumor cell positivity.
1032
1033 **Figure 7. Tumor microenvironment profiling shows subtype differences and immune**
1034 **evasion characteristics in both immune-cold and immune-hot tumors.**
1035 (A) A heatmap of immune signatures including copy number status, mRNA, and protein level. In
1036 general, the ImmuneScore positively correlates with CD3 count, antigen presenting molecules,
1037 and immune-related pathways. Bars on the right side of the heatmap denote Spearman's
1038 correlation coefficients between ImmuneScores and each gene/signature.
1039 (B) Boxplot showing the ImmuneScore by proteome subtypes for cohort 1.
1040 (C) Boxplot showing the ImmuneScore between paired primary and metastatic tumors.
1041 (D) Volcano plot showing differential protein expression results between immune-cold and
1042 immune-hot. Proteins with statistical significance ($q<0.05$) are shown in red (log2(fold change)>1)
1043 and blue (log2(fold change) < -1). The horizontal dashed line denotes $q=0.05$.
1044 (E) GSE plot of antigen presenting pathway-related KEGG and REACTOME terms, showing
1045 antigen presenting processes are significantly downregulated in immune-cold tumors compared
1046 to immune-hot tumors.
1047 (F) A bar plot of APM protein expression ratios between immune-cold and immune-hot in cohort
1048 1. * denotes statistical significance in the differential expressional analysis shown in (D).
1049 (G) Boxplot of IRF1 protein expression between immune-hot and immune-cold tumors.
1050 (H) A correlation plot (colored by Spearman correlation coefficient) including the ImmuneScore,
1051 IRF1, and antigen processing machinery protein expression. *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$.
1052 IFNgamma_response scores are ssGSEA scores of the
1053 "Hallmark_Interferon_Gamma_Reponse" term.
1054 (I) Boxplot of immune-suppressive molecules between immune-cold and immune-hot tumors
1055 showing only TIM3 is significantly upregulated in immune-hot tumors vs. immune-cold tumors.
1056 (J) Boxplot of LGALS9 protein expression by subtypes (left side) and representative IHC image
1057 (right side). P2 shows a trend for higher expression of LGALS9 compared to P1 or P3. Bar, 50
1058 μm.
1059 (K) Boxplots of CD68-positive or CD163 -sitive cell counts by subtypes.
1060 (L) Kaplan-Meier analyses of recurrence-free survival in combined cohorts 1 and 2, stratified by
1061 median values of CD68-positive or CD163-positive cell counts.
1062
1063

1064 **Figure 8. IHC-based classifier development and integrated summary of proteome-defined**
1065 **STLMS subtypes.**
1066 (A) Workflow for development of an IHC-based classifier for proteome-defined STLMS subtype
1067 prediction. Twenty-three candidate IHC markers were evaluated, including 18 markers assessed
1068 in this study and 5 additional subtype-informative markers identified through subtype-specific
1069 differential expression analyses. All possible 3-, 4-, 5-, and 6-marker combinations were tested
1070 in the training cohort using multinomial logistic regression with 10-fold cross-validation. The final
1071 selected classifier included CD133, MYC, CDK2, AURKA, AURKB, and HLA-ABC.
1072 (B) Heatmap of the 6 selected IHC markers in cohort 1 and cohort 2.
1073 (C) Recurrence-free survival according to inferred proteomic subtype in the independent
1074 validation cohort 2 classified by the locked IHC model.
1075 (D) A schematic overview of the proteome-defined STLMS subtypes summarizing major genomic
1076 alterations, pathway and kinome features, DNA repair states, tumor microenvironment
1077 characteristics, prognostic associations, and potential therapeutic vulnerabilities. APM, antigen-
1078 processing machinery; EMT, epithelial-mesenchymal transition; HRD, homologous
1079 recombination deficiency; LOH, loss of heterozygosity; NHEJ, nonhomologous end joining; TME,
1080 tumor microenvironment.

1081 **References**

1082 1. Soft Tissue and Bone Tumours. (2020). WHO Classification of Tumours, 5th edition
1083 Edition (IARC Publications).

1084 2. Devaud, N., Vornicova, O., Abdul Razak, A.R., Khalili, K., Demicco, E.G., Mitric, C.,
1085 Bernardini, M.Q., and Gladdy, R.A. (2022). Leiomyosarcoma: Current Clinical
1086 Management and Future Horizons. Surg Oncol Clin N Am *31*, 527-546.
1087 10.1016/j.soc.2022.03.011.

1088 3. Serrano, C., and George, S. (2013). Leiomyosarcoma. Hematol Oncol Clin North Am *27*,
1089 957-974. 10.1016/j.hoc.2013.07.002.

1090 4. Acem, I., van Houdt, W.J., Grünhagen, D.J., van der Graaf, W.T.A., Rueten-Budde, A.J.,
1091 Gelderblom, H., Verhoef, C., and van de Sande, M.A.J. (2022). The role of perioperative
1092 chemotherapy in primary high-grade extremity soft tissue sarcoma: a risk-stratified
1093 analysis using PERSARC. European journal of cancer (Oxford, England : 1990) *165*, 71-80.
1094 10.1016/j.ejca.2022.01.013.

1095 5. Gronchi, A., Miah, A.B., Dei Tos, A.P., Abecassis, N., Bajpai, J., Bauer, S., Biagini, R.,
1096 Bielack, S., Blay, J.Y., Bolle, S., et al. (2021). Soft tissue and visceral sarcomas: ESMO-
1097 EURACAN-GENTURIS Clinical Practice Guidelines for diagnosis, treatment and follow-
1098 up(☆). Ann Oncol *32*, 1348-1365. 10.1016/j.annonc.2021.07.006.

1099 6. Pisters, P.W., Leung, D.H., Woodruff, J., Shi, W., and Brennan, M.F. (1996). Analysis of
1100 prognostic factors in 1,041 patients with localized soft tissue sarcomas of the extremities.
1101 Journal of clinical oncology : official journal of the American Society of Clinical Oncology
1102 *14*, 1679-1689. 10.1200/jco.1996.14.5.1679.

1103 7. Lewis, J.J., Leung, D., Heslin, M., Woodruff, J.M., and Brennan, M.F. (1997). Association
1104 of local recurrence with subsequent survival in extremity soft tissue sarcoma. Journal of
1105 clinical oncology : official journal of the American Society of Clinical Oncology *15*, 646-
1106 652. 10.1200/jco.1997.15.2.646.

1107 8. Trovik, C.S., Bauer, H.C., Alvegård, T.A., Anderson, H., Blomqvist, C., Berlin, O., Gustafson,
1108 P., Saeter, G., and Wallöe, A. (2000). Surgical margins, local recurrence and metastasis in
1109 soft tissue sarcomas: 559 surgically-treated patients from the Scandinavian Sarcoma
1110 Group Register. European journal of cancer (Oxford, England : 1990) *36*, 710-716.
1111 10.1016/s0959-8049(99)00287-7.

1112 9. Linch, M., Miah, A.B., Thway, K., Judson, I.R., and Benson, C. (2014). Systemic treatment
1113 of soft-tissue sarcoma-gold standard and novel therapies. Nat Rev Clin Oncol *11*, 187-
1114 202. 10.1038/nrclinonc.2014.26.

1115 10. Savina, M., Le Cesne, A., Blay, J.Y., Ray-Coquard, I., Mir, O., Toulmonde, M., Cousin, S.,
1116 Terrier, P., Ranchere-Vince, D., Meeus, P., et al. (2017). Patterns of care and outcomes of
1117 patients with METAstatic soft tissue SARComa in a real-life setting: the METASARC
1118 observational study. BMC Med *15*, 78. 10.1186/s12916-017-0831-7.

1119 11. Chibon, F., Lagarde, P., Salas, S., Pérot, G., Brouste, V., Tirode, F., Lucchesi, C., de Reynies,
1120 A., Kauffmann, A., Bui, B., et al. (2010). Validated prediction of clinical outcome in
1121 sarcomas and multiple types of cancer on the basis of a gene expression signature related
1122 to genome complexity. Nature medicine *16*, 781-787. 10.1038/nm.2174.

1123 12. TCGA (2017). Comprehensive and Integrated Genomic Characterization of Adult Soft
1124 Tissue Sarcomas. Cell *171*, 950-965.e928. 10.1016/j.cell.2017.10.014.
1125 13. Nacev, B.A., Sanchez-Vega, F., Smith, S.A., Antonescu, C.R., Rosenbaum, E., Shi, H., Tang,
1126 C., Socci, N.D., Rana, S., Gularte-Merida, R., et al. (2022). Clinical sequencing of soft tissue
1127 and bone sarcomas delineates diverse genomic landscapes and potential therapeutic
1128 targets. Nat Commun *13*, 3405. 10.1038/s41467-022-30453-x.
1129 14. Italiano, A., Kind, M., Stoeckle, E., Jones, N., Coindre, J.M., and Bui, B. (2011).
1130 Temsirolimus in advanced leiomyosarcomas: patterns of response and correlation with
1131 the activation of the mammalian target of rapamycin pathway. Anticancer Drugs *22*, 463-
1132 467. 10.1097/CAD.0b013e3283442074.
1133 15. Schwartz, G.K., Tap, W.D., Qin, L.X., Livingston, M.B., Undevia, S.D., Chmielowski, B.,
1134 Agulnik, M., Schuetze, S.M., Reed, D.R., Okuno, S.H., et al. (2013). Cixutumumab and
1135 temsirolimus for patients with bone and soft-tissue sarcoma: a multicentre, open-label,
1136 phase 2 trial. The Lancet. Oncology *14*, 371-382. 10.1016/s1470-2045(13)70049-4.
1137 16. Burns, J., Wilding, C.P., R, L.J., and P, H.H. (2020). Proteomic research in sarcomas -
1138 current status and future opportunities. Semin Cancer Biol *61*, 56-70.
1139 10.1016/j.semcancer.2019.11.003.
1140 17. Gillette, M.A., Satpathy, S., Cao, S., Dhanasekaran, S.M., Vasaikar, S.V., Krug, K., Petralia,
1141 F., Li, Y., Liang, W.W., Reva, B., et al. (2020). Proteogenomic Characterization Reveals
1142 Therapeutic Vulnerabilities in Lung Adenocarcinoma. Cell *182*, 200-225.e235.
1143 10.1016/j.cell.2020.06.013.
1144 18. Zhang, B., Wang, J., Wang, X., Zhu, J., Liu, Q., Shi, Z., Chambers, M.C., Zimmerman, L.J.,
1145 Shaddox, K.F., Kim, S., et al. (2014). Proteogenomic characterization of human colon and
1146 rectal cancer. Nature *513*, 382-387. 10.1038/nature13438.
1147 19. Burns, J., Wilding, C.P., Krasny, L., Zhu, X., Chadha, M., Tam, Y.B., Ps, H., Mahalingam,
1148 A.H., Lee, A.T.J., Arthur, A., et al. (2023). The proteomic landscape of soft tissue
1149 sarcomas. Nat Commun *14*, 3834. 10.1038/s41467-023-39486-2.
1150 20. Tang, S., Wang, Y., Luo, R., Fang, R., Liu, Y., Xiang, H., Ran, P., Tong, Y., Sun, M., Tan, S., et
1151 al. (2024). Proteomic characterization identifies clinically relevant subgroups of soft
1152 tissue sarcoma. Nat Commun *15*, 1381. 10.1038/s41467-024-45306-y.
1153 21. AACR Project GENIE: Powering Precision Medicine through an International Consortium.
1154 (2017). Cancer Discov *7*, 818-831. 10.1158/2159-8290.Cd-17-0151.
1155 22. Ucar, D., and Lin, D.I. (2015). Amplification of the bromodomain-containing protein 4
1156 gene in ovarian high-grade serous carcinoma is associated with worse prognosis and
1157 survival. Mol Clin Oncol *3*, 1291-1294. 10.3892/mco.2015.622.
1158 23. Brahmi, M., Lesluyes, T., Dufresne, A., Toulmonde, M., Italiano, A., Mir, O., Le Cesne, A.,
1159 Valentin, T., Chevreau, C., Bonvalot, S., et al. (2021). Expression and prognostic
1160 significance of PDGF ligands and receptors across soft tissue sarcomas. ESMO Open *6*,
1161 100037. 10.1016/j.esmoop.2020.100037.
1162 24. Peng, L., Li, J., Wu, J., Xu, B., Wang, Z., Giamas, G., Stebbing, J., and Yu, Z. (2021). A Pan-
1163 Cancer Analysis of SMARCA4 Alterations in Human Cancers. Front Immunol *12*, 762598.
1164 10.3389/fimmu.2021.762598.

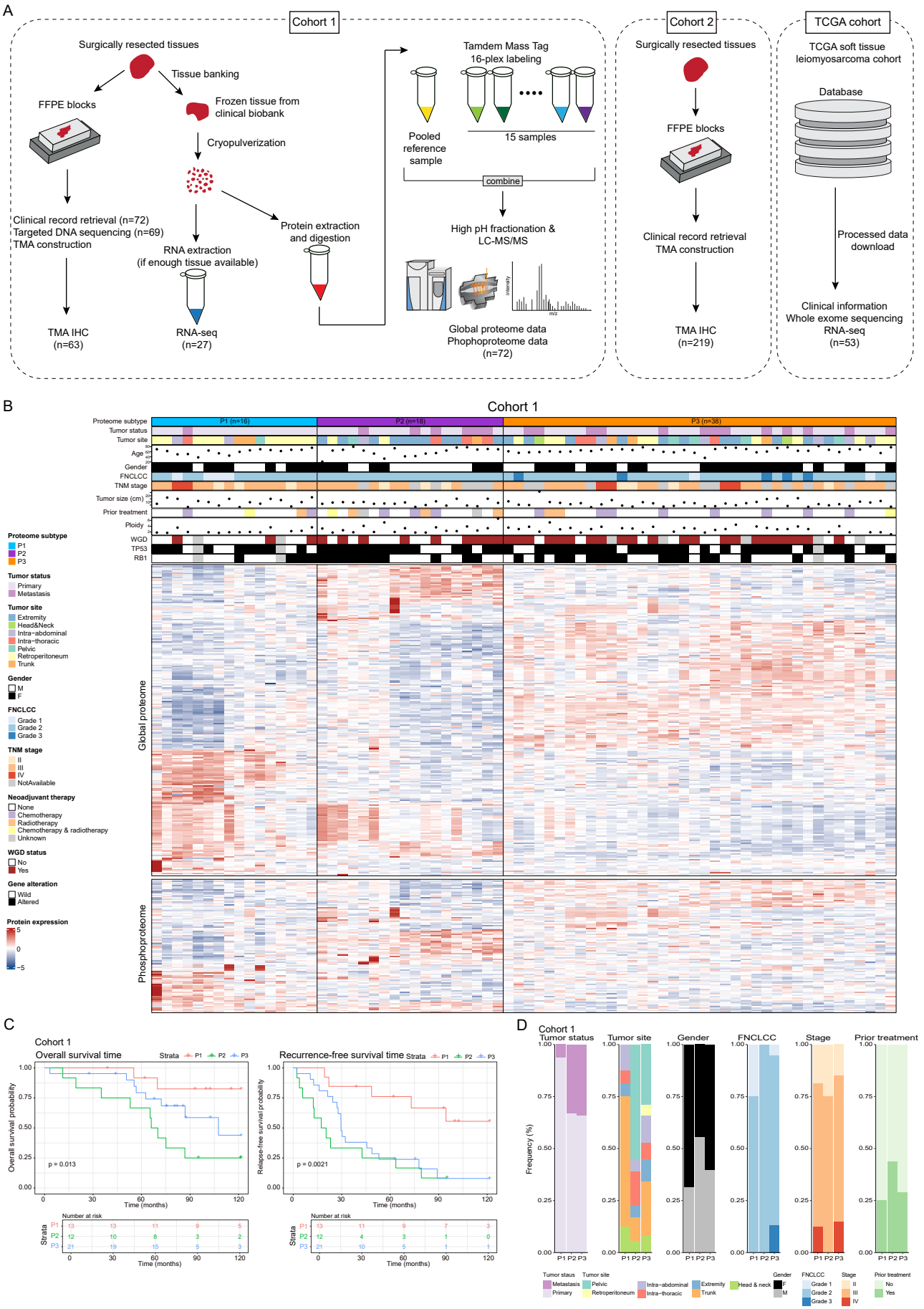
1165 25. Liang, J., Li, B., Yuan, L., and Ye, Z. (2015). Prognostic value of IGF-1R expression in bone
1166 and soft tissue sarcomas: a meta-analysis. Onco Targets Ther *8*, 1949-1955.
1167 10.2147/ott.S88293.
1168 26. Li, Y., Li, Y., Liu, Y., Xie, P., Li, F., and Li, G. (2014). PAX6, a novel target of microRNA-7,
1169 promotes cellular proliferation and invasion in human colorectal cancer cells. Dig Dis Sci
1170 *59*, 598-606. 10.1007/s10620-013-2929-x.
1171 27. Ooki, A., Dinalankara, W., Marchionni, L., Tsay, J.J., Goparaju, C., Maleki, Z., Rom, W.N.,
1172 Pass, H.I., and Hoque, M.O. (2018). Epigenetically regulated PAX6 drives cancer cells
1173 toward a stem-like state via GLI-SOX2 signaling axis in lung adenocarcinoma. Oncogene
1174 *37*, 5967-5981. 10.1038/s41388-018-0373-2.
1175 28. Jing, N., Du, X., Liang, Y., Tao, Z., Bao, S., Xiao, H., Dong, B., Gao, W.Q., and Fang, Y.X.
1176 (2024). PAX6 promotes neuroendocrine phenotypes of prostate cancer via enhancing
1177 MET/STAT5A-mediated chromatin accessibility. J Exp Clin Cancer Res *43*, 144.
1178 10.1186/s13046-024-03064-1.
1179 29. Ballotti, R., Cheli, Y., and Bertolotto, C. (2020). The complex relationship between MITF
1180 and the immune system: a Melanoma ImmunoTherapy (response) Factor? Molecular
1181 cancer *19*, 170. 10.1186/s12943-020-01290-7.
1182 30. Du, R., Huang, C., Liu, K., Li, X., and Dong, Z. (2021). Targeting AURKA in Cancer: molecular
1183 mechanisms and opportunities for Cancer therapy. Molecular cancer *20*, 15.
1184 10.1186/s12943-020-01305-3.
1185 31. Yilmaz, S., Ayati, M., Schlatzer, D., Cicek, A.E., Chance, M.R., and Koyuturk, M. (2021).
1186 Robust inference of kinase activity using functional networks. Nat Commun *12*, 1177.
1187 10.1038/s41467-021-21211-6.
1188 32. Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the
1189 genome integrates intrinsic and environmental signals. Nature genetics *33 Suppl*, 245-
1190 254. 10.1038/ng1089.
1191 33. Li, X., and Heyer, W.D. (2008). Homologous recombination in DNA repair and DNA
1192 damage tolerance. Cell Res *18*, 99-113. 10.1038/cr.2008.1.
1193 34. Lord, C.J., and Ashworth, A. (2016). BRCAness revisited. Nature reviews. Cancer *16*, 110-
1194 120. 10.1038/nrc.2015.21.
1195 35. Turajlic, S., and Swanton, C. (2016). Metastasis as an evolutionary process. Science (New
1196 York, N.Y.) *352*, 169-175. 10.1126/science.aaf2784.
1197 36. Bakhoum, S.F., Ngo, B., Laughney, A.M., Cavallo, J.A., Murphy, C.J., Ly, P., Shah, P., Sriram,
1198 R.K., Watkins, T.B.K., Taunk, N.K., et al. (2018). Chromosomal instability drives metastasis
1199 through a cytosolic DNA response. Nature *553*, 467-472. 10.1038/nature25432.
1200 37. Planas-Paz, L., Pliego-Mendieta, A., Hagedorn, C., Aguilera-Garcia, D., Haberecker, M.,
1201 Arnold, F., Herzog, M., Bankel, L., Guggenberger, R., Steiner, S., et al. (2023). Unravelling
1202 homologous recombination repair deficiency and therapeutic opportunities in soft tissue
1203 and bone sarcoma. EMBO Mol Med *15*, e16863. 10.15252/emmm.202216863.
1204 38. Baumann, P., and West, S.C. (1998). Role of the human RAD51 protein in homologous
1205 recombination and double-stranded-break repair. Trends Biochem Sci *23*, 247-251.
1206 10.1016/s0968-0004(98)01232-8.

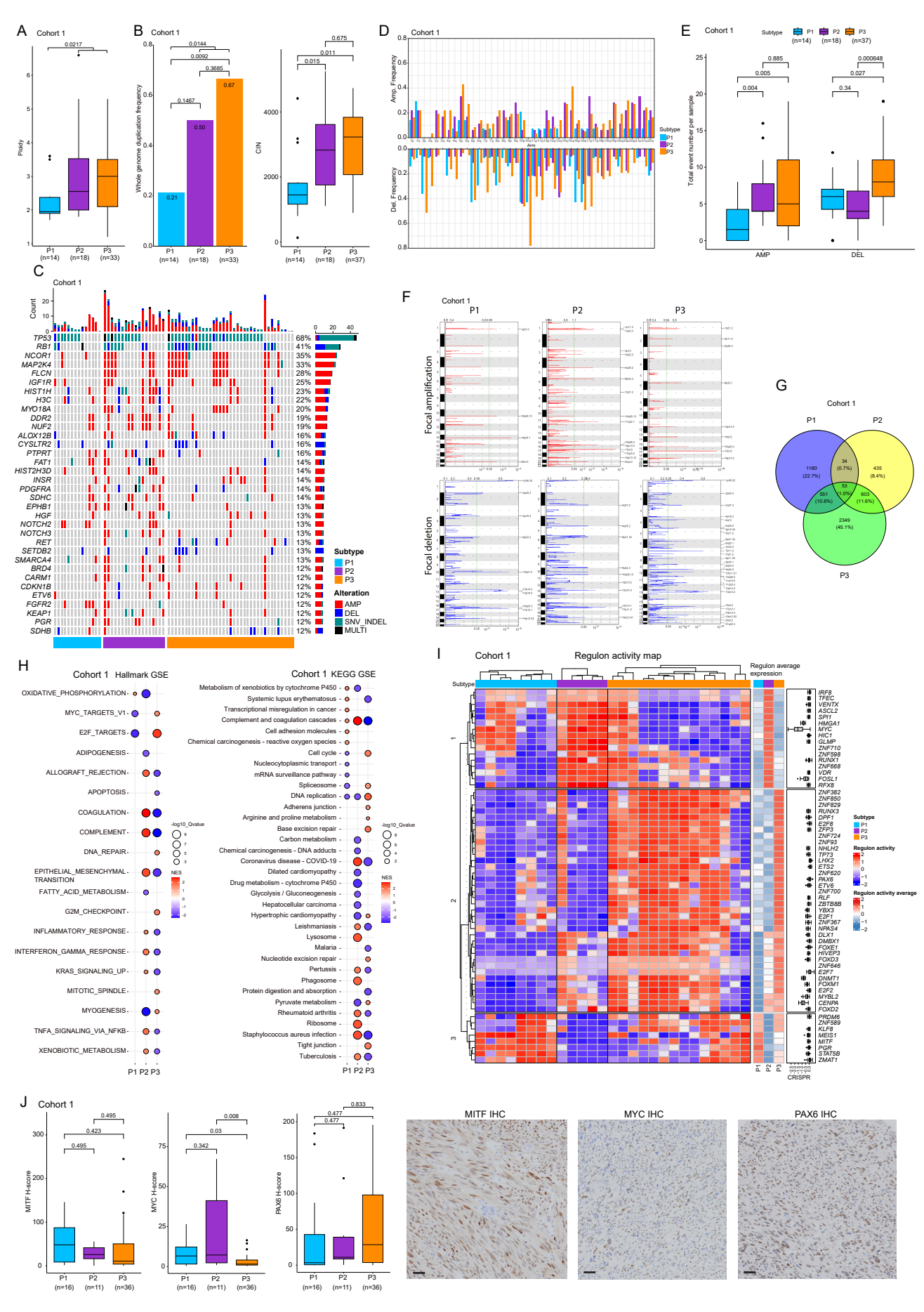
1207 39. Mansour, W.Y., Rhein, T., and Dahm-Daphi, J. (2010). The alternative end-joining
1208 pathway for repair of DNA double-strand breaks requires PARP1 but is not dependent
1209 upon microhomologies. Nucleic Acids Res *38*, 6065-6077. 10.1093/nar/gkq387.
1210 40. Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W.,
1211 Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and
1212 stromal and immune cell admixture from expression data. Nat Commun *4*, 2612.
1213 10.1038/ncomms3612.
1214 41. Zhou, F. (2009). Molecular mechanisms of IFN-gamma to up-regulate MHC class I antigen
1215 processing and presentation. Int Rev Immunol *28*, 239-260.
1216 10.1080/08830180902978120.
1217 42. Der, S.D., Zhou, A., Williams, B.R., and Silverman, R.H. (1998). Identification of genes
1218 differentially regulated by interferon alpha, beta, or gamma using oligonucleotide arrays.
1219 Proceedings of the National Academy of Sciences of the United States of America *95*,
1220 15623-15628. 10.1073/pnas.95.26.15623.
1221 43. Kikutake, C., and Suyama, M. (2023). Pan-cancer analysis of whole-genome doubling and
1222 its association with patient prognosis. BMC cancer *23*, 619. 10.1186/s12885-023-11132-
1223 6.
1224 44. Baker, T.M., Lai, S., Lynch, A.R., Lesluyes, T., Yan, H., Ogilvie, H.A., Verfaillie, A., Dentro,
1225 S., Bowes, A.L., Pillay, N., et al. (2024). The History of Chromosomal Instability in Genome-
1226 Doubled Tumors. Cancer Discov *14*, 1810-1822. 10.1158/2159-8290.Cd-23-1249.
1227 45. Nikhil, K., and Shah, K. (2024). The significant others of aurora kinase a in cancer:
1228 combination is the key. Biomark Res *12*, 109. 10.1186/s40364-024-00651-4.
1229 46. Jamalzadeh, S., Dai, J., Lavikka, K., Li, Y., Jiang, J., Huhtinen, K., Virtanen, A., Oikkonen, J.,
1230 Hietanen, S., Hynninen, J., et al. (2024). Genome-wide quantification of copy-number
1231 aberration impact on gene expression in ovarian high-grade serous carcinoma. BMC
1232 cancer *24*, 173. 10.1186/s12885-024-11895-6.
1233 47. Yamamoto, H., Williams, E.G., Mouchiroud, L., Cantó, C., Fan, W., Downes, M., Héligon,
1234 C., Barish, G.D., Desvergne, B., Evans, R.M., et al. (2011). NCoR1 is a conserved
1235 physiological modulator of muscle mass and oxidative function. Cell *147*, 827-839.
1236 10.1016/j.cell.2011.10.017.
1237 48. Kohn, E.C., Lee, J.M., and Ivy, S.P. (2017). The HRD Decision-Which PARP Inhibitor to Use
1238 for Whom and When. Clin Cancer Res *23*, 7155-7157. 10.1158/1078-0432.Ccr-17-2186.
1239 49. Dall, G., Vandenberg, C.J., Nesic, K., Ratnayake, G., Zhu, W., Vissers, J.H.A., Bedő, J.,
1240 Penington, J., Wakefield, M.J., Kee, D., et al. (2023). Targeting homologous
1241 recombination deficiency in uterine leiomyosarcoma. J Exp Clin Cancer Res *42*, 112.
1242 10.1186/s13046-023-02687-0.
1243 50. Slade, D., and Loizou, J.I. (2023). Leveraging homologous recombination repair deficiency
1244 in sarcoma. EMBO Mol Med *15*, e17453. 10.15252/emmm.202317453.
1245 51. Cheng, D.T., Mitchell, T.N., Zehir, A., Shah, R.H., Benayed, R., Syed, A., Chandramohan,
1246 R., Liu, Z.Y., Won, H.H., Scott, S.N., et al. (2015). Memorial Sloan Kettering-Integrated
1247 Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-
1248 Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology.
1249 The Journal of molecular diagnostics : JMD *17*, 251-264. 10.1016/j.jmoldx.2014.12.006.

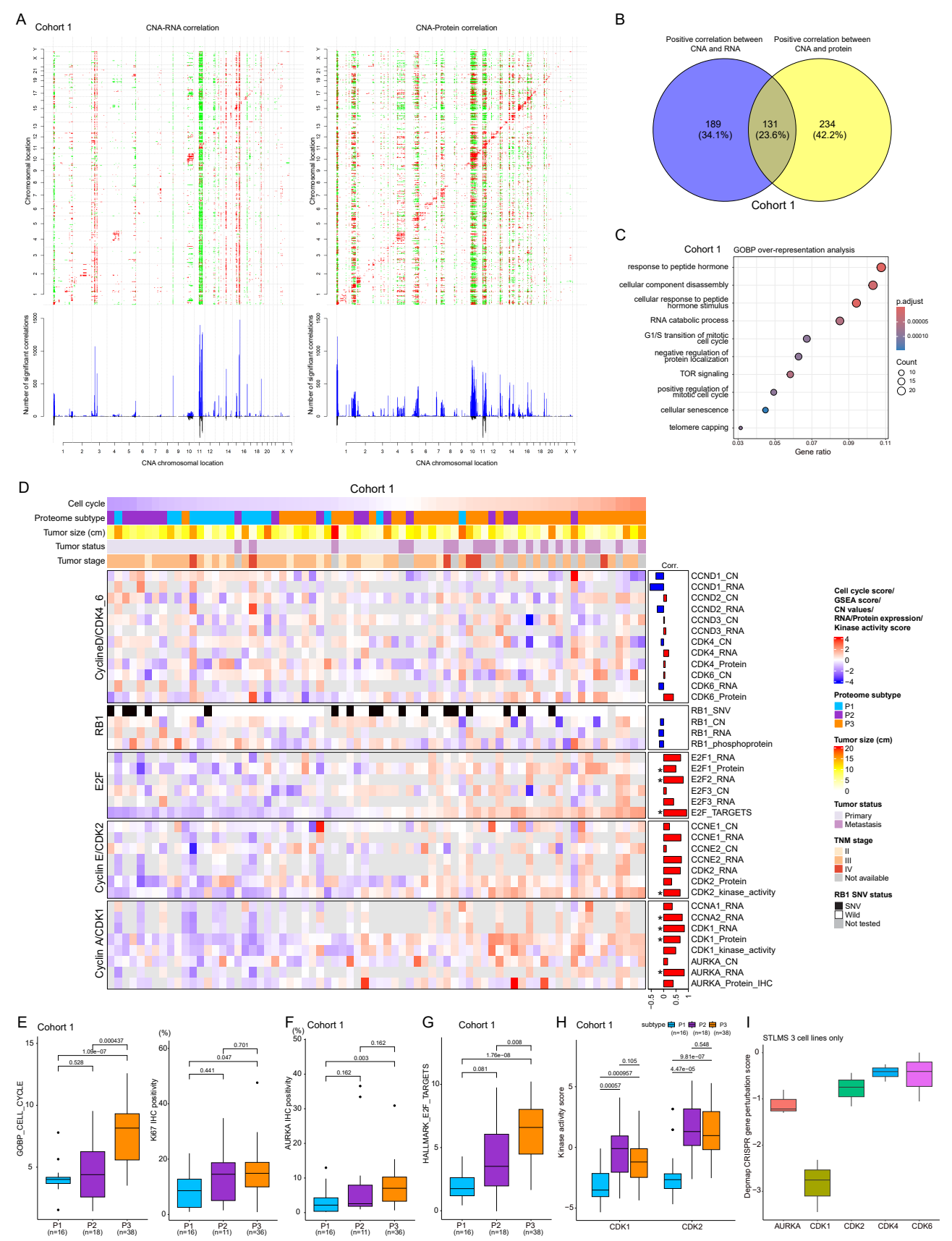
1250 52. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-
1251 Wheeler transform. Bioinformatics (Oxford, England) *25*, 1754-1760.
1252 10.1093/bioinformatics/btp324.
1253 53. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis,
1254 A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation
1255 discovery and genotyping using next-generation DNA sequencing data. Nature genetics
1256 *43*, 491-498. 10.1038/ng.806.
1257 54. Mose, L.E., Wilkerson, M.D., Hayes, D.N., Perou, C.M., and Parker, J.S. (2014). ABRA:
1258 improved coding indel detection via assembly-based realignment. Bioinformatics
1259 (Oxford, England) *30*, 2813-2815. 10.1093/bioinformatics/btu376.
1260 55. Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel,
1261 S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point
1262 mutations in impure and heterogeneous cancer samples. Nature biotechnology *31*, 213-
1263 219. 10.1038/nbt.2514.
1264 56. Ye, K., Schulz, M.H., Long, Q., Apweiler, R., and Ning, Z. (2009). Pindel: a pattern growth
1265 approach to detect break points of large deletions and medium sized insertions from
1266 paired-end short reads. Bioinformatics (Oxford, England) *25*, 2865-2871.
1267 10.1093/bioinformatics/btp394.
1268 57. Lai, Z., Markovets, A., Ahdesmaki, M., Chapman, B., Hofmann, O., McEwen, R., Johnson,
1269 J., Dougherty, B., Barrett, J.C., and Dry, J.R. (2016). VarDict: a novel and versatile variant
1270 caller for next-generation sequencing in cancer research. Nucleic Acids Res *44*, e108.
1271 10.1093/nar/gkw227.
1272 58. Shen, R., and Seshan, V.E. (2016). FACETS: allele-specific copy number and clonal
1273 heterogeneity analysis tool for high-throughput DNA sequencing. Nucleic Acids Res *44*,
1274 e131. 10.1093/nar/gkw520.
1275 59. Bielski, C.M., Zehir, A., Penson, A.V., Donoghue, M.T.A., Chatila, W., Armenia, J., Chang,
1276 M.T., Schram, A.M., Jonsson, P., Bandlamudi, C., et al. (2018). Genome doubling shapes
1277 the evolution and prognosis of advanced cancers. Nature genetics *50*, 1189-1195.
1278 10.1038/s41588-018-0165-1.
1279 60. Colaprico, A., Silva, T.C., Olsen, C., Garofano, L., Cava, C., Garolini, D., Sabedot, T.S.,
1280 Malta, T.M., Pagnotta, S.M., Castiglioni, I., et al. (2016). TCGAbiolinks: an R/Bioconductor
1281 package for integrative analysis of TCGA data. Nucleic Acids Res *44*, e71.
1282 10.1093/nar/gkv1507.
1283 61. Sztupinszki, Z., Diossy, M., Krzystanek, M., Reiniger, L., Csabai, I., Favero, F., Birkbak, N.J.,
1284 Eklund, A.C., Syed, A., and Szallasi, Z. (2018). Migrating the SNP array-based homologous
1285 recombination deficiency measures to next generation sequencing data of breast cancer.
1286 NPJ Breast Cancer *4*, 16. 10.1038/s41523-018-0066-6.
1287 62. Shi, Z., Chen, B., Han, X., Gu, W., Liang, S., and Wu, L. (2023). Genomic and molecular
1288 landscape of homologous recombination deficiency across multiple cancer types.
1289 Scientific reports *13*, 8899. 10.1038/s41598-023-35092-w.
1290 63. Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., and Getz, G.
1291 (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal

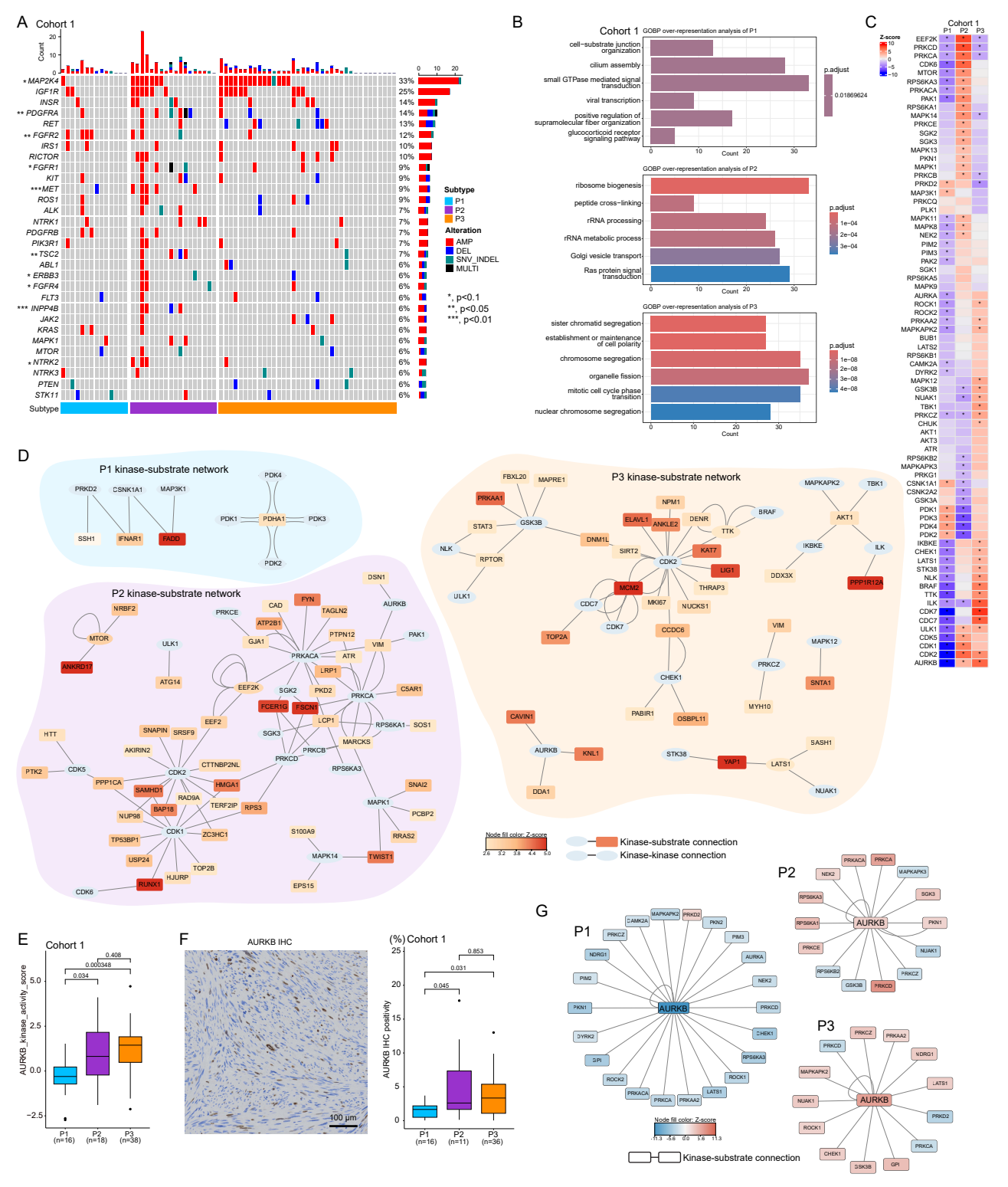
1292 somatic copy-number alteration in human cancers. Genome Biol *12*, R41. 10.1186/gb-
1293 2011-12-4-r41.
1294 64. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson,
1295 M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics
1296 (Oxford, England) *29*, 15-21. 10.1093/bioinformatics/bts635.
1297 65. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq
1298 data with or without a reference genome. BMC bioinformatics *12*, 323. 10.1186/1471-
1299 2105-12-323.
1300 66. McAlister, G.C., Nusinow, D.P., Jedrychowski, M.P., Wuhr, M., Huttlin, E.L., Erickson, B.K.,
1301 Rad, R., Haas, W., and Gygi, S.P. (2014). MultiNotch MS3 Enables Accurate, Sensitive, and
1302 Multiplexed Detection of Differential Expression across Cancer Cell Line Proteomes.
1303 Analytical Chemistry *86*, 7150-7158. 10.1021/ac502040v.
1304 67. Navarrete-Perea, J., Yu, Q., Gygi, S.P., and Paulo, J.A. (2018). Streamlined Tandem Mass
1305 Tag (SL-TMT) Protocol: An Efficient Strategy for Quantitative (Phospho)proteome
1306 Profiling Using Tandem Mass Tag-Synchronous Precursor Selection-MS3. J Proteome Res
1307 *17*, 2226-2236. 10.1021/acs.jproteome.8b00217.
1308 68. Liu, H., Sadygov, R.G., and Yates, J.R., 3rd (2004). A model for random sampling and
1309 estimation of relative protein abundance in shotgun proteomics. Analytical chemistry *76*,
1310 4193-4201. 10.1021/ac0498563.
1311 69. Plubell, D.L., Wilmarth, P.A., Zhao, Y., Fenton, A.M., Minnier, J., Reddy, A.P., Klimek, J.,
1312 Yang, X., David, L.L., and Pamir, N. (2017). Extended Multiplexing of Tandem Mass Tags
1313 (TMT) Labeling Reveals Age and High Fat Diet Specific Proteome Changes in Mouse
1314 Epididymal Adipose Tissue. Mol Cell Proteomics *16*, 873-890.
1315 10.1074/mcp.M116.065524.
1316 70. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package
1317 for differential expression analysis of digital gene expression data. Bioinformatics
1318 (Oxford, England) *26*, 139-140. 10.1093/bioinformatics/btp616.
1319 71. Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix
1320 factorization. BMC bioinformatics *11*, 367. 10.1186/1471-2105-11-367.
1321 72. Kim, H., and Park, H. (2007). Sparse non-negative matrix factorizations via alternating
1322 non-negativity-constrained least squares for microarray data analysis. Bioinformatics
1323 (Oxford, England) *23*, 1495-1502. 10.1093/bioinformatics/btm134.
1324 73. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for
1325 comparing biological themes among gene clusters. Omics *16*, 284-287.
1326 10.1089/omi.2011.0118.
1327 74. Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C.,
1328 Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that
1329 oncogenic KRAS-driven cancers require TBK1. Nature *462*, 108-112.
1330 10.1038/nature08460.
1331 75. Hornbeck, P.V., Zhang, B., Murray, B., Kornhauser, J.M., Latham, V., and Skrzypek, E.
1332 (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res *43*,
1333 D512-520. 10.1093/nar/gku1267.

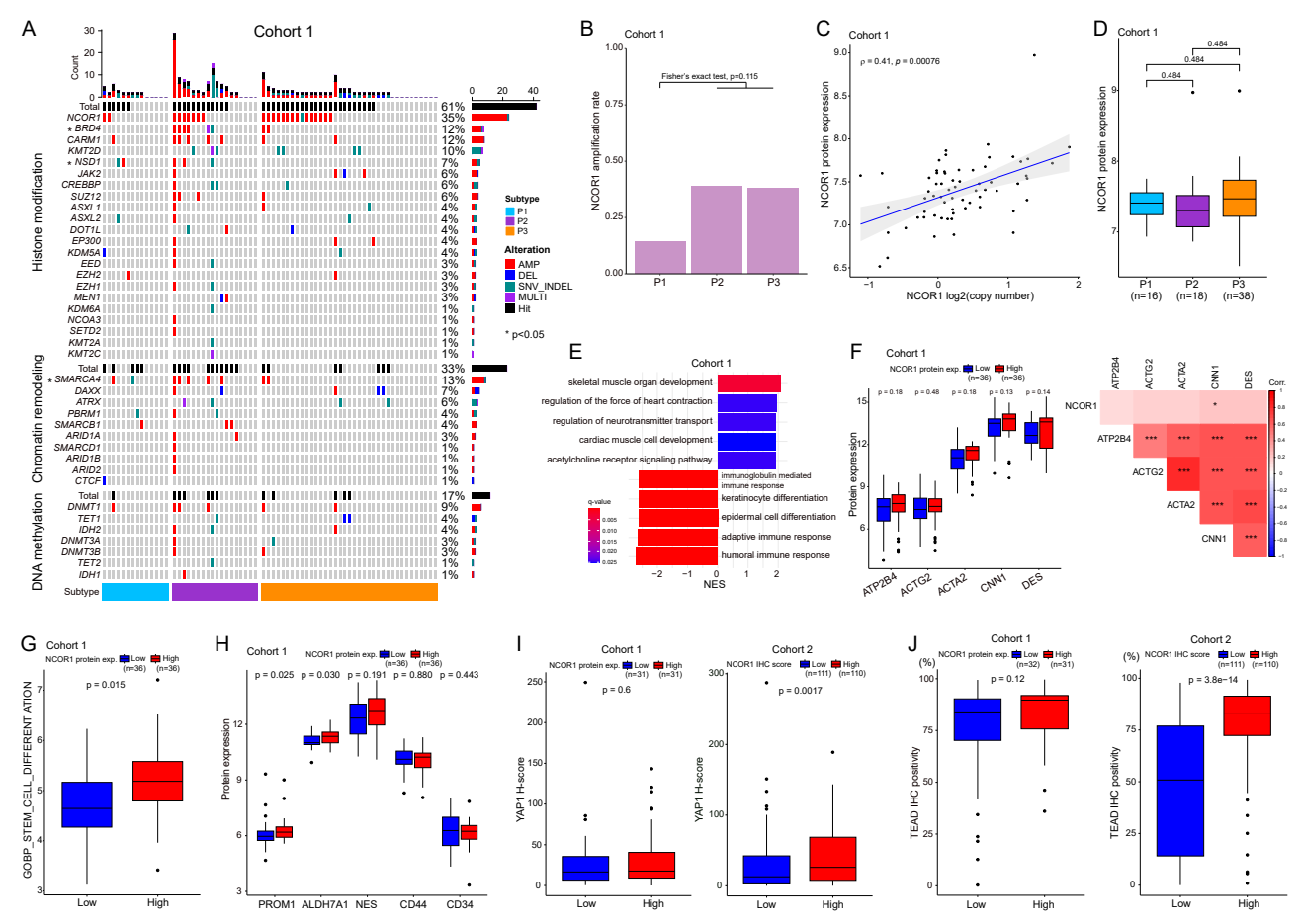
1334 76. Castro, M.A., de Santiago, I., Campbell, T.M., Vaughn, C., Hickey, T.E., Ross, E., Tilley,
1335 W.D., Markowetz, F., Ponder, B.A., and Meyer, K.B. (2016). Regulators of genetic risk of
1336 breast cancer identified by integrative network analysis. Nature genetics *48*, 12-21.
1337 10.1038/ng.3458.
1338 77. Fletcher, M.N., Castro, M.A., Wang, X., de Santiago, I., O'Reilly, M., Chin, S.F., Rueda,
1339 O.M., Caldas, C., Ponder, B.A., Markowetz, F., and Meyer, K.B. (2013). Master regulators
1340 of FGFR2 signalling and breast cancer risk. Nat Commun *4*, 2464. 10.1038/ncomms3464.
1341 78. Robertson, A.G., Kim, J., Al-Ahmadie, H., Bellmunt, J., Guo, G., Cherniack, A.D., Hinoue,
1342 T., Laird, P.W., Hoadley, K.A., Akbani, R., et al. (2017). Comprehensive Molecular
1343 Characterization of Muscle-Invasive Bladder Cancer. Cell *171*, 540-556.e525.
1344 10.1016/j.cell.2017.09.007.
1345 79. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J.,
1346 Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell *172*, 650-
1347 665. 10.1016/j.cell.2018.01.029.
1348 80. Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., and
1349 Califano, A. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory
1350 networks in a mammalian cellular context. BMC bioinformatics *7 Suppl 1*, S7.
1351 10.1186/1471-2105-7-s1-s7.
1352 81. Carro, M.S., Lim, W.K., Alvarez, M.J., Bollo, R.J., Zhao, X., Snyder, E.Y., Sulman, E.P., Anne,
1353 S.L., Doetsch, F., Colman, H., et al. (2010). The transcriptional network for mesenchymal
1354 transformation of brain tumours. Nature *463*, 318-325. 10.1038/nature08712.
1355 82. Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill,
1356 S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a Cancer
1357 Dependency Map. Cell *170*, 564-576.e516. 10.1016/j.cell.2017.06.010.
1358 83. Kuhn M, W.H. (2020). Tidymodels: a collection of packages for modeling and machine
1359 learning using tidyverse principles.
1360

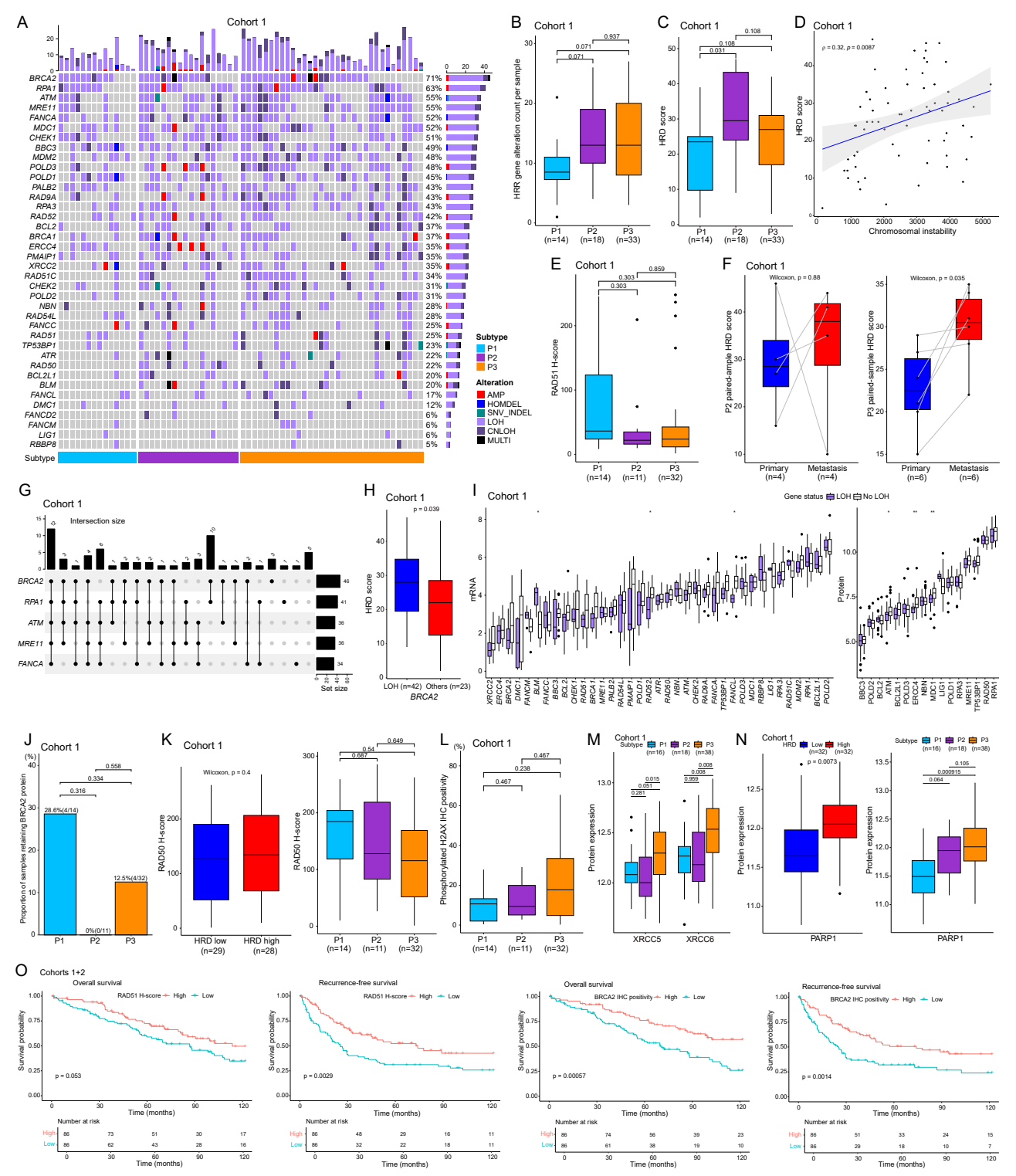

A
Cohort 1
Surgically resected tissues
Tissue banking
FFPE blocks
Frozen tissue from clinical biobank
Cryopulverization
Clinical record retrieval (n=72)
Targeted DNA sequencing (n=69)
TMA construction
RNA extraction (if enough tissue available)
Protein extraction and digestion
TMA IHC (n=63)
RNA-seq (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=219)
TCGA cohort
TCGA soft tissue leiomyosarcoma cohort
Database
Processed data download
Clinical information
Whole exome sequencing
RNA-seq
(n=53)
B
Cohort 1
Proteome subtype
P1 (n=16)
P2 (n=18)
P3 (n=38)
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
P1 13 13 11 9 5
P2 12 10 8 3 2
P3 21 19 15 5 3
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

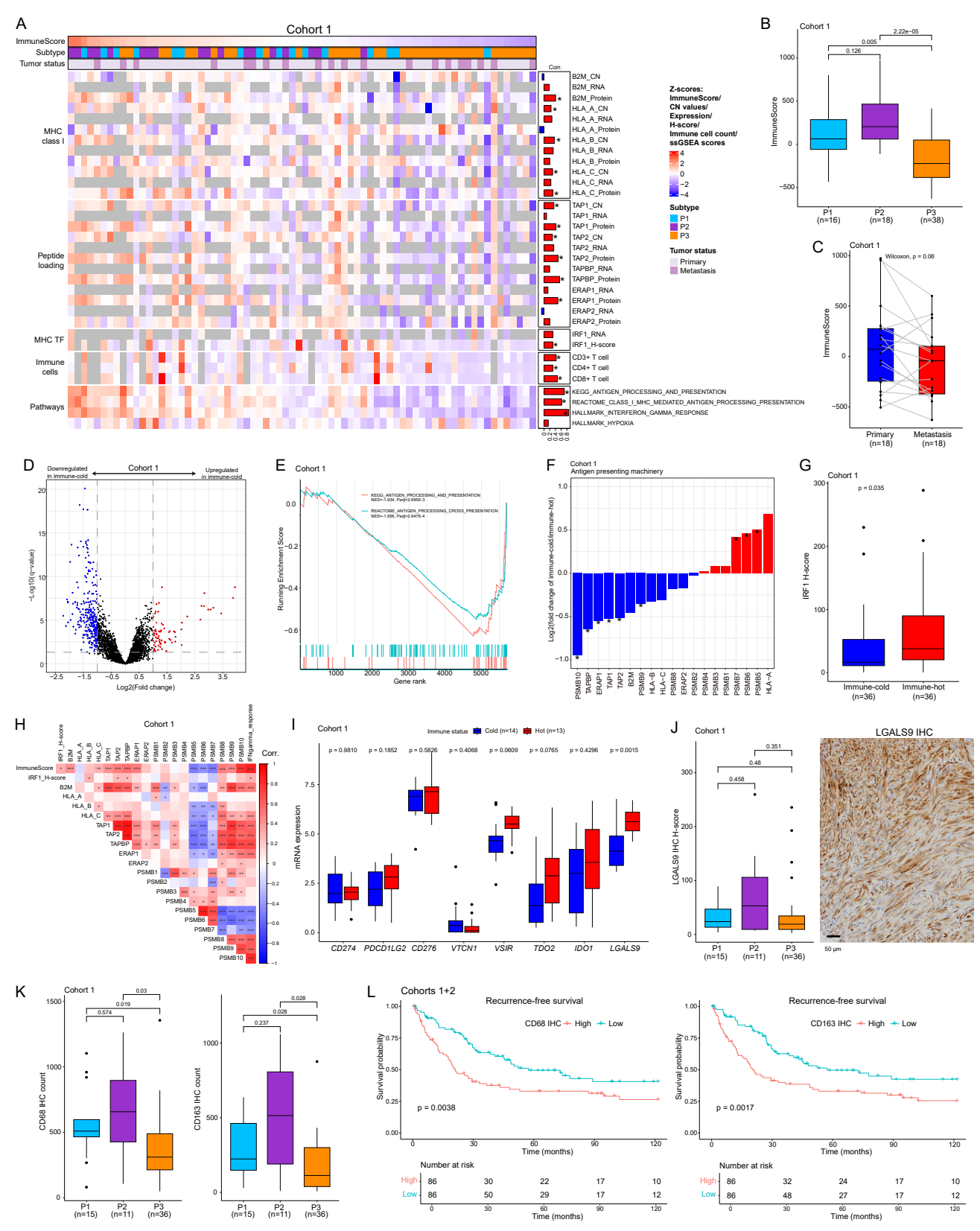

A
Cohort 1
Ploidy
0.0217
P1 (n=14)
P2 (n=18)
P3 (n=33)
B
Cohort 1
Whole genome duplication frequency
0.0144
0.0092
0.3685
0.1467
0.21
0.50
0.67
CIN
0.675
0.011
0.015
C
Cohort 1
Count
TP53
RB1
NCOR1
MAP2K4
FLCN
IGF1R
HIST1H
H3C
MYO18A
DDR2
NUF2
ALOX12B
CYSLTR2
PTPRT
FAT1
HIST2H3D
INSR
PDGFRA
SDHC
EPHB1
HGF
NOTCH2
NOTCH3
RET
SETDB2
SMARCA4
BRD4
CARM1
CDKN1B
ETV6
FGFR2
KEAP1
PGR
SDHB
Subtype
P1
P2
P3
Alteration
AMP
DEL
SNV_INDEL
MULTI
D
Cohort 1
Amp. Frequency
Del. Frequency
Arm
Subtype
P1
P2
P3
E
Cohort 1
Subtype
P1 (n=14)
P2 (n=18)
P3 (n=37)
Total event number per sample
0.885
0.005
0.004
0.000648
0.027
0.34
AMP
DEL
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
ZNF620
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
P1 P2 P3
CRISPR
Subtype
P1
P2
P3
Regulon activity
Regulon activity average
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 (n=16)
P2 (n=11)
P3 (n=36)
MITF IHC
MYC IHC
PAX6 IHC


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
K67 IHC positivity
F
Cohort 1
AURKA IHC positivity
G
Cohort 1
HALLMARK_E2F_TARGETS
H
Cohort 1
Kinase activity score
CDK1
CDK2
I
STLMS 3 cell lines only
Depmap CRISPR gene perturbation score
AURKA
CDK1
CDK2
CDK4
CDK6


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
PRKCA
C5AR1
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
Node fill color: Z-score
Kinase-substrate connection
Kinase-kinase connection
P3 kinase-substrate network
FBXL20
MAPRE1
PRKAA1
GSK3B
STAT3
NLK
RPTOR
ULK1
DNM1L
NPM1
ELAVL1
ANKLE2
DENR
BRAF
TTK
SIRT2
CDK2
KAT7
LIG1
THRAP3
MCM2
MKI67
NUCKS1
CDC7
CDK7
TOP2A
CCDC6
CHEK1
PABIR1
OSBPL11
CAVIN1
AURKB
KNL1
DDA1
MAPKAPK2
TBK1
AKT1
IKBKE
ILK
DDX3X
PPP1R12A
VIM
MAPK12
PRKCZ
SNTA1
MYH10
SASH1
STK38
YAP1
LATS1
NUAK1
E
Cohort 1
AURKB_kinase_activity_score
0.408
0.000348
0.034
P1 (n=16)
P2 (n=18)
P3 (n=38)
F
AURKB IHC
100 μm
(%) Cohort 1
AURKB IHC positivity
0.853
0.031
0.045
P1 (n=16)
P2 (n=11)
P3 (n=36)
G
P1
AURKB
CAMK2A
MAPKAPK2
PRKD2
PRKN2
PIM3
AURKA
NEK2
PRKCD
CHEK1
RPS6KA3
ROCK1
LATS1
PRKAA2
PRKCA
PRKACA
ROCK2
GPI
DYRK2
PKN1
PIM2
NDRG1
PRKCZ
P2
AURKB
NEK2
PRKACA
PRKCA
MAPKAPK3
SGK3
PKN1
NUAK1
PRKCZ
PRKCD
GSK3B
PRKCE
RPS6KB2
RPS6KA3
RPS6KA1
P3
AURKB
PRKCZ
PRKAA2
NDRG1
LATS1
PRKD2
PRKCA
GPI
CHEK1
ROCK1
NUAK1
MAPKAPK2
PRKCD
Node fill color: Z-score
Kinase-substrate connection


A
Cohort 1
Count
Histone modification
Chromatin remodeling
DNA methylation
Total 61%
NCOR1 35%
* BRD4 12%
CARM1 12%
KMT2D 10%
* NSD1 7%
JAK2 6%
CREBBP 6%
SUZ12 6%
ASXL1 4%
ASXL2 4%
DOT1L 4%
EP300 4%
KDM5A 4%
EED 3%
EZH2 3%
EZH1 3%
MEN1 3%
KDM6A 1%
NCOA3 1%
SETD2 1%
KMT2A 1%
KMT2C 1%
Total 33%
* SMARCA4 13%
DAXX 7%
ATRX 6%
PBRM1 4%
SMARCB1 4%
ARID1A 3%
SMARCD1 1%
ARID1B 1%
ARID2 1%
CTCF 1%
Total 17%
DNMT1 9%
TET1 4%
IDH2 4%
DNMT3A 3%
DNMT3B 3%
TET2 1%
IDH1 1%
Subtype
Subtype
P1
P2
P3
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
0.484
0.484
0.484
NCOR1 protein expression
P1 (n=16)
P2 (n=18)
P3 (n=38)
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 (n=36) High (n=36)
p = 0.18
p = 0.48
p = 0.18
p = 0.13
p = 0.14
Protein expression
ATP2B4
ACTG2
ACTA2
CNN1
DES
NCOR1
Corr.
G
Cohort 1
NCOR1 protein exp. Low (n=36) High (n=36)
p = 0.015
GOBP_STEM_CELL_DIFFERENTIATION
Low
High
H
Cohort 1
NCOR1 protein exp. Low (n=36) High (n=36)
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 (n=31) High (n=31)
p = 0.6
YAP1 H-score
Low
High
Cohort 2
NCOR1 IHC score Low (n=111) High (n=110)
p = 0.0017
YAP1 H-score
Low
High
J
Cohort 1
NCOR1 protein exp. Low (n=32) High (n=31)
(%)
p = 0.12
TEAD IHC positivity
Low
High
Cohort 2
NCOR1 IHC score Low (n=111) High (n=110)
(%)
p = 3.8e−14
TEAD IHC positivity
Low
High


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 (n=14)
P2 (n=18)
P3 (n=33)
C
Cohort 1
HRD score
0.031
0.108
0.108
P1 (n=14)
P2 (n=18)
P3 (n=33)
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 (n=14)
P2 (n=11)
P3 (n=32)
F
Cohort 1
Wilcoxon, p = 0.88
P2 paired-ample HRD score
Primary (n=4)
Metastasis (n=4)
Wilcoxon, p = 0.035
P3 paired-sample HRD score
Primary (n=6)
Metastasis (n=6)
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 (n=42)
Others (n=23)
BRCA2
I
Cohort 1
Gene status LOH No LOH
mRNA
Protein
J
Cohort 1
Proportion of samples retaining BRCA2 protein
0.316
0.334
0.558
28.6%(4/14)
0%(0/11)
12.5%(4/32)
P1
P2
P3
K
Cohort 1
Wilcoxon, p = 0.4
RAD50 H-score
HRD low (n=29)
HRD high (n=28)
0.687
0.54
0.649
P1 (n=14)
P2 (n=11)
P3 (n=32)
L
Cohort 1
Phosphorylated H2AX IHC positivity
0.467
0.238
0.467
P1 (n=14)
P2 (n=11)
P3 (n=32)
M
Cohort 1
Subtype P1 (n=16) P2 (n=18) P3 (n=38)
Protein expression
0.281
0.051
0.015
0.959
0.008
0.008
XRCC5
XRCC6
N
Cohort 1
HRD Low (n=32) High (n=32)
p = 0.0073
Protein expression
PARP1
Subtype P1 (n=16) P2 (n=18) P3 (n=38)
0.064
0.000915
0.105
PARP1
O
Cohorts 1+2
Overall survival
RAD51 H-score High Low
p = 0.053
Recurrence-free survival
RAD51 H-score High Low
p = 0.0029
Overall survival
BRCA2 IHC positivity High Low
p = 0.00057
Recurrence-free survival
BRCA2 IHC positivity High Low
p = 0.0014
Survival probability
Time (months)
Number at risk
High 86 73 51 30 17
Low 86 62 43 28 16
High 86 48 29 16 11
Low 86 32 22 18 11
High 86 74 56 39 23
Low 86 61 38 19 10
High 86 51 33 24 15
Low 86 29 18 10 7


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 (n=16)
P2 (n=18)
P3 (n=38)
C
Cohort 1
Wilcoxon, p = 0.06
ImmuneScore
Primary (n=18)
Metastasis (n=18)
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=2.688E-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 (n=36)
Immune-hot (n=36)
H
Cohort 1
Corr.
ImmuneScore
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
IFNgamma_response
I
Cohort 1
Immune status
Cold (n=14)
Hot (n=13)
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
0.458
0.48
0.351
LGALS9 IHC H-score
P1 (n=15)
P2 (n=11)
P3 (n=36)
LGALS9 IHC
50 μm
K
Cohort 1
0.574
0.019
0.03
CD68 IHC count
0.237
0.028
0.028
CD163 IHC count
P1 (n=15)
P2 (n=11)
P3 (n=36)
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
CD163 IHC
p = 0.0017
High 86 32 24 17 10
Low 86 48 27 17 12

A

Workflow for development and validation of the TMA-based IHC subtype classifier

Marker selection

IHC markers from Fig. 1-7
Markers with characteristic subtype patterns identified in the main IHC analyses

+

Additional subtype-specific markers
Five markers selected from one-versus-others comparison

↓

Candidate IHC marker set (n=23)
Combined feature pool for classifier development

Classifier development

Cohort 1 TMA IHC data
Quantified IHC scores for candidate markers

↓

Data preprocessing

↓

Model construction with 10-fold CV
Testing all 3-, 4-, 5-, and 6-marker combinations from 23 IHC candidates

↓

Model selection
Choose optimal marker set and parameters

Final model and assignment rule

Final model
Multinominal logistic regression model
Markers: CD133, MYC, CDK2, AURKA, AURKB, HLA-ABC

↓

Assign subtype if top probability > subtype threshold

Yes → Classified case
Label as specific subtype

No → Low-confidence case
Label as unclassified

Validation

Parameter-locked final IHC-based classifier

↓

Validation cohort
Cohort 2 TMA IHC data

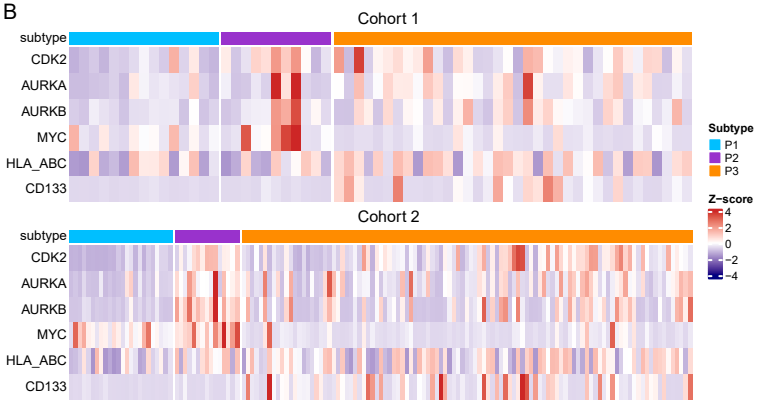
↓

Subtype assignment

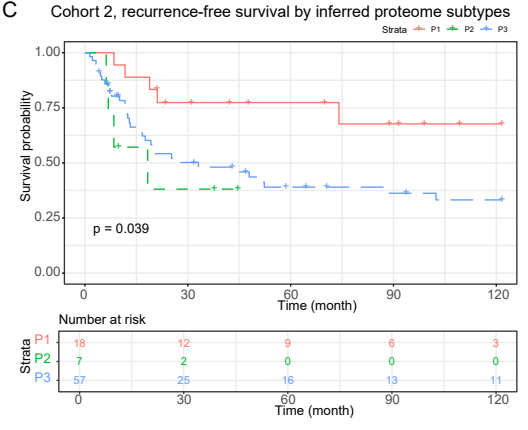
↓

Downstream evaluation
Kaplan-Meier analysis

B



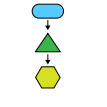
C

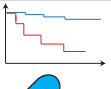


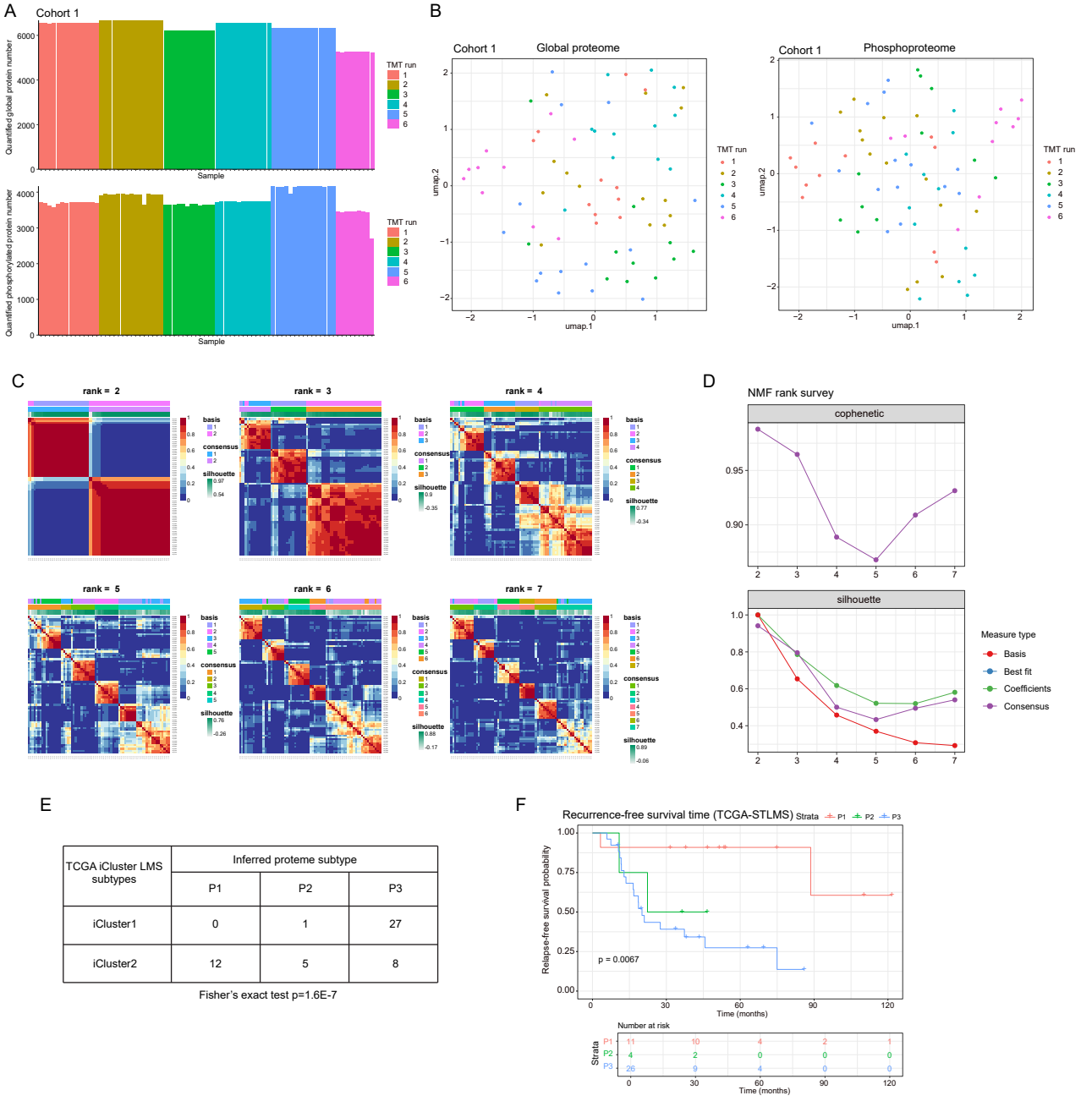
D

Proteome subtypes of STLMS

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







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.

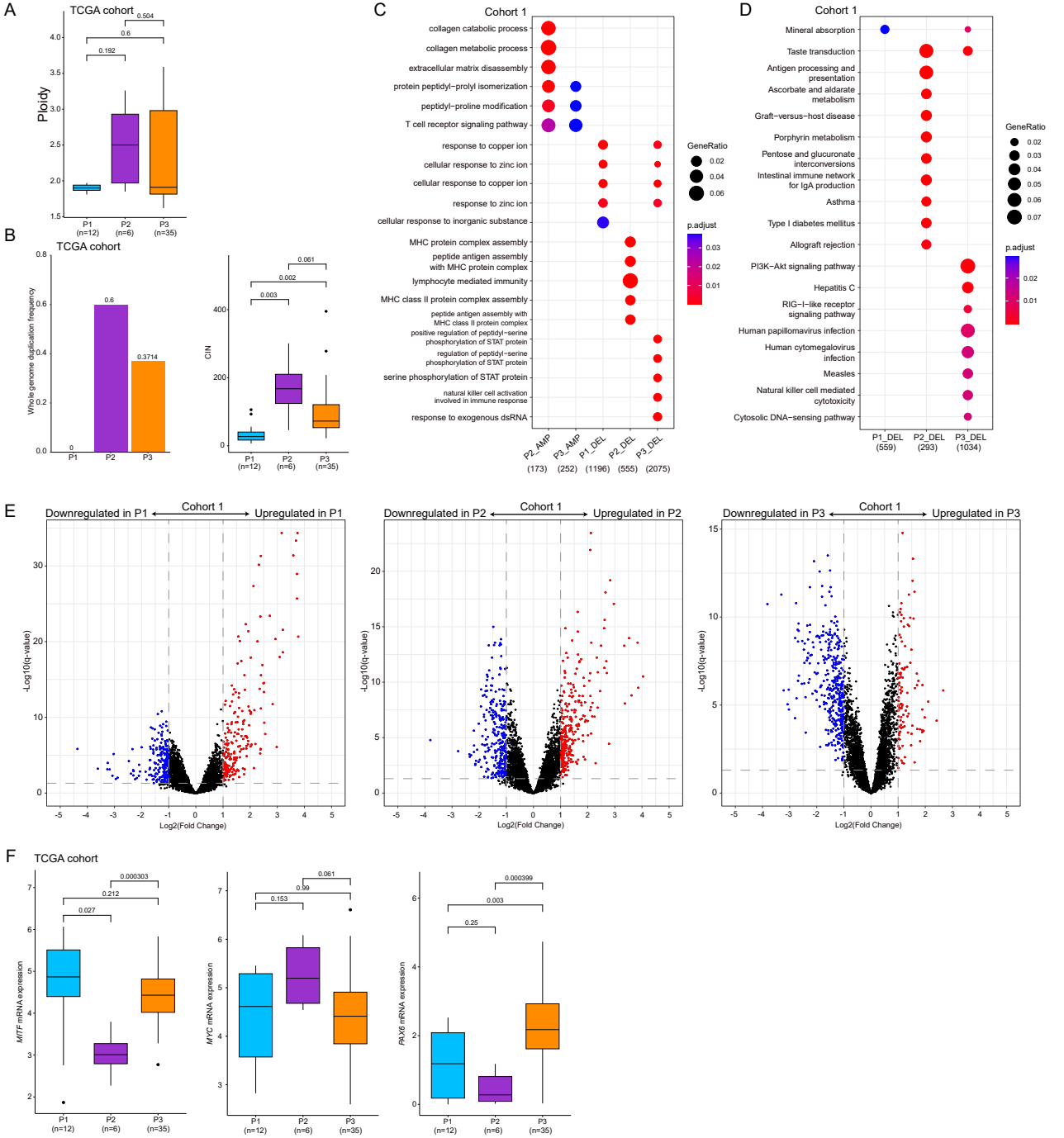


Figure S2. Genomic instability features and regulator profiles in the TCGA STLMS cohort and differential expression analyses of cohort 1, related to Figure 2

(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.

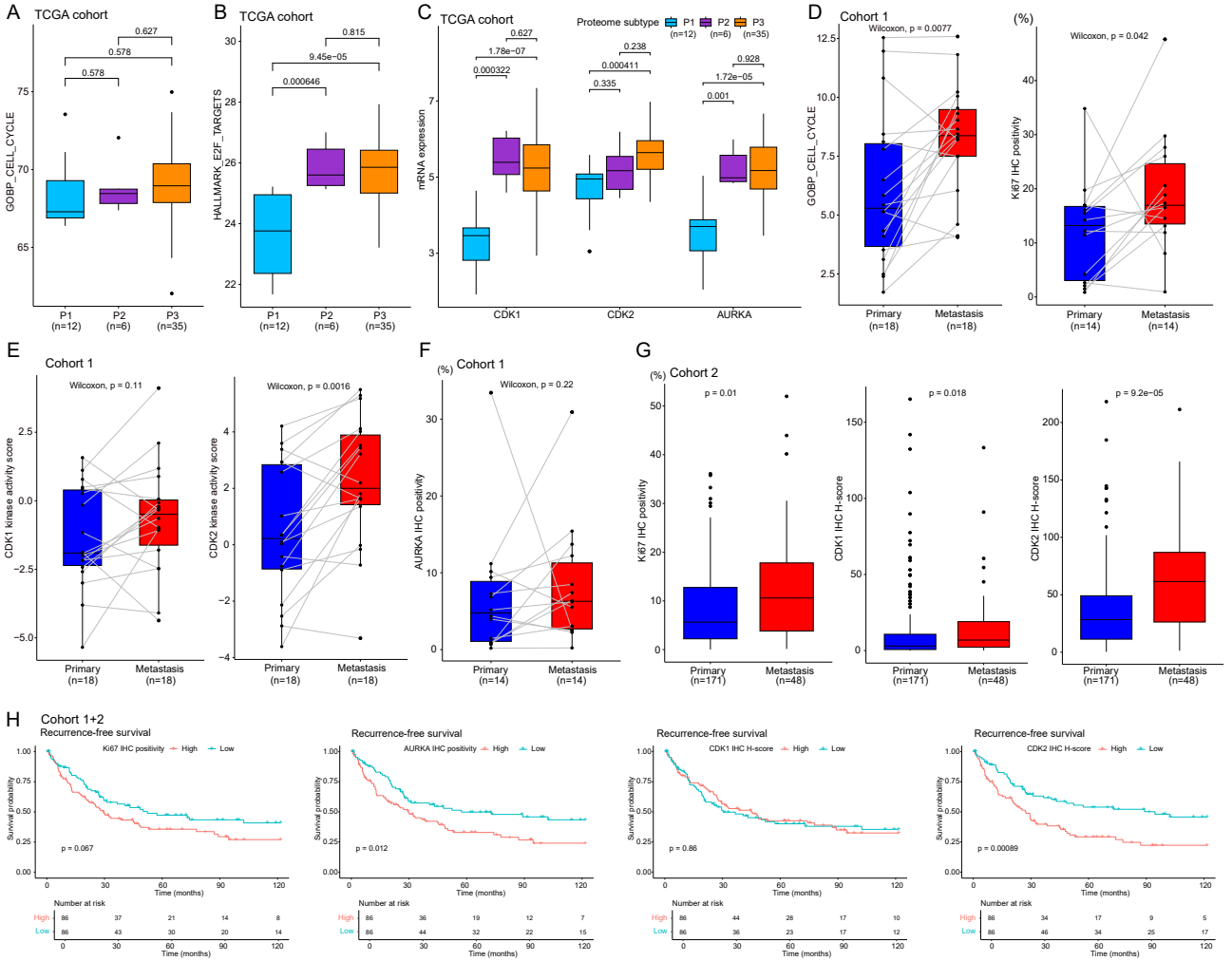


Figure S3. Cell cycle feature differences in the TCGA STLMS transcriptome data and comparison of metastatic samples in cohort 1 against matched primary samples, related to Figure 3

(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).

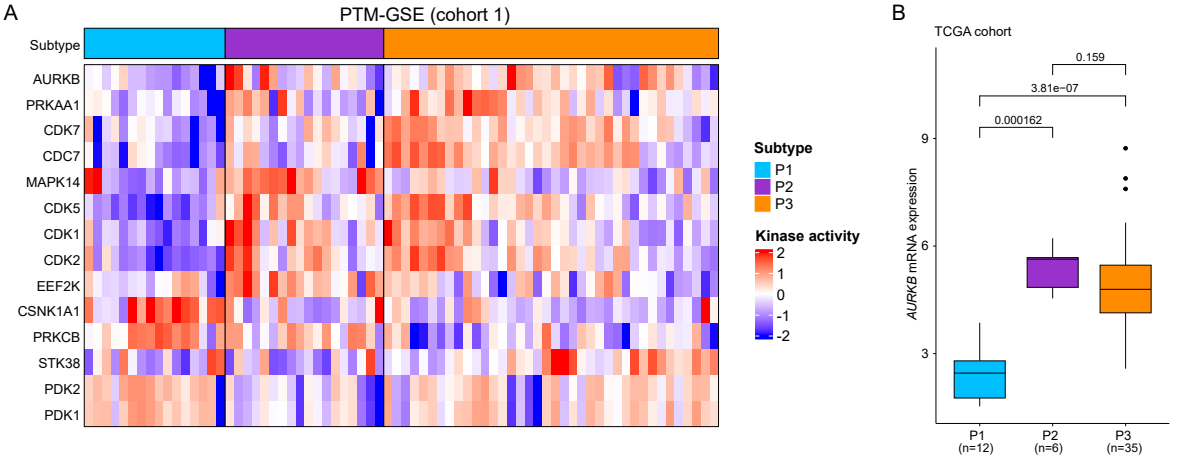


Figure S4. Phosphorylation landscape of cohort 1 and validation of findings in the independent cohort 2 and the external TCGA STLMS cohort, related to Figure 4
(A) A heatmap of PTM-GSE scores. Terms with significant differences between subtypes (ANOVA q<0.01) are shown. AURKB activity is high in P2/P3,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

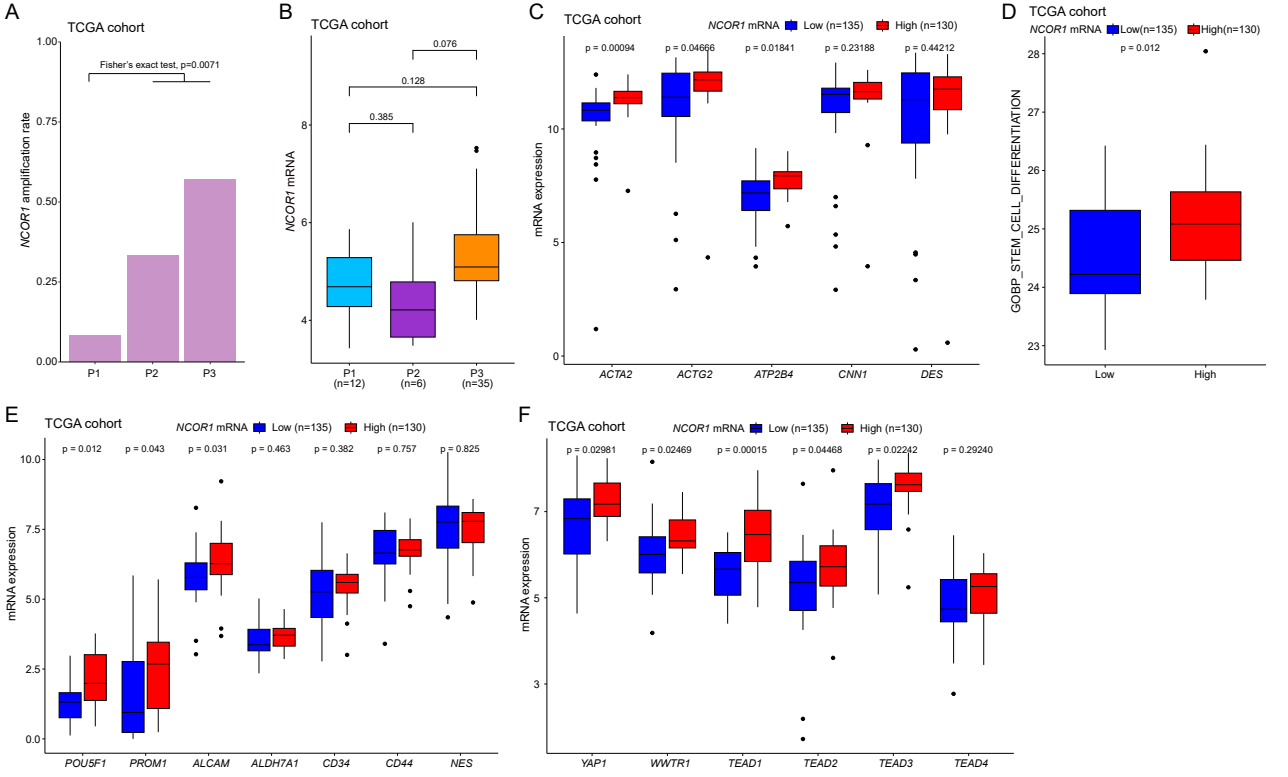


Figure S5. NCOR1 characteristics in the external TCGA STLMS dataset, related to Figure 5

(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.

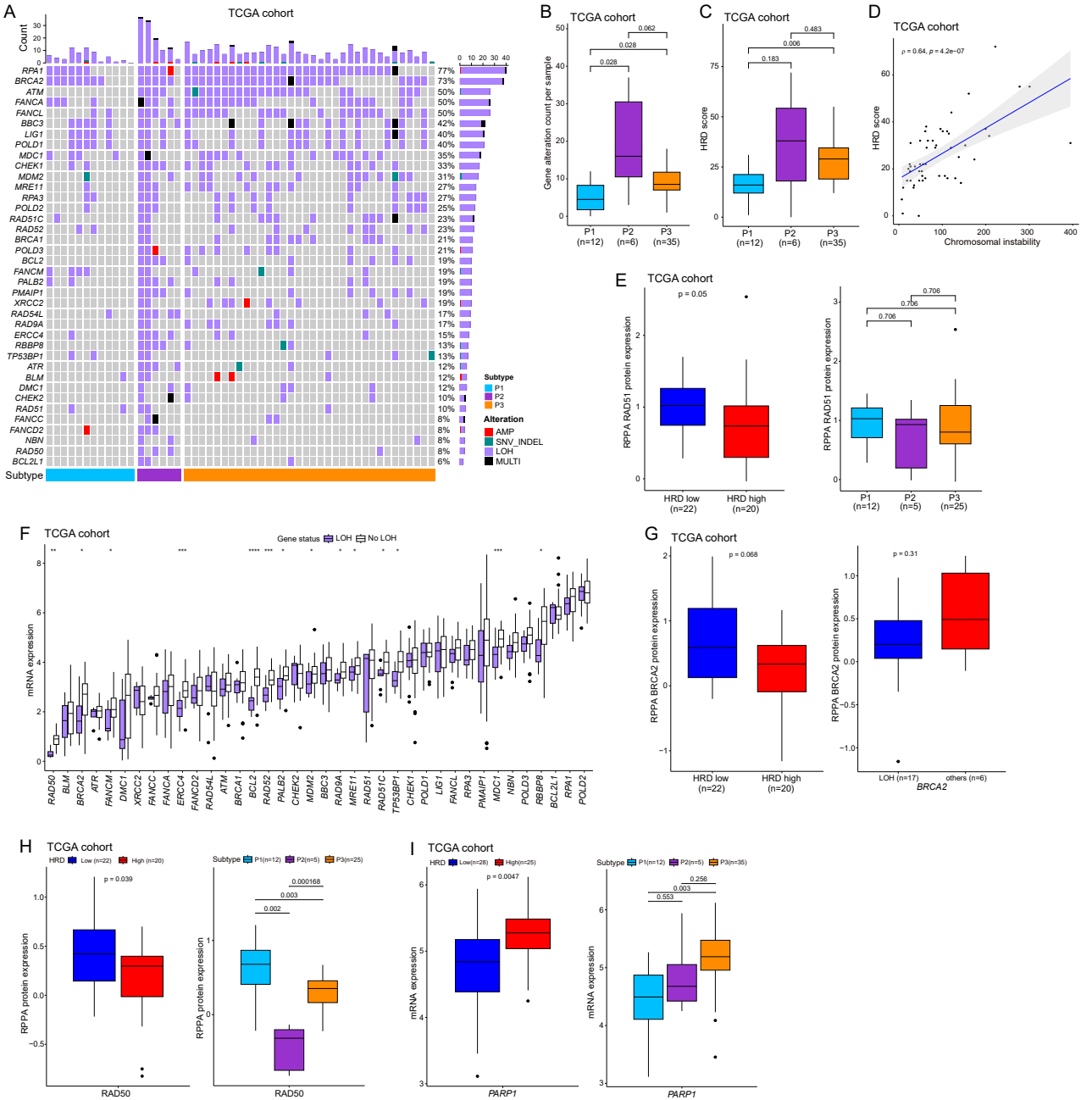


Figure S6. HRR pathway in validation cohort 2 and the external TCGA cohort, related to Figure 6

(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.

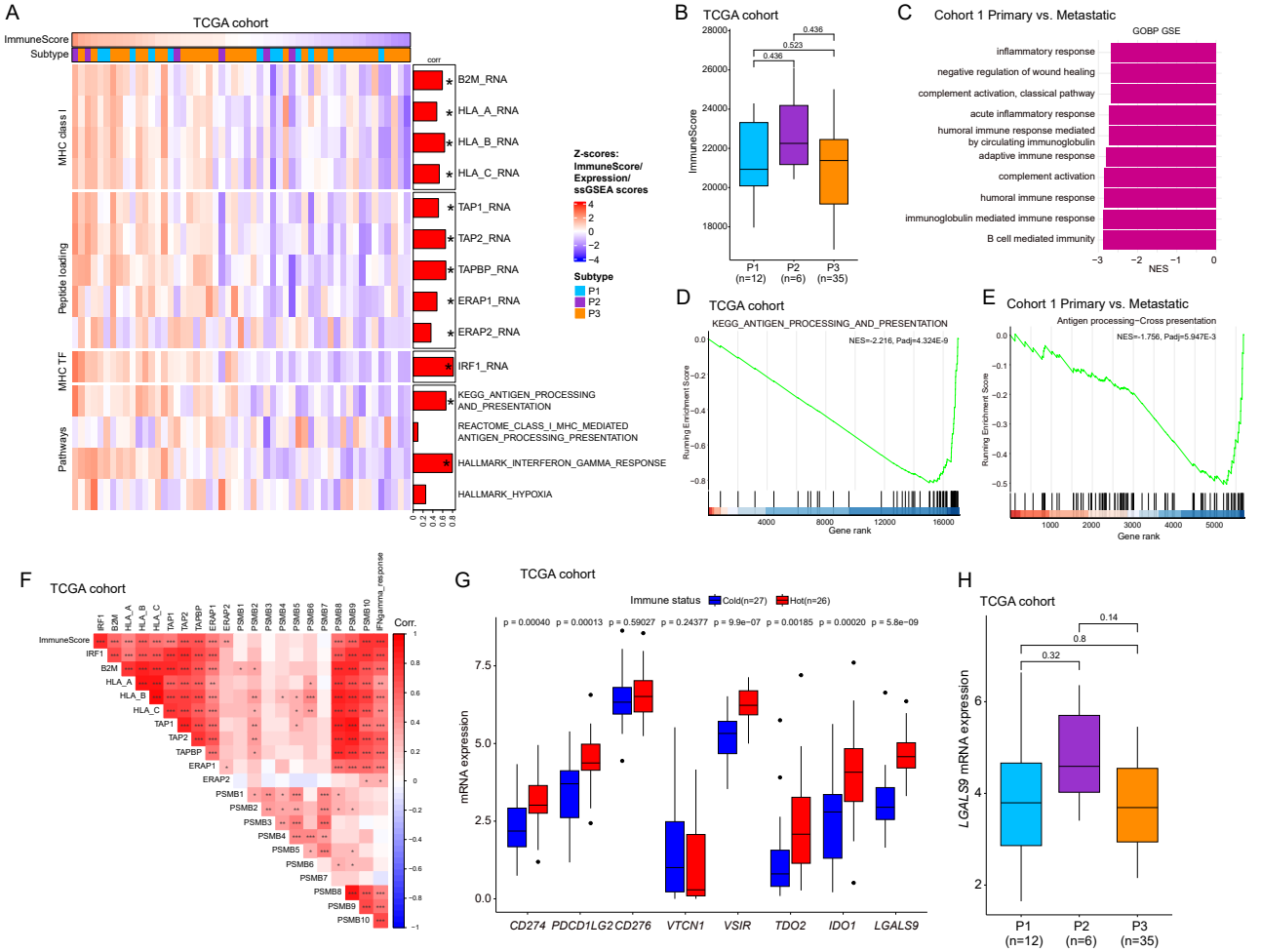


Figure S7. Tumor microenvironment analysis of the external TCGA STLMS dataset and comparison of primary vs. metastatic lesions, related to Figure 7

(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.