Single cell analysis of early metastasis identifies targetable tumor subpopulation and mechanisms of immune evasion in squamous cell cancers

Hong Sheng Quah

National Cancer Centre Singapore   https://orcid.org/0000-0001-8430-8528

Elaine Yiqun Cao

Duke-NUS Medical School

Lisda Suteja

National Cancer Centre Singapore

Hui Leong

National Cancer Centre Singapore

Fui Chong

National Cancer Centre Singapore

Constance Li

National Cancer Centre Singapore

Shilpi Gupta

Singapore Immunology Network

Camille Arcinas

National Cancer Centre   https://orcid.org/0000-0001-5374-9232

John Ouyang

Duke-NUS Medical School   https://orcid.org/0000-0002-1239-1577

Vivian Ang

Singapore Immunology Network

Daniel Tan

Division of Medical Oncology, National Cancer Centre Singapore   https://orcid.org/0000-0002-6514-6786

Subhra BISWAS

Human Innate Immunity Lab

Owen Rackham

University of Bristol   https://orcid.org/0000-0002-4390-0872

N. Gopalakrishna Iyer (✉ gopaliyer@singhealth.com.sg)

National Cancer Centre Singapore   https://orcid.org/0000-0002-8812-6219

Article

1   **Single cell analysis of early metastasis identifies targetable tumor subpopulation and**

2   **mechanisms of immune evasion in squamous cell cancers**

3   **Authors:** Hong Sheng Quah[1,2*], Elaine Yiqun Cao[3*], Lisda Suteja[1], Hui Sun Leong[1], Fui Teen Chong[1], Constance H

4   Li[1], Shilpi Gupta[4], Camille Arcinas[1,2], John F Ouyang[3], Vivian Ang[4], Daniel SW Tan[1,2,5], Subhra K Biswas[4], Owen JL

5   Rackham[3] and N Gopalakrishna Iyer[1,2,6,7#]

6   **Affiliation:**

7   [1]Cancer Therapeutics Research Laboratory, National Cancer Centre Singapore, Singapore

8   [2]Academic Clinical Program in Oncology, Duke-NUS Medical School, Singapore

9   [3]Program in Cardiovascular and Metabolic Disorders, Duke-NUS Medical School, Singapore

10   [4]Singapore Immunology Network, Singapore

11   [5]Division of Medical Oncology, National Cancer Centre Singapore, Singapore

12   [6]Department of Head and Neck Surgery, National Cancer Centre Singapore, Singapore

13   [7]Division of Medical Sciences, National Cancer Centre Singapore, Singapore

14

15   [*]These authors contributed equally to this study

16   [#]Corresponding author

17   **Correspondence to:**

18   N Gopalakrishna Iyer, 11 Hospital Crescent, National Cancer Centre Singapore, Singapore 169610.

19   Email: gopaliyer@singhealth.com.sg, Tel: +65-64368000, Fax: +65-62257559

20   **Keywords:** pre-metastatic, single-cell genomics, targeted therapy, t-cell receptor, cytotoxic T-lymphocytes, EMT

21 **Abstract**

22 Profiling tumors at single-cell resolution provides an opportunity to understand complexities underpinning

23 lymph-node metastases in head and neck squamous-cell carcinoma. Single-cell RNAseq (scRNAseq) analysis of

24 cancer-cell trajectories identified a sub-population of pre-metastatic cells, driven by actionable pathways

25 including AXL and AURK. Blocking these two proteins blunted tumor invasion in patient-derived cultures.

26 Furthermore, scRNAseq analyses of tumor-infiltrating CD8+ T-lymphocytes showed two distinct trajectories to

27 T-cell dysfunction, corroborated by their clonal architecture based on single-cell T-cell receptor sequencing. By

28 determining key modulators of these trajectories, followed by validation using external datasets and functional

29 experiments, we uncovered a novel role for SOX4 in mediating T-cell exhaustion. Finally, interactome-analyses

30 between pre-metastatic tumor-cells and CD8+ T-lymphocytes uncovered a putative role for the Midkine

31 pathway in immune-modulation; this was confirmed by scRNAseq of tumors from humanized mice. Aside from

32 specific findings, this study demonstrates the importance of tumor heterogeneity analyses in identifying key

33 vulnerabilities during early metastasis.

34

Introduction

In most solid tumors development of lymph node metastasis portends poor outcomes, pre-dating distant metastasis[1-3]. In head and neck squamous cell cancers (HNSCC), these patients are treated with curative intent by surgery and radiation therapy with the prime objective of eradicating existing and future disease by depleting clones with a metastatic potential[4,5]. Metastasis is a continuum of phenotypes ranging from pre-metastatic features (*eg* lympho-vascular invasion), circulating tumor cells/emboli, microscopic lymph node deposits, gross nodal involvement and adjacent soft-tissue invasion, oligo-metastasis and finally, full blown distant metastasis[6]. Most studies focus on the terminal event, highlighting the role of definitive epithelial-mesenchymal transition (EMT); however bulk analyses in HNSCC suggests that EMT does not appear to be a pre-requisite for lymph node dissemination[7-11]. Recent studies have also highlighted that EMT itself exists as a spectrum, and tumor cells exhibit a significant amount of plasticity which may account for the range of clinical manifestations observed[12,13]. Single-cell analyses have the ability to resolve both issues: identification of rare clones with true metastatic potential and identifying pathways and vulnerabilities that can be exploited in the clinical setting to prevent further dissemination of these.

The role of the immune system during the metastatic cascade is gaining clinical relevance with current advancements in checkpoint blockade therapies[14]. This is especially pertinent in the context of lymph node metastasis, as lymph nodes are believed to be the main organ for T-cell priming, expansion and trafficking[15]. Understanding the mechanisms by which tumors evade immune-based killing within lymph nodes is critical to target early metastases[16-19]. Again, this can be addressed by single-cell analyses by defining the immune landscape, and in-depth dissection of interactions involved during immune evasion at the primary and nodal sites.

Here, we profiled primary and early (nodal) metastatic HNSCC tumors using single-cell RNAseq (scRNAseq) and TCRseq (scTCRseq) with two major objectives: to identify metastatic tumor subpopulations and identification of targetable vulnerabilities, and to determine the evolutionary trajectory of tumor-targeting T-cells as well as dissecting pathways employed by tumors to evade immune destruction during nodal dissemination.

Results

Single-cell transcriptional states of primary and lymph node metastasis in HNSCC

To delineate 'whole-tumor' single-cell landscapes in primary tumors and lymph node metastases, we developed a protocol to rapidly process freshly resected tissue for single-cell RNA sequencing (scRNAseq) and establishing primary cultures **(Figure 1A)**[20,21]. Tumors were harvested from fourteen treatment-naïve patients with locally advanced, HPV-negative HNSCC from primary and cervical lymph nodes (**Supplementary Table S1 and S2**). Seven pairs were processed for scRNAseq and single-cell T-cell receptor sequencing (scTCRseq), while primary cultures were successfully established for seven.

scRNAseq data for fresh tumors describes 53,459 cells (3,553–11,308 per patient) and 23,148 genes, with a median of 776 genes per cell (details on quality controls steps in **Methods** and **Supplementary Figure 1A-B**). Using Seurat v3.0, the data was normalized, pooled, and clustered **(Figure 1B)**. Canonical markers were used to broadly annotate these populations into: epithelial (*KRT7, KRT17*), salivary (*STATH*), fibroblasts (*COL1A2*), endothelial (*PECAM*) and immune *(PTPRC)* cells **(Figure 1B and Supplementary Figure 1C)**. Fibroblasts were further subdivided into cancer associated fibroblasts (CAFs; *MMP2*) and myofibroblasts (*ACTA2*), while immune cells were organized into T- (*CD3E, NKG7*), NK- (*NKG7, XCL2*), B- (*CD79A),* plasma- (*IGHG1*), mast- (*TPSAB1*), conventional (*LAMP3*) and plasmacytoid (*LILR4*) dendritic cells, as well as macrophages/monocytes (*CD163*). These were well-distributed across samples from all patients, apart from salivary cells, which were only observed in one patient, likely due to harvest of adjacent parotid gland tissue (HN263). However, there were differences in composition between primary and metastatic sites **(Figure 1C)**, with higher proportions of CAFs and TAMs in the primary tumor, versus more B-cells, plasma cells and dendritic cells at the metastatic sites, typical of a lymph node**.** These were similar to cellular composition proportions derived from bulk data from TCGA **(Supplementary Figure 1D)**. Inferred copy number variant analyses on the epithelial population showed that aneuploidy was evident in >95 % of cells validating that this population comprised cancer cells **(Figure 1D and Supplementary Figure 1E)**. Copy number alterations (CNAs) were further analyzed using the CopyKat algorithm[22], and identified those frequently observed in HNSCC[23], including gains across chromosomes 7 and 8q and loss of 3p and 5q **(Supplementary Figure 1F)**. Significant overlap of CNAs was also noted between the primary and metastatic sites in each patient **(Supplementary Figure 1G)**.

Tumor cells demonstrate varying degree of epithelial-mesenchymal transition during metastasis

We next focused on tumor cells (total of 6,115 cells & 17,784 genes) by extracting only the epithelial population with aneuploidy. Using Seurat 3.0, we pooled and re-analysed this subset, visualized as distinct clusters for each individual patient, with varying degree of overlap across cells from primary and nodal sites **(Figure 2A and Supplementary 2A)**. Tumor cell data can be accessed and interrogated as an interactive web application via the following Shiny app (http://hnc.ddnetbio.com/). Tumors from patients HN242, HN257 and HN272 show

significant overlap in tumor cells derived from both sites, while patients HN251 and HN279 show distinct site-specific sub-clusters. When comparing EMT gene markers in primary *vs* nodal metastases populations, nodal tumor cells had higher EMT scores compared to the primary in all patients except HN257 **(Figure 2B)**.

To identify the pre-nodal metastases subpopulation in primary tumors, we built trajectories using Monocle 2.0, and labelled the origin and direction based on the ground truth of site (*ie* primary tumor presumed to pre-date nodal disease), incorporating EMT-scores, and CytoTRACE (see Methods). The latter is a tool to determine degrees of differentiation, assuming de-differentiation co-occurs with the metastatic phenotype[24,25]. This approach was effective in identifying pre-nodal cells in patients HN242, HN251, HN272 and HN279 **(Figure 2C-D and Supplementary Figure 2C, 2G and 2H)**. For patients HN251 and HN279, pseudo-time ordering demonstrated an ordered, progressive, step-wise transition from primary to nodal disease. Nodal tumor cells largely dominate the end of the trajectory with higher CytoTrace scores. Major pathways over-represented across pseudotime include epithelial de-differentiation, oxidative phosphorylation and EMT (**Figure 2E**). Even in more complex trajectories such as HN272, the same approach was used to determine the likely trajectory to lymph node metastases, and identify sub-populations of primary cells (pre-nodal cells) that are similar to and likely gave rise to the metastatic phenotype (**Supplementary Figure 2C).** We next applied GeneSwitches[26] to identify actionable genes associated with the trajectory from primary to pre-nodal cells; these identified AXL, Aurora kinase, TYMS and STAT2 at potentially critical genes in this process (**Figure 2F and Supplementary Figure 2D-F**). This approach was validated on an external dataset comprising scRNAseq data from 5 tumors from primary and nodal sites available for analyses (2076 cells) (**Supplementary Figure 2H-P**)[12]. In three of these (p25, p26 and p28), EMT was higher in nodal tumor cells compared to the primary, hence could be resolved using the method described to identify a pre-nodal subpopulation (**Supplementary Figure 2P-R**). Several actionable genes identified through GeneSwitches appear to be implicated in this dataset as well: AXL (p25, p26, P28), STAT2 (p25, p26) and AURK (p26, p28) (**Supplementary Figure 2U**).

In contrast, analyses of patient HN257 was more complicated as the primary tumor had higher EMT scores than nodal tumor cells, and tumor trajectories were haphazard with no directionality **(Supplementary Figure 2H).** Cytotrace showed a distinct de-differentiated sub-population in the primary tumor that had high EMT scores and expression of *SNAI2* **(Figure 2G and Supplementary Figure 2I-J)**. We hypothesized that this was an aggressive, rapidly evolving tumor subpopulation. Differential expression analyses identified a panel of 132 up-regulated and 45 down-regulated genes in this subpopulation involved in oxidative phosphorylation and tumor metabolism, and immune evasion, respectively **(Figure 2H, Supplementary Table S3)**. Based on these gene sets, tumors in TCGA with the same signature (based on RNAseq data) had significantly poorer outcomes (**Figure 2I and Supplementary Figure 2K)**. In the validation scRNAseq dataset above, two of the tumors (p5 and p20) also showed a similar trend, with specific subpopulations in the primary tumor with high EMT scores (**Supplementary Figure 2S-T)**. Therefore, we postulate that in these tumors, distinct sub-populations in the primary tumor showed a more aggressive phenotype, that likely evolved after nodal dissemination had occurred.

Identifying vulnerabilities to target pre-metastatic tumor cells

We then proceed to test whether targets identified in this manner presented an opportunity for therapeutic intervention. scRNAseq using the C1 platform was performed on patient-derived cultures (PDCs) from primary and nodal metastatic sites (n=7 pairs). The data was processed using Seurat 3.0 and PAGODA (pathway and gene set overdispersion analysis) (**Figure 3A and Supplementary Figure 3A-B**). We derived scRNAseq data for a total of 1,317 cells and 55,216 genes. Similar to above, tumor-cell clusters were based on individual patients. However, PDCs demonstrated distinct separation between primary and metastatic cells, with EMT as one of the major differentiating principal component pathways **(Figure 3B and Supplementary Figure 3A-B**). Here, pre-nodal cells in HN137, HN159 and HN220 were identified as small primary subpopulations that clustered with metastatic cells.

Differential expression analyses for these pre-nodal populations identified *AXL* (in HN137) and *AURKB* (in HN159 and HN220) as putative actionable targets **(Figure 3C and Supplementary Table S4-6)**. Expression of these genes was validated using immunohistochemistry or immunofluorescence in both PDCs and respective tumor tissue, and this was recapitulated on flow cytometry for AXL (HN137) and AURK (HN159 and HN220), respectively **(Figure 3D and Supplementary Figure 3C-D)**. In HN137, expression of protein and transcript AXL was detected in a majority of metastatic cells compared with only a small sub-population of primary cells. Similarly, for HN159 and HN220, AURKB expression was significantly lower in metastatic cells, compared to primary cells. We focused on AXL and AURKB because both have specific inhibitors: BGB324 targeting cells with high AXL expression, and barasertib (pan-AURK inhibitor) targeting cells with limiting AURKA/AURKB levels. There were no differences in clonogenicity between primary and metastatic cultures from patient HN137 treated with BGB324, nor HN159 and HN220 treated with barasertib **(Supplementary Figure 3E-G)**. In contrast, all three metastatic lines HN137, HN159 and HN220 (treated with their respective drugs) demonstrated lower cell migration/invasion compared to untreated cultures, measured by scratch and Boyden chamber invasion assays **(Figure 3E-G):** AXL-inhibition significantly reduced invasive potential of both primary and metastatic cells of HN137 (**Figure 3E**) while AURK-inhibition significantly reduced the invasive potential of only metastatic cells of HN159 and HN220 **(Figure 3F and G)**. As AXL is a surface membrane protein, primary cells were sorted into AXL low-, medium- and high-expressing cells. As predicted, BGB324 specifically inhibited invasion only in the AXL-high primary subpopulation compared to AXL-low cells **(Figure 3H).** These data indicate AXL and AURKB play major roles in invasion and provide an opportunity for specific anti-metastatic therapy.

Evolution of CD8+ T-cells derived from analysis of primary tumor and lymph node metastasis

CD3+ T-cells form one of the major subpopulations sequenced at both primary and nodal sites. Data from 10,168 cells (covering 13,729 genes) were pooled, analyzed using Seurat, and visualized as ten distinct T-cell clusters (**Figure 4A**). The identity of each cluster was delineated based on differential gene expression of known T-cell markers (**Figure 4B and Supplementary Figure 4A-B)**. Some were distinct for CD4+ cells (Tregs and Tfh) and CD8+ cells (Pre-dysfunctional, Dysfunctional, Proliferative), while others comprise both CD4+ and CD8+ lineages

167   (Naïve-like and Transitional). Majority of naïve-like cells were derived from nodal tissue while the remaining

168   clusters appear to have equal representation from the primary and nodal metastatic sites **(Figure 4B and**

169   **Supplementary Figure 4C)**.

170   CD8+ T-cells (total of 3,387 cells, 11,847 genes) were extracted from this pooled T-cell dataset and re-analyzed

171   after regression for cell cycle-driven artefacts to identify lineage-based clusters. CD8+ T-cell data can be accessed

172   and interrogated as an interactive web application using the following Shiny app (http://hnc.ddnetbio.com/).

173   Six distinct clusters were labelled as naïve, transitional, tissue-resident memory, pre-dysfunctional, proliferative

174   and late dysfunctional based on canonical markers **(Figure 4C-D).** Using Slingshot, we performed trajectory

175   analyses on the CD8+ T-cells using the *CXCL13*-high, *LAYN*-high exhausted/senescent population as the end-

176   point[27], and this identified two convergent trajectories (**Figure 4E**). Expression plots across Trajectory 1 showed

177   a progressive loss of naïve markers, gradual gain of dysfunctional (and senescent) markers and an intervening

178   proliferative 'burst', that likely reflects expanding clones of tumor targeting CD8+ cells **(Figure 4F)**. Specifically,

179   this lineage suggests a scenario where naïve CD8+ T-cells from lymph nodes or circulation were trafficking into

180   the primary tumor with loss of circulating markers *KLF2*, *SELL* and *CCR7*, gain of tissue resident marker

181   *CD103*/*ITGAE*, progressive decline in the expression of naïve genes *TCF7*, *IL7R*, *CCR7*, and gradual gain of

182   dysfunctional markers (*TIM3*, *CTLA4*, *TIGIT*, *CXCL13*, *LAYN*) with an intermediary proliferative burst with high

183   levels of *MKI67*, *TOP2A*, *TYMS* **(Figure 4B, 4E-F)**. This is also reflected by progressive increase from *GZMK* to

184   *GZMB*, *PRF1*, and *IFNG* in pre-dysfunctional to dysfunctional cells*. In contrast, the trajectory of tissue-resident

185   memory (TRM) to dysfunctional cells (Trajectory 2) shows fewer genes being activated as the expression level

186   of many of the tissue resident (*ITGAE*), dysfunctional (*CTLA4*) and granzymes (*GZMs*) genes were already

187   upregulated **(Figure 4B)**. The Geneswitches algorithm was applied to trajectory 1 (naïve-to-dysfunction) to

188   predict key gene expression changes across pseudotime and identify factors that could account for these **(Figure**

189   **4G)**[26]. Our results indicate the major nodes appear to be an early loss of *KLF2*, intermediate increase in *NKG7*

190   and late increase in *SOX4*, *DUSP4* and *RBPJ* (**Figure 4G-H**).

191

192   **Modulating genes driving tumor-targeting cells dysfunction/exhaustion**

193   Based on the data above, expression of *SOX4*, *DUSP4* and *RBPJ* appears to coincide with the transition between

194   dysfunction and exhaustion, but whether these genes modulate the process remains untested. We attempted

195   to validate these findings in two separate datasets. Re-analysis of data from Puram et al (scRNAseq from 542

196   CD8+ T cells) showed that expression levels of SOX4 and RBPJ were higher in dysfunctional CD8 cell populations,

197   while DUSP4 expression was more generalized **(Figure 5A and Supplementary Figure 5A-C)**[12]. The second

198   scRNAseq dataset comprised T-cells obtained from cutaneous squamous-cell carcinoma patients before and

199   after treatment with PD1-blockade (**Supplementary Figure 5D**)[28]. Here, all three genes showed higher

200   expression in the exhausted CD8 subpopulation in this dataset **(Figure 5B and Supplementary Figure 5E)**.

201   However, only levels of *SOX4* and *DUSP4* were reduced after PD1 blockade, where there is expected re-

202   activation of tumor-targeting clones and reduction in the exhaustion phenotype **(Figure 5C)**. Combining these

203 results, SOX4 appears to be the most likely gene associated during the transition between pre-dysfunction to

204 dysfunction/exhaustion. To test whether *SOX4* plays a causative role in T-cell dysfunction, we performed RNAi-

205 based knock-down on activated PBMCs. Cells were transfected with Accell pooled siRNA against *SOX4*, *DUSP4*

206 or non-targeting siRNA as controls, activated with anti-CD3/CD28 microbeads and harvested for flow cytometry.

207 Remarkably, *SOX4* knockdown resulted in a reduction in senescent CD57+ and dysfunctional PD1+ and CD39+

208 populations, compared to *DUSP4* and control siRNAs **(Figure 5D and Supplementary Figure 5F-G)**. Taken

209 together, these data provide functional validation for our CD8+ T-cell trajectory mapping and implicates SOX4

210 as an important driver of T-cell dysfunction/exhaustion.

211

212 **Establishing clonal architecture in CD8+ T-cells using single-cell T-cell receptor sequencing**

213 Clonal identifiers obtained by TCR analysis allows for elucidation of CDR3 sequences as well as providing a unique

214 dataset to infer the lineage structure of T-cells. Specifically, our current dataset can be used to model clonal

215 selection and amplification across the CD8+ T-cell subpopulations and trajectories. We recovered productive

216 TCR-alpha and TCR-beta sequences from 1,461 and 1,948 cells, respectively, and identified 1,590 unique TCR

217 sequences. No shared clones were found between patients, with unique TCRs for each patient. Clonal expansion

218 was seen in 17.39% of CD8+cells, and clone size ranged from 2 to 60 cells per clone **(Figure 5E, Supplementary**

219 **Figure 5H and 5I**). Clonal overlap between the two different sites for each tumor (primary and lymph node) was

220 demonstrated in patients HN257 and HN272 **(Figure 5F)**. There was a progressive increase in clonality across the

221 dysfunctional gradient, with evidence of single naïve or TRM-derived clones subsequently expanding to give rise

222 to multiple dysfunctional clones that span these trajectories **(Figure 5F and 5G).**

223 There appeared to be patient-specific biases for one trajectory over the other. For example, there are CD8+ T-

224 cell clones in patient HN272 that followed a naïve-dysfunction trajectory (Trajectory 1), with expansion of lymph

225 node derived naïve clonotypes, migrating to the primary site and captured there along a dysfunctional gradient

226 (pre-dysfunctional, proliferative and then late-dysfunction) **(Figure 5F)**. This supports a previous observation

227 which suggests that circulation is one of the major sources of tumor-targeting dysfunctional cells, which in this

228 case is the regional lymphatics draining nodal tissue[28]. In contrast, in patient HN263 and selected CD8+ T-cell

229 clones in patient HN272, the dysfunctional gradient appears to comprise of tissue resident memory (TRM) cells

230 derived from the primary tumor, which amplified into putative tumor-targeting clonotypes **(Figure 5F)**. This is

231 consistent with a model of ongoing differentiation and proliferation of dysfunctional T-cells at the tumor site

232 itself[29]. It is likely that both mechanisms contribute to the dysfunction gradient, sometimes even within the same

233 patient. For example, lineage tracing in HN257 and HN272 demonstrates extensive trafficking and interplay

234 between the primary site and lymph node: there is evidence of lymph node-derived naïve cells expanded in the

235 primary site as expected, but also surprisingly TRM cells expanding and subsequently migrating to the lymph

236 node **(Figure 5F and 5G)**. This scTCR data adds intriguing complexity to concepts of clonal expansion and lineage

237 structure in a treatment naïve setting.

238

239 **Pre-nodal cells and immune micro-environment**

240 Our analyses identified a pre-nodal sub-population in primary tumors with intrinsic properties of invasion and

241 migration. However, metastasis also requires acquisition of an immune evasion phenotype. To test whether the

242 pre-nodal cells identified above demonstrated specific immune-modulatory phenotypes, we subjected three

243 tumors (from our study) and two tumors (from the Puram dataset) each with a minimum RNAseq dataset to

244 interactome analyses using Cellchat. To do this, we divided primary tumor cells into two subpopulations (primary

245 and pre-nodal) and analyzed the interactions of these two tumor subpopulations with CD8+, CD4+ and T-reg

246 lymphocytes and TAMs. For HN251, HN272 and HN279, the analysis showed similar trends in primary to pre-

247 nodal malignant cells, with increasing interactions between the pre-nodal subpopulation and T-lymphocytes,

248 specifically with CD8+ cells **(Figure 6A)**. The analyses implicated a number of pathways that were differentially

249 modulated by primary versus pre-nodal populations on T-lymphocytes **(Supplementary Figure 6A-C)**. In

250 particular, the interaction between Midkine (MDK, secreted by tumor cells) and a number of MDK-receptors

251 (ITGA4, ITGA6, ITGB1, NCL, LRP1) on CD8+ T-cells appears to be a recurrent immunosuppressive pathway seen

252 across all three patients **(Figure 6B)**. Applying the same approach to the external dataset also implicated the

253 MDK pathway as being differentially activated by the pre-nodal population in one (p17) out of two tumors tested

254 **(Figure 6B and Supplementary Figure 6D-E)**.

255 Recent published data suggest that MDK-driven modulation is important for immune evasion in melanomas with

256 activation of NFKB and its downstream pathways[30]. To test whether MDK-driven immune suppression dampens

257 the effect of immune checkpoint blockade (ICB) therapy, we developed a humanized mouse model engrafted

258 with pre-nodal cells from the tumor of patient HN279, and treated these with PD1-blockade. As expected, the

259 majority of cancer-cells expressed *MDK* **(Figure 6C-6D and Supplementary Figure 6F-G)**, together with a number

260 of genes associated with the pre-nodal phenotype (*eg SNAI2, AXL, STAT2*) that were unaffected by ICB **(Figure**

261 **6E and Supplementary Figure 6H)**. In contrast, expression of *AURKB* and *TOP2A* (cell cycle genes) in cancer cells

262 was significantly downregulated after pembrolizumab treatment **(Figure 6E)**, indicating a reduction in cancer

263 cell proliferation.

264 Analyses of the CD8+ T-cell fraction revealed naïve, TRM, transitional, proliferative and dysfunctional/exhausted

265 subpopulations, with an additional cytotoxic populations (likely bystander) **(Figure 6F and Supplementary Figure**

266 **6I)**. CD8+ cells from mice treated with pembrolizumab showed reduction in naïve, dysfunctional and memory

267 with concomitant increase in proliferative, cytotoxic/bystander, tissue resident subpopulations compared to

268 untreated mice **(Figure 6G)**. These changes suggest a re-invigoration and re-activation of dysfunctional and

269 memory, respectively, into tumor-targeting cells[29]. Remarkably, analyses of MDK receptor-expressing CD8 cells

270 (*ITGA4*, *ITGB1, NCL*) showed the opposite trend, with an increase in dysfunctional and reduction in the

271 proliferative (tumor-targeting) populations **(Figure 6H and Supplementary Figure 6J)**. These findings suggest

272 MDK-signaling promotes immune-suppression, that abrogates re-invigoration by PD1-blockade. Indeed, these

273 changes were also associated with NFKB1 activation which is significantly higher in the dysfunctional CD8

274 population after pembrolizumab treatment **(Figure 6I)**. Moreover, plotting the expression levels of several MDK-

275 receptors (*ITGA4, ITGB1*, NCL) with *NFKB1* show a good correlation in gene expression in CD8+ T-cells where the

276   RNA could be quantified **(Figure 6J).** Taken together, these results implicate MDK-signaling as a pathway through

277   which pre-nodal cells evade CD8-mediated immune-editing.

278

279

280

281

Discussion

Currently available algorithms analyzing single-cell data have the ability to construct evolutionary trajectories, which are especially powerful in studying specific events in space (*eg* relationships between different tumor sites, eg primary vs lymph node metastasis) and time (*eg* pre- and post- treatment analysis)[12,28]. Here, we applied these to explore early lymph node metastasis across tumor and immune sub-compartments within the tumor. Analysis of tumor cells shows that nodal metastasis is an early event, where canonical epithelial-to-mesenchymal transition is less apparent than postulated. Our findings support previous studies that suggest EMT is not an all-or-none phenomenon, but instead occurs in graded levels [31,32]. This is in contrast to *in vitro* systems (including our own) where cultured tumor cells from lymph nodes display more canonical features of EMT[33]. Despite overlap between tumor cells derived from primary and nodal sites, trajectory mapping could define evolutionary pathways at individual tumor levels, although this process require a combination of trajectory algorithms, scoring for aggressiveness (based on EMT and stemness) and knowledge of the ground truth. These have expanded the results of previous studies in the identification of a pre-nodal or metastatic population [12], and importantly identified actionable drivers of that could be targeted for anti-metastatic therapy, in this case AXL and AURK. Targeting AXL would not only prevent pathways involved in dissemination, but presumably reduce tumor heterogeneity by targeting the specific clones[34]. The role of aurora kinases is less clear; rather than impacting the metastatic process, it is possible that this vulnerability reflects a generalized reduction in cell cycling that occurs during EMT with a concomitant sensitivity to all cell cycle inhibitors. We recently demonstrated the same phenomenon during drug resistance: reduction in cell proliferation, limited AURK expression and sensitivity to inhibitors of AURK and other cell cycle targets[35]. Nevertheless, the ability to profile tumors and identify vulnerabilities in metastasis-inducing clones is an attractive notion, with increasing interest in low-dose, long term anti-metastatic therapy.

Alignment of CD8+ T lymphocyte populations is driven by existing knowledge on T-cell maturation. The fact that we could pool data across different patients increased the number of cells available and in itself was a form of validation. The alignment was further supported by single-cell VDJ sequencing, which reinforced trajectories from naïve or memory populations, towards clonally expanded, dysfunctional and potentially tumor targeting CD8+ subpopulations These supported the notion that both naïve CD8+ cells from adjacent lymph nodes and tissue resident CD8+ T-cells were sources of expanded tumor-infiltrating CD8+ T cells, and trafficking was bidirectional. Strikingly, this trajectory could be used to identify novel modulators of T-cell dysfunction by studying gene expression changes along pseudotime, and was used to identify SOX4 as novel driver of dysfunction in CD8 cells. Interactome analyses performed to identify signaling networks within CD8+ T cells during early metastasis converged onto the MDK pathway. Remarkably, in a humanized mouse model, MDK signaling was associated with a reduced ability to reinvigorate exhausted T-cells. This is supported by a recent publication which identified that the MDK pathway could abrogate immune reactivation by ICB therapy in melanoma, and this could be reversed using MDK-specific inhibitors[30]. In a similar context, MDK-inhibition could be explored in the prevention and treatment of tumor metastasis in HNSCC and add synergy to PD1-blockade which is the current standard of care in metastatic HNSCC.

319   In conclusion, we applied single-cell genomics to uncover pathways and mechanisms that mediate early nodal

320   metastasis in HNSCC. The data presented here shows that early metastasis is a much more nuanced process

321   than previously presumed. Collectively these indicate the discovery potential of single cell studies and existing

322   computational tools, when applied to specific clinical contexts and questions. Future studies will focus on more

323   specific tumor subpopulations including CD8+ cells and the impact of treatment on tumor recurrence and

324   metastasis.

325

326

Methods

Tumor collection and processing

Patient tumors were harvested in the operating room and transported to the lab for processing within 30 minutes. Tumors were *a priori* confirmed histologically to be HNSCC and patients were consented prior to surgery. This study is approved by SingHealth Centralized Institutional Review Board (CIRB: 2014/2093, 2018/2512, 2016/2757). All tumors were dissociated using the gentleMACS™ Octo system (Miltenyi Biotech, Bergisch Gladbach, Germany) as described in manufacturer's protocol. These were subjected to filtration, washing and magnetic bead separation, where required, prior to single cell capturing (details in Supplementary methods).

Patient-derived cell cultures

Cultures were established as previously described[20,21]. Cells were maintained in complete RPMI (C/RPMI) containing 10% FBS, 1% pen-strep, 1% anti-mycotic and a humidified incubator at 37 °C with 5% CO2. All lines were tested and confirmed to be free of mycoplasma using an EZ-PCR Mycoplasma Detection Kit (Biological Industries, Kibbutz Beit Haemek, Israel) at the time of experiments. Cells were processed for scRNAseq and for immunostaining as described in Supplementary methods. Invasion assays cells were treated with or without 0.25 μM of bemcentinib (BGB324) or 0.25 μM of barasertib (both from Selleck Chem, Houston, TX), then seeded on an 8μm filter membrane within a 24-well transwell insert (Corning, New York City, NY), with C/RPMI at bottom of wells of 24-well Falcon TC Companion Plate (Corning, New York City, NY). After 72hrs, the bottom of each inserts was fixed and stained for quantification of invaded cells. Cell invasion area was determined by quantifying the area with crystal violet staining using the ImageJ software.

Humanized mouse model

Sixteen NOG-EXL (hGM-CSF/hIL-3 NOG) mice (hNOG-EXL), pre-engrafted with human CD34+ hematopoietic stem cells, were procured from CIEA-SIgN. At 16 weeks post-engraftment, mice were injected subcutaneously with cells from HN279, and treated intraperitoneally with 12.5mg/kg of pembrolizumab or with phosphate buffered saline (PBS) on day 17, 19, 21 and 24. Mice were euthanized on day 25 and tumors harvested for dissociation and preparation for scRNAseq as described.

Small-interfering RNA knock-down of SOX4 and DUSP4

Peripheral blood mononuclear cells (PBMCs) from healthy donors were cultured at a density of 0.5-1 x 10^6 cells/ml in 24-well plate (Corning, New York City, NY), containing TexMACS Medium and T-Cell TransAct (both from Miltenyi Biotech, Bergisch Gladbach, Germany) at 1:200 dilution. A final concentration of 1μM of Accell pooled small-interfering RNA (siRNA) targeting human SOX4 (Gene ID 6659) or DUSP4 (Gene ID 1846), or non-

361   targeting siRNA (all from Dharmacon, Lafayette, CO) was added into respective wells. After 5 days of incubation,

362   cells were harvested for flow cytometry.

363

364   *Flow cytometry*

365   For AXL surface staining, trypsinized cells were stained with fluorochrome-conjugated antibody recognizing AXL

366   (#108724; R&D systems, Minneapolis, MN) or with mouse IgG1 isotype antibody (MOPC-21; BD Biosciences,

367   Franklin Lakes, NJ). For intracellular AURKB staining, trypsinized cells were fixed and permeabilized with a

368   Foxp3/Transcription Factor Staining Buffer Set (eBioscience, San Diego, CA) according to the manufacturer

369   protocol. After fixation, cells were stained with primary antibody recognizing AURKB (clone RM278; Invitrogen,

370   Carlsbad, CA) or rabbit IgG1 isotype antibody (DA1E; R&D systems, Minneapolis, MN), and subsequently with

371   goat anti-rabbit IgG secondary antibody conjugated to Alex Fluor 647 (#A32733; Waltham, MA). For siRNA

372   knock-down PBMC experiments, harvested cells were stained with fluorochrome-conjugated antibodies

373   recognizing CD57 (HNK-1), LAG3 (11C3C65), CD39 (A1) and CD4 (OKT4) all from Biolegend, San Diego, CA; PD1

374   (J105) and CD8 (SK1) from eBioscience, San Diego, CA; and CD4 (SK3) from BD Biosciences, Franklin Lakes, NJ.

375   These cells were stained for 30mins on ice in the dark with 2% BSA in PBS. Live/dead cells were distinguished

376   using a Fixable Live Dead Blue Dead Cell Stain Kit (Thermofisher, Waltham, MA). Cells were acquired and

377   analyzed using a BD FACSCanto II instrument and FlowJo v10.5.3 software (both from BD Biosciences, Franklin

378   Lakes, NJ) respectively.

379

380   *Generation of single cell gene expression and TCR libraries by droplet-based (10x system) and microfluidic-based*

381   *technologies*

382   The 5' gene expression (GEX) and TCR single cell RNA libraries from tumors were prepared using the 10x

383   Chromium Single Cell V(D)J Reagent Kits (10x Genomics, Pleasanton, CA), as described in the manufacturer's

384   protocol. Briefly, freshly dissociated tumor cells were sorted into CD45+ and CD45- fractions, mixed at a 1:1 ratio

385   and loaded into the Single Cell A Chip for gel bead-in-emulsion (GEM) generation and barcoding, targeting for a

386   cell recovery of 4000-7000 cells per sample. Reverse transcription, cDNA amplification, GEX and TCR library

387   construction were performed as described. For C1, single cell suspensions were loaded and captured using

388   medium-sized (10-17um) Fluidigm Integrated Fluidic Circuit (IFC) and a Fluidigm C1 instrument (Fluidigm, South

389   San Francisco, CA), according to the manufacturer's protocol. cDNA product was harvested from the IFC,

390   barcoded for individual cell identity and pooled. Sequencing was performed by an Illumina Hiseq 4000 (Illumina,

391   San Diego, CA) with 151-bp single-ended or pair-ended reads.

392

393

394   *Data processing of single-cell RNA-seq libraries and clustering*

395   scRNAseq reads were aligned to the GRCh38 reference genome and quantified using Cellranger count (10x

396   Genomics, version 2.2.0). Downstream analyses were performed using Seurat (version 3.1.5). For malignant-cell

397   analysis, we isolated subsets of cells identified as malignant cells based on broad clustering and reprocessed

398   using Seurat without patient alignment, since tumor cells tend to be patient specific. For T-cell clustering, we

399   isolated subsets of cells identified as T-cells based on broad clustering. Cells were then re-clustered using Seurat

400   alignment across patients similar as with previous analysis. For CD8+ T-cell clustering, CD8+ T-cells were

401   extracted from the T-cell clustering based on the following two criteria: 1) in Pre-dysfunctional, Dysfunctional

402   and Proliferative clusters, and with zero CD4 expression, 2) in Naïve-like, Memory and Transitional clusters, with

403   zero CD4 and positive CD8 (either CD8A or CD8B) expression. TCR reads were mapped to

404   vdj_GRCh38_alts_ensembl-3.1.0-3.1.0 reference genome and quantified using cellranger count (10x Genomics,

405   version 3.1.0). Further details on Seurat analysis, UMAP visualization and use of the following algorithms:

406   InferCNV, Monocle, PAGODA, Slingshot, functional annotation, Geneswitches, Cytotrace and other analyses

407   tools, are described in detail in Supplementary methods.

408

409   *Statistical analysis*

410   Statistical analysis was performed using GraphPad Prism software (GraphPad Software, Inc., San Diego, CA), or

411   otherwise indicated in the figure legends and Supplementary methods.

412

References

433   1   Steeg, P. S. Targeting metastasis. *Nat Rev Cancer* **16**, 201-218, doi:10.1038/nrc.2016.25 (2016).

434   2   Chaffer, C. L. & Weinberg, R. A. A perspective on cancer cell metastasis. *Science* **331**, 1559-1564, doi:10.1126/science.1203543 (2011).

436   3   Gupta, G. P. & Massague, J. Cancer metastasis: building a framework. *Cell* **127**, 679-695, doi:10.1016/j.cell.2006.11.001 (2006).

438   4   Chow, L. Q. M. Head and Neck Cancer. *N Engl J Med* **382**, 60-72, doi:10.1056/NEJMra1715715 (2020).

439   5   Iyer, N. G. *et al.* Randomized trial comparing surgery and adjuvant radiotherapy versus concurrent chemoradiotherapy in patients with advanced, nonmetastatic squamous cell carcinoma of the head and neck: 10-year update and subset analysis. *Cancer* **121**, 1599-1607, doi:10.1002/cncr.29251 (2015).

443   6   Mehlen, P. & Puisieux, A. Metastasis: a question of life or death. *Nat Rev Cancer* **6**, 449-458, doi:10.1038/nrc1886 (2006).

445   7   Heerboth, S. *et al.* EMT and tumor metastasis. *Clin Transl Med* **4**, 6, doi:10.1186/s40169-015-0048-3 (2015).

447   8   Brabletz, T., Kalluri, R., Nieto, M. A. & Weinberg, R. A. EMT in cancer. *Nat Rev Cancer* **18**, 128-134, doi:10.1038/nrc.2017.118 (2018).

449   9   Williams, E. D., Gao, D., Redfern, A. & Thompson, E. W. Controversies around epithelial-mesenchymal plasticity in cancer metastasis. *Nat Rev Cancer* **19**, 716-732, doi:10.1038/s41568-019-0213-x (2019).

451   10   Vettore, A. L. *et al.* Mutational landscapes of tongue carcinoma reveal recurrent mutations in genes of therapeutic and prognostic relevance. *Genome Med* **7**, 98, doi:10.1186/s13073-015-0219-2 (2015).

453   11   Tan, D. S. *et al.* Tongue carcinoma infrequently harbor common actionable genetic alterations. *BMC Cancer* **14**, 679, doi:10.1186/1471-2407-14-679 (2014).

455   12   Puram, S. V. *et al.* Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **171**, 1611-1624 e1624, doi:10.1016/j.cell.2017.10.044 (2017).

457   13   McFaline-Figueroa, J. L. *et al.* A pooled single-cell genetic screen identifies regulatory checkpoints in the continuum of the epithelial-to-mesenchymal transition. *Nat Genet* **51**, 1389-1398, doi:10.1038/s41588-019-0489-5 (2019).

460   14   Wei, S. C., Duffy, C. R. & Allison, J. P. Fundamental Mechanisms of Immune Checkpoint Blockade Therapy. *Cancer Discov* **8**, 1069-1086, doi:10.1158/2159-8290.CD-18-0367 (2018).

462   15   Mempel, T. R., Henrickson, S. E. & Von Andrian, U. H. T-cell priming by dendritic cells in lymph nodes occurs in three distinct phases. *Nature* **427**, 154-159, doi:10.1038/nature02238 (2004).

464   16   van der Leun, A. M., Thommen, D. S. & Schumacher, T. N. CD8(+) T cell states in human cancer: insights from single-cell analysis. *Nat Rev Cancer*, doi:10.1038/s41568-019-0235-4 (2020).

466   17   Dunn, G. P., Old, L. J. & Schreiber, R. D. The three Es of cancer immunoediting. *Annu Rev Immunol* **22**, 329-360, doi:10.1146/annurev.immunol.22.012703.104803 (2004).

468   18   Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J. & Schreiber, R. D. Cancer immunoediting: from immunosurveillance to tumor escape. *Nat Immunol* **3**, 991-998, doi:10.1038/ni1102-991 (2002).

470   19   Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).

472   20   Chia, S. *et al.* Phenotype-driven precision oncology as a guide for clinical decisions one patient at a time. *Nat Commun* **8**, 435, doi:10.1038/s41467-017-00451-5 (2017).

474   21   Leong, H. S. *et al.* Targeting cancer stem cell plasticity through modulation of epidermal growth factor and insulin-like growth factor receptor signaling in head and neck squamous cell cancer. *Stem Cells Transl Med* **3**, 1055-1065, doi:10.5966/sctm.2013-0214 (2014).

477   22   Gao, R. *et al.* Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat Biotechnol* **39**, 599-608, doi:10.1038/s41587-020-00795-2 (2021).

479   23   Yang, J., Chen, Y., Luo, H. & Cai, H. The Landscape of Somatic Copy Number Alterations in Head and Neck Squamous Cell Carcinoma. *Front Oncol* **10**, 321, doi:10.3389/fonc.2020.00321 (2020).

481   24   Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-386, doi:10.1038/nbt.2859 (2014).

483   25   Gulati, G. S. *et al.* Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* **367**, 405-411, doi:10.1126/science.aax0249 (2020).

485   26   Cao, E. Y., Ouyang, J. F. & Rackham, O. J. L. GeneSwitches: ordering gene expression and functional
486   events in single-cell experiments. *Bioinformatics* **36**, 3273-3275, doi:10.1093/bioinformatics/btaa099
487   (2020).
488   27   Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC*
489   *Genomics* **19**, 477, doi:10.1186/s12864-018-4772-0 (2018).
490   28   Yost, K. E. *et al.* Clonal replacement of tumor-specific T cells following PD-1 blockade. *Nat Med* **25**,
491   1251-1259, doi:10.1038/s41591-019-0522-3 (2019).
492   29   Li, H. *et al.* Dysfunctional CD8 T Cells Form a Proliferative, Dynamically Regulated Compartment within
493   Human Melanoma. *Cell*, doi:10.1016/j.cell.2018.11.043 (2018).
494   30   Cerezo-Wallis, D. *et al.* Midkine rewires the melanoma microenvironment toward a tolerogenic and
495   immune-resistant state. *Nat Med* **26**, 1865-1877, doi:10.1038/s41591-020-1073-3 (2020).
496   31   Revenco, T. *et al.* Context Dependency of Epithelial-to-Mesenchymal Transition for Metastasis. *Cell*
497   *Rep* **29**, 1458-1468 e1453, doi:10.1016/j.celrep.2019.09.081 (2019).
498   32   Chen, Y. *et al.* Dual reporter genetic mouse models of pancreatic cancer identify an epithelial-to-
499   mesenchymal transition-independent metastasis program. *EMBO Mol Med* **10**,
500   doi:10.15252/emmm.201809085 (2018).
501   33   Sharma, A. *et al.* Longitudinal single-cell RNA sequencing of patient-derived primary cells reveals
502   drug-induced infidelity in stem cell hierarchy. *Nat Commun* **9**, 4931, doi:10.1038/s41467-018-07261-3
503   (2018).
504   34   Antony, J. & Huang, R. Y. AXL-Driven EMT State as a Targetable Conduit in Cancer. *Cancer Res* **77**,
505   3725-3732, doi:10.1158/0008-5472.CAN-17-0392 (2017).
506   35   Low, J. L. *et al.* A chemical genetic screen identifies Aurora kinases as a therapeutic target in EGFR
507   T790M negative, gefitinib-resistant head and neck squamous cell carcinoma (HNSCC). *EBioMedicine*
508   **64**, 103220, doi:10.1016/j.ebiom.2021.103220 (2021).

509

510
511

Figure 1. Tumor samples for single cell RNAseq (A) Workflow of sample acquisition, processing, and analyses for single cell transcriptome and TCR clonality of tumors (and patient-derived cultures) from primary and metastatic lymph nodes of HNSCC patients. Diagram was created with BioRender.com. (B) Uniform manifold approximation and projection (UMAP) of scRNAseq data from all cells within primary tumors and metastatic lymp nodes from 7 patients. Clusters are denoted by colors and labelled according to inferred cell types. Violin plots show the expression of selected genes used to define the inferred cell types. (C) Distribution of different cell types (color) for each patient sample (top) and comparing primary and metastatic samples (bottom) as indicated on the y-axis. (D) Chromosomal gains and losses prediction for malignant epithelial cells by inferCNV using non-malignant cells from respective samples as controls. Cyan indicates primary malignant epithelial; yellow indicates lymph node malignant epithelial; sample identities on the y-axis, chromosome numbers on the x-axis.

Figure 2. scRNAseq analysis of malignant epithelial cells and identification of pre-metastatic sub-population. (A) UMAP of malignant epithelial cells only, clustered by Seurat clusters (left), patients (middle), and tissue origin (primary/metastatic) (right). (B) Boxplot showing epithelial–mesenchymal transition (EMT) scores across patients and tissue origin (primary versus metastasis). Line represents mean scores, while box represents 2 standard deviations. (C) and (D) Monocle plots demonstrating the derivation of pre-metastatic populations in HN251 (C) and HN279 (D) based on (from left to right) tissue origin, monocle clusters, EMT scores, CytoTRACE scores to derive trajectory. (E) Gene ontology pathways that are significantly altered across pseudotime derived in C and D. (F) Potentially actionable genes identified to be increased in pre-metastatic population. (G) t-SNE plot of tumor cells in HN257 showing a highly aggressive sub-population in the primary tumor with high CytoTRACE scores and expression of SNAI2. (H) Gene set enrichment analysis (GSEA) showing normalized enrichment scores and (I) Kaplan-Meier plot of TCGA data showing overall survival in patients with high versus low scores based on genes expressed by the specific subpopulation in (G). Shaded area shows 95% confidence interval and p-value as indicated based on log-rank test.

Figure 3. Functional analysis of actionable genes enriched in pre-metastatic population in patient-derived cultures (PDCs). (A) Dimension reduction plots based on PAGODA for PDCs derived from matched primary and metastatic lymph nodes (nodal metastatic; M). Clusters are denoted by patient identity and site of origin (left), and Seurat clusters (right). (B) Heatmap of differentially expressed pathways (rows) across samples and tumor origin (columns), showing selected Hallmark and Gene Ontology (GO) gene sets. Bars on the top of the heat map indicate the site of sample origins, clusters and patient samples corresponding to those of (A). (C) Boxplot showing the gene expression level of *AXL* (left) and *AURKB* (right) of malignant cells from primary and metastatic PDCs for the indicated patients. Line represents mean expression, while box represents 2 standard deviations; colors and cluster numbers of the bars correspond to (A). (D) Immunocytochemistry of AXL in HN137 and AURKB in HN159 and HN220 of primary and metastatic PDCs. Scale bar indicates 100 µm. (E) Representative micrographs from Boyden chamber assays of invaded cells (purple) (top), and quantification of invaded cells (bottom) in barplots from primary and metastatic cell cultures treated with or without BGB324 or barasertib.

19

550   **p < 0.01, ***p < 0.001, ****p < 0.0001 (significant difference) using student t-test compared with untreated

551   at corresponding site of origin. Error bars represent one standard deviation. (H) Flow cytometry dot plots

552   representing anti-AXL (left) and mouse IgG1 isotype control (right) staining of primary and metastatic PDCs of

553   HN137. (I) Gating used for identification and isolation of AXL^{hi}, AXL^{mid} and AXL$^{neg/low}$ from HN137 primary PDC by

554   FACS sorting (left). Micrographs representing isolated AXL-based subpopulations treated with or without

555   BGB324 and their respective invasive potential in Boyden chamber assays (right).

556

557   **Figure 4. scRNAseq analysis of tumor infiltrating T-cells and establishing a trajectory for tumor-targeting CD8+**

558   **lymphocytes.** (A) UMAP of tumor infiltrating T-cells from primary and metastatic tumors with clusters denoted

559   by colors and labelled with inferred cell identities. (B) Heatmap of differentially expressed genes (rows) between

560   cells classified into inferred T-cell subsets. Bars on the top of the heatmap indicate the site of origin and cell type

561   corresponding to those of (A) with selected genes indicated. (C) UMAP of all CD8 T-cells from primary and

562   metastatic tumors. Clusters are denoted by colours and labelled with inferred cell identities based on (D)

563   expression of selected genes used for CD8 T-cell subset annotation for. (E) Slingshot analysis of CD8 T-cells

564   showing two potential trajectories giving rise to tumor-targeting CD8+ cells: Trajectory 1 (top)- from naïve to

565   dysfunctional and Trajectory 2 (bottom)- memory to dysfunctional. (F) Graphs showing the estimate scores of

566   curated genes related to naïve-like (*IL7R*, *TXNIP*, *SELL*, *CCR7*, *TCF7*), proliferative (*MKI67*, *HMGB2*, *TYMS*),

567   dysfunctional (*GZMB*, *GNYL*, *CTLA4*, *LAYN*, *LAG3*, *TIGIT*) populations, and expression of *CXCL13* during the

568   development of CD8 T-cell along the naïve-proliferation-dysfunction axis in Trajectory 1. (G) Geneswitches

569   output showing ordering of the top switching genes along the naïve to dysfunctional (Trajectory 1) CD8 T-cell

570   axis using. Key genes are highlighted with enlarged font size. (H) UMAP projections of expression levels for genes

571   highlighted in (G).

572

573   **Figure 5. Functional analysis of genes involved in CD8 dysfunction and T-cell receptor sequencing analysis.**

574   Violin plots showing expression of *SOX4*, *DUSP4* and *RBPJ* in CD8 T-cell subpopulations derived from published

575   cohorts of scRNAseq meta-dataset from (A) HNSCC and (B) skin squamous cell cancer[12,28]. (C) Boxplots showing

576   expression of *SOX4*, *DUSP4* and *RBPJ* in CD8 T-cell subpopulations from (B), grouped by pre- and post-

577   pembrolizumab treatment. *p < 0.05 and ****p < 0.0001 denotes a significant difference compared with pre-

578   treatment of corresponding CD8 T-cell subsets by paired t-test. (B-C) X-axis labels: CD8_mem = CD8 memory;

579   CD8_eff = CD8 effector; CD8_act = CD8 activated; CD8_ex_act = CD8 exhausted/activated; CD8_ex = CD8

580   exhausted. (D) Bar graph showing percentage of CD8 T-cells expressing CD39, CD57, LAG3 or PD1 from PBMCs

581   that were activated and cultured with siNT, siSOX4 or siDUSP4 for 5 days (n = 4). Black lines and error bars

582   represent mean ± SEM. *p < 0.05, **p < 0.01, ****p < 0.0001 (significant difference) by paired t-test compared

583   with siNT of respective markers. (E) Barplots of the percentage of TCR clone(s) detected once (n=1), twice (n=2)

584   or more than two times (n>2) across the CD8 T-cell subpopulations of all patients with HNSCC subjected to

585   scRNAseq. (F) UMAP projection of CD8 T-cells from HN272, HN263 and HN257 colored by selected TCR

586   clonotypes. (G) Schematic diagram summarizing the development and trafficking of CD8 T-cell clones between

587 primary tumor, lymph node and metastasis, and bloodstream of HN272, HN263 and HN257 based on the

588 clonotype data from (F). Diagram was created with BioRender.com.

589

590 **Figure 6. Determining the interaction between pre-metastatic malignant cells and CD8+ T lymphocyte**

591 **populations.** (A) Hierarchical plot derived from Cellchat analyses showing ligand-receptor interactions between

592 tumor cells (primary and pre-nodal subpopulations) with T-lymphocytes (CD8+, CD4+ and Treg cells) and TAMs.

593 Circle sizes are proportional to the number of cells in each cell group available for and edge width represents

594 the communication probability with number of potential ligand-receptor pair as indicated. (B) Dot (bubble) plots

595 showing significant MDK ligand-receptor pairs contributing to the signaling from primary or pre-metastatic

596 cancer cells (epithelial) to Treg, CD4 or CD8 T-cells. The dot color and size represent the calculated

597 communication probability, and p-values determined from one-sided permutation test. (C) UMAP of cells

598 derived from tumors of humanized NOG-EXL mice treated with or without anti-PD1. Clusters are denoted by

599 colors labelled with inferred cell types, with a 2D projection of MDK gene expression (inset). (D) Frequency of

600 MDK+ (blue) and MDK- (orange) malignant cells in control or anti-PD1 treated mice. (E) Expression level of

601 selected genes involved in tumor cell proliferation in malignant cells from control or anti-PD1-treated mice. *p

602 < 0.05 and **p < 0.01 indicate significant difference by unpaired t test when compared to control. (F) UMAP of

603 tumor infiltrating CD8 T-cells only extracted from (C). Clusters are denoted by colors labelled with inferred cell

604 identities. (G) Distribution of CD8 T-cell subpopulations in control vs anti-PD1 treated mice. (H) Delta (Δ)

605 percentage of CD8 T-cells expressing the specific MDK-receptors *ITGA4*, *ITGB1* or *NCL* showing changes in

606 dysfunctional, transitional and proliferating subpopulations, comparing untreated versus anti-PD1 treated mice.

607 Delta percentage is determined by the percentage of MDK receptor+ CD8+ T-cells from anti-PD1 treated mice

608 minus that of the control mice. (I) Expression of *NFKB1* in the three CD8 subpopulations in controls and anti-PD1

609 treated mice. *p < 0.05 indicates significant difference by unpaired t test when compared to control. (J)

610 Scatterplot showing the correlation of expression between *NFKB1* with the following MDK receptor(s): *ITGA4*,

611 *ITGB1* and/or *NCL* in the dysfunctional CD8 T-cells subpopulation. Each dot represents one dysfunctional CD8 T-

612 cell from control (red) or anti-PD1 (blue) treated mice. The R and p values were determined using Pearson
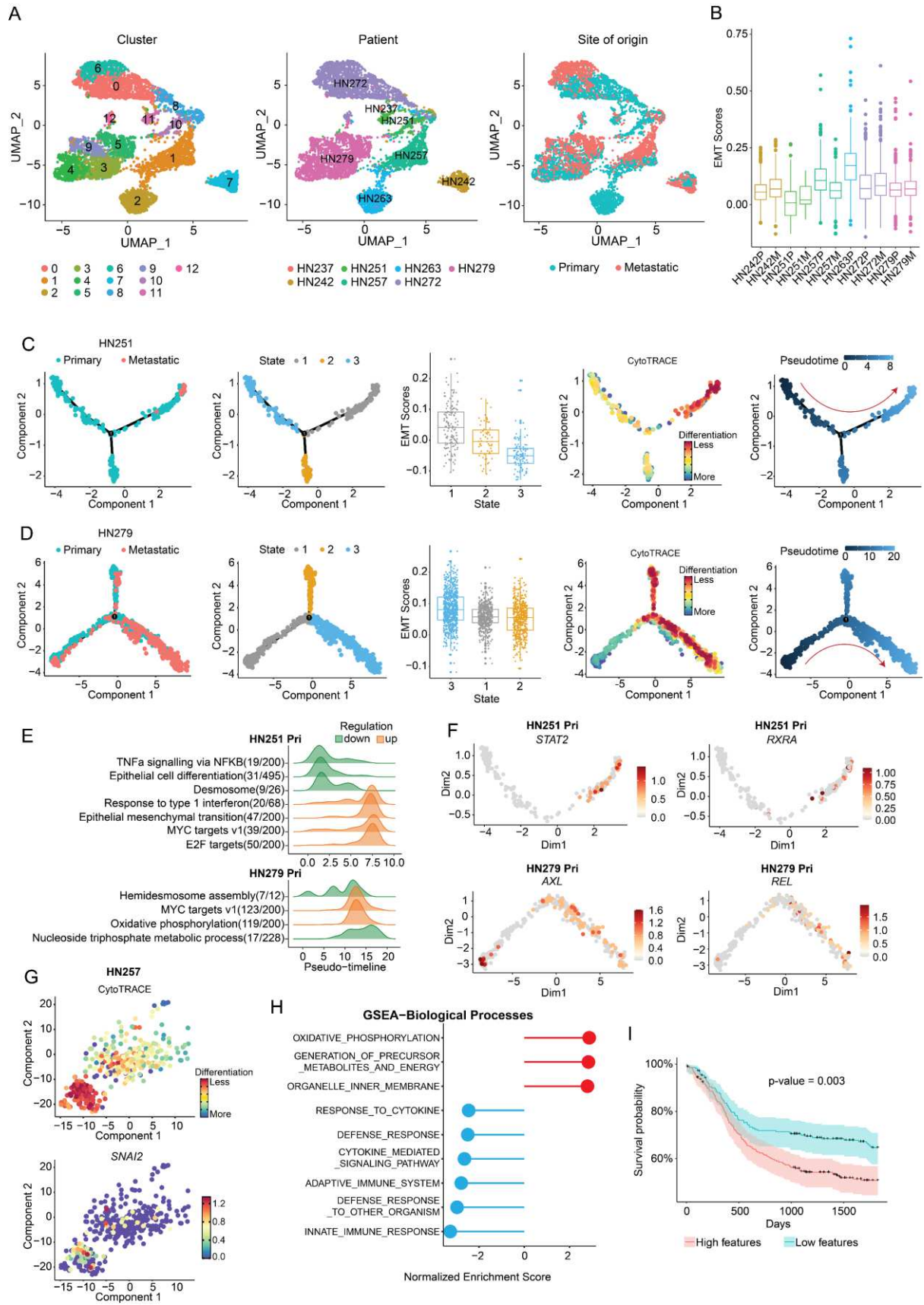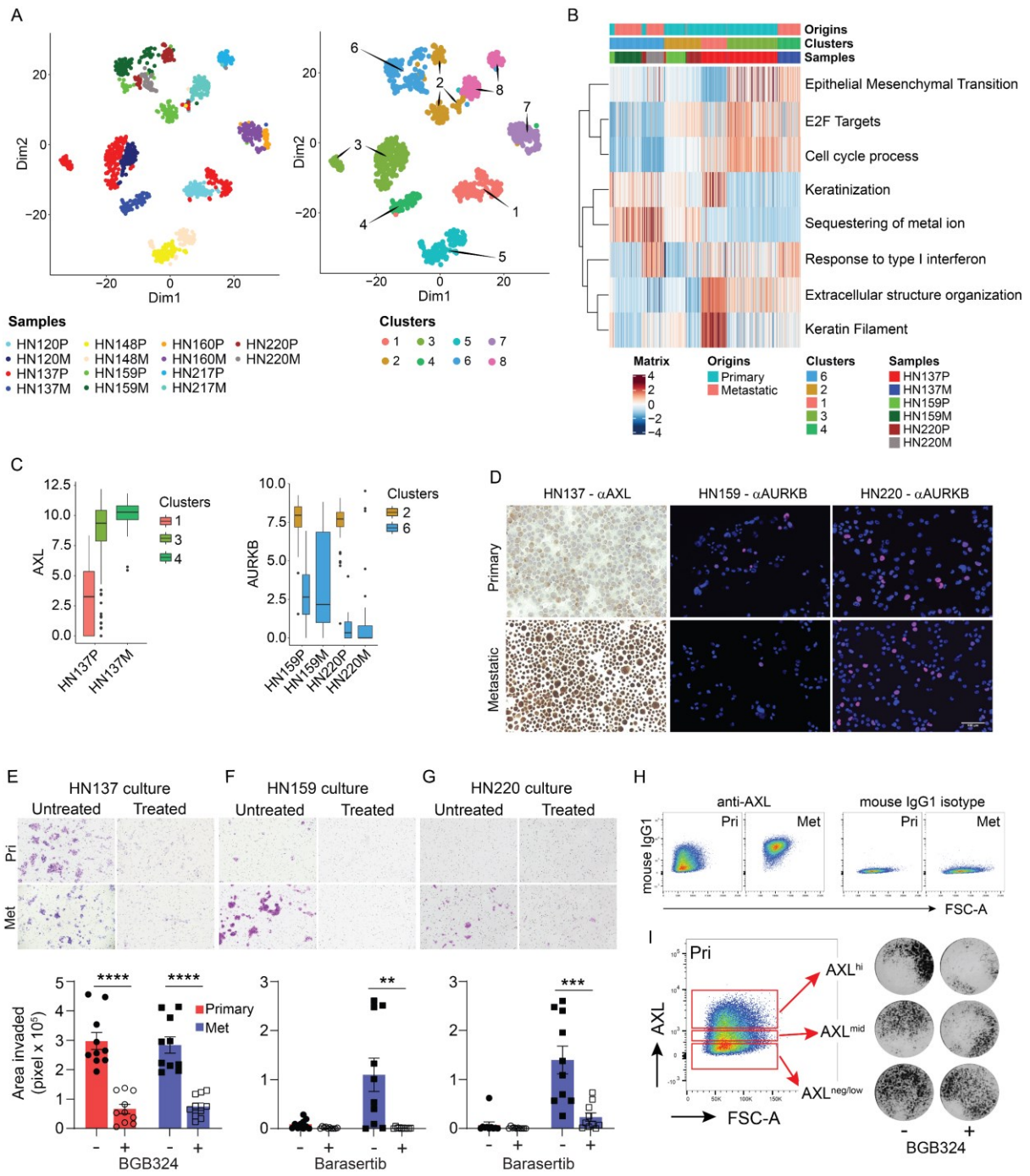
613 correlation statistical analysis.
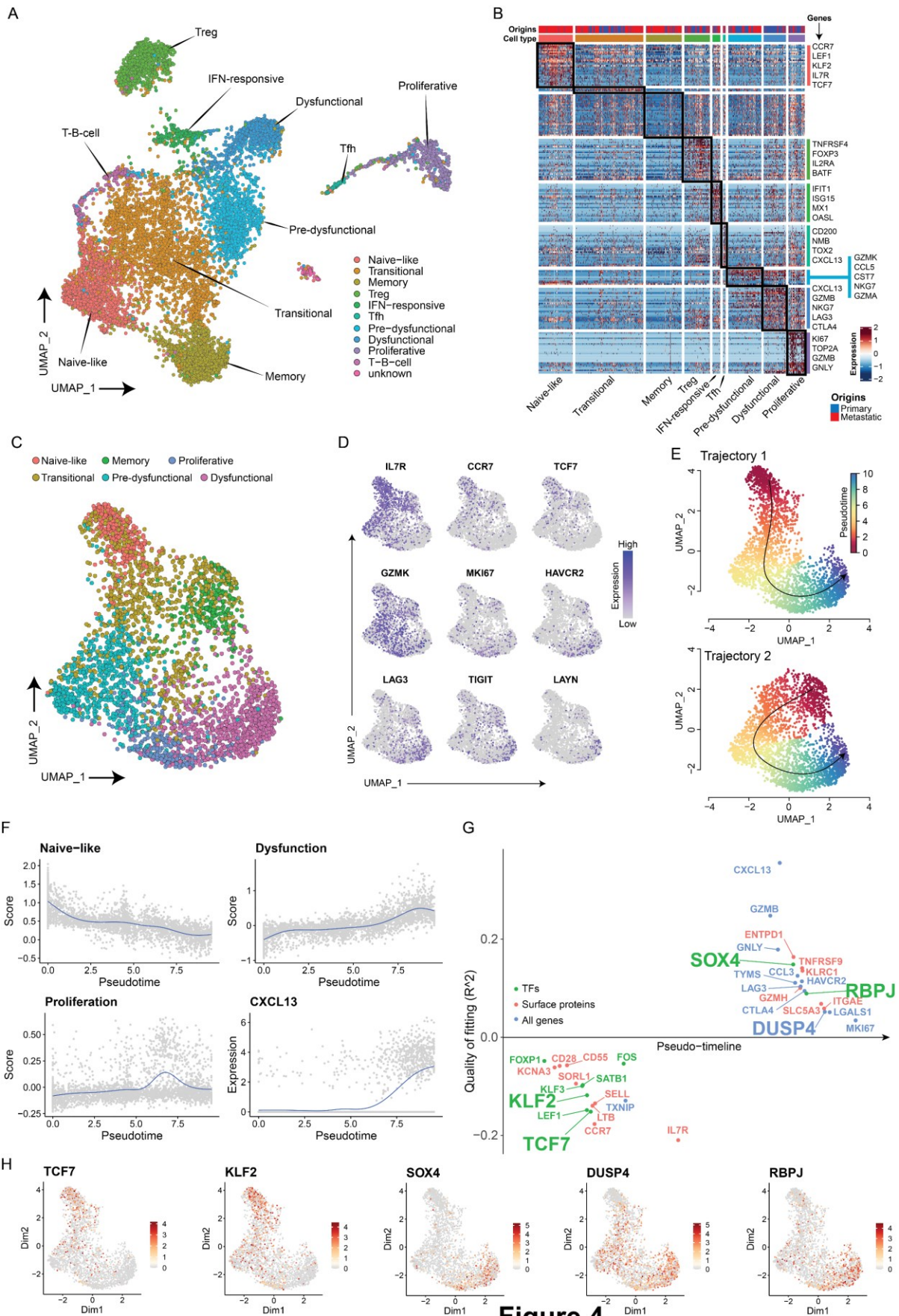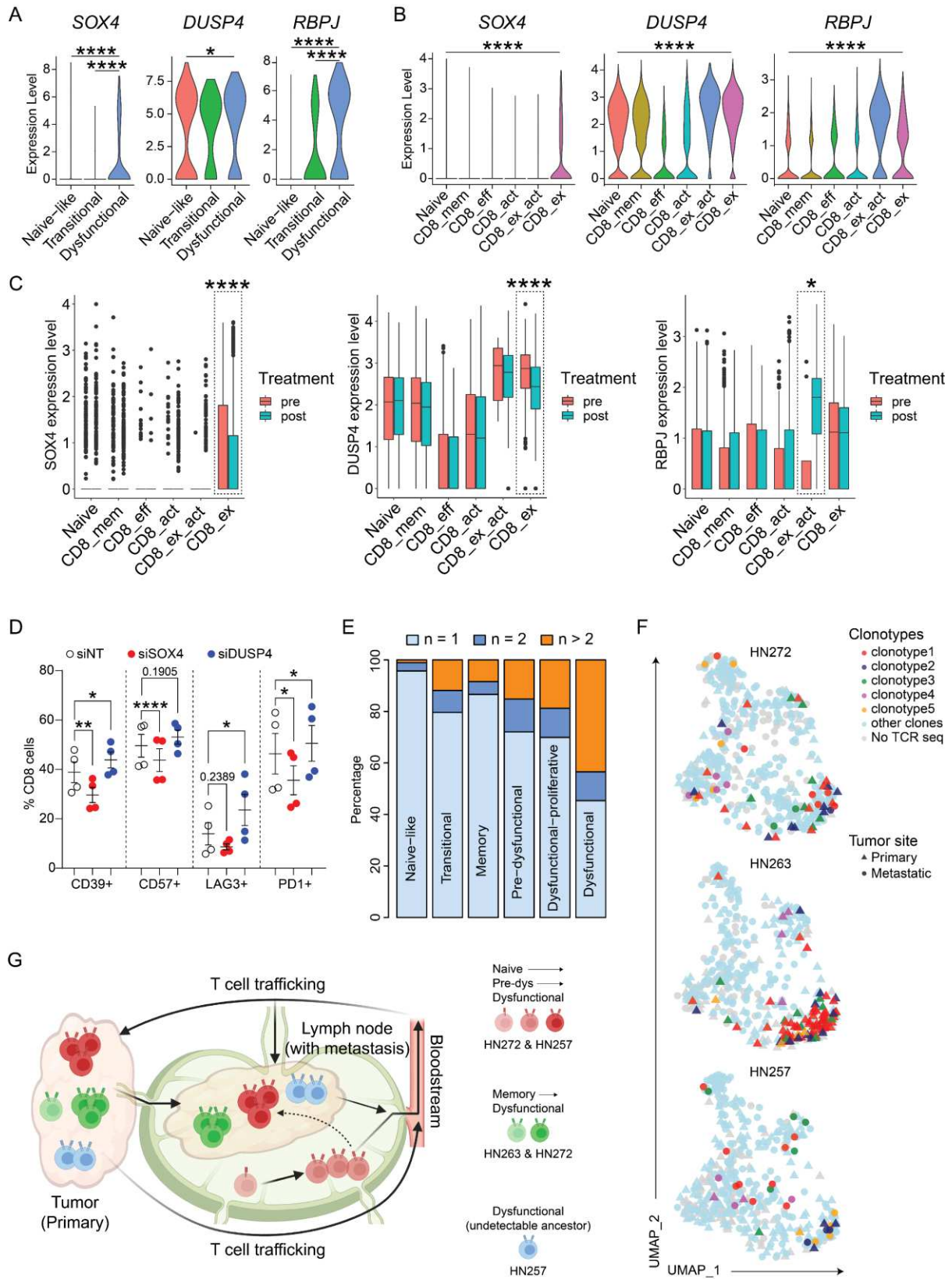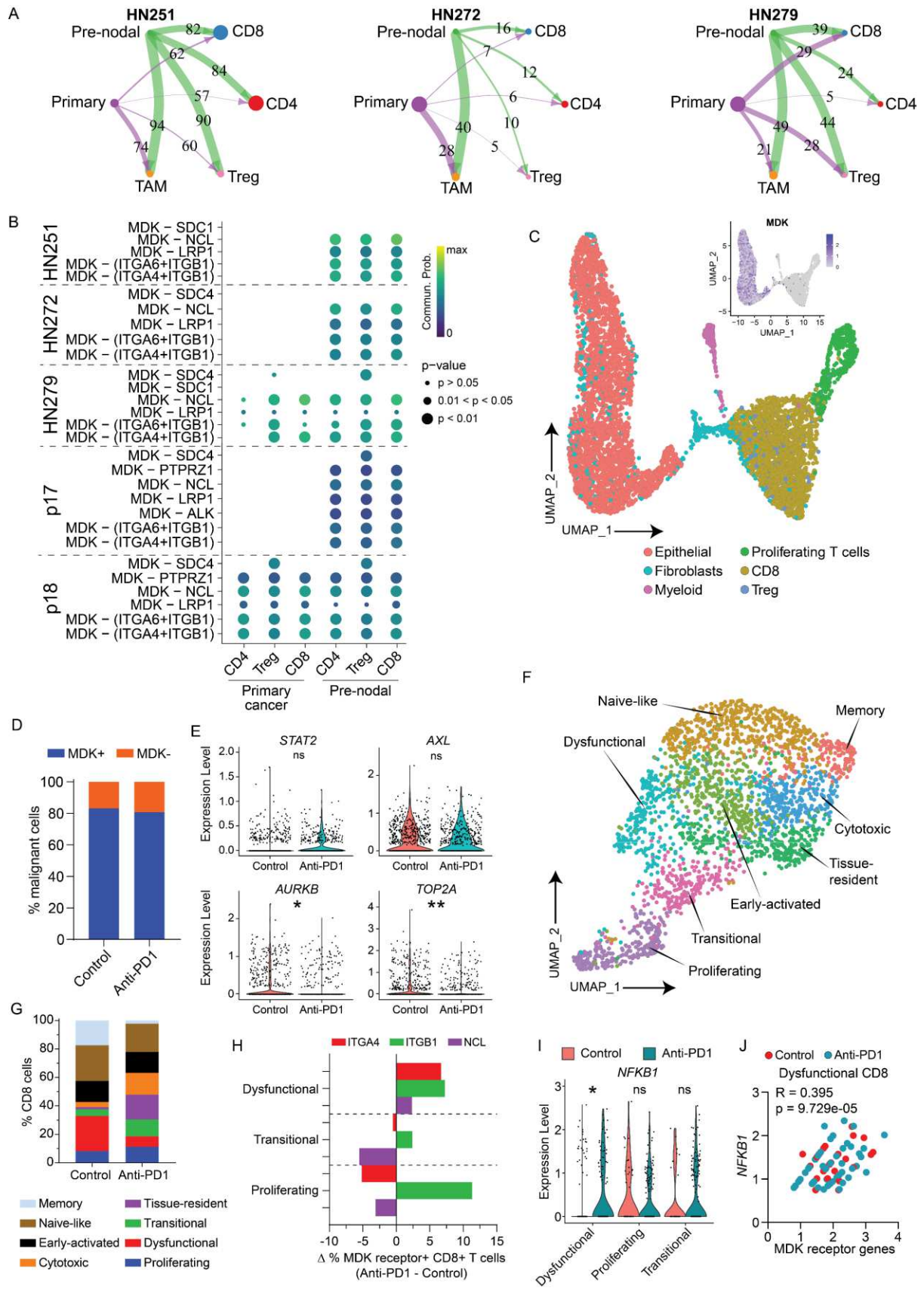
614
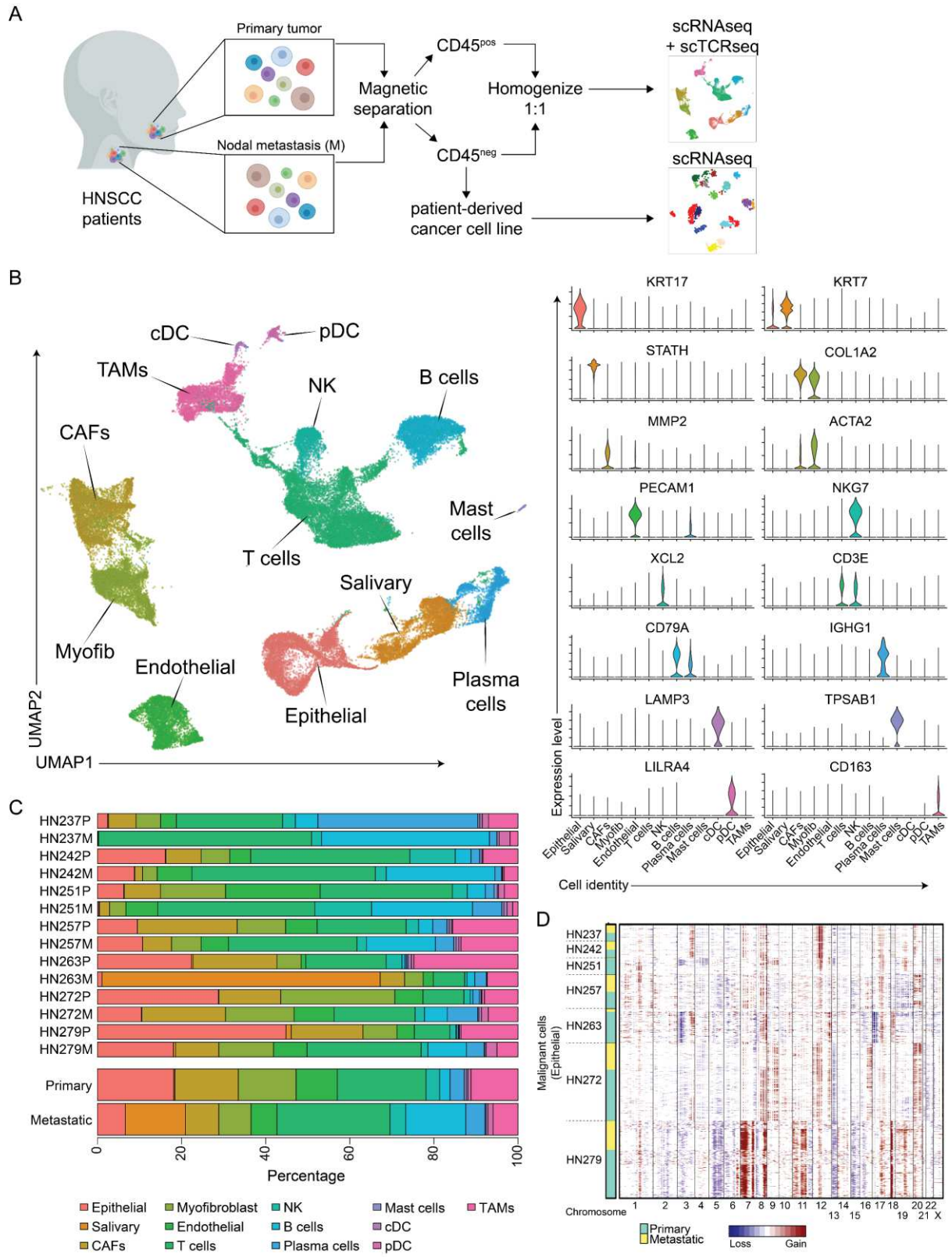
Figure 1

Figure 3

Figure 4

Figure 5

Figure 6

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigure16v3.pdf
- SupplementaryTables.pdf
- scSupplementaryMethodsv2.pdf