Multi-omics evaluation of cell lines as models for metastatic prostate cancer

Xueying Liu[1], Yugang Wang[2,*], Ke Liu[1,*]

Affiliations:

1. Department of Medical Dataology, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan, China
2. Department of Biochemistry and Molecular Biology, School of Basic Medicine, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

* To whom correspondence should be addressed

Abstract

Cancer cell lines are indispensable for prostate cancer research, yet their ability to model metastatic prostate cancer remains insufficiently characterized. In this study, we performed a systematic multi-omics evaluation integrating genomic and transcriptomic data from large public datasets. We identified VCaP as the most representative model for metastatic prostate adenocarcinoma, while highlighting the limited fidelity of the widely used PC3 cell line. Our analysis underscores the importance of selecting hypermutated cell lines, which display distinct tumor microenvironment characteristics. Moreover, overexpression of *MET* in VCaP induced partial stem-like traits but did not fully recapitulate the MSPC transcriptomic profile. Notably, stem-like prostate organoids demonstrated higher similarity to MSPC and may serve as more suitable MSPC models. Our study provides guidance for the selection of prostate cancer cell lines, assisting researchers in choosing more representative metastatic prostate cancer cell line models.

Introduction

Cancer cell lines are indispensable preclinical models in oncology research. Their immortality, cost-effectiveness, and capacity for long-term cultivation and cloning make them vital for reproducible and scalable cancer research and drug development[1, 2, 3, 4, 5]. However, prolonged *in vitro* culture can lead to the accumulation of additional genetic and epigenetic alterations, resulting in substantial discrepancies compared to patient-derived tumor samples[6, 7, 8, 9]. Therefore, systematic evaluation of the molecular fidelity of cancer cell lines relative to patient-derived cancer cells is essential to ensure the translational relevance of cell line–based studies.

The widespread availability of large-scale open cancer genomics data has enabled researchers to comprehensively assess the suitability of cell lines as models in cancer research. For example, Domcke et al. compared genomic profiles between ovarian cancer cell lines and patient tumor samples and found several commonly used cell lines poorly recapitulated high-grade serous ovarian cancer[10]; Yu et al. performed a pan-cancer transcriptomic comparison between cell lines and patient tumor samples and proposed the cell line panel TCGA-110-CL[11]; Jin et al. conducted similar pan-cancer analysis with their own computational approaches to select appropriate cell line models for different cancer types; We previously performed multi-omics comparison between breast cancer cell lines and metastatic tumor samples and identified cell lines that faithfully represent distinct breast cancer subtypes[12]. These

studies have deepened our understanding of cancer cell lines and highlighted the importance of selecting appropriate models based on molecular fidelity.

Prostate cancer is among the most common and lethal malignancies affecting men's health worldwide, posing a substantial public health challenge[13, 14]. Similar to other cancer types, cell lines have been widely utilized to explore the underlying biological mechanisms of prostate cancer. Although several studies have evaluated the molecular representativeness of prostate cancer cell lines relative to tumor samples[11, 15, 16], three major limitations persist. First, metastasis is the leading cause of prostate cancer–related mortality[13, 17]; however, most of the evaluation studies mainly focus on primary tumors. Second, existed studies were performed in a pan-cancer manner, which applies uniform analytical criteria across various cancer types and overlooks tumor-specific characteristics[3]. Third, the majority of analyses rely exclusively on transcriptomic data, with limited integration of multi-omics information for a more comprehensive evaluation.

In this study, we leveraged the publicly available bulk and single-cell omics data to comprehensively evaluate the suitability of prostate cancer cell lines as models for metastatic prostate cancer. We proposed principles of cell line selection for different research scenarios and meanwhile point the strength and limitation of widely used established cell line. Our study provides a framework for selecting prostate cancer cell

lines tailored to specific biological contexts and highlight the need to develop more representative models for understudied subtypes.

Results

Systematic comparison of genomic profiles

We first compared the somatic mutation profiles between metastatic prostate cancer samples and prostate cancer cell lines, focusing on genes that were either highly mutated in metastatic prostate cancer or exhibited differential mutation rates between metastatic and primary prostate cancer. The somatic mutation data of primary prostate cancer, metastatic prostate cancer, and prostate cancer cell lines were downloaded from the Cancer Genome Atlas (TCGA), Stand Up to Cancer (SU2C), and the Cancer Cell Line Encyclopedia (CCLE), respectively. A total of 68 highly mutated genes were identified using SU2C samples, along with 13 differentially mutated genes between SU2C and TCGA samples (**Supplementary. Fig. 1a-c**). The integration of these two lists identified 73 unique genes (**Fig. 1a**). The median mutation rate of the 73 genes across the CCLE prostate cancer cell lines was 0.3, suggesting most of them could be recapitulated by only a few cell lines. *MUC16*, *SYNE2*, *SYNE1*, *OBSCN* have a mutation rate higher than 50%, with *MUC16* being the most frequently mutated gene (mutation rate = 0.62). *TP53*, a key tumor suppressor gene, displayed a mutation rate of 0.38. Surprisingly, there were four genes (*APC*, *SPOP*, *CDK12*,

106 and *FAM186A)* which were entirely unmutated in all analyzed prostate cell lines,
107 highlighting the limitation of using the analyzed cell lines in researching the role of
108 them in metastasis-related study.
109
110 Cell lines were ranked based on the number of mutations present in a set of 73 genes
111 associated with metastatic prostate cancer, and found the top four cell lines were
112 LNCaP, DU145, 22RV1, and MDA-PCa-2b, respectively **(Fig 1b)**. Interestingly, we
113 noticed these four cell lines were all inferred with microsatellite instability (MSI)[18]
114 and had significantly higher mutation burden than other cell lines **(Fig. 1c)**. To adjust
115 such bias, we re-ranked the cell lines according to a normalized ratio value (computed
116 as number of mutated genes / mutation burden) and found the top three cell lines were
117 PC3, VCaP, and NCI-H660 (**Fig. 1d**).
118
119 We next investigated whether those recurrent single-nucleotide alterations (which we
120 call "mutation hotspot") identified in metastatic prostate cancer samples were also
121 presented in prostate cancer cell lines. We defined mutation hotspots as
122 non-synonymous mutations occurring in at least three SU2C samples. Based on this
123 criterion, 58 mutation hotspots were identified (**Supplementary Table 1**). These
124 mutation hotspots were distributed in 45 genes, with *TP53* and *AR* being the two
125 genes harboring most mutation hotspots (**Fig. 1e and Supplementary. Fig. 1d)**. *TP53*
126 mutation hotspots were predominantly located within the DNA-binding domain, a

127 functionally critical region required for the transcriptional regulation of tumor
128 suppressor genes[19]. For *AR*, mutation hotspots were primarily clustered within the
129 ligand-binding domain, which is pivotal for androgen receptor-targeted therapies[20].
130
131 Strikingly, only seven mutation hotspots were detected in at least one prostate cancer
132 cell line (**Fig. 1f**), and all were presented exclusively in 22RV1, MDA-PCa-2b, and
133 LNCaP. Of them, three were single nucleotide point mutations of *AR* and the other
134 four were all frameshift mutations. Of the *AR* mutation hotspots, T878A and L702H
135 mutations have been reported to be associated with resistance to AR inhibition
136 therapy[21]. Notably, MDA-PCa-2b harbors both mutations (T878A and L702H),
137 reflecting the clinical phenomenon of co-occurring multiple AR resistance mutations
138 within individual tumors[21]. Among the frameshift mutation hotspots, K16fs mutation
139 in *RPL22* was observed in all three cell lines, whereas the other three frameshift
140 mutations were exclusive to 22RV1.
141
142 Besides somatic mutation profiles, we also compared copy number variation (CNV)
143 profiles between metastatic and primary prostate cancer samples. Among the
144 differential CNV genes between SU2C and TCGA samples, *AR* exhibited the highest
145 level of amplification (**Fig. 1g**). A similar pattern was observed in site-specific
146 analyses of differential CNV genes (**Supplementary. Fig. 1e-g**). This finding may not
147 be surprising since *AR* is a critical driver of prostate cancer development and

148 progression[22, 23, 24]. Notably, VCaP exhibited the highest level of *AR* amplification
149 among all prostate cancer cell lines analyzed (**Fig. 1h**).
150
151 **Hypermutation of prostate cancer cells influenced the tumor microenvironment**
152 Previous studies have demonstrated hypermutated colorectal tumor samples
153 exhibiting increased tumor-infiltrating, elevated neoantigen burden, and potentially
154 enhanced sensitivity to immunotherapy[25, 26]. Motivated by these findings, we aim to
155 assess whether hypermutation should be taken into consideration when selecting cell
156 lines for studying the metastatic mechanisms of prostate cancer.
157
158 To identify hypermutated prostate cancer samples, we integrated samples from TCGA,
159 SU2C, CCLE together and determined a mutation burden threshold above which
160 samples were classified as hypermutated (**Fig. 2a**). Based on our criteria, we
161 identified four hypermutated samples in TCGA, ten in SU2C, and four in CCLE. As
162 expected, the aforementioned four hypermutated cell lines LNCaP, DU145, 22RV1,
163 and MDA-PCa-2b were successfully identified. Notably, three of them harbor *RPL22*
164 frameshift mutations, which is frequently observed in MSI tumors[27, 28]. To further
165 validate our strategy, we computed the mutation frequencies of mismatch repair
166 (MMR) genes[29] and found hypermutated samples exhibiting a significantly higher
167 mutation frequency (**Fig. 2b and Supplementary Fig. 2a**). These results
168 demonstrated the validity of the determined threshold value for identifying

169 hypermutated prostate cancer samples.

170

171 We next performed differential gene expression analysis between hypermutated and
172 non-hypermutated TCGA samples and identified 125 upregulated and 85
173 downregulated genes, respectively (**Supplementary Table 2**). Notably, the T cell
174 marker *CD8A*, the immune checkpoint molecule *CTLA-4*, and the cytotoxic effector
175 *GZMA* were significantly upregulated in hypermutated samples (**Fig. 2c and**
176 **Supplementary. Fig. 2b**), suggesting the increased CD8+ T cell infiltration level; in
177 contrast, the gamma-delta T cell marker *TRGV9* was downregulated, suggesting a
178 potential reduction in gamma-delta T cell infiltration in hypermutated tumors. We
179 performed Gene Ontology (GO) enrichment analysis on upregulated genes and found
180 pathways associated with T cell activation (e.g., the T cell receptor signaling pathway,
181 antigen receptor-mediated signaling pathway) and natural killer (NK) cell-mediated
182 cytotoxicity (e.g., NK cell-mediated immunity, NK cell-mediated cytotoxicity) were
183 significantly enriched (**Fig. 2d and Supplementary Table 3**). The tumor
184 microenvironment (TME) has been demonstrated to play pivotal roles in the distant
185 metastasis of primary cancer[30, 31, 32] and our analysis suggests notable TME
186 differences between hypermutated and non-hypermutated primary prostate cancer
187 samples. Therefore, we propose that hypermutated cell lines (such as LNCaP) should
188 be more appropriate for investigating the metastatic mechanisms of hypermutated
189 prostate cancer. To the best of our knowledge, such consideration has not been

190 addressed in previous cell line evaluation studies.

191

192 **Correlating cell lines with metastatic prostate cancer using bulk and single-cell**

193 **RNA-seq data**

194 Transcriptome correlation analysis (TC analysis) has been proven to be an effective

195 method for assessing the utility of cell lines[11, 12, 33]. Consequently, we ranked all 1,019

196 CCLE cell lines based on their transcriptomic similarity to MET500 prostate cancer

197 samples and the top three cell lines were LNCaP, VCaP, and MDA-PCa-2b,

198 respectively (**Fig. 3a**). We next examined whether the metastatic site influenced TC

199 analysis results and found that cell line rankings were highly consistent among

200 various metastatic sites (bone, liver, and lymph node) (**Fig. 3b and Supplementary.**

201 **Fig. 3a-c**).

202

203 To double confirm the analysis results derived from bulk RNA-seq data, we further

204 ranked 1,019 CCLE cell lines according to their transcriptomic similarity with the

205 malignant cells from a scRNA-seq dataset of metastatic prostate cancer[22]. Consistent

206 with previous results, the top three cell lines with the highest similarity were VCaP,

207 LNCaP, and MDA-PCa-2b (**Fig. 3c**). We further performed metastatic-site-specific

208 transcriptomic similarity analysis (bone and lymph node) and the results were highly

209 correlated (**Fig. 3d and Supplementary Fig. 3d-e**).

210

211 **Strength and limitation of the PC3 cell line**

212 PC3 is a widely used cell line in studies investigating metastatic prostate cancer[34, 35].
213 We conducted a PubMed search of studies related to metastatic prostate cancer and
214 found PC3 was the most frequently cited cell line, followed by LNCaP and DU145
215 (**Supplementary Table 4 and Supplementary. Fig. 4a**). However, despite its
216 widespread use, transcriptomic correlation analysis revealed that PC3 shows limited
217 similarity to metastatic prostate cancer samples (ranked 444 among 1,019 CCLE cell
218 lines) (**Fig .3a**). Prior to our research, several studies had revealed the heterogeneity
219 of prostate cancer[36, 37, 38]. Considering that the metastatic prostate cancer samples used
220 in the transcriptome correlation analysis mainly exhibited adenocarcinoma histology
221 (in which malignant cells primarily show luminal differentiation)[39, 40, 41], we further
222 investigated whether PC3 could serve as a representative model for other types of
223 prostate cancer.

224

225 We performed principal component analysis (PCA) on CCLE prostate cancer cell
226 lines and found they could be briefly classified into three clusters (**Fig. 4a**). Cluster 1
227 consisted of luminal lineage cell lines (such as VCaP), as indicated by expression of
228 luminal lineage markers such as *AR* and *KLK4*; cluster 2 only included NCI-H660, a
229 small cell carcinoma cell line of neuroendocrine origin expressing *SYP*; cluster 3
230 comprised DU145 and PC3. Interestingly, our PCA results align with Han et al 's
231 research, which classified prostate cancer into three intrinsic subtypes: androgen

232 receptor positive prostate cancer (ARPC), neuroendocrine prostate cancer (NEPC),
233 and mesenchymal and stem-like prostate cancer (MSPC)[42]. Since PC3 and DU145
234 were separated with other cell lines on the first principal component, we performed
235 gene set enrichment analysis on the top 200 genes with the highest positive loadings
236 along the PC1 axis using MSigDB Hallmark gene sets (**Fig. 4b**). The
237 "Epithelial-Mesenchymal Transition" gene set exhibited the highest level of
238 enrichment; in addition, the enrichment of "*IL-2/STAT5* Signaling" and
239 "*IL-6/JAK/STAT3* Signaling" gene sets suggests activation of the *JAK/STAT* pathway,
240 a known driver of cellular plasticity[41, 43]. Furthermore, PC3 exhibited relatively higher
241 expression of *CD44* (a marker of stemness), along with lower expression of the
242 luminal marker *AR* and the neuroendocrine marker *SYP* (**Fig. 4c-e**). We double
243 confirmed the expression of these marker genes using a scRNA-seq dataset of PC3
244 (**Supplementary. Fig. 4b**). Collectively, these results support the classification of
245 PC3 as a cell line of the MSPC subtype.
246
247 We next evaluated whether PC3 could adequately recapitulate the transcriptomic
248 profile of MSPC malignant cells. MSPC has been reported to be enriched in
249 castration-resistant prostate cancer (CRPC) patients[42]. To this end, we analyzed a
250 CRPC scRNA-seq dataset and found the malignant cells could be classified into
251 ARPC, NEPC, and MSPC subtypes based on the expression of *AR*, *SYP*, and *CD44*,
252 respectively (**Fig. 4f and Supplementary. Fig. 4c)**. We then assembled pseudo-bulk

253 transcriptomic profiles for each subtype and ranked 1,019 CCLE cell lines according
254 to transcriptomic similarity in a subtype-specific manner. As expected, MDA-PCa-2b,
255 VCaP, and LNCaP showed the highest similarity to ARPC, while NCI-H660 exhibited
256 a high similarity to NEPC (**Supplementary Fig. 4d-e**). Surprisingly, for the MSPC
257 subtype, although PC3 ranked highest among prostate cancer cell lines, its overall
258 similarity to MSPC malignant cells remained modest (ranked 260 among 1,019 CCLE
259 cell lines) (**Fig. 4g**). Moreover, the transcriptomic similarity between individual
260 MSPC malignant cells and PC3 was significantly lower than that between ARPC
261 malignant cells and VCaP, suggesting limited fidelity of PC3 as a model of MSPC
262 (**Fig. 4h**).
263
264 Interestingly, we found MSPC malignant cells expressed basal lineage markers such
265 as *CAV2* and *TP63* (**Supplementary Fig. 4f**), suggesting their basal differentiation.
266 Consistent with this, when compared with three normal prostate epithelial subtypes
267 (luminal, basal and neuroendocrine), they showed the highest similarity to basal
268 epithelial cells (**Fig. 4i and Supplementary. Fig. 4g**). However, PC3 lacked
269 expression of key basal markers (such as *TP63*), which may explain its low
270 transcriptomic similarity with MSPC malignant cells **(Supplementary Fig. 4h).**
271
272 Taken together, our findings highlight both the strength and limitation of PC3 cell
273 line.

274 PC3 exhibits molecular features consistent with stemness, epithelial–mesenchymal
275 transition, and cellular plasticity, and among currently available prostate cancer cell
276 lines, it most closely resembles the MSPC subtype. However, its overall
277 transcriptomic similarity to primary MSPC malignant cells remains limited. Further
278 studies are warranted to identify or develop more representative cell line models that
279 faithfully recapitulate the molecular characteristics of MSPC prostate cancer.
280

281 **Patient-derived organoids better resemble the transcriptome of MSPC than PC3**

282 Since PC3 does not fully recapitulate the transcriptomic features of MSPC, we next
283 investigated other type of models that could be more representative. We focus on two
284 categories of models: engineered prostate cancer cell lines and patient-derived
285 organoids.
286

287 Previous studies have reported that AR inhibition by enzalutamide or *TP53/RB1*
288 double-knockout can promote LNCaP (a hypermutated cell line of ARPC) into MSPC
289 state[42, 43]. Therefore, we included these two engineered LNCaP-derived models in our
290 analysis. In addition, we generated a VCaP-derived cell line by overexpressing *MET*
291 (VCaP-*MET*-oe), as *MET* has been reported to promote EMT and sustain stemness in
292 cancer cells[44, 45, 46]. As expected, bulk RNA-seq analysis suggested that *MET*
293 overexpression induced upregulation of *CD44* (log2fc = 0.251, p-value = 0.028) and
294 increased *JAK/STAT* pathway activity, while downregulating luminal markers *KLK2*,

295 *KLK3*, and *KLK4*, suggesting the occurrence of de-differentiation (**Fig. 5a-b and**
296 **Supplementary Table 5**).
297
298 We next computed the transcriptomic similarity with MSPC malignant cells for the
299 above three engineered cell lines, respectively (**Fig. 5c-e**). Surprisingly, all of them
300 exhibited lower similarity to MSPC than PC3. Actually, when comparing the
301 transcriptomic profiles of these engineered cell lines to the 1,019 CCLE cell lines, the
302 highest correlations were consistently observed with their respective parental cell
303 lines (**Supplementary Fig. 5a–c)**, suggesting that these manipulations did not fully
304 trigger a lineage transition. Given that MSPC malignant cells exhibit basal
305 differentiation, this likely explains why the engineered cell lines failed to recapitulate
306 the MSPC transcriptional landscape.
307
308 We next explored whether patient-derived organoids could better recapitulate the
309 MSPC transcriptome. We computed the transcriptomic similarity between ten PDOs
310 (all of which were claimed as MSPC according to Tang's study)[47] and MSPC
311 malignant cells (**Fig. 5f**). Notably, we found six PDOs had higher similarity to MSPC
312 malignant cells than PC3; of them, MSKPCa12 demonstrated the highest correlation
313 and appeared to be the most representative (**Supplementary. Fig**. **5d**). Further lineage
314 analysis suggested that this organoid exhibits a basal lineage (**Supplementary Fig.**
315 **5e**), with expression of several basal specific markers such as *TP63*, *CAV2*, and *KRT5*

316 (**Supplementary. Fig. 5f-h**).

317

318 These efforts to engineer existing prostate cancer cell lines motivated us to explore
319 alternative models. In recent years, patient-derived organoids have gained increasing
320 attention as promising platforms for modeling tumor heterogeneity and
321 subtype-specific biology. Their ability to maintain key molecular and
322 histopathological features of the original tumors underscores their potential value in
323 studies of MSPC biology and therapeutic response.

324

325

326

327

328

329

330

331

332

333

334

335

336

Discussion

Cell lines are widely utilized in cancer research, particularly as models for metastatic cancers. Prior studies have demonstrated that cell lines derived from the same tumor type exhibit considerable molecular heterogeneity, with certain cell lines more accurately recapitulating the characteristics of metastatic tumors than others[10, 33, 48, 49]. This underscores the critical importance of selecting cell line models that most faithfully represent the biology of metastatic cancers. Our study proposed useful guidance for cell line selection in the study of metastatic prostate cancer.

Adenocarcinoma is the predominant histological subtype of prostate cancer[50, 51]. Based on our multi-omics analysis results, VCaP appears to be the most suitable model for studying metastatic prostate cancer according to its relatively higher concordance with metastatic prostate adenocarcinoma patient samples in multiple aspects (for example, the high transcriptomic similarity with patient samples, and the highest level of *AR* amplification). However, according to Pubmed search results, VCaP is rarely used in metastasis-related research, suggesting a gap between model suitability and actual utilization in the field.

Although VCaP represents a suitable model for metastatic prostate adenocarcinoma, it may not be appropriate for all contexts. One particular situation is hypermutated patient samples. Our analysis suggests that the microenvironment in hypermutated

358 prostate cancer samples differs markedly from that in non-hypermutated tumors,
359 particularly with respect to increased immune infiltration. This indicates that the
360 hypermutation phenomenon should be considered when selecting cell line models.
361 Given its high mutation burden and strong transcriptomic similarity to metastatic
362 prostate adenocarcinoma samples, LNCaP was identified as a promising model for
363 studying hypermutated prostate cancer.
364
365 Another situation is to research AR resistance. We found that the prostate cancer cell
366 lines LNCaP and MDA-PCa-2b harbor *AR* mutation hotspots associated with clinical
367 resistance, showing a certain degree of concordance with patient observations.
368 Specifically, MDA-PCa-2b carries both the T878A and L702H mutations, which are
369 linked to resistance to flutamide and glucocorticoids, respectively. The T878A
370 mutation was also detected in LNCaP, while 22RV1 harbors another recurrent *AR*
371 mutation hotspot, H875Y. Notably, despite harboring these resistance-associated
372 mutations, all three cell lines remain AR-dependent, suggesting that they are valuable
373 *in vitro* models for studying AR-driven resistance mechanisms and evaluating novel
374 antiandrogen therapies.
375
376 According to our PubMed search results, PC3 is the most widely used prostate cancer
377 cell line; however, it exhibits low transcriptomic similarity to metastatic prostate
378 adenocarcinoma. Although further analysis classified PC3 within the MSPC subtype,

379 it does not faithfully model MSPC at the transcriptomic level, either. This discrepancy
380 may arise from its lack of basal lineage marker, whereas MSPC malignant cells
381 typically show high expression of basal markers and share transcriptomic features
382 with normal prostate basal cells. Surprisingly, several bladder cancer cell lines (such
383 as RT4 and KMBC2, derived from urinary tract) display greater transcriptomic
384 resemblance to MSPC malignant cells. While this observation may reflect random
385 variation, we favor the interpretation of a true underlying biological similarity, as
386 urinary tract–derived cell lines consistently exhibit higher basal marker expression
387 than PC3. Collectively, these findings suggest that alternative cell lines may better
388 recapitulate the molecular characteristics of the MSPC subtype and serve as more
389 effective models for elucidating its underlying biology.
390
391 Mixed subtypes have been reported in prostate cancer patients[41, 52, 53, 54]. These
392 findings underscore the importance of employing multiple cell lines simultaneously
393 when investigating mixed subtype phenomena, in order to more accurately capture
394 and model intratumoral heterogeneity. According to scRNA-seq analysis of metastatic
395 prostate cancer samples, the majority of malignant cells exhibited an ARPC
396 phenotype, whereas MSPC cells were relatively rare. This observation raises
397 important biological questions regarding the role of MSPC malignant cells in
398 metastasis. Specifically, it remains unclear whether MSPC cells facilitate the
399 dissemination of ARPC cells, or whether ARPC and MSPC represent independently

metastasizing malignant populations. Further studies are needed to elucidate the functional interplay between these subtypes and their respective contributions to metastatic progression.

MSPC phenotype remains insufficiently characterized, and existing prostate cancer cell lines do not reliably model this subtype. In this study, we attempted to induce an MSPC phenotype in ARPC-derived cell lines. However, this strategy resulted only in the partial acquisition of stem-like features. Notably, certain prostate organoids with intrinsic stem-like properties appear to represent more appropriate models for recapitulating the MSPC subtype. Future research should focus on identifying or engineering prostate cancer cell lines that more faithfully reproduce the molecular and transcriptomic characteristics of MSPC.

In summary, by leveraging the molecular characteristics of metastatic prostate cancer and prostate cancer cell lines, we identified suitable metastatic prostate cancer cell lines for specific research applications. We anticipate that the adoption of more appropriate metastatic prostate cancer cell lines will enhance the translational relevance of *in vitro* studies and accelerate the development of clinically meaningful therapeutic strategies.

421 # Materials and Methods

422 **Selection of genes used for genomic profile comparison**

423 The somatic mutation data of metastatic and primary prostate cancer used in this
424 study were from SU2C and TCGA, respectively. In our analysis, non-adenocarcinoma
425 samples were excluded. Given a gene, we performed Fisher's exact to evaluate
426 whether its mutation frequency is significantly different between primary and
427 metastatic prostate cancer samples. Since SU2C contains metastatic samples from
428 three sites (liver, bone, and lymph node), the analysis was performed in site-specific
429 manner. Only genes which are differently mutated (adjusted p-value < 0.05) in at least
430 two sites were considered as differentially mutated genes.

431

432 Genes whose mutation frequency exceeded 5% in at least two metastatic sites (in the
433 SU2C cohort) were defined as highly mutated genes. The non-synonymous mutations
434 occurring in at least three SU2C samples were defined as mutation hotspot.

435

436 **Identification of differential CNV genes between metastatic and primary**
437 **prostate cancer**

438 We downloaded CNV segment data and then utilized CNVTools[55] to generate
439 gene-level copy number. For each gene, the Wilcoxon rank-sum test was applied to
440 compare its CNV profile between SU2C and TCGA samples. Genes which are
441 differently copy number variation (adjusted p-value < 0.05) in at least two sites and an

442 absolute median CNV difference ≥ 0.3 were considered as differentially CNV genes.

443

444 **Transcriptome correlation analysis with bulk RNA-seq and scRNA-seq data**

445 In our previous research, we identified 1,000 genes which are highly varied across
446 CCLE cell lines and proved their utility in transcriptome correlation analysis[12]. Given
447 a cell line and a patient sample (or single cell), we defined their transcriptome
448 similarity as the Spearman rank correlation across the 1,000 genes. The transcriptome
449 similarity between a cell line and several patient samples (or single cells) was
450 computed as the median correlation value.

451

452 **Differential gene expression analysis**

453 DESeq2[56] was used to identify differentially expressed genes (adjusted p-value < 0.05
454 and abs(log2FC) > 1) and the geneset enrichment analysis was conducted using
455 Enrichr[57]. Terms with a p-value < 0.05 were considered statistically significant.

456

457 **Pubmed search**

458 Our Pubmed search was performed on March 11th, 2025. For each prostate cancer cell
459 line, we searched the database using the keyword "[cell line name] + metastasis" and
460 counted the number of returned citations.

461

462 **Analysis of scRNA-seq data**

463 In the analysis of CRPC scRNA-seq dataset, we used inferCNV[58, 59] to infer copy
464 number variation of epithelial cells (immune cells as reference). CNV values were
465 extracted and subjected to k-means clustering to identify malignant cells. The marker
466 genes used in our analysis are: *PTPRC* (immune cells) and *EPCAM* (epithelial cells).
467

468 **Lineage analysis**

469 We used a normal prostate epithelial scRNA-seq dataset to identify three epithelial
470 cell types: luminal (*KLK2*, *KLK4*, *ACPP*), basal (*KRT5*, *TP63*, *KRT17*, *KRT15*), and
471 neuroendocrine (*ASCL1*, *FOXA2*, *MYCN*, *POU3F2*, *SIAH2*, *NCAM1*, *CHGA*, *CHGB*,
472 *SYP*, *ENO2*)[60, 61]. For each epithelial cell type, we generated pseudo-bulk
473 transcriptomes to create a centroid matrix. We then identified the top 5,000 highly
474 variable genes across the three cell types and used these genes to calculate
475 transcriptomic similarity between each malignant cell and the reference centroids.
476 Each malignant cell was assigned to the epithelial lineage with the highest similarity
477 score.
478

479 **Software tools and statistical methods**

480 All of the analysis was conducted in R. The ggplot2[62] and ComplexHeatmap[63]
481 packages were used for data visualization. Seurat[64] package was used for processing
482 scRNA-seq data. GSVA was used for ssGSEA analysis[65, 66]. If not specified, the two-sided
483 Wilcoxon rank-sum test was used to compute p-value and Benjamini-Hochberg

484 procedure was applied to compute adjusted p-value.

485

486 **Datasets**

487 The genomic and transcriptomic data of CCLE cell lines were obtained from the
488 DepMap data portal (https://depmap.org/portal/). The somatic mutation and copy
489 number variation data of TCGA and SU2C samples were downloaded from
490 cBioPortal (https://www.cbioportal.org/)[67, 68]. The bulk RNA-seq data of TCGA
491 samples was downloaded from UCSC Xena data portal (https://xena.ucsc.edu/)[69]. The
492 bulk RNA-seq data of metastatic prostate adenocarcinoma patients were from
493 MET500 cohort. The prostate adenocarcinoma scRNA-seq data was downloaded from
494 Single Cell Portal (https://singlecell.broadinstitute.org/single_cell/study/SCP1244).
495 The scRNA-seq data of normal human prostate, cell lines and organoids were
496 downloaded from GEO (accession number: GSE157220, GSE117403, GSE175975,
497 GSE162225 and GSE181374)[42, 43, 61, 70]. The scRNA-seq data of CRPC was
498 downloaded from SRA (accession number: PRJNA699369)[71].

499

500 **Data Availability**

501 The authors declare that all data supporting the findings of this study are available
502 within the article and its supplementary information files or from the corresponding
503 author upon reasonable request.

504

Acknowledgments

The research is supported by National Natural Science Foundation of China (Fund 32370715), Science Fund Program for Distinguished Oversea Young Scholars of Shandong Province (2023HWYQ-015), and the Cheeloo Young Scholar Program of Shandong University. The content is solely the responsibility of the authors and does not necessarily represent the official views of sponsors.

Authors' contribution

X.Y.L. and K.L. conceived the study. X.Y.L. performed the majority of computational analysis, Y.G.W. performed all the experimental assays, and K.L. supervised the study. All authors contributed to writing, reviewing, and editing the manuscript and approved the manuscript.

Competing Interests

The researchers declare no competing interests.

References

1. Kaur G, Dufour JM. Cell lines: Valuable tools or useless artifacts. *Spermatogenesis* **2**, 1-5 (2012).

2. Goodspeed A, Heiser LM, Gray JW, Costello JC. Tumor-Derived Cell Lines as Molecular Models of Cancer Pharmacogenomics. *Molecular cancer research : MCR* **14**, 3-13 (2016).

3. Bose B, Bozdag S. Identifying cell lines across pan-cancer to be used in precl inical research as a proxy for patient tumor samples. *Commun Biol* **7**, 1101 (2 024).

4. Dutil J, Chen Z, Monteiro AN, Teer JK, Eschrich SA. An Interactive Resource to Probe Genetic Diversity and Estimated Ancestry in Cancer Cell Lines. *Canc er research* **79**, 1263-1273 (2019).

5. Mouradov D*, et al.* Colorectal cancer cell lines are representative models of th e main molecular subtypes of primary cancer. *Cancer research* **74**, 3238-3247 (2014).

6. Gillet JP, Varma S, Gottesman MM. The clinical relevance of cancer cell lines. *Journal of the National Cancer Institute* **105**, 452-458 (2013).

7. Weinstein JN. Drug discovery: Cell lines battle cancer. *Nature* **483**, 544-545 (2 012).

8. Ertel A, Verghese A, Byers SW, Ochs M, Tozeren A. Pathway-specific differen ces between tumor cell lines and normal and tumor tissue cells. *Molecular can cer* **5**, 55 (2006).

9. Salvadores M, Fuster-Tormo F, Supek F. Matching cell lines with cancer type and subtype of origin via mutational, epigenomic, and transcriptomic patterns. *Science advances* **6**, (2020).

10. Domcke S, Sinha R, Levine DA, Sander C, Schultz N. Evaluating cell lines a s tumour models by comparison of genomic profiles. *Nat Commun* **4**, 2126 (2

013).

11. Yu K, *et al.* Comprehensive transcriptomic analysis of cell lines as models of primary tumors across 22 tumor types. *Nat Commun* **10**, 3574 (2019).

12. Liu K, *et al.* Evaluating cell lines as models for metastatic breast cancer through integrative analysis of genomic data. *Nat Commun* **10**, 2138 (2019).

13. Xin S, *et al.* ScRNA-seq revealed an immunosuppression state and tumor microenvironment heterogeneity related to lymph node metastasis in prostate cancer. *Exp Hematol Oncol* **12**, 49 (2023).

14. Rebello RJ, *et al.* Prostate cancer. *Nature reviews Disease primers* **7**, 9 (2021).

15. Jin H, *et al.* Systematic transcriptional analysis of human cell lines for gene expression landscape and tumor representation. *Nat Commun* **14**, 5417 (2023).

16. Peng D, *et al.* Evaluating the transcriptional fidelity of cancer models. *Genome medicine* **13**, 73 (2021).

17. Hao M, *et al.* HIC1 loss promotes prostate cancer metastasis by triggering epithelial-mesenchymal transition. *The Journal of pathology* **242**, 409-420 (2017).

18. Ghandi M, *et al.* Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503-508 (2019).

19. Hernández Borrero LJ, El-Deiry WS. Tumor suppressor p53: Biology, signaling pathways, and therapeutic targeting. *Biochimica et biophysica acta Reviews on cancer* **1876**, 188556 (2021).

20. Hunter I, Jamieson C, McEwan IJ. The androgen receptor amino-terminal domain: structure, function and therapeutic potential. *Endocrine oncology (Bristol, England)* **5**, e240061 (2025).

21. Robinson D, *et al.* Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215-1228 (2015).

22. He MX, *et al.* Transcriptional mediators of treatment resistance in lethal prostate cancer. *Nat Med* **27**, 426-433 (2021).

23. Augello MA, Den RB, Knudsen KE. AR function in promoting metastatic prostate cancer. *Cancer metastasis reviews* **33**, 399-411 (2014).

24. Heinlein CA, Chang C. Androgen receptor in prostate cancer. *Endocrine reviews* **25**, 276-308 (2004).

25. Lin A, Zhang J, Luo P. Crosstalk Between the MSI Status and Tumor Microenvironment in Colorectal Cancer. *Frontiers in immunology* **11**, 2039 (2020).

26. Seo MK, Kang H, Kim S. Tumor microenvironment-aware, single-transcriptome prediction of microsatellite instability in colorectal cancer using meta-analysis. *Scientific reports* **12**, 6283 (2022).

27. Ferreira AM, *et al.* High frequency of RPL22 mutations in microsatellite-unstable colorectal and endometrial tumors. *Human mutation* **35**, 1442-1445 (2014).

28. Weinstein HNW, *et al.* RPL22 is a tumor suppressor in MSI-high cancers and a splicing regulator of MDM4. *Cell Rep* **43**, 114622 (2024).

29. Cortes-Ciriano I, Lee S, Park WY, Kim TM, Park PJ. A molecular portrait of microsatellite instability across multiple cancers. *Nat Commun* **8**, 15180 (2017).

30. Lambert AW, Zhang Y, Weinberg RA. Cell-intrinsic and microenvironmental determinants of metastatic colonization. *Nature cell biology* **26**, 687-697 (2024).

31. Biray Avci C, Goker Bagca B, Nikanfar M, Takanlou LS, Takanlou MS, Nourazarian A. Tumor microenvironment and cancer metastasis: molecular mechanisms and therapeutic implications. *Frontiers in pharmacology* **15**, 1442888 (2024).

32. Joyce JA, Pollard JW. Microenvironmental regulation of metastasis. *Nature reviews Cancer* **9**, 239-252 (2009).

634 33. Chen B, Sirota M, Fan-Minogue H, Hadley D, Butte AJ. Relating hepatocellul
635 ar carcinoma tumor samples and cell lines using gene expression data in transl
636 ational research. *BMC medical genomics* **8 Suppl 2**, S5 (2015).

637

638 34. Moya L, *et al.* Characterisation of cell lines derived from prostate cancer patie
639 nts with localised disease. *Prostate Cancer Prostatic Dis* **26**, 614-624 (2023).

640

641 35. Molter CW, Muszynski EF, Tao Y, Trivedi T, Clouvel A, Ehrlicher AJ. Prostate
642 cancer cells of increasing metastatic potential exhibit diverse contractile forces,
643 cell stiffness, and motility in a microenvironment stiffness-dependent manner.
644 *Frontiers in cell and developmental biology* **10**, 932510 (2022).

645

646 36. Erickson A, *et al.* A Systematic Review of Prostate Cancer Heterogeneity: Und
647 erstanding the Clonal Ancestry of Multifocal Disease. *European urology oncolo*
648 *gy* **4**, 358-369 (2021).

649

650 37. Løvf M, *et al.* Multifocal Primary Prostate Cancer Exhibits High Degree of G
651 enomic Heterogeneity. *European urology* **75**, 498-505 (2019).

652

653 38. Song H, *et al.* Single-cell analysis of human primary prostate cancer reveals th
654 e heterogeneity of tumor-associated epithelial cell states. *Nat Commun* **13**, 141
655 (2022).

656

657 39. Guo W, *et al.* Single-cell transcriptomics identifies a distinct luminal progenitor
658 cell type in distal prostate invagination tips. *Nature genetics* **52**, 908-918 (202
659 0).

660

661 40. Zhang D, Zhao S, Li X, Kirk JS, Tang DG. Prostate Luminal Progenitor Cells
662 in Development and Cancer. *Trends in cancer* **4**, 769-783 (2018).

663

664 41. Chan JM, *et al.* Lineage plasticity in prostate cancer depends on JAK/STAT in
665 flammatory signaling. *Science* **377**, 1180-1191 (2022).

666

667 42. Han H, *et al.* Mesenchymal and stem-like prostate cancer linked to therapy-ind
668 uced lineage plasticity and metastasis. *Cell Rep* **39**, 110595 (2022).

669

670 43. Deng S, *et al.* Ectopic JAK-STAT activation enables the transition to a stem-li

ke and multilineage state conferring AR-targeted therapy resistance. *Nat Cancer* **3**, 1071-1087 (2022).

44. Jeon HM, Lee J. MET: roles in epithelial-mesenchymal transition and cancer stemness. *Annals of translational medicine* **5**, 5 (2017).

45. van Leenders GJ, *et al.* Activation of c-MET induces a stem-like phenotype in human prostate cancer. *PloS one* **6**, e26753 (2011).

46. Dong J, Li J, Li Y, Ma Z, Yu Y, Wang CY. Transcriptional super-enhancers control cancer stemness and metastasis genes in squamous cell carcinoma. *Nat Commun* **12**, 3974 (2021).

47. Tang F, *et al.* Chromatin profiles classify castration-resistant prostate cancers suggesting therapeutic targets. *Science* **376**, eabe1505 (2022).

48. Virtanen C, *et al.* Integrated classification of lung tumors and cell lines by expression profiling. *Proc Natl Acad Sci U S A* **99**, 12357-12362 (2002).

49. Sinha R, *et al.* Analysis of renal cancer cell lines from two major resources enables genomics-guided cell line selection. *Nat Commun* **8**, 15165 (2017).

50. Xing P, *et al.* Treatment strategies and drug resistance mechanisms in adenocarcinoma of different organs. *Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy* **71**, 101002 (2023).

51. Raychaudhuri R, Lin DW, Montgomery RB. Prostate Cancer: A Review. *Jama* **333**, 1433-1446 (2025).

52. Cejas P, *et al.* Subtype heterogeneity and epigenetic convergence in neuroendocrine prostate cancer. *Nat Commun* **12**, 5775 (2021).

53. Bluemn EG, *et al.* Androgen Receptor Pathway-Independent Prostate Cancer Is Sustained through FGF Signaling. *Cancer Cell* **32**, 474-489.e476 (2017).

54. Lindh C, *et al.* Ductal and acinar components of mixed prostatic adenocarcinoma frequently have a common clonal origin. *The Prostate* **82**, 576-583 (2022).

55. The CNTools package. https://bioconductor.org/packages/release/bioc/html/CNTools.html.).

56. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

57. Kuleshov MV, *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97 (2016).

58. Patel AP, *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396-1401 (2014).

59. Puram SV, *et al.* Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **171**, 1611-1624.e1624 (2017).

60. Dong B, *et al.* Single-cell analysis supports a luminal-neuroendocrine transdifferentiation in human prostate cancer. *Commun Biol* **3**, 778 (2020).

61. Henry GH, *et al.* A Cellular Anatomy of the Normal Adult Human Prostate and Prostatic Urethra. *Cell Rep* **25**, 3530-3542.e3535 (2018).

62. The ggplot2 package. https://cran.r-project.org/web/packages/ggplot2/index.html.).

63. The ComplexHeatmap package. https://bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html.).

64. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495-502 (2015).

65. Subramanian A, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).

66. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC bioinformatics* **14**, 7 (2013).

745
746 67. Gao J, *et al.* Integrative analysis of complex cancer genomics and clinical prof
747 iles using the cBioPortal. *Science signaling* **6**, pl1 (2013).
748
749 68. Cerami E, *et al.* The cBio cancer genomics portal: an open platform for explo
750 ring multidimensional cancer genomics data. *Cancer discovery* **2**, 401-404 (201
751 2).
752
753 69. Vivian J, *et al.* Toil enables reproducible, open source, big biomedical data an
754 alyses. *Nat Biotechnol* **35**, 314-316 (2017).
755
756 70. Kinker GS, *et al.* Pan-cancer single-cell RNA-seq identifies recurring programs
757 of cellular heterogeneity. *Nature genetics* **52**, 1208-1218 (2020).
758
759 71. Cheng Q, *et al.* Pre-existing Castration-resistant Prostate Cancer-like Cells in P
760 rimary Prostate Cancer Promote Resistance to Hormonal Therapy. *European uro*
761 *logy* **81**, 446-455 (2022).
762
763