1   **_In silico_ Computed Clusters of Prostate Luminal Progenitors Match FACS-Enriched LSCmed cells**

2

3   [1,*]Emilia Puig Lombardi, [2]Manon Baures, [2]Charles Dariane, [2]Jacques-Emmanuel Guidotti,

4   [2,*]Vincent Goffin

5

6   [1]Bioinformatics Core Platform, INSERM UMR1163, Imagine Institute/Paris Descartes University, Paris,

7   France

8   [2]Inserm U1151/Institut Necker Enfants Malades (INEM), Université de Paris, Paris, France

9

10   *Corresponding authors (maria-emilia.puig-lombardi@institutimagine.org ;

11   vincent.goffin@inserm.fr)

12

13   **Abstract**

14   Several groups recently published single-cell (sc) expression atlases of the adult mouse prostate cells

15   based on RNA sequencing (scRNA-seq) data. All studies identified one computerized cluster of non-

16   secretory luminal progenitor cells enriched in luminal and stemness-related gene transcripts. The

17   actual correspondence between these luminal progenitor cell clusters has not been investigated. In

18   addition, the presence of _Krt4_ (encoding cytokeratin 4) in these _in silico_-identified luminal progenitors

19   suggested the overlap with FACS-enriched LSCmed luminal progenitor cells earlier identified as a stem-

20   like, castration-tolerant and tumor-initiating cell population. Here, we used a unified bioinformatics

21   pipeline to re-analyze published prostate scRNA-seq datasets and perform various pan-transcriptomic

22   comparisons including the LSCmed cell signature. Our study demonstrates that i) the mouse prostate

23   luminal progenitor cell clusters identified in the different scRNA-seq studies largely overlap and can be

24   defined by a common 15-gene signature including _Krt4_, ii) mouse LSCmed cells match both mouse and

25   human luminal progenitors identified by scRNA-seq analysis. Bridging these _in silico_-identified and _ex_

26   _vivo_-characterized prostate luminal progenitor subsets should benefit our understanding of their

27   actual involvement in prostate diseases.

Introduction

Prostate diseases include benign (prostatitis, benign prostate hyperplasia) and malignant (prostate cancer) pathologies. The cells-of-origin of these pathologies are still poorly understood and their identification is the object of intense investigations to improve treatments. The prostate gland contains a glandular compartment surrounded by a stroma. The glandular epithelium contains two main cell types named basal and luminal according to their position in the pseudostratified epithelium, and very rare neuroendocrine cells (Abate-Shen and Shen 2000). Adult prostate stem cells have been hypothesized to mainly reside within the basal cell layer (English et al. 1987, Evans and Chandler 1987, Lawson et al. 2007). Subsequent studies suggested that rare luminal cells could also display stemness and initiate prostate tumors (Wang et al. 2009, Wang et al. 2014). In 2014, we discovered and isolated by cell sorting a rare population of luminal cells exhibiting stem/progenitor properties. We named these cells LSC^{med} according to their FACS profile ($\underline{Lin}^-/\underline{Sca}-1^+$/$\underline{CD}$49f$^{\underline{med}}$) using stem cell antigen-1 (Sca-1) and CD49f (integrin α6) as cell surface markers (Sackmann-Sala et al. 2014). LSC^{med} cells express typical luminal genes (e.g. cytokeratin (CK) 8 and androgen receptor) and exhibit stem/progenitor-like properties being able to generate spheres and organoids *in vitro*, and glandular structures when engrafted into host mice (Sackmann-Sala et al. 2014, Kwon et al. 2016, Sackmann Sala et al. 2017). The transcriptomic signature of LSC^{med} (*versus* other epithelial cells) identified them as a distinct epithelial cell entity, and CK4 was validated as a specific protein biomarker of LSC^{med} progenitor cells on prostate sections from various mouse models (Sackmann Sala et al. 2017). Within the past 12 months, several groups published single cell (sc) atlases of the adult mouse prostate cells based on RNA sequencing (scRNA-seq) data (Crowley et al. 2020, Guo et al. 2020, Joseph et al. 2020, Karthaus et al. 2020, Mevel et al. 2020). All studies identified one computerized cluster of non-secretory luminal progenitor cells enriched in luminal and stemness-related gene transcripts. The correspondence between the luminal progenitor cell clusters described in each scRNA-seq study was not investigated beyond the qualitative overlap of a few common markers. Furthermore, the presence of *Krt4* (encoding CK4) among the latter raised the question of the overlap between *in silico*-identified luminal progenitors (Crowley et al. 2020, Guo et al. 2020, Joseph et al. 2020, Karthaus et al. 2020, Mevel et al. 2020) and FACS-enriched LSC^{med} cells (Sackmann Sala et al. 2017). To address these questions, we used a unified bioinformatics pipeline to re-analyze the published prostate scRNA-seq datasets and perform various pan-transcriptomic comparisons. The data presented here demonstrate that i) the mouse prostate luminal progenitor cell clusters identified in the different scRNA-seq studies largely overlap and can be defined by a common 15-gene signature, ii) mouse LSC^{med} cells match both mouse and human luminal progenitors identified by scRNA-seq analysis.

62 **Experimental procedures**

63 **scRNA-seq data retrieval**. Data re-analyzed as part of this study was retrieved from public databases
64 with accession numbers GSE145861, GSE145865, GSE146811, GSE150692, GSE151944 (GEO, Gene
65 Expression Omnibus) and OEP000825 (NODE, The National Omics Data Encyclopedia) (**Table 1**). We
66 only retained samples taken from intact (i.e. non-castrated) mice. In particular, data retrieved from
67 GEO repository GSE151944 were available as MULTI-seq sample barcodes; these were demultiplexed
68 using the *MULTIseqDemux* function implemented in *Seurat* (Stuart et al. 2019). For datasets
69 GSE145861 and GSE145865 (Joseph et al. 2020), the data from the prostate and urethral regions were
70 aggregated during the analysis step. Finally, data retrieved from the OEP000825 repository were raw
71 fastq files which were processed (read alignment, generation of feature-barcode matrices) with Cell
72 Ranger (10x Genomics) prior to any data analysis step.

73 **Data analysis**. All datasets were processed using the same analytical pipeline. Low quality cells were
74 filtered-out by consecutively filtering each sample individually (i.e. prior to any aggregation step) based
75 on unique molecular identifiers (UMI) counts, percentage mitochondrial content and number of genes
76 (in that precise order). No filter was applied on the percentage of ribosomal gene content. Filtering
77 was performed as previously described (Henry et al. 2018, Joseph et al. 2020): filter thresholds were
78 chosen dynamically for samples based on the distribution of each parameter, with code adapted from
79 sc-TissueMapper (v2.0.0) (Henry and Strand 2020). Upper and lower filters were applied on UMIs (the
80 lower bound of the UMI filter was strictly set to 200 whilst the *RenyiEntropy* thresholding technique
81 was applied to determine the upper bound after binning the data), while the percentage mitochondrial
82 content had only an upper filter (abnormally high percentages of mitochondrial content were
83 determined using the *Triangle* filter on binned data) and feature number had only lower filters
84 (determined using the *MinErrorI* filter on binned data). *RenyiEntropy*, *Triangular*, and *MinErrorI*
85 thresholding were applied using functions from the *autothresholdr* (v1.3.9) R package. For a given
86 dataset, if multiple samples were available, samples were aggregated by normalizing with the
87 *sctransform* (version 0.3.2) method and using *Seurat*'s reciprocal PCA method. Furthermore, cells
88 displaying high stress signatures associated with the tissue dissociation experimental step were
89 removed as described (Henry et al. 2018): aggregated cells were scored for stress using *Seurat*'s
90 *AddModuleScore* method and a mouse geneset enriched for stressed cells earlier described (van den
91 Brink et al. 2017). Finally, principal component analysis (PCA) was performed on the data and graph-
92 based clustering was performed using the principal components representing 90% of the associated
93 cumulative variance.

94 **LSC^med score calculation**. LSC^med-specific genes (n = 111 genes, assayed in WT mice) were retrieved
95 from Sackmann Sala *et al*. (2017). The LSC^med similarity score was based on the calculation of the

96   average mRNA levels of the 111 signature genes for each single cell (to which is subtracted the

97   averaged mRNA levels of 50 randomly chosen control genes) using *Seurat*'s *AddModuleScore* method.

98   For calculation in the human prostate dataset, mouse stable gene IDs were matched to human stable

99   gene IDs, retaining one-to-one ortholog matches. The mouse-human orthology table was generated

100   with Ensembl BioMart.

101   **Marker gene identification**. Differentially expressed genes for the identified cell subpopulations were

102   determined using Wilcoxon rank sum tests on genes present in at least 20% of cells in the population

103   of interest, only retaining positive gene markers. Testing was limited to genes which showed, on

104   average, at least 0.2-fold difference (on a log-scale) between the different groups. Finally, genes

105   displaying an adjusted *p*-value inferior to 5% ($P_{adj} < 0.05$) were retained.

106   **Data access**. The scRNA-seq datasets re-analyzed in the scope of this project were stored as processed

107   R objects in .rds format and are available from the corresponding author on reasonable request.

108

109   **Results**

110   To investigate the overlap between the luminal progenitor clusters described in published single cell

111   studies (Crowley et al. 2020, Guo et al. 2020, Joseph et al. 2020, Karthaus et al. 2020, Mevel et al.

112   2020), we reanalyzed their publicly available scRNA-seq transcriptomic data (**Table 1**) to identify the

113   top biomarkers of luminal progenitors in each individual study and compare their expression level in

114   each cell population across studies. Importantly, we applied the same analytical pipeline on count data

115   retrieved for each of the studies, allowing notably to identify and remove stressed cells from the

116   analyses (see **Experimental procedures**, *Data analysis*), enabling comparisons between the different

117   datasets. Starting from 4,624, 13,688, 5,288, 34,444 and 90,345 single cells for each study (**Table 1**),

118   we retained 1,213, 5,158, 2,362, 19,503 and 45,432 single cells after quality filtering, respectively (see

119   Legend to **Figure 1A**), with a median of 2,205 genes assayed per cell.

120   Subsequently, clustering was followed by dimensional reduction for visualization using UMAP (Uniform

121   Manifold Approximation and Projection) plots, which depicts cell populations with distinct

122   transcriptional signatures (**Figure 1A**). For each dataset, we identified between 7 and 14 *in silico*

123   computed clusters expressing different phenotypic markers (epithelial, immune and stromal cells) as

124   described (Crowley et al. 2020, Guo et al. 2020, Joseph et al. 2020, Karthaus et al. 2020, Mevel et al.

125   2020), which were used to match previously labeled sub-populations of interest to our re-analyzed

126   datasets (**Figure 1A**). In particular, we were able to recapitulate the Lum D (Mevel et al. 2020), Lum 2

127   (Karthaus et al. 2020), Lum P (Crowley et al. 2020), Lum C (Guo et al. 2020) and Ur Lum (Joseph et al.

128   2020) subsets of non-secretory luminal progenitor cells as clear clusters. Marker gene identification

4

129 for these subpopulations (**Table S1**) allowed to identify 21 different genes common to at least 4

130 studies, of which 15 were common to all studies (**Figure 2**). These included *Krt4*, *Psca, Clu, Wfdc2,*

131 *Cyp2f2, Tspan8, Gsta4 and Tacstd2.* Of note, the latter has also been used as a surface protein

132 biomarker (Trop2) to enrich luminal progenitors by cell sorting (Crowell et al. 2019, Crowley et al. 2020,

133 Guo et al. 2020). This analysis demonstrates that the various luminal progenitor clusters identified by

134 scRNA-seq exhibit a high degree of similarity and can be defined by a common transcriptomic

135 signature.

136 All above-mentioned genes are part of the LSC^{med} signature (Sackmann Sala et al. 2017). To assess the

137 actual overlap of LSC^{med} cells with luminal progenitors identified by scRNA-seq, we sought for the

138 enrichment of the LSC^{med} cell signature (111 genes) in the various prostate cell populations identified

139 *in silico* (see **Experimental procedures**, *LSCmed score calculation*). As expected, significant enrichment

140 of the LSC^{med} cell signature was observed in each scRNA-seq dataset for the sole population identified

141 as luminal progenitors (**Figure 1A,B**). Altogether, this indicates that LSC^{med} progenitor cells enriched by

142 cell sorting (Sackmann-Sala et al. 2014, Kwon et al. 2016, Sackmann Sala et al. 2017) largely overlap

143 with luminal progenitor cell clusters identified by scRNA-seq.

144 Single cell analyses of the human prostate have also been recently reported (Henry et al. 2018, Crowley

145 et al. 2020, Joseph et al. 2020, Karthaus et al. 2020). The original report by the group of D. Strand

146 identified two closely-related clusters of luminal progenitor-like cells showing transcriptomic

147 similarities with Club and Hillock progenitor cells described in the lung (Henry et al. 2018) (**Figure 1C**).

148 Prostate Club/Hillock cells share typical markers with mouse luminal progenitors including as *KRT4*,

149 *TACSTD2* and *PSCA.* As observed above for mouse luminal progenitors (**Figure 1A,B**), the LSC^{med} cell

150 signature was significantly enriched in Club and Hillock progenitors, but not in other cells of the human

151 prostate (**Figure 1C**).

152 Together, the full transcriptomic profile comparisons reported above provide unbiased evidence that

153 mouse prostate LSC^{med} luminal progenitors enriched by cell sorting correspond to luminal progenitor

154 clusters identified in mouse and human prostates by scRNA-seq.

155

156 **Discussion**

157 Luminal progenitors have recently emerged as key players of prostate pathogenesis including

158 inflammation (Wang et al. 2015), benign prostate hyperplasia (Crowell et al. 2019, Joseph et al. 2020)

159 and prostate cancer (Korsten et al. 2009, Wang et al. 2014, Sackmann Sala et al. 2017, Guo et al. 2020).

160 These findings have generated a particular interest for this particular cell subset (Joseph et al. 2021).

161 However, one recurrent issue in the stem cell field is to evaluate to which extent cell populations

5

162   identified in different studies, often by different experimental approaches, actually correspond to

163   equivalent cell entities. While the probable overlap of *in silico* computed clusters of prostate luminal

164   progenitors in the various scRNA-seq studies was suggested (Crowley et al. 2020), to our knowledge,

165   no *bona fide* transcriptomic analysis covering all scRNA-seq data available was provided to support this

166   hypothesis. Furthermore, despite of the fact that CK4 was stressed as a biomarker of the luminal

167   progenitor cluster in at least two studies (Guo et al. 2020, Mevel et al. 2020), the potential similarity

168   with LSC^{med} cells was ignored in all but one scRNA-seq report (Joseph et al. 2020). The present study

169   involving pan-transcriptomic comparative analysis of human and mouse luminal progenitor clusters

170   identified *in silico versus* enriched *ex vivo* by cell sorting definitely confirms their molecular equivalence

171   which is materialized by the common 15-gene signature. We and others demonstrated that mouse

172   prostate LSC^{med} luminal progenitors can initiate tumors in reconstitution assays and are castration-

173   tolerant in both normal and pathological states (Kwon et al. 2016, Sackmann Sala et al. 2017, Karthaus

174   et al. 2020), suggesting they may contribute to resistance to treatments targeting androgen signaling.

175   In keeping with this, some scRNA-seq studies highlighted the transcriptional plasticity of epithelial

176   luminal cells upon castration (Karthaus et al. 2020, Mevel et al. 2020). Bridging such *in silico*

177   transcriptomic information with functional characterization of FACS-enriched luminal progenitor cells

178   knowing they apply to the same cell entity should speed up our understanding of their biology in health

179   and disease.

180

181

187

188   **Legends**

189   **Figure 1. Transcriptomic similarity between FACS-enriched WT mouse LSC^med cells and luminal**

190   **progenitor cell clusters identified by scRNA-seq analyses of mouse (A,B) and human (C) healthy**

191   **prostates.**

192   **(A)** UMAP projections based on linear dimensionality reduction by principal component analysis (PCA)

193   for, from left to right, 45,432 (Joseph et al. 2020), 5,158 (Karthaus et al. 2020), 2,362 (Crowley et al.

194   2020), 1,213 (Mevel et al. 2020) and 19,503 (Guo et al. 2020) single-cell transcriptomes. **(B)** In each

195   study, a single subpopulation matched LSC^med-like cells (Sackmann Sala et al. 2017), as shown by high

196   calculated LSC^med gene signature scores. The violin plots show the calculated LSC^med scores per cluster

197   and per study (*** , Tukey multiple comparisons of means P_{adj} < 0.001 for all pairwise comparisons

198   performed). **(C)** The same analysis was performed for human prostate single-cell transcriptomes

199   (Joseph et al. 2020).

200

201   **Figure 2. Common markers of luminal progenitors identified by scRNA-seq study of WT mouse**

202   **prostate.**

203   Dot plot representation of marker genes for LSC^med-like cell subpopulations found in the five scRNA-

204   seq studies discussed in the text (as indicated), showing their relative expression across all detected

205   clusters. Each dot depicts both detection rate and average gene expression in detected cells for a gene

206   in a cluster. Darker red colors indicate higher average gene expression, and a larger dot diameter

207   indicates that the gene was detected in greater proportion of cells from the cluster. Stars (*) depict

208   LSC^med-like cell marker genes detected in all datasets. Squares indicate marker genes that were

209   significantly (black) or nearly significantly (white) identified in the LSC^med cell signature (Sackmann Sala

210   et al. 2017). The other genes were all expressed in LSC^med cells at similar levels as in basal and/or

211   luminal cells, therefore they were not identified as LSC^med cell markers.

212 **References**

213   Abate-Shen, C. and M. M. Shen (2000). "Molecular genetics of prostate cancer." Genes Dev **14**(19):
214   2410-2434. DOI 10.1101/gad.819500.

215   Crowell, P. D., J. J. Fox, T. Hashimoto, J. A. Diaz, H. I. Navarro, G. H. Henry, B. A. Feldmar, M. G. Lowe,
216   A. J. Garcia, Y. E. Wu, D. P. Sajed, D. W. Strand and A. S. Goldstein (2019). "Expansion of Luminal
217   Progenitor Cells in the Aging Mouse and Human Prostate." Cell Rep **28**(6): 1499-1510 e1496. DOI
218   10.1016/j.celrep.2019.07.007.

219   Crowley, L., F. Cambuli, L. Aparicio, M. Shibata, B. D. Robinson, S. Xuan, W. Li, H. Hibshoosh, M. Loda,
220   R. Rabadan and M. M. Shen (2020). "A single-cell atlas of the mouse and human prostate reveals
221   heterogeneity and conservation of epithelial progenitors." Elife **9**. DOI 10.7554/eLife.59465.

222   English, H. F., R. J. Santen and J. T. Isaacs (1987). "Response of glandular versus basal rat ventral
223   prostatic epithelial cells to androgen withdrawal and replacement." Prostate **11**(3): 229-242. DOI
224   10.1002/pros.2990110304.

225   Evans, G. S. and J. A. Chandler (1987). "Cell proliferation studies in the rat prostate: II. The effects of
226   castration and androgen-induced regeneration upon basal and secretory cell proliferation."
227   Prostate **11**(4): 339-351. DOI 10.1002/pros.2990110406.

228   Guo, W., L. Li, J. He, Z. Liu, M. Han, F. Li, X. Xia, X. Zhang, Y. Zhu, Y. Wei, Y. Li, R. Aji, H. Dai, H. Wei, C. Li,
229   Y. Chen, L. Chen and D. Gao (2020). "Single-cell transcriptomics identifies a distinct luminal
230   progenitor cell type in distal prostate invagination tips." Nat Genet **52**(9): 908-918. DOI
231   10.1038/s41588-020-0642-1.

232   Henry, G. H., A. Malewska, D. B. Joseph, V. S. Malladi, J. Lee, J. Torrealba, R. J. Mauck, J. C. Gahan, G.
233   V. Raj, C. G. Roehrborn, G. C. Hon, M. P. MacConmara, J. C. Reese, R. C. Hutchinson, C. M. Vezina
234   and D. W. Strand (2018). "A Cellular Anatomy of the Normal Adult Human Prostate and Prostatic
235   Urethra." Cell Rep **25**(12): 3530-3542 e3535. DOI 10.1016/j.celrep.2018.11.086.

236   Henry, G. H. and D. W. Strand (2020). "Strand Lab analysis of single-cell RNA sequencing." Zenodo. DOI
237   doi:10.5281/zenodo.3687064.

238   Joseph, D. B., G. H. Henry, A. Malewska, N. S. Iqbal, H. M. Ruetten, A. E. Turco, L. L. Abler, S. K. Sandhu,
239   M. T. Cadena, V. S. Malladi, J. C. Reese, R. J. Mauck, J. C. Gahan, R. C. Hutchinson, C. G. Roehrborn,
240   L. A. Baker, C. M. Vezina and D. W. Strand (2020). "Urethral luminal epithelia are castration-
241   insensitive cells of the proximal prostate." Prostate **80**(11): 872-884. DOI 10.1002/pros.24020.

242   Joseph, D. B., A. E. Turco, C. M. Vezina and D. W. Strand (2021). "Progenitors in prostate development
243   and disease." Dev Biol **473**: 50-58. DOI 10.1016/j.ydbio.2020.11.012.

244   Karthaus, W. R., M. Hofree, D. Choi, E. L. Linton, M. Turkekul, A. Bejnood, B. Carver, A. Gopalan, W.
245   Abida, V. Laudone, M. Biton, O. Chaudhary, T. Xu, I. Masilionis, K. Manova, L. Mazutis, D. Pe'er, A.

246 Regev and C. L. Sawyers (2020). "Regenerative potential of prostate luminal cells revealed by
247 single-cell analysis." Science **368**(6490): 497-505. DOI 10.1126/science.aay0267.

248 Korsten, H., A. Ziel-van der Made, X. Ma, T. van der Kwast and J. Trapman (2009). "Accumulating
249 progenitor cells in the luminal epithelial cell layer are candidate tumor initiating cells in a Pten
250 knockout mouse prostate cancer model." PLoS One **4**(5): e5662. DOI
251 10.1371/journal.pone.0005662.

252 Kwon, O. J., L. Zhang and L. Xin (2016). "Stem Cell Antigen-1 Identifies a Distinct Androgen-Independent
253 Murine Prostatic Luminal Cell Lineage with Bipotent Potential." Stem Cells **34**(1): 191-202. DOI
254 10.1002/stem.2217.

255 Lawson, D. A., L. Xin, R. U. Lukacs, D. Cheng and O. N. Witte (2007). "Isolation and functional
256 characterization of murine prostate stem cells." Proc. Natl. Acad. Sci. U. S. A **104**(1): 181-186. DOI
257 10.1073/pnas.0609684104.

258 Mevel, R., I. Steiner, S. Mason, L. C. Galbraith, R. Patel, M. Z. Fadlullah, I. Ahmad, H. Y. Leung, P. Oliveira,
259 K. Blyth, E. Baena and G. Lacaud (2020). "RUNX1 marks a luminal castration-resistant lineage
260 established at the onset of prostate development." Elife **9**. DOI 10.7554/eLife.60225.

261 Sackmann-Sala, L., A. Chiche, N. Mosquera-Garrote, F. Boutillon, C. Cordier, I. Pourmir, L. Pascual-
262 Mathey, K. Kessal, N. Pigat, P. Camparo and V. Goffin (2014). "Prolactin-Induced Prostate
263 Tumorigenesis Links Sustained Stat5 Signaling with the Amplification of Basal/Stem Cells and
264 Emergence of Putative Luminal Progenitors." Am. J. Pathol **184**: 3105-3119. DOI
265 10.1016/j.ajpath.2014.07.020.

266 Sackmann Sala, L., F. Boutillon, G. Menara, A. De Goyon-Pelard, M. Leprevost, J. Codzamanian, N.
267 Lister, J. Pencik, A. Clark, N. Cagnard, C. Bole-Feysot, R. Moriggl, G. P. Risbridger, R. A. Taylor, L.
268 Kenner, J. E. Guidotti and V. Goffin (2017). "A rare castration-resistant progenitor cell population
269 is highly enriched in Pten-null prostate tumors." J Pathol **243**: 54-64. DOI 10.1002/path.4924.

270 Stuart, T., A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, 3rd, Y. Hao, M. Stoeckius,
271 P. Smibert and R. Satija (2019). "Comprehensive Integration of Single-Cell Data." Cell **177**(7): 1888-
272 1902 e1821. DOI 10.1016/j.cell.2019.05.031.

273 van den Brink, S. C., F. Sage, A. Vertesy, B. Spanjaard, J. Peterson-Maduro, C. S. Baron, C. Robin and A.
274 van Oudenaarden (2017). "Single-cell sequencing reveals dissociation-induced gene expression in
275 tissue subpopulations." Nat Methods **14**(10): 935-936. DOI 10.1038/nmeth.4437.

276 Wang, L., M. Zoetemelk, B. R. Chitteti, T. L. Ratliff, J. D. Myers, E. F. Srour, H. Broxmeyer and T. J. Jerde
277 (2015). "Expansion of prostate epithelial progenitor cells after inflammation of the mouse
278 prostate." Am J Physiol Renal Physiol **308**(12): F1421-1430. DOI 10.1152/ajprenal.00488.2014.

279 Wang, X., M. Kruithof-de Julio, K. D. Economides, D. Walker, H. Yu, M. V. Halili, Y. P. Hu, S. M. Price, C.

280   Abate-Shen and M. M. Shen (2009). "A luminal epithelial stem cell that is a cell of origin for

281   prostate cancer." Nature **461**(7263): 495-500. DOI 10.1038/nature08361.

282 Wang, Z. A., R. Toivanen, S. K. Bergren, P. Chambon and M. M. Shen (2014). "Luminal cells are favored

283   as the cell of origin for prostate cancer." Cell Rep **8**(5): 1339-1346.

284

Table 1

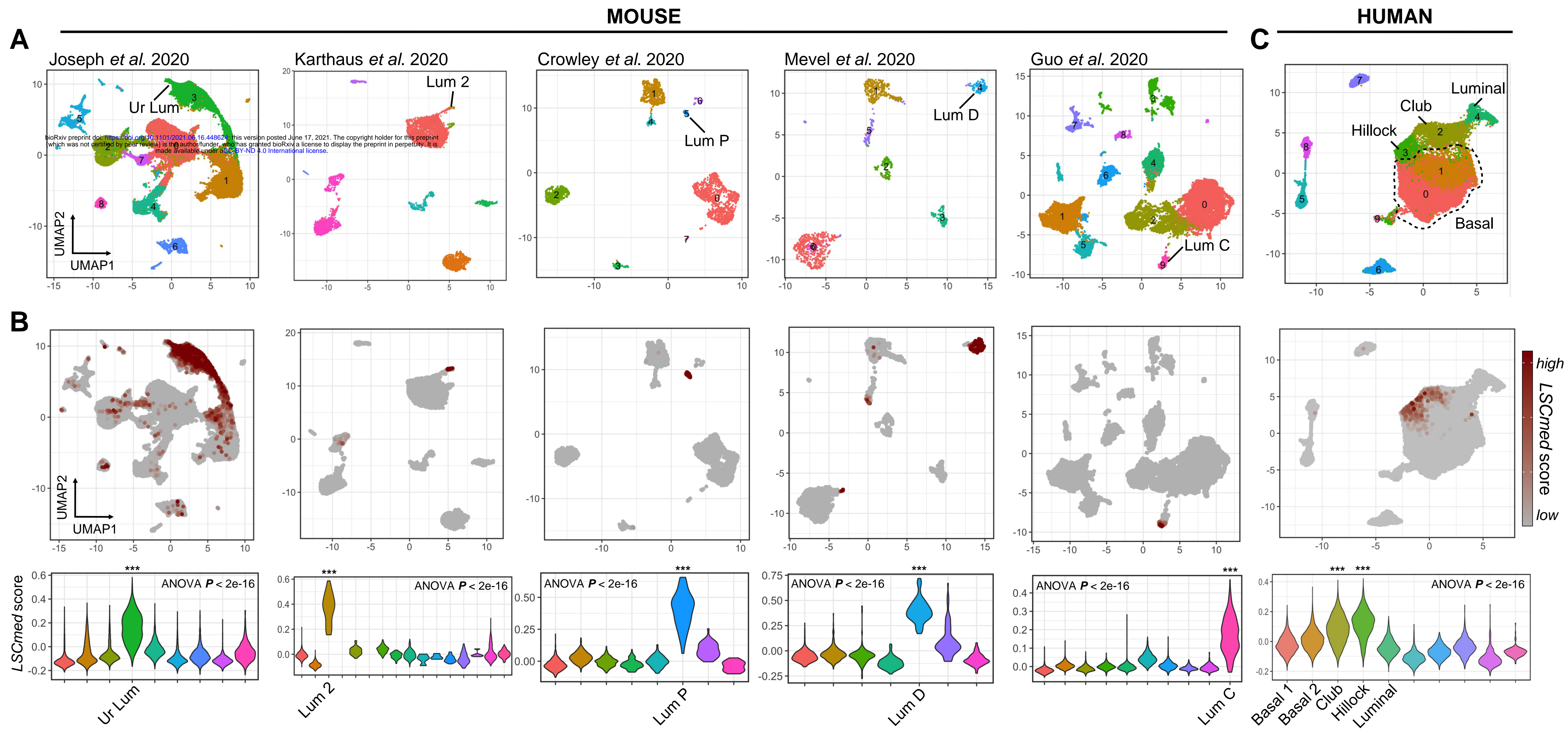| Repository	Accession	Data format	Cell count	Reference	DOI
GEO (Gene Expression Omnibus)	GSE145861 GSE145865	Count matrices in h5 format for each sample	90345	Joseph et al. (The Prostate, 2020)	10.1002/pros.24020
GEO (Gene Expression Omnibus)	GSE146811	Pooled Count matrices in h5 format	13688	Karthaus et al. (Science, 2020)	10.1126/science.aay0267
NODE (National Omics Data Encyclopedia)	OEP000825	Raw fastq files	34444	Guo et al. (Nat Genetics, 2020)	10.1038/s41588-020-0642-1
GEO (Gene Expression Omnibus)	GSE150692	Raw counts matrices in tsv format for each sample	5288	Crowley et al. (eLife, 2020)	10.7554/eLife.59465
GEO (Gene Expression Omnibus)	GSE151944	MULTI-seq outputs as raw count matrices per sample	4624	Mevel et al. (eLife, 2020)	10.7554/eLife.60225

Figure 1

MOUSE

A

Joseph *et al.* 2020 — Ur Lum

Karthaus *et al.* 2020 — Lum 2

Crowley *et al.* 2020 — Lum P

Mevel *et al.* 2020 — Lum D

Guo *et al.* 2020 — Lum C

HUMAN

C

Hillock, Club, Luminal, Basal

B

LSCmed score: high / low

ANOVA *P* < 2e-16

***

Ur Lum, Lum 2, Lum P, Lum D, Lum C

Basal 1, Basal 2, Club, Hillock, Luminal

Figure 2