Single-cell analysis supports a luminal-neuroendocrine trans-differentiation in human prostate cancer

Authors: Baijun Dong[b#], Juju Miao[a,c#], Yanqing Wang[b#], Wenqin Luo[a], Zhongzhong Ji[a,c], Huadong Lai[a,c], Man Zhang[a,c], Xiaomu Cheng[a,c], Jinming Wang[b], Yuxiang Fang[a,b], Helen He Zhu[a,b], Chee Wai Chua[a,b], Liancheng Fan[b], Yinjie Zhu[b], Jiahua Pan[b], Jia Wang[a,b]*, Wei Xue[b]*, Wei-Qiang Gao[a,c]*

Affiliations:

[a] State Key Laboratory of Oncogenes and Related Genes, Renji-Med-X Stem Cell Research Center, Department of Urology, Ren Ji Hospital, School of Medicine and School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, 200127, China

[b] Department of Urology, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200127, China;

[c] School of Biomedical Engineering & Med-X Research Institute, Shanghai Jiao Tong University, Shanghai 200030, China;

[#]**These authors contributed equally to this work.**

***Correspondence author**: Jia Wang (wj860520@163.com), Wei Xue (xuewei@renji.com) or Wei-Qiang Gao (gao.weiqiang@sjtu.edu.cn).

Tel: 86-21-68383917, Fax: 86-21-68383916.

Address:

[a,c] Stem Cell Research Center, Ren Ji Hospital, 160 Pujian Rd., School of Medicine, Shanghai Jiao Tong University, Shanghai, 200127, China.

[b] Department Urology, Ren Ji Hospital, 160 Pujian Rd., School of Medicine, Shanghai Jiao Tong University, Shanghai, 200127, China.

Keywords: Castration-resistant prostate cancer, Neuroendocrine prostate cancer, Neuroendocrine differentiation, Single-cell RNA sequencing,

Running title:

Single-cell analysis of NEPC

Abstract

Neuroendocrine prostate cancer is one of the most aggressive subtypes of prostate tumor. Although much progress has been made in understanding the development of neuroendocrine prostate cancer, the cellular architecture associated with neuroendocrine differentiation in human prostate cancer remain incompletely understood. Here, we use single-cell RNA sequencing to profile the transcriptomes of 21,292 cells from needle biopsies of 6 castration-resistant prostate cancers. Our analyses reveal that all neuroendocrine tumor cells display a luminal-like epithelial phenotype. In particular, lineage trajectory analysis suggests that focal neuroendocrine differentiation exclusively originate from luminal-like malignant cells rather than basal compartment. Further tissue microarray analysis validates the generality of the luminal phenotype of neuroendocrine cells. Moreover, we uncover neuroendocrine differentiation-associated gene signatures that may help us to further explore novel intrinsic molecular mechanisms deriving neuroendocrine prostate cancer. In summary, our single-cell study provides direct evidence into the cellular states underlying neuroendocrine transdifferentiation in human prostate cancer.

Introduction

Lineage plasticity endows cancer cells with the ability to switch their cellular phenotype [1] and is often associated with more aggressive stages of cancers [2]. In prostate cancer, lineage plasticity contributes to the acquisition of the neuroendocrine (NE) phenotype [3–5], with the emergence of a highly aggressive variant, termed neuroendocrine prostate cancer (NEPC) [6]. Current studies support that NEPC tumors arise clonally from prostate adenocarcinoma (PCA) [7], accompanying with a phenotypic transition from acini epithelial tumor cells to NE-like tumor cells [8]. This lineage transition enables tumor cells to evade androgen receptor (AR) pathway inhibitors such as enzalutamide by shedding their dependence on the AR pathway [4,9]. Consequently, tumors develop resistant to the traditional androgen dependent therapy (ADT) and thus became one subtype of castration-resistant prostate cancer (CRPC), leading to the most lethal stages of this disease [10,11]. Therefore, understanding the cellular and molecular basis underlying neuroendocrine differentiation (NED) of prostatic tumor cells is of important clinical significance.

In the last decade, the molecular features of NEPC have gradually come to light, including genomic loss of RB1 and TP53 [12], and amplification of MYCN [9,13,14]. In addition, reprogramming to an NEPC state is also linked to overexpression of neural progenitor-associated genes such as *SOX2* [5,15], *MYCN* [9,13,14], *EZH2* [16], *POU3F2* [17], *FOXA2* [18,19] and *SIAH2* [19]. Many NEPC drivers such as *SOX2* and *MYCN* have also been reported to be es-

sential for maintaining cell stemness [20,21], raising the possibility that the emergence of NEPC is associated with acquisition of stem-like properties. Interestingly, we previously showed that *SOX2* is normally expressed in prostatic basal epithelial cells and a small population of luminal cells [22], highlighting a potential role of normal SOX2-expressing epithelial cells in deriving NED. Other findings also suggest that acquisition of NE phenotype of the prostate cancer cell is likely to link with epithelial-to-mesenchymal transition (EMT) state that could be both induced upon down-regulation of AR signalling [4,23]. On the other hand, several groups have attempted to uncover the cell of origins of focal NED and even NEPC. The normal prostate gland consists predominantly of cells of the luminal and the basal compartment with a small minority of NE cells that are scattered between the luminal and the basal cell compartment [24]. As normal NE cells share many features with malignant NE cells (for example, expressing *SYP* and *CHGA*), it has been proposed that NEPC might arise from transformed NE cells [25]. However, genomic studies seem move supportive of an epithelial origin of NEPC, given that NEPC showed significant genomic overlap with PCA, such as TMPRSS2-ERG fusion [26,27]. Within the prostatic epithelial cell compartments, both luminal and basal epithelial cells have been shown to be capable of generating prostate cancer and even NEPC. For example, Zou et. al. [28] have demonstrated that focal NED, as well as eventual well-differentiated neuroendocrine disease directly arises via transdifferentiation from luminal adenocarcinoma cells. In

contrast, Lee et. al. [29] have recently reported that basal cells can directly give rise to NE cells during prostatic tumorigenesis without undergoing an intermediate luminal state. In addition, there are some studies have suggested that NE cells derived from basal cells exhibit a loss of basal features and up-regulation of luminal features during NED [13,30]. Overall, there is no consensus on the cellular characteristics during the transition from epithelial tumor cells to neuroendocrine (NE) tumor cells.

Gene expression is a key determinant of cellular phenotypes. Previous population-based RNA sequencing (RNA-seq) method has been performed to compare the transcriptional similarity between prostatic basal and luminal epithelial cells and suggested that metastatic NEPC molecularly resembled stem cell in basal compartment [31,32]. Recent advance in single-cell RNA sequencing (scRNA-seq) technology has greatly empowered us to gain a more comprehensive understanding of the transcriptional signatures of distinct subpopulations of epithelial cells in human and mouse prostate [33–36]. However, a detailed analysis of the cellular states of NED in primary human prostate cancer at single-cell resolution is still lacking. Herein, we apply scRNA-seq technology to determine the cellular identity associated with NED in human prostate cancer. Our datasets reveal that a luminal epithelial state is highly linked with of NED of prostate cancer cells. Furthermore, we show by intra-tumoral RNA velocity analysis that the NE cells are directly generated by luminal-like adenocarcinoma cells. Finally, we dissect the transcriptomic

landscape underlying NED and validate single-cell derived NED-related gene signatures in bulk RNA samples. Altogether, our results support the epithelial-NE transdifferentiation model regarding the NED in human prostate cancer and offer fresh insights into cellular states and molecular features associated with this process.

Results

Single-cell transcriptional profiling of biopsies from 6 CRPC

Given that focal NED can be more frequently detected in patients with advanced prostate cancer undergoing ADT but not in primary prostatic adenocarcinoma [37–39], we sought to perform scRNA-seq on tumor biopsies from CRPC patients. In this study, we isolated fresh cells from six CRPC patients, four out of whom were found to have low PSA levels (<20 ng/ml; Table 1, Fig. 1A and supplementary fig. 1), indicating a higher likelihood of having NED. In these patients, three had received the first-line therapy of the LHRH analogue goserelin coupled with the AR inhibitor bicalutamide, two had underwent surgical castration coupled with bicalutamide, while the remaining one was diagnosed as small-cell NEPC at the beginning and treated with chemotherapeutic drug docetaxel. By pathological examination, biopsy tissues from three patients (#2, #5 and #6) displayed cellular morphology resembling small cell carcinoma and biopsies from patient #1 and #4 presented a classical PCA phenotype (Fig. 1B). However, biopsy from patient #3 was characterized as

prostatic intraepithelial neoplasia, which may due to the inaccuracy of the biopsy procedure. The clinical and pathologic features of the biopsy samples are summarized in Table1.

Then, single-cell suspension from each tissue was subjected to scRNA-seq by a 10x Genomics-based single-tube protocol with exclusive transcript counting through barcoding with unique molecular identifiers [40]. After exclusion of red blood cells as well as cells not passing quality controls, we obtained a total of 21,292 high-quality cells at ~2884 genes detected on average per cell (Supplementary fig. 2A and supplementary table1). Using an unsupervised graph-based clustering strategy, we manually classified different cell clusters into eight major cell types with canonical markers curated from the literatures, including epithelial cells, immune cells (T cells, B cells, myeloid cells and mast cells), stromal cells (fibroblasts and myofibroblasts) and endothelial cells (Supplementary fig. 2A, 2B and supplementary table2).

NE cells present an epithelial phenotype

Next, in keeping with our aim to characterize NED, we defined a NE index using 14 well-known NE markers that have been previously characterized as biomarker or driver genes of NEPC, such as *ASCL1*, *CHGA/B*, *FOXA2*, *SOX2* [4,14,16,18,19,41,42]. We scored each cell using this NE gene set (Supplementary table3). In line with the pathological results, this analysis identified obvious NED in three patients (patient #2, #5 and #6; Fig. 1C). Notably, we found that NE^{high} cell population detected in these three patients all belong to

the epithelial cells instead of the non-epithelial cell compartments (Fig. 1C and supplementary fig. 2C, 2D), supporting an epithelial origin of NED. In addition, we noticed that majority of epithelial cells from patient #2 and #5 were scored for a NE phenotype, while only part of epithelial cells from patient #6 have a NE phenotype (Fig. 1C), manifesting different extent of NED among these three patients. Taken together, single-cell analysis showed that three patients likely have NED and suggested an epithelial origin of NED in human prostate cancer.

NE cells present a malignant luminal-like phenotype

Having characterizing an epithelial phenotype of NED, we next focused on epithelial compartment by computationally removing all non-epithelial cells. In order to gain more insight into the molecular features of NED in each patient, we then scored each cell for epithelial basal/luminal lineage markers [35], AR signature genes [41,43,44], EMT as well as stem cell genes expression [45]. Pairwise correlation analysis of all epithelial cells revealed that nearly all of epithelial cell from two patients (patient #2 and #5) represents an obvious NED phenotype (Fig.2A, 2B and supplementary table4). Epithelial cells of patient #6 were divided into two main groups: a small population of NE-like cells and the remaining majority of $NE^{-}AR^{high}$ cells, illustrating significant intra-tumoral heterogeneity regarding NED (Fig. 2A-2C). By analysing cellular phenotypes/states, we found that all NE cells prominently exhibited a luminal phenotype rather than basal phenotype (Fig.2A, 2B). Of note, AR scores were ex-

tremely low in NE cells, which is consistent with previous findings that AR signalling activity is downregulated in NEPC [46]. As introduced earlier, the EMT process and stem cell state have also been proposed to participate in NEPC formation [23]. However, our analysis demonstrated that only NE cells from patient #5 displayed higher EMT and stemness signature scores.

We further interrogated malignant identity of NE cells by performing inferred copy number variation (CNV) analysis on the basis of the average expression of 101 genes in each chromosomal region [47,48]. We used the normal prostate epithelial cells from Henry dataset as "reference" cells [35], such that their average CNV value was subtracted from all cells. Inferred CNV analysis showed that most NE cells exhibited remarkable CNVs, indicating their malignant identity (Fig.2D and supplementary fig. 3). Interestingly, most basal-like epithelial cells lacked CNVs, and thus likely representing a group of normal epithelial cells (Fig. 2E). In addition, we noticed that the epithelial cells of patient #3 had very few CNVs consistent with its histologically intraepithelial neoplasia characteristic (Supplementary fig. 3). Taken together, preliminary analyses revealed a malignant feature of NE cells and suggested a link between NED and a luminal-like $AR^{low/-}$ phenotype.

Intra-tumoral analyses identify different extent of focal NED

To better understand the extent of NED in each individual tumor, we next investigated intra-tumoral epithelial diversity. Re-clustering epithelial cells from each tumor combined with heatmap analysis showed that epithelial cell

sub-clusters from each sample highly expressed luminal cell markers such as *KRT8* and *KRT18*, while the expression of basal, NE and AR signature genes exhibited significant intra- and inter-tumor heterogeneity (Fig. 3A, 3B and supplementary table5). We thus annotated all epithelial clusters into basal, luminal and NE subtypes, respectively, according to their transcriptional landscapes. For example, in patient #1, we identified a cluster of basal cells ($KRT17^{+}$; cluster 1 and 7), several clusters of luminal cells ($KRT5^{-}$ $KRT8^{+}$, cluster 0, 2, 3, 4, 5, and 6). This analysis also confirmed the NE phenotype of patients #2 and #5 and showed that most clusters in these two patients uniformly expressed NE markers, manifesting a pure NE phenotype. In addition, epithelial cells of patient #6 consisted of a group of NE cells (cluster 4, expressing *ASCL1*, *CHGA* and *CHGB*), a group of basal cells (cluster 5, expressing *KRT5*, *KRT14* and *KRT15*) as well as the remaining AR^{high} luminal cells, presenting mixed features of both adenocarcinoma and NEPC (Fig. 3B). The most interesting observation was from patient #4, a histologically diagnosed adenocarcinoma, in which we found that when compared with other epithelial clusters, cluster 5 preferentially expressed NE markers *CHGA* and *SYP* (Fig. 3B), probably representing a population of early NE precursors. These observations were further validated by IHC assays for two NE markers (SYP and SOX2) and AR in sections from five samples (Fig. 3C). For instance, we detected a minority of scattered SYP^{+} NE cells in section from patient #4，which may corresponded to the cells of cluster 4 revealed by single-cell anal-

ysis. In addition, IHC analyses of patient #5 samples also showed an overall good concordance with the single-cell transcriptional profiles that SOX2 was intensively expressed, while another NE marker SYP was almost undetectable (Fig. 3C). Thus, intra-tumoral analysis confirmed NED in three patients (patient #2, #5 and #6) and enabled us to detect NE cells in a PCA (patient #4).

Epithelial cellular relationships in patient #4

We next paid specific attention to patient #4, given that the NE-subpopulation detected in this PCA may represent an early state of transdifferentiation from epithelial towards NE fate. Epithelial cells in patient #4 were partitioned into four main subtypes: basal cells (cluster 6, expressing *KRT5* and *TP63*), urothelial-like cells (cluster 4, expressing *UPK1A* and *GATA3*), NE cells (cluster 5, expressing *SYP* and *EZH2*) and luminal cells with a $KRT5^{-}UPK1A^{-}SYP^{-}KRT8^{+}$ feature (clusters 0-3; Fig. 4A, 4B). Uniform Manifold Approximation and Projection (UMAP) visualization suggested that NE cells were transcriptionally closer to luminal cells than basal or urothelial-like cells. Immunofluorescence analysis of SYP and KRT8 further validated a luminal phenotype of SYP-expressing cells (Fig. 4C). Interestingly, the early NED cells and luminal cells shared almost the same CNV pattern, indicating that they had a common clonal origin (Fig. 4A and supplementary fig. 4A). In contrast, basal cells in this sample displayed very few CNVs. Thus, the separation of different epithelial subtypes may reflect their marked genomic differences.

A closer relationship between NE cells and luminal-like malignant cells was

further supported by visualization using Partition-based approximate graph abstraction (PAGA) [49] (Fig. 4D). To deepen our understanding of the dynamics of epithelial to NE transition, we next applied RNA velocity analysis that predicts the future state of an individual cell by leveraging the fact that newly transcribed, unspliced pre-mRNAs and mature, spliced mRNAs can be distinguished in common single-cell RNA-seq protocols [50]. Notably, unlike many other existing computational methods [51], RNA velocity analysis doesn't rely upon a priori knowledge to set up the starting cell for inferring the trajectory and thus enable us to more unbiasedly and accurately predict the cellular differentiation trajectory. Given the heterogeneous epithelial composition, we utilized scVelo, a likelihood-based dynamical model that has recently be introduced to solves the full gene-wise transcriptional dynamics [52]. This analysis clearly showed positive velocity from luminal malignant cells (cluster 3) towards early NED cells (cluster 5; Fig. 4E). In contrast, $KRT5^+$ basal and $UPK1A^+$ urothelial-like cells were clustered far from NED cells and did not show a tendency to progress into SYP^+ cells. Therefore, this finding suggested that luminal-like malignant cells may serve as the direct progenitor cells responsible for early NED in this patient we analysed here.

TMA analysis confirms the prevalence of luminal-like NED phenotype

We next validated the generality of this observation in a large population using clinical PC TMAs, which contained 297 cancer tissues (280 PCA, 10 CRPC and 7 NEPC) (Supplementary Table6). We carried out triple IF staining for K18,

K5 and SYP to evaluate the basal/luminal phenotypes of NE cells (Fig. 4F). Consequently, we detected SYP-positive cells in 102 tumors, of which 81% were $K18^{+}K5^{-}SYP^{+}$, and 5% exhibited both $K18^{+}K5^{-}SYP^{+}$ and $K18^{-}K5^{-}SYP^{+}$ characteristics (Fig. 4G). Notably, no $K18^{-}K5^{+}SYP^{+}$ cells were found in any of the 297 cancer tissues. This analysis therefore verified that NED precursors in human prostate cancer had a prevalent luminal phenotype. Of interest, a substantial number of the SYP-expressing tumor specimens came from patients who had not received any therapy (96/102), demonstrating that NED in fact occurred much earlier than the development of castration resistance, which is in line with previous findings that neuroendocrine differentiation is present in 10% to 100% of localized PCAs and increases with disease progression [53,54]. Taken together, TMA analysis confirmed the single-cell results, demonstrating that NED in human prostate cancer was primarily presented as a luminal feature.

Epithelial cellular relationships in patient #6

Similar to patient #4, epithelial cells of patient #6 also showed intra-tumoral NED heterogeneity, which was composed of a small population of NE cells (cluster 4), a small population of basal cells (cluster 8) and the vast majority of luminal epithelial cells (Fig. 5A, 5B). Interestingly, like patient #4, basal epithelial cells in patient #6 epithelial cells also displayed relatively fewer CNVs compared with luminal compartment as well as NE cells (Fig. 5A and supplementary fig. 4B), indicating that basal epithelial cells were less likely to be the

direct progenitors of NE cells. The cellular relationship was further indicated by PAGA (Fig. 5C), showing that NE cells in this sample still connected with luminal-like tumor cells. We next inferred cellular dynamics using RNA velocity, which predicated similar cellular processes that NED in this sample was exclusively branched from luminal cells (Fig. 5D). We further sought to identify genes that display pronounced dynamic expression patterns linked to the transition state toward a NE fate (Supplementary table7). As expected, signatures of AR signalling such as *KLK2* and *KLK3* were notably downregulated along with the emergence of NE phenotype (Fig. 5E). We then paid particular attention to genes that were positively correlated with NED. Within the top-ranked likelihood genes, we found *ASCL1*, a key transcription factor for neuronal differentiation [55], which has also been associated with NED in prostate cancer [56] (Fig. 5E). In addition, this analysis also illustrated many unknown genes that might serve as the potential drivers or biomarkers of the NED transdifferentiation, for example, *VGF*, *SCGN* and *PAPPA2*, the roles of which in NEPC have not been reported. Altogether, deeper analyses of epithelial cell relationships in this sample also suggested that malignant cells with a luminal phenotype fuels the development of NE cells.

Identifying NED-associated gene meta-programs

We next sought to understand the underlying molecular features associated with NED. For this purpose, we applied non-negative matrix factorization (NMF) to define underlying transcriptional programs specific to the epithelial cells

from each tumor [57,58] (Fig. 6A and supplementary table8). To relate these meta-programs to cell phenotypes, we scored these ordered cells according to basal, luminal, NE, EMT, AR and cell cycle marker genes (Fig. 6B). This analysis revealed three meta-programs highly associated with NED (P1, P2, P4). For example, meta-program P1 was characterized by neuroendocrine markers such as *CHGB* and *CHGA* and meta-program P2 contained NE-related transcriptional factor (TF) *EZH2* and *DLX5*, a homeobox transcription factor gene. *DLX5* has been recently reported to mark delaminating neural crest cells during development [59]. Of note, neural crest cells can differentiate into numerous derivatives including neuroendocrine cells [60,61], implying a potential role of this gene in participating NED of prostate cancer cells. Moreover, we identified a cell cycle-related meta-program (P3) that was obviously upregulated in NE cells of patient #2 and #5)，likely reflecting well-differentiated NE state of these two tumors. More interestingly, meta-program P2 was specifically associated with patient #2, while meta-program P4 was preferentially expressed in patient #5, suggesting two kinds of NED features.

We next asked whether the NE-related gene meta-programs derived from single-cell data could robustly reflect the NED in bulk expression profiles. Thus, we used three bulk transcriptomic datasets [7,41,62] , which included both CRPC and NEPC patients. We first performed correlation analysis between the expression of all genes from three meta-programs (P1, P2 and P4) and the

NE score defined by the average expression of well-established NE markers to screen out genes that were most relevant to NED. This analysis identified 121 genes highly correlated with the NE score (Pearson R ≥ 0.3; Fig. 6C and supplementary table9). Consistently, we found that by plotting their expression in the 5 groups of samples that was defined by the expression patterns of NE and AR activity genes [41], most genes displayed significantly higher expression in the $AR^{-}NE^{+}$ group than in NE^{-} groups (Fig. 6D). Thus, NED-associated gene signatures derived from single-cell data can provide reliable clues for distinguishing human NEPC and searching for new drivers involved in NED.

Identifying NED-associated transcription factor regulatory network

The above NMF analysis revealed that two well-differentiated NEPC displayed distinct NED signatures. To explore the underlying molecular mechanisms driving the distinct NE differentiation phenotypes, we next used single-cell regulatory network inference and clustering (SCENIC) to identify the co-expressed transcription factors and their putative target genes, as an indicator of transcription factor regulatory activity [63]. SCENIC analysis showed that NED from different patients could upregulate the expression of different transcription-factor networks (Fig. 7A). For instance, *DLX6* and *ASCL1* regulons were highly active in NE cell of patient #2, whereas expression of *FOXA2* and *SOX21* network was restricted in NE cells of patient #5. In line with reports that *SOX2* are essential for NED in prostate cancer, we found that *SOX2* regulon was upregulated across almost all NE cells from patient #2 and #5 (Fig.

7A). Thus, single-cell regulatory network analysis provided an explanation for the divergence of NED from our patient cohort. In addition to many well-established NE-related TFs identified in this analysis (such as *SOX2* and *ASCL1*), it also predicted many neuronal differentiation related TFs that might also be involved in NED. For instance, expression of *LHX2* has previously been showed to confer neuronal competency for activity-dependent dendritic development of cortical neurons [64], but its role in NED of prostate cancer remains undetermined and need future studies to clarify their specific roles.

Next, we analysed TF regulons of epithelial cells from two patients with intra-tumoral NED heterogeneity. Analysis of patient #4 revealed that NE subpopulation specifically up-regulated transcriptional activities of *NKX2-2*, *HES6*, *FOXA2* and *ASCL1*, all of which have been previously reported to be essential for a variety of neural cell types' differentiation (Fig. 7B, 7C). The intra-tumoral heterogeneity in term of TF activity was also observed in patient #6, showing that NE subpopulation have obviously higher TF activities of *SOX2* and *FOXA2* (Fig. 7D, 7E). In addition, NE subpopulation strongly upregulated activities of *UNCX* and *CELF5* regulatory networks, which have been both reported to involve in maintaining neural cells survive or promoting some neuron diseases [65,66]. Overall, TF network analysis revealed both known and unknown NED-associated TFs and offered more insight into both inter-tumoral and intra-tumoral heterogeneity regarding NED.

Discussion

In this study, we generated 21,292 single-cell transcriptomes from 6 CRPC patients with a focus on the cellular phenotypes associated with NED. We detected NED in four tumors, in which all of the NE cells exhibited a luminal rather than basal epithelial phenotype. It is important to note that in two tumors that contain both NE cells and non-NE epithelial cells (patient #4 and #6), there is clear cell fate transition tendency from luminal-like adenocarcinoma cells towards NE cells (Fig. 4E and Fig. 5D). Thus, our finding has identified the transdifferentiation process that has been proposed for a long time in explaining NED in prostate cancer. Although previous genomic analyses have suggested that NEPC are clonally derived from PCA that usually present luminal-like phenotype [7,26,27], this is the first study to our knowledge that has shown the cellular diversity in human CPRC as well as the cellular phenotypes associated with NED at single-cell resolution.

Our current study is limited regarding the total number of samples that contain NED for analyses. To unbiasedly evaluate the cellular phenotypes associated with NED, we next performed triple IF staining against KRT5, KRT8 and SYP on our large cohort of PC TMAs. We found that the luminal-like malignant phenotype of NE cells ($K5^{-}K18^{+}SYP^{+}$) is mainly detected in adenocarcinomas (Fig. 4G). Therefore, this results further confirmed a closer relationship between NED and a luminal state rather than basal state in human prostate cancer. It should be noted that our results don't exclude the probability that

basal cells serve as cell of origin of NEPC. According to *in vivo* cell lineage tracing studies, both basal and luminal cells are capable of initiating prostate tumorigenesis [67]. In particular, prostate cancer originated from human basal cells gradually loss basal features and upregulation of luminal hallmarks [13,68]. Based on these findings and our current results, we propose a model that PCA can be initiated from both basal and luminal cells, while focal and eventual NEPC is more likely to be made by NE precursors with luminal phenotype (Fig. 8). We also consider the possibility that a direct basal-NE transdifferentiation may happen. If NE cells are directly transdifferentiated from basal cells, we would expect to see hybrid cells with both basal and NE phenotypes more frequently. However, our analysis in PC TMAs reveals that only about 1% of patients (1/102) contain SYP^+ cells that express both K8 and K5 in adenocarcinoma tissues argue strong for the notion that such direct basal-NE transdifferentiation is likely rare in human prostate cancer, but rather luminal-NE transdifferentiation is fundamentally responsible for phenotypic transition from acinar adenocarcinomas towards NEPC. Interestingly, a recent cell lineage tracing study using TRAMP mouse models (p63-CreERT2;Rosa-LoxP-STOP-LoxP-tdRFP;TRAMP and K8-CreERT2;Rosa-LoxP-STOP-LoxP-tdRFP;TRAMP) has demonstrated that NEPC is directly originated from basal progenitor cells but not luminal cells or pre-existing $KRT8^+$ adenocarcinoma cells [29]. This observation is different from the results obtained from the double *p53* and *Pten* knockout-induced PCA

mouse model in which all NEPC cells were transdifferentiated from *NKX3.1*-expressing luminal cells. According to our results, we are inclined to think that a transformed basal cell would firstly differentiate to a luminal-like tumor cells and then execute NED process. Nevertheless, future single-cell studies of serial tumor samples from individuals will be needed in principle to map the cellular dynamic involved in NED process as much as possible.

Our next aim is to explore the signature genes driving NE transdifferentiation. By performing NMF analysis, we further identified three gene meta-programs consisting of many genes highly correlated with NED (Fig. 6A). The bulk datasets analysis has validated the robustness of this result, showing that most genes are significantly expressed in patients with NED (Fig. 6D). Interestingly, we found that two well-differentiated NEPCs (patient #2 and #5) seem to have distinct NED programs. SCENIC analysis highlighted that the heterogenous NED might be determined by distinct TF networks. Nevertheless, the exact role of many identified genes in prostate cancer, especially NEPC, is unknown and needs further comprehensive investigation.

In summary, our single-cell study has disentangled both intra- and inter-tumoral heterogeneity regarding to NED in human prostate cancer and characterized both cellular phenotypes and molecular features linked with the luminal to NE transdifferentiation. Understanding the progressive trajectory of NED will benefit the development of early diagnosis and even therapeutic treatments for human NEPC.

Methods & materials

Patient selection

With a focus on neuroendocrine prostate cancer, the participating patients were required to meet the following requirements: 1) the patients must have developed resistance to castration therapy; 2) CT imaging showed an apparent prostate tumor (Supplementary fig. 1A). In addition, we preferentially selected patients whose circulating PSA level was lower than 20 ng/ml. The patient information is described in detail in Table 1. The present study was approved by the Institutional Ethics Review Board of Ren Ji Hospital, Shanghai Jiao Tong University School of Medicine, and written informed consent was obtained from every patient.

Isolation of single cells

Prostate biopsies were transported to the research laboratory on ice in DMEM/F12 (Gibco, 11320033) with 3% FBS (Gibco, 10099-141) within 30 min of collection. Each specimen was equally separated into two fragments. One fragment was processed for histopathological assessment, and the remainder of the provided tissues was processed for scRNA-seq. In brief, fresh tumor samples were minced and place in a 1.5 ml Eppendorf tube, where they were enzymatically digested with collagenase IV (Gibco) and DNase I (Sigma) for 1 h at 37 °C with agitation. After digestion, samples were sieved through a 70-μm cell strainer, washed with 1% BSA and 2 mM EDTA in PBS, and cen-

trifuged for 5 min at 350×g. Single cell suspensions were subjected to Lympholyte-H separation (Cedarlane, CL5020) to remove RBCs and debris according to the manufacturer's specifications. Pelleted cells were then resuspended in DMEM/F12 with 3% BSA and were assessed for viability and size using a Countess instrument (Thermo).

Single-cell library preparation and sequencing

A total of 5,000 cells per sample were targeted for capture. Then, the cell suspension of each sample was run in the Chromium Controller with appropriate reagents to generate single cell gel bead-in-emulsions (GEMs) for sample and cell barcoding. The libraries were then pooled and sequenced on a NovaSeq 6000 (Illumina) at a depth of ~400 M reads per sample.

Single-cell data preprocessing and quality control (QC)

Raw sequencing data were converted to FASTQ files with Illumina bcl2fastq, version 2.19.1, and data were aligned to the human genome reference sequence (GRCH38). The CellRanger (10X Genomics, 2.1.1 version) analysis pipeline was used for sample demultiplexing, barcode processing and single-cell 3′ gene counting to generate a digital gene-cell matrix from these data. Of note, Cell Ranger filters any barcode that contains less than 10% of the 99th percentile of total UMI counts per barcode, as these are considered to be associated with low quality cell barcodes. This processing resulted in an average of 160,233 reads per cell, and an average of 2,884 genes were detected

per cell (Supplementary table1). The gene expression matrix was then processed and analysed by Seurat (version 3.0) and an R toolkit (https://github.com/satijalab/seurat), using the software R (version 3.6.0). We performed Seurat-based filtering of cells based on the number of detected genes per cell (500 to 7000) and the percentage of mitochondrial genes expressed (<10%). The mitochondrial genes and ribosomal genes were also removed from the gene expression matrix. Following quality control, 21,292 high-quality cells were retained with an average of 2419 genes were detected per cell (Supplementary table1). Each single-cell dataset was then processed by *SCTransform* from the *Seurat* package, which contained the function of normalization, regression and identification of variable genes.

UMAP visualization and cell type annotation

We used UMAP [69] to visualize the clusters of cells that passed quality control for each sample. Clusters were associated with cell types based on the scores of differential expression of well-established marker genes for each cell type: T cells (*CD2*, *CD3D*, *CD3E* and *CD3G*), B cells (*CD79A*, *CD79B*, *CD19* and *MS4A1*), myeloid cells (*CD14*, *CD68*, *AIF1* and *CSF1R*), mast cells (*MS4A2*, *ENPP3*, *PCER1A* and *KIT*), fibroblasts (*DCN*, *TNFAIP6*, *APOD* and *FBLN1*), myofibroblasts (*MYH11*, *GJA4*, *RGS5* and *MT1A*), endothelial cells (*ENG*, *CLDN5*, *VWF* and *CDH5*) and epithelial cells (*EPCAM*, *KRT8*, *KRT5* and *CDH1* [35,70–74].

Defining Cell Scores

We used cell scores to evaluate the degree to which individual cells express a certain pre-defined expression program as described previously [74]. We defined gene sets corresponding to basal, luminal, NE, AR pathway, EMT state and cell stemness from previous literatures [4,14,16,18,19,35,41–45]. The detailed gene list can be found in Supplementary table2. Given a set of pre-defined genes (Gj), we calculate for each cell i, a score, SCj(i), quantifying the relative expression of Gj in cell i, as the average relative expression (Er) of the genes in Gj. To control for this effect we also add a control gene-set (Gjcont); we calculate a similar cell score with the control gene-set and subtract it from the initial cell scores: SCj(i) = average[Er(Gj,i)] – average[Er(Gjcont,i)]. The control gene set contains 100 genes with the most similar aggregate expression level.

Inferred CNV analysis from scRNA-seq

Large-scale CNVs inferred from single-cell gene expression profiles using a previously described approach (https://github.com/broadinstitute/inferCNV/wiki) [47,48]. To identify the distinct chromosomal gene expression pattern of epithelial cells in comparison to putative noncarcinoma cells, we set normal prostate epithelial cells from a dataset which contains 78,286 prostate epithelial cells of 3 health men, which were captured by Henry et. Al. [35], as the reference "normal" cells. In addition, those genes expressed in fewer than 200 cells were removed from the count matrix. Average expression was calculated

using the log transformed data (log2[1 + UMI]), and absolute values of fold change were bound by 3. All genes were sorted by their chromosome number and start position. The chromosomal expression patterns were estimated from the moving averages of 101 genes to determine the window size, and they were adjusted as central values across genes. The average CNV signal was estimated by averaging the CNV modification for 22 autosomes.

Multiple datasets integration and Batch correcting

For merging multiple datasets, we applied Harmony integration [75], which has been showed to reduce technical batch effects while preserving biological variation for multiple batch integration. RunHarmony returns a Seurat object, updated with the corrected Harmony coordinates. The manifold was subjected to re-clustering use the corrected Harmony embeddings rather than principal components (PCs), set reduction = 'harmony', with parameters of Seurat analysis.

Differential gene expression analysis

DEGs in a given cell type compared with all other cell types were determined with the FindAllMarkers function from the Seurat package (one-tailed Wilcoxon rank sum test, P values adjusted for multiple testing using the Bonferroni correction). For computing DEGs, all genes were probed provided they were expressed in at least 25% of cells in either of the two populations compared and the expression difference on a natural log scale was at least 0.25.

RNA velocity

RNA velocities were predicted using velocyto in R program (http://velocyto.org, version 0.6) [50,52]. Briefly, spliced/unspliced reads were annotated by velocyto.py with CellRanger (version 2.2.0), generating BAM files and an accompanying GTF; then, they were saved in .loom files. The .loom files were then loaded to R (version 3.6.0) using the read.loom.matrices function, and they generated count matrices for spliced and unspliced reads. Next, the count matrices were size-normalized to the median of total molecules across cells. The top 3,000 highly variable genes are selected out of those that pass a minimum threshold of 10 expressed counts commonly for spliced and unspliced mRNA. For velocity estimation, we use the default procedures in scVelo (n_neighbors=30, n_pcs=30). In consideration that the assumptions of a common splicing rate and the observation of the full splicing dynamics with steady-state mRNA levels were often violated, we used the function recover_dynamics, a likelihood-based dynamical model, to break these restrictions. Finally, the directional flow is visualized as single-cell velocities or streamlines in the UMAP embedding with the Seurat cluster annotations.

Connectivity of cell clusters

To identify potential developmental relationships of cell clusters in patient #4 and #6, we utilized the partition-based graph abstraction (PAGA) [49] to esti-

mate any potential developmental relationships among the three prostate lineages. The computations were performed on the same subset of variable genes as for clustering, using the default parameters.

Identification of epithelial gene meta-programs

Transcriptional programs were determined by applying NMF as previously described [57,58]. Analysis was performed for the epithelial cells only. We set the number of factors to 10 for each tumor. For each of the resulting factors, we considered the 30 genes with the highest NMF scores as characteristics of that given factor (Supplementary table8). All single cells were then scored according to these NMF programs. Hierarchical clustering of the scores for each program using Pearson correlation coefficients as the distance metric and Ward's linkage revealed four correlated sets of programs with our focus.

SCENIC

In order to further investigate the gene regulatory networks (GRNs) in process of NED, we applied SCENIC [63] workflow to reconstruction the GRNs. The input matrices for SCENIC of every single sample was the corrected UMI counts in “SCT assay” of Seurat, in which we removed the variation of mitochondrial mapping percentage. For the combined sample (epithelial cells of 6 patients), Combat [76] were run to correct for “patient of origin” as source of batch effect. Following the standard procedure of SCENIC, we used GENIE3 (for single sample) and GRNBoost (for combined sample) to identify potential

TF targets. Besides, the activity of each regulon in each cell is evaluated using AUCell, which calculates the Area Under the recovery Curve, integrating the expression ranks across all genes in a regulon. Finally, we used the default "AUCCellThreshholds" for each regulon as the threshold to binarize the regulon activity scores and created the "Binary regulon activity matrix". The motifs database for Homo sapiens was downloaded from the website https://pyscenic.readthedocs.io/en/latest/.

Bulk dataset analysis

Bulk-transcriptomic data were collected from Morrissey et. al. (GEO:GSE126078) [41], Beltran et. al. (https://www.cbioportal.org/study/summary?id=nepc_wcm_2016) [7] and Charles L. Sawyers et. al. (https://github.com/cBioPortal/datahub/tree/master/public/prad_su2c_2019) [62]. To estimate the correlation of the P1, P2 and P4 meta-program with NED, we first defined an NE score by gene set variation analysis (GSVA) [77,78] for every sample in these bulk RNA-Seq data, and the NE markers we used are listed in the Supplementary materials. Then, we filtered cell cycle-related genes from the gene list of the three meta-program and performed Pearson correlation coefficient analysis of the remaining genes.

Tissue microarrays

Tissue specimens from 297 patients who underwent radical prostatectomy

were collected for the construction of tumor microarrays (TMAs), and then the specimens were cut into 5μm thick sections using a standard microtome. These tissue cores were assessed by uropathologists to determine tumor stages according to the haematoxylin and eosin staining results (Supplementary table6).

Immunohistochemistry (IHC) and immunofluorescence (IF)

Formalin-fixed and paraffin embedded tissue sections (5μm) were deparaffinized and rehydrated. Antigen retrieval was carried out using 10 mM sodium citrate (pH 6.0) in a microwave oven. For DAB staining, endogenous peroxidase activity was blocked with 0.3% hydrogen peroxide for 10 min and 5% BSA in PBS for 1 h. Slides were incubated overnight at 4°C with a primary antibody, which was followed by incubation with an HRP-linked secondary antibody (CST) at room temperature (30 min). Diaminobenzidine (DAB) was used as chromogen, and the sections were counterstained with haematoxylin. For immunofluorescence staining, the sections were washed with PBS and transferred to a blocking solution (10% normal donkey serum in PBS) for 1 h at room temperature. After blocking, specimens were incubated overnight at 4°C with diluted primary antibodies. The next day, slides were washed with PBS three times for 10 min each, and then they were incubated for 1 h at room temperature with secondary antibodies conjugated to Alexa-488, -555 or -647, which were diluted with PBS containing 1% normal donkey serum (1:1000). Then, the secondary antibody was rinsed, and the slides were washed three

times with PBS before being mounted with Vector Shield mounting medium containing DAPI (Vector Laboratories, H-1200).

Image acquisition

IF images were acquired using a Zeiss LSM 710 confocal microscope and were processed by ZEN Imaging Software. IHC images were acquired using an Olympus BX53 System Microscope.

Primary antibodies

The following antibodies were used in these studies: anti-SOX2 (Abcam, ab236557), anti-AR (Abcam, EPR1535(2)), anti-KLK3/PSA (Cell Signaling Technology, #2475), anti-NKX3-1 (Cell Signaling Technology, #92998), anti-Cytokeratin 5 (Abcam, ab52635; For IHC), anti-P63 (Santa Cruz Biotechnology, sc-8431), anti-CK5 (Biolegend, 905904; For IF), anti-SYP (Cell Signaling Technology, #36406), and anti-K18 (ProteinTech, 66187-1-Ig).

Statistical analysis

Statistical analysis was performed using R (version 3.6.0) and GraphPad Prism (version 8). Wilcoxon rank-sum tests were used in this study and are described in each figure. * $p< 0.05$, ** $p< 0.01$, *** $p< 0.001$, **** $p< 0.0001$ and ns, not significant.

Data availability

The scRNA-seq data were deposited in the NCBI Gene Expression Omnibus

(GEO) database under accession number GSE137829.

Acknowledgments

This work was supported by funds to WQG from Ministry of Science and Technology of the People's Republic of China (2017YFA0102900), National Natural Science Foundation of China (81872406 and 81630073), KC Wong foundation and the Shanghai Young Eastern Scholar Funds (QD2018021) to JW.

Author contributions

B.D. and W.Q.G. conceived and designed the study. J.M., W.L., and H.L. performed scRNA- Seq analysis. Z.J., M.Z., X.C., J.M.W., and Y.W. carried out data analysis. H.H.Z., C.W.C., and Y.F. provided technical help. L.F., Y.Z. and J.P. gave advice for analysis and J.W. and W.X. wrote the manuscript with input from all authors. W.Q.G. supervised the project.

Competing interests: The authors declare no competing financial interests.

References

1. Graf T, Enver T (2009) Forcing cells to change lineages. *Nature* **462**: 587–594.
2. Le Magnen C, Shen MM, Abate-Shen C (2018) Lineage Plasticity in Cancer Progression and Treatment. *Annu Rev Cancer Biol* **2**: 271–289.
3. Beltran H, Hruszkewycz A, Scher HI, Hildesheim J, Isaacs J, Yu EY, Kelly K, Lin D, Dicker AP, Arnold JT, et al. (2019) The role of lineage plasticity in prostate cancer therapy resistance. *Clin Cancer Res* clincanres.1423.2019.
4. Davies AH, Beltran H, Zoubeidi A (2018) Cellular plasticity and the neuroendocrine phenotype in prostate cancer. *Nat Rev Urol* **15**: 271–286.
5. Mu P, Zhang Z, Benelli M, Karthaus WR, Hoover E, Chen CC, Wongvipat J, Ku SY, Gao D, Cao Z, et al. (2017) SOX2 promotes lineage plasticity and antiandrogen resistance in TP53-and RB1-deficient prostate cancer. *Science (80-)* **355**:.
6. Beltran H, Tomlins S, Aparicio A, Arora V, Rickman D, Ayala G, Huang J, True L, Gleave ME, Soule H, et al. (2014) Aggressive variants of castration-resistant prostate cancer. *Clin Cancer Res* **20**: 2846–2850.
7. Beltran H, Prandi D, Mosquera JM, Benelli M, Puca L, Cyrta J, Marotz C, Giannopoulou E, Chakravarthi BVSK, Varambally S, et al. (2016) Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat Med* **22**: 298–305.
8. Puca L, Vlachostergios PJ, Beltran H (2019) Neuroendocrine differentiation in prostate cancer: Emerging biology, models, and therapies. *Cold Spring Harb Perspect Med* **9**: 1–20.
9. Berger A, Brady NJ, Bareja R, Robinson BD, Conteduca V, Augello MA, Puca L, Ahmed A, Dardenne E, Lu X, et al. (2019) N-Myc-mediated epigenetic reprogramming drives lineage plasticity in advanced prostate cancer. *J Clin Invest*.
10. Alanee S, Moore A, Nutt M, Holland B, Dynda D, El-Zawahry A, McVary KT (2015) Contemporary incidence and mortality rates of neuroendocrine prostate cancer. *Anticancer Res* **35**: 4145–4150.
11. Aggarwal R, Zhang T, Small EJ, Armstrong AJ (2014) Neuroendocrine prostate cancer: Subtypes, biology, and clinical outcomes. *JNCCN J Natl Compr Cancer Netw* **12**: 719–726.
12. Ku SY, Rosario S, Wang Y, Mu P, Seshadri M, Goodrich ZW, Goodrich MM, Labbé DP, Gomez EC, Wang J, et al. (2017) Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. **83**: 78–83.
13. Lee JK, Phillips JW, Smith BA, Park JW, Stoyanova T, McCaffrey EF, Baertsch R, Sokolov A, Meyerowitz JG, Mathis C, et al. (2016) N-Myc Drives Neuroendocrine Prostate Cancer Initiated from Human Prostate Epithelial Cells. *Cancer Cell* **29**: 536–547.
14. Beltran H, Rickman DS, Park K, Chae SS, Sboner A, MacDonald TY, Wang Y, Sheikh KL, Terry S, Tagawa ST, et al. (2011) Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets. *Cancer Discov* **1**: 487–495.
15. Bishop JL, Thaper D, Vahid S, Davies A, Ketola K, Kuruma H, Jama R, Nip KM, Angeles A, Johnson F, et al. (2016) The Master Neural Transcription Factor BRN2 Is an Androgen Receptor – Suppressed Driver of Neuroendocrine Differentiation in Prostate Cancer.

16. Dardenne E, Beltran H, Benelli M, Gayvert K, Berger A, Puca L, Cyrta J, Sboner A, Noorzad Z, MacDonald T, et al. (2016) N-Myc Induces an EZH2-Mediated Transcriptional Program Driving Neuroendocrine Prostate Cancer. *Cancer Cell* **30**: 563–577.
17. Bishop JL, Thaper D, Vahid S, Davies A, Ketola K, Kuruma H, Jama R, Nip KM, Angeles A, Johnson F, et al. (2017) The master neural transcription factor BRN2 is an androgen receptor–suppressed driver of neuroendocrine differentiation in prostate cancer. *Cancer Discov* **7**: 54–71.
18. Park JW, Lee JK, Witte ON, Huang J (2017) FOXA2 is a sensitive and specific marker for small cell neuroendocrine carcinoma of the prostate. *Nat Publ Gr* **30**: 1262–1272.
19. Qi J, Nakayama K, Cardiff RD, Borowsky AD, Kaul K, Williams R, Krajewski S, Mercola D, Carpenter PM, Bowtell D, et al. (2010) Siah2-Dependent Concerted Activity of HIF and FoxA2 Regulates Formation of Neuroendocrine Phenotype and Neuroendocrine Prostate Tumors. *Cancer Cell* **18**: 23–38.
20. Mei W, Lin X, Kapoor A, Gu Y, Zhao K, Tang D (2019) The Contributions of Prostate Cancer Stem Cells in Prostate Cancer Initiation and Metastasis. *Cancers (Basel)* **11**: 434.
21. Kerosuo L, Neppala P, Hsin J, Mohlin S, Vieceli FM, Török Z, Laine A, Westermarck J, Bronner ME (2018) Enhanced expression of MycN/CIP2A drives neural crest toward a neural stem cell-like fate: Implications for priming of neuroblastoma. *Proc Natl Acad Sci U S A* **115**: E7351–E7360.
22. Wang J, Zhu HH, Chu M, Liu Y, Zhang C, Liu G, Yang X, Yang R, Gao WQ (2014) Symmetrical and asymmetrical division analysis provides evidence for a hierarchy of prostate epithelial cell lineages. *Nat Commun* **5**: 1–13.
23. Soundararajan R, Paranjape AN, Maity S, Aparicio A, Mani SA (2018) EMT, stemness and tumor plasticity in aggressive variant neuroendocrine prostate cancers. *Biochim Biophys Acta - Rev Cancer* **1870**: 229–238.
24. Abate-Shen C, Shen MM (2000) Molecular genetics of prostate cancer. *Genes Dev* **14**: 2410–2434.
25. Bonkhoff H (1998) Neuroendocrine cells in benign and malignant prostate tissue: Morphogenesis, proliferation, and androgen receptor status. *Prostate* **36**: 18–22.
26. Guo CC, Dancer JY, Wang Y, Aparicio A, Navone NM, Troncoso P, Czerniak BA (2011) TMPRSS2-ERG gene fusion in small cell carcinoma of the prostate. *Hum Pathol* **42**: 11–17.
27. Williamson SR, Zhang S, Yao JL, Huang J, Lopez-beltran A, Shen S, Osunkoya AO, Maclennan GT, Montironi R (2011) ERG – TMPRSS2 rearrangement is shared by concurrent prostatic adenocarcinoma and prostatic small cell carcinoma and absent in small cell carcinoma of the urinary bladder : evidence supporting monoclonal origin. *Mod Pathol* 1120–1127.
28. Zou M, Toivanen R, Mitrofanova A, Floch N, Hayati S, Sun Y, Le Magnen C, Chester D, Mostaghel EA, Califano A, et al. (2017) Transdifferentiation as a mechanism of treatment resistance in a mouse model of castration-resistant prostate cancer. *Cancer Discov* **7**: 736–749.
29. Lee D-K, Liu Y, Liao L, Li W, Danielpour D, Xu J (2019) Neuroendocrine prostate

carcinoma cells originate from the p63-expressing basal cells but not the pre-existing adenocarcinoma cells in mice. *Cell Res* **29**: 420–422.

30. Park JW, Lee JK, Sheu KM, Wang L, Balanis NG, Nguyen K, Smith BA, Cheng C, Tsai BL, Cheng D, et al. (2018) Reprogramming normal human epithelial tissues to a common, lethal neuroendocrine cancer lineage. *Science (80-)* **362**: 91–95.
31. Smith BA, Sokolov A, Uzunangelov V, Baertsch R, Newton Y, Graim K, Mathis C, Cheng D, Stuart JM, Witte ON (2015) A basal stem cell signature identifies aggressive prostate cancer phenotypes. *Proc Natl Acad Sci* **112**: E6544–E6552.
32. Zhang D, Park D, Zhong Y, Lu Y, Rycaj K, Gong S, Chen X, Liu X, Chao HP, Whitney P, et al. (2016) Stem cell and neurogenic gene-expression profiles link prostate basal cells to aggressive prostate cancer. *Nat Commun* **7**: 1–15.
33. Wang X, Xu H, Cheng C, Ji Z, Zhao H, Sheng Y, Li X, Wang J, Shu Y, He Y, et al. (2020) Identification of a Zeb1 expressing basal stem cell subpopulation in the prostate. *Nat Commun* **11**:.
34. Karthaus WR, Hofree M, Choi D, Linton EL, Turkekul M, Chaudhary O, Xu T, Masilionis I, Manova K, Mazutis L, et al. (2020) Regenerative potential of prostate luminal cells revealed by single-cell analysis. **505**: 497–505.
35. Henry GH, Malewska A, Joseph DB, Malladi VS, Lee J, Torrealba J, Mauck RJ, Gahan JC, Raj G V., Roehrborn CG, et al. (2018) A Cellular Anatomy of the Normal Adult Human Prostate and Prostatic Urethra. *Cell Rep* **25**: 3530-3542.e5.
36. Barros-Silva JD, Linn DE, Steiner I, Guo G, Ali A, Pakula H, Ashton G, Peset I, Brown M, Clarke NW, et al. (2018) Single-Cell Analysis Identifies LY6D as a Marker Linking Castration-Resistant Prostate Luminal Cells to Prostate Progenitors and Cancer. *Cell Rep* **25**: 3504-3518.e6.
37. Vashchenko N, Abrahamsson PA (2005) Neuroendocrine differentiation in prostate cancer: Implications for new treatment modalities. *Eur Urol* **47**: 147–155.
38. Berruti A, Mosca A, Porpiglia F, Bollito E, Tucci M, Vana F, Cracco C, Torta M, Russo L, Cappia S, et al. (2007) Chromogranin A Expression in Patients With Hormone Naïve Prostate Cancer Predicts the Development of Hormone Refractory Disease. *J Urol* **178**: 838–843.
39. Nelson EC, Cambio AJ, Yang JC, Ok JH, Lara PN, Evans CP (2007) Clinical implications of neuroendocrine differentiation in prostate cancer. *Prostate Cancer Prostatic Dis* **10**: 6–14.
40. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, et al. (2017) Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**:.
41. Labrecque MP, Coleman IM, Brown LG, True LD, Kollath L, Lakely B, Nguyen HM, Yang YC, Gil da Costa RM, Kaipainen A, et al. (2019) Molecular profiling stratifies diverse phenotypes of treatment-refractory metastatic castration-resistant prostate cancer. *J Clin Invest*.
42. Ramalingam S, Eisenberg A, Foo WC, Freedman J, Armstrong AJ, Moss LG, Harrison MR (2016) Treatment-related neuroendocrine prostate cancer resulting in Cushing's syndrome. *Int J Urol* **23**: 1038–1041.
43. Sahu B, Laakso M, Ovaska K, Mirtti T, Lundin J, Rannikko A, Sankila A, Turunen JP,

Lundin M, Konsti J, et al. (2011) Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer. *EMBO J* **30**: 3962–3976.

44. Gao D, Vela I, Sboner A, Iaquinta PJ, Karthaus WR, Gopalan A, Dowling C, Wanjala JN, Undvall EA, Arora VK, et al. (2014) Organoid cultures derived from patients with advanced prostate cancer. *Cell* **159**: 176–187.
45. Zhu H, Lee RJ, Desai N, Trautwein J, Arora KS, Shioda T, Broderick KT, Zheng Y, Fox DB, Toner M, et al. (2015) RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. *Science (80-)* **349**: 1351–1356.
46. Bonkhoff H, Stein U, Remberger K (1993) Androgen receptor status in endocrine-paracrine cell types of the normal, hyperplastic, and neoplastic human prostate. *Virchows Arch A Pathol Anat Histopathol* **423**: 291–294.
47. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, Cahill DP, Nahed B V., Curry WT, Martuza RL, et al. (2014) Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science (80-)*.
48. Puram S V., Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, Rodman C, Luo CL, Mroz EA, Emerick KS, et al. (2017) Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **171**: 1611-1624.e24.
49. Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Göttgens B, Rajewsky N, Simon L, Theis FJ (2019) PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol* **20**: 1–9.
50. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastriti ME, Lönnerberg P, Furlan A, et al. (2018) RNA velocity of single cells. *Nature* **560**: 494–498.
51. Saelens W, Cannoodt R, Todorov H, Saeys Y (2019) A comparison of single-cell trajectory inference methods. *Nat Biotechnol* **37**: 547–554.
52. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ (2019) Generalizing RNA velocity to transient cell states through dynamical modeling. *bioRxiv* 820936.
53. Hirano D, Okada Y, Minei S, Takimoto Y, Nemoto N (2004) Neuroendocrine Differentiation in Hormone Refractory Prostate Cancer Following Androgen Deprivation Therapy. *Eur Urol* **45**: 586–592.
54. Rubin MA Neuroendocrine Expression in Metastatic Prostate Cancer : Evaluation of High Throughput Tissue Microarrays to Detect Heterogeneous Protein Expression. 406–414.
55. Park NI, Guilhamon P, Desai K, McAdam RF, Langille E, O'Connor M, Lan X, Whetstone H, Coutinho FJ, Vanner RJ, et al. (2017) ASCL1 Reorganizes Chromatin to Direct Neuronal Fate and Suppress Tumorigenicity of Glioblastoma Stem Cells. *Cell Stem Cell* **21**: 209-224.e7.
56. Rapa I, Ceppi P, Bollito E, Rosas R, Cappia S, Bacillo E, Porpiglia F, Berruti A, Papotti M, Volante M (2008) Human ASH1 expression in prostate cancer with neuroendocrine differentiation. *Mod Pathol* **21**: 700–707.
57. Hovestadt V, Smith KS, Bihannic L, Filbin MG, Shaw MKL, Baumgartner A, DeWitt JC, Groves A, Mayr L, Weisman HR, et al. (2019) Resolving medulloblastoma cellular architecture by single-cell genomics. *Nature* **572**: 74–79.
58. Zhu X, Ching T, Pan X, Weissman SM, Garmire L (2017) Detecting heterogeneity in

single-cell RNA-Seq data by non-negative matrix factorization. *PeerJ* **2017**: 1–20.

59. Soldatov R, Kaucka M, Kastriti ME, Petersen J, Chontorotzea T, Englmaier L, Akkuratova N, Yang Y, Häring M, Dyachuk V, et al. (2019) Spatiotemporal structure of cell fate decisions in murine neural crest. *Science (80-)* **364**:.
60. Szczyrba J, Niesen A, Wagner M, Wandernoth PM, Aumüller G, Wennemuth G (2017) Neuroendocrine cells of the prostate derive from the neural crest. *J Biol Chem* **292**: 2021–2031.
61. Simões-Costa M, Bronner ME (2015) Establishing neural crest identity: a gene regulatory recipe. *Development* **142**: 242–257.
62. Abida W, Cyrta J, Heller G, Prandi D, Armenia J, Coleman I, Cieslik M, Benelli M, Robinson D, Van Allen EM, et al. (2019) Genomic correlates of clinical outcome in advanced prostate cancer. *Proc Natl Acad Sci* 201902651.
63. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine JC, Geurts P, Aerts J, et al. (2017) SCENIC: Single-cell regulatory network inference and clustering. *Nat Methods* **14**: 1083–1086.
64. Wang CF, Hsing HW, Zhuang ZH, Wen MH, Chang WJ, Briz CG, Nieto M, Shyu BC, Chou SJ (2017) Lhx2 Expression in Postmitotic Cortical Neurons Initiates Assembly of the Thalamocortical Somatosensory Circuit. *Cell Rep* **18**: 849–856.
65. Sammeta N, Hardin DL, McClintock TS (2010) Uncx regulates proliferation of neural progenitor cells and neuronal survival in the olfactory epithelium. *Mol Cell Neurosci* **45**: 398–407.
66. Kline RA, Kaifer KA, Osman EY, Carella F, Tiberi A, Ross J, Pennetta G, Lorson CL, Murray LM (2017) Comparison of independent screens on differentially vulnerable motor neurons reveals alpha-synuclein as a common modifier in motor neuron diseases. *PLoS Genet* **13**: 1–20.
67. Wang ZA, Mitrofanova A, Bergren SK, Abate-Shen C, Cardiff RD, Califano A, Shen MM (2013) Lineage analysis of basal epithelial cells reveals their unexpected plasticity and supports a cell-of-origin model for prostate cancer heterogeneity. *Nat Cell Biol* **15**: 274–283.
68. Goldstein AS, Huang J, Guo C, Garraway IP, Witte ON (2010) Identification of a cell of origin for human prostate cancer. *Science* **329**: 568–571.
69. Mcinnes L, Healy J, Melville J (2018) UMAP : Uniform Manifold Approximation and Projection for Dimension Reduction arXiv : 1802 . 03426v2 [stat . ML] 6 Dec 2018.
70. van Galen P, Hovestadt V, Wadsworth MH, Hughes TK, Griffin GK, Battaglia S, Verga JA, Stephansky J, Pastika TJ, Lombardi Story J, et al. (2019) Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell* **176**: 1265-1281.e24.
71. Inage E, Kasakura K, Yashiro T, Suzuki R, Baba Y, Nakano N, Hara M, Tanabe A, Oboki K, Matsumoto K, et al. (2014) Critical Roles for PU.1, GATA1, and GATA2 in the Expression of Human FcεRI on Mast Cells: PU.1 and GATA1 Transactivate FCER1A , and GATA2 Transactivates FCER1A and MS4A2. *J Immunol* **192**: 3936–3946.
72. Cells M, Tsai SH, Kinoshita M, Kusu T, Kayama H, Okumura R, Ikeda K (2015) The Ectoenzyme E-NPP3 Negatively Regulates ATP- Dependent Chronic Allergic Responses by Basophils Article The Ectoenzyme E-NPP3 Negatively Regulates

ATP-Dependent Chronic Allergic Responses by Basophils and Mast Cells. 279–293.

73. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, Rotem A, Rodman C, Lian C, Murphy G, et al. (2016) Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science (80-)* **352**: 189–196.
74. Puram S V., Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, Rodman C, Luo CL, Mroz EA, Emerick KS, et al. (2017) Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **171**: 1611-1624.e24.
75. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh P ru, Raychaudhuri S (2019) Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**: 1289–1296.
76. Johnson WE, Li C, Rabinovic A (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**: 118–127.
77. Hänzelmann S, Castelo R, Guinney J (2013) GSVA: Gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics* **14**:.
78. Chung W, Eum HH, Lee HO, Lee KM, Lee HB, Kim KT, Ryu HS, Kim S, Lee JE, Park YH, et al. (2017) Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat Commun* **8**: 1–12.

Figure legends

Fig. 1 Single-Cell Transcriptomic Profiling of 6 CRPC Tumors

(A) Workflow for single-cell extraction, sequencing and analysis. (**B**) Haematoxylin and eosin (H&E) staining for 6 CRPC patients. The scale bars represent 25μm. (**C**) UMAP plots of cells from 6 patients with cells colored based on the cell types (upper row) and NE scores using the well-established NE marker genes (lower row). The minimum score is indicated by light grey and the maximum score is indicated by blue. The red arrows pointed to high NE score cell population.

Fig.2 NE cells present an epithelial phenotype

(**A**) Pairwise correlations between the expression profiles of 12,861 epithelial cells (rows, column) from 6 CRPC samples (color bar). (**B**) Enrichment scores for basal, luminal, NE, AR, stemness and EMT pathway associated genes. (**C**) Inferred CNV profiles. Black indicates the CNV (Supplementary fig. 3). (**D**) UMAP visualization of all 12，861 epithelial cells from 6 patients with color-coded for the sample origin. (**E**) UMAP visualization of all 12,861 epithelial cells for the 6 patients with cells colored based on the basal score (left) and average CNV signal (right). The minimum score is indicated by light grey and the maximum score is indicated by blue (left) or black (right).

Fig.3 Intratumor heterogeneity analyses reveal different extents of NE differentiation

(**A**) UMAP visualization of epithelial cell sub-clusters from each sample. (**B**) Heatmap depicting prostate lineage marker genes and AR pathway gene expression levels in epithelial cell sub-clusters from each sample. Those highlighted in yellow frame showed cluster 5 in patient #4 and cluster 4 in patient #6 was NE sub-clusters. (**C**) Immunohistochemistry (IHC) staining for AR, SYP and SOX2 in sections from 5 samples. Scale bars represent 50μm.

Fig. 4 Epithelial cellular relationships in patient #4

(**A**) UMAP visualization of epithelial cells from patient #4 with color-coded for the corresponding sub-cluster (left) and the average inferred CNVs signals (right; grey to black). (**B**) Violin plots of the expression level of NE, urothelial-like, basal and luminal lineage markers across the populations shown in Fig 4A. (**C**) Immunofluorescence (IF) co-staining for K18 (red) and SYP (green) in sections for patient #4. Scale bar represents 100μm. (**D**) The PAGA graph and connectivity scores of the populations shown in Fig. 4A. (**E**) Velocities of epithelial cells from patient #4 are visualized as streamlines in a UMAP-based embedding, in which color-coded for the corresponding populations shown in Fig. 4A. (**F**) Representative confocal fluorescence microscopy of triple co-staining of SYP (green), K18 (grey) and K5 (red) in PC TMA sections. The SYP+ NE cells have 3 subtypes: $K18^{+}K5^{-}SYP^{+}$, $K18^{-}K5^{-}SYP^{+}$ and $K18^{+}K5^{+}SYP^{+}$. Scale bars represent 25μm. (**G**) Pie chart of statistics for PC

TMA co-stainning results showing that the major part of prostate cancers contain NE cells with exclusive luminal phnotype ($K18^+SYP^+$,83/102).

Fig. 5 Epithelial cellular relationships in patient #6

(**A**) UMAP visualization of epithelial cells from Patient #6 with color-coded for the corresponding sub-cluster (top) and the average inferred CNV signal (bottom; grey to black). (**B**) Violin plots of the expression level of NE, basal and luminal lineage markers across the populations shown in Fig. 5A. (**C**) The PAGA graph and connectivity scores of the populations shown in Fig. 5A. (**D**) Velocities of epithelial cells from patient #6 are visualized as streamlines in a UMAP-based embedding, in which color-coded for the corresponding Seurat cluster in Fig. 5A. (**E**) Phase portraits (upper row) and expression dynamics along latent time (lower row) for specific genes selected from top-ranked likelihood gene set (gene likelihood >0.2).

Fig.6 Intra-tumoral meta-programs underlying NED

(**A**) Heatmap showing scores of 12861 epithelial cells (column, from 6 CRPC patients) for each of 60 programs (rows) derived from NMF analysis of individual sample. Cells and programs are hierarchically clustered, and 3 NE-related meta-programs (P1, P2 and P4) and a cell cycle-related meta-program (P3) are highlighted. (**B**) Enrichment scores of prostate lineages: basal, luminal, NE marker genes and AR, stemness, EMT and cell cycle pathway genes in cells ordered as in Fig. 6A, with the color-coding for the

corresponding CRPC sample. (**C**) Pearson correlation between the expression of genes of P1, P2 and P4 and the NE score, as measured by the average expression of 14 known NE markers. Three previously published bulk RNA-Seq datasets were used in this analysis, as described in method part. Highlighted in red are some known NED genes. (**D**) Heatmap depicting strong expression of 120 genes (Pearson R ≥ 0.3, as measured by Pearson correlation analysis shown in Fig. 6C) in $AR^{-}NE^{+}$ group of Morrissey dataset. Total samples are divided into five groups as previously suggested in ref.40.

Fig.7 Transcription factor regulatory networks underlying NED

(**A**) Heatmap of SCENIC binary regulon activities (row) and NE scores (row) of 12,861 epithelial cells (column). Three TF regulatory networks with high activities in NE cells were highlighted. (**B**) Heatmap of the mean regulon activities (row) that differentially expressed on epithelial clusters (column) of patient #4. (**C**) t-SNE on the SCENIC regulon activity matrix and the representative regulon activities on epithelial cells from patient #4. Cells are colored by the corresponding cluster and gradient of regulon activity (grey to red). **(D)** Heatmap of the mean regulon activities (row) that differentially expressed on epithelial clusters (column) of patient #6. **(E)** t-SNE on the SCENIC regulon activity matrix and the representative regulon activities on epithelial cells from patient #6. Cells are colored by the corresponding cluster and gradient of regulon activity (grey to red).

Fig.8 Cellular relationship and disease progression model of NEPC

Schematic illustration of tumor evolution toward the neuroendocrine phenotype, in which dotted arrows indicate the potential relationship between cell lineages and the solid arrows indicate that NEPC is directly originated from AR-dependent tumor cells. In this model, we suppose that the NE precursor, AR-independent tumor cell, is directly transdifferentiate from the luminal-like tumor cell, and that is the precursor, which will next evolve in forming the focal NEPC and finally progress to small-cell (pure) NEPC. The extent of AR and NE signature scores vary over the spectrum of adenocarcinoma to neuroendocrine transdifferentiation (orange indicates a high level of AR signal and green indicates a high level of NE signal).

Fig.S1 CT images of 5 prostate tumors

CT images showing the tumor sites of five CRPC patients. Red arrows indicate the biopsy sites in each patient.

Fig.S2 Single-Cell Transcriptomic Profiling of 6 CRPC Tumors

(**A**) UMAP visualization of the 21,292 cells from 6 CRPC patients colored by clusters and sample origin, respectively. (**B**) UMAP plots of the 21,291 cells from 6 patients with cells colored by the score of marker gene sets for particular cell types (marker genes and associated cell types are indicated next to each plot). The minimum score is indicated by light grey and the maximum score is indicated by red. (**C**) Heatmap shows the expression level of epithelial

lineage markers and NE markers. (**D**) Bar plot for NE index in cells that grouped by cell type (Data are mean ± SD).

Fig.S3 Cell identity determination by inferred copy number variation (CNV) analysis

Chromosomal landscape of inferred large-scale CNVs for tumor cells from 6 patients, in which epithelial cells from 3 healthy men were set as the reference "normal" cells. The color bar in row indicates for 22 chromosomes while which in column indicates for corresponding patient.

Fig.S4 Clonal analysis of patient #4 and #6 by inferred copy number variation (CNV) analysis

(**A** and **B**) Chromosomal landscape (left) of inferred large-scale CNVs for epithelial cells from patient #4 (A) and #6 (B), in which epithelial cells from 3 healthy men were set as the reference "normal" cells. The color bar in row indicates for 22 chromosomes while which in column indicates for corresponding Seurat cluster. UMAP visualization of average inferred CNV signals (right).

Table 1 - Clinical characteristics of the 6 CRPC patients

Patient ID	Age	PSA level at diagnosis of Pca (ng/ml)	The Gleason score at diagnosis of Pca	The TNM stage at diagnosis of Pca	First-line therapy	Second-line therapy	Time from treatment start to CRPC (mo)	Time from CRPC to now (mo)	PSA level at present (ng/ml)	The TNM stage at present
Patient #1	82	8.88	4+5=9	cT2cN0M0	Goserelin, bicalutamide	None	9	0.5	5.85	cT4N0M0
Patient #2	82	56.53	4+3=7	cT3bN1M0	Goserelin, bicalutamide	None	15.9	3.2	18.57	cT3bN0M0
Patient #3	86	55.16	4+4=8	cT2cN0M0	Goserelin, bicalutamide	None	15.9	3.7	11.76	cT2cN0M0
Patient #4	78	> 149	4+3=7	cT4N0M0	Bilateral orchidectomy, bicalutamide	Docetaxel, Abiraterone	12.5	18.6	117.95	cT4N0M1
Patient #5	65	15.6	4+4=8	cT3bN1M1a	Bilateral orchidectomy, bicalutamide	None	27.8	2.6	0.9	cT4N1M1
Patient #6	70	72.72	small-cell NEPC	cT4N1M1	Docetaxel	None	14.8	3.7	7.18	cT3bN0M1

PSA, prostate specific antigen; Pca, prostate cancer; TNM, tumor node metastasis; CRPC, castration-resistant prostate cancer; NEPC, neuroendocrine prostate cancer.

Fig. 1

Fig. 2

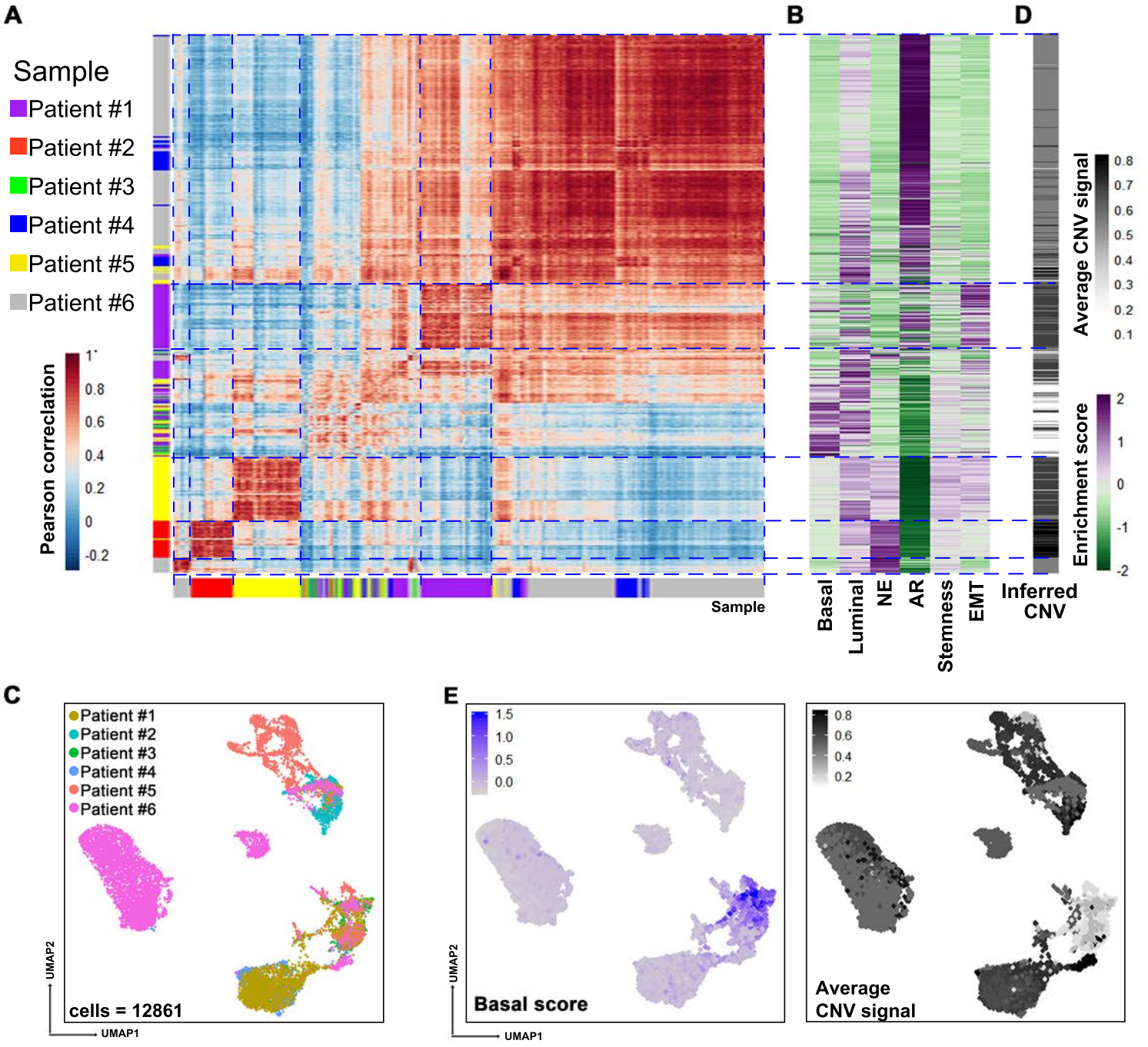


Fig. 3

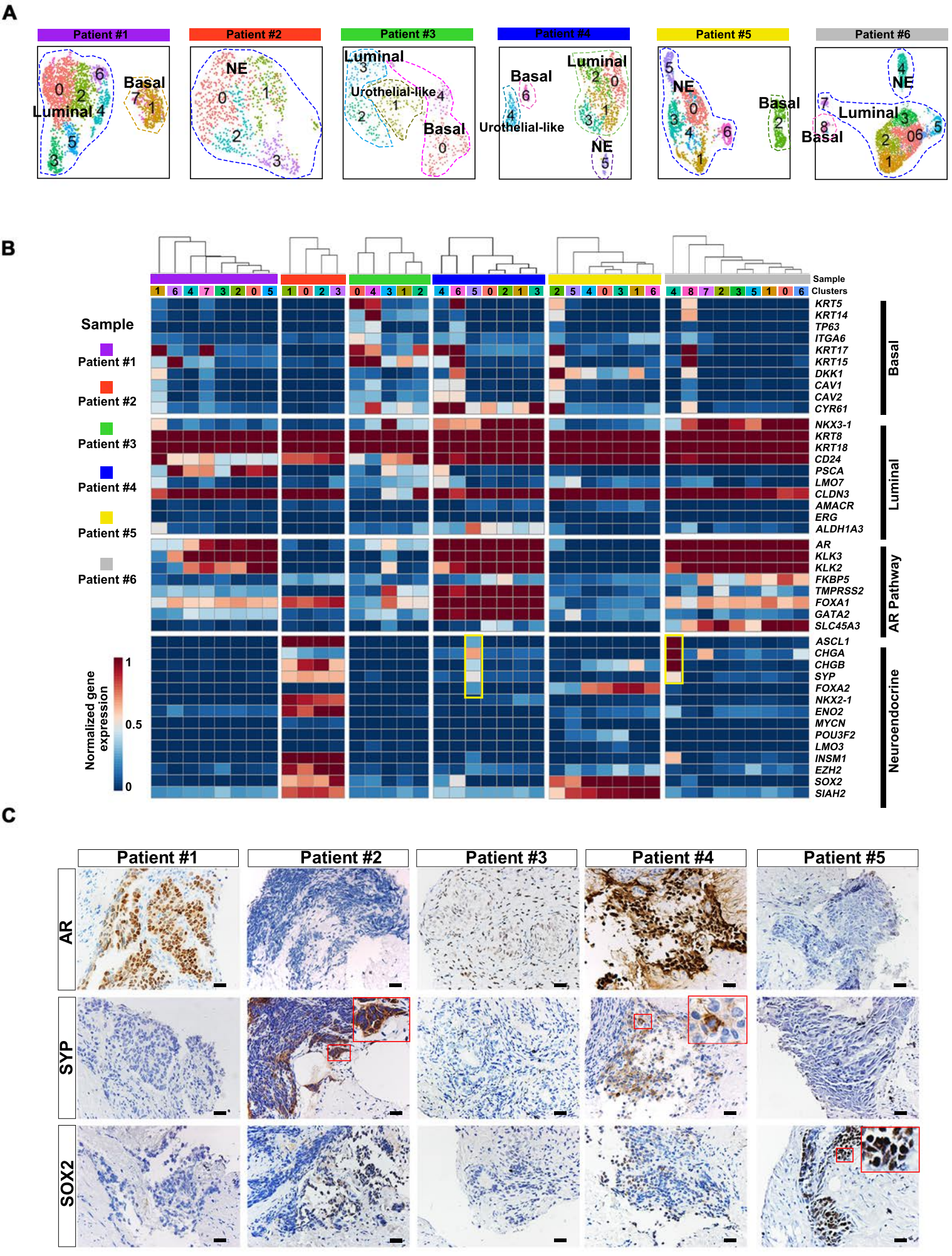


Fig. 4

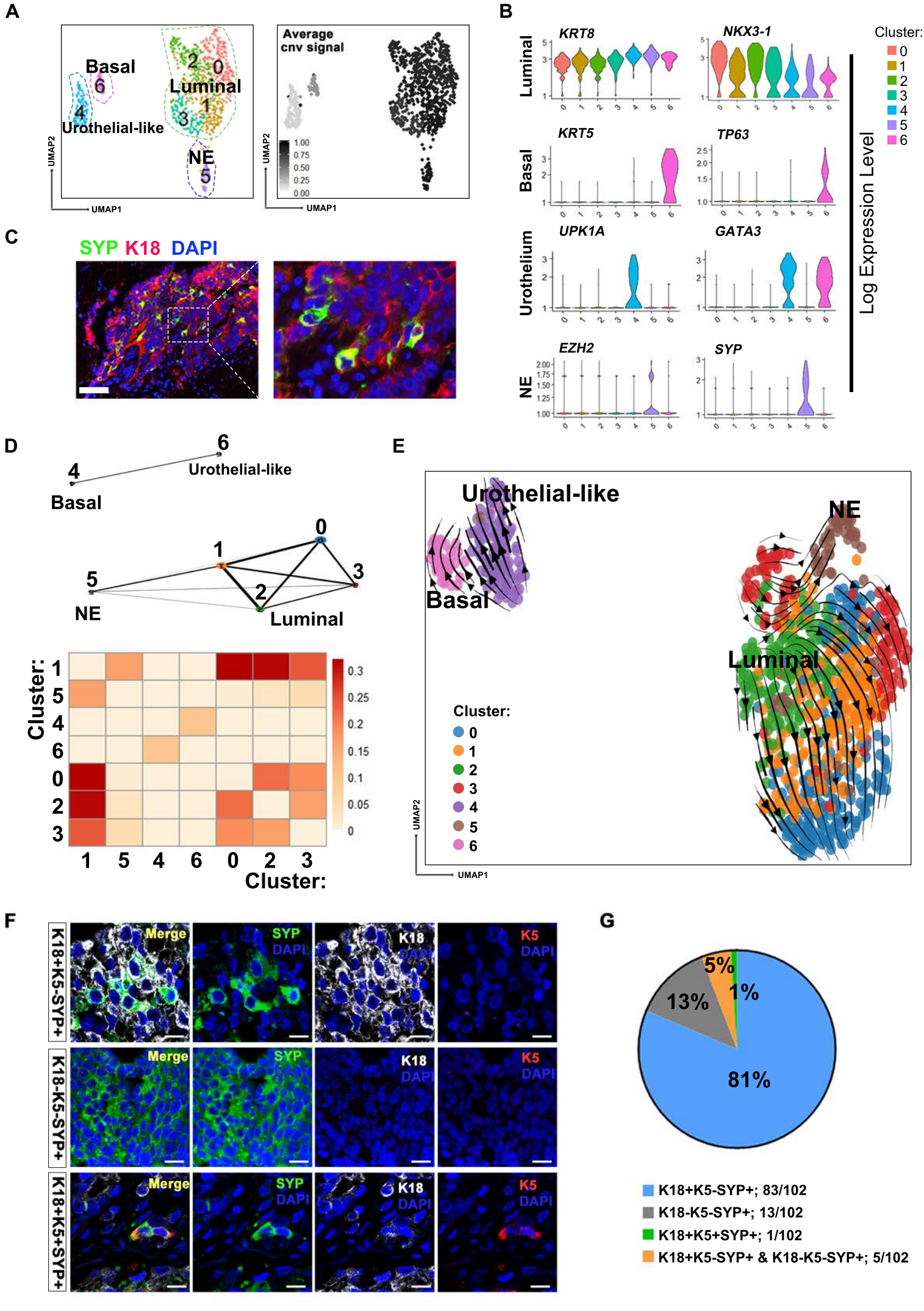


Fig. 5

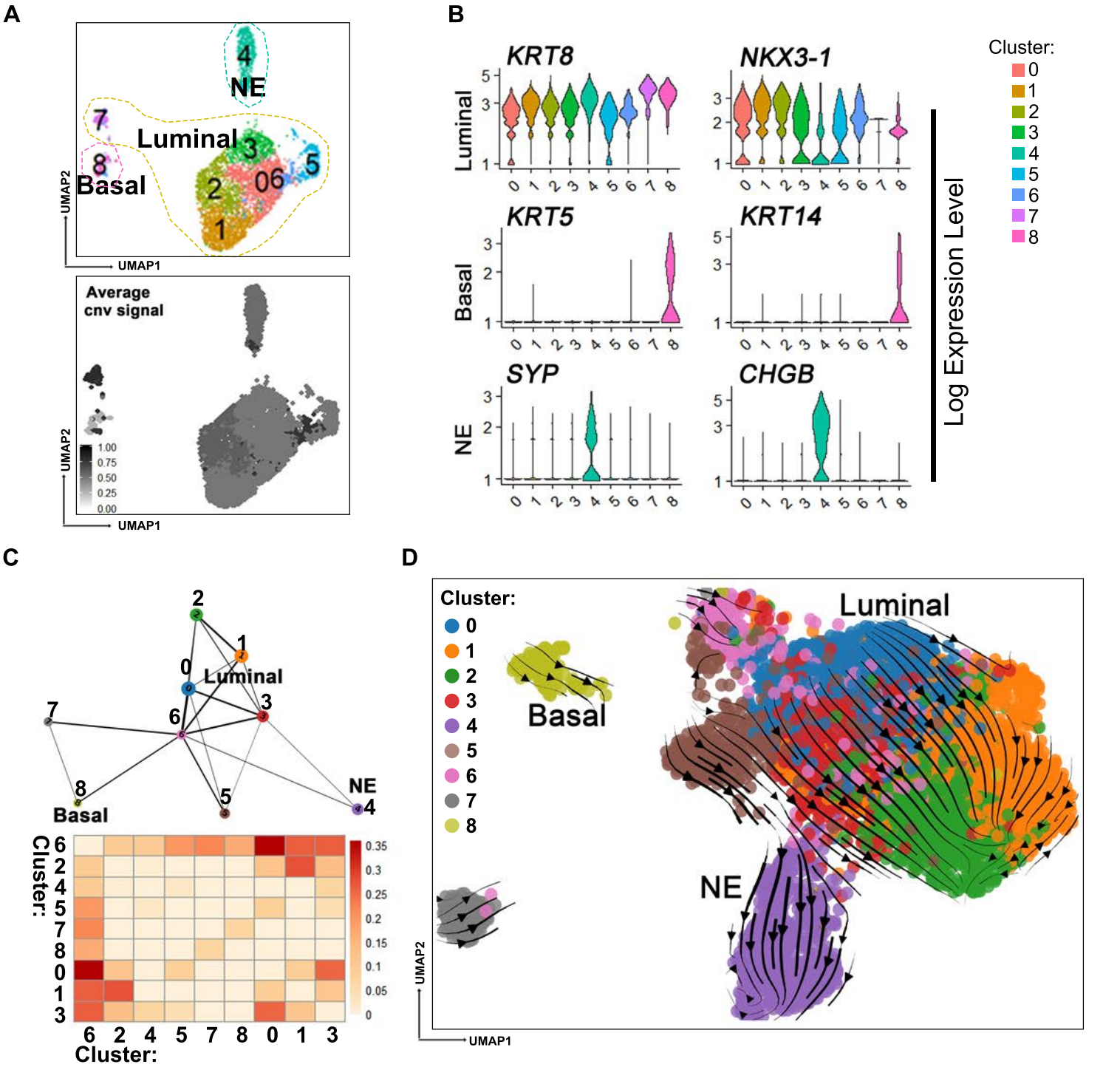


Fig. 6

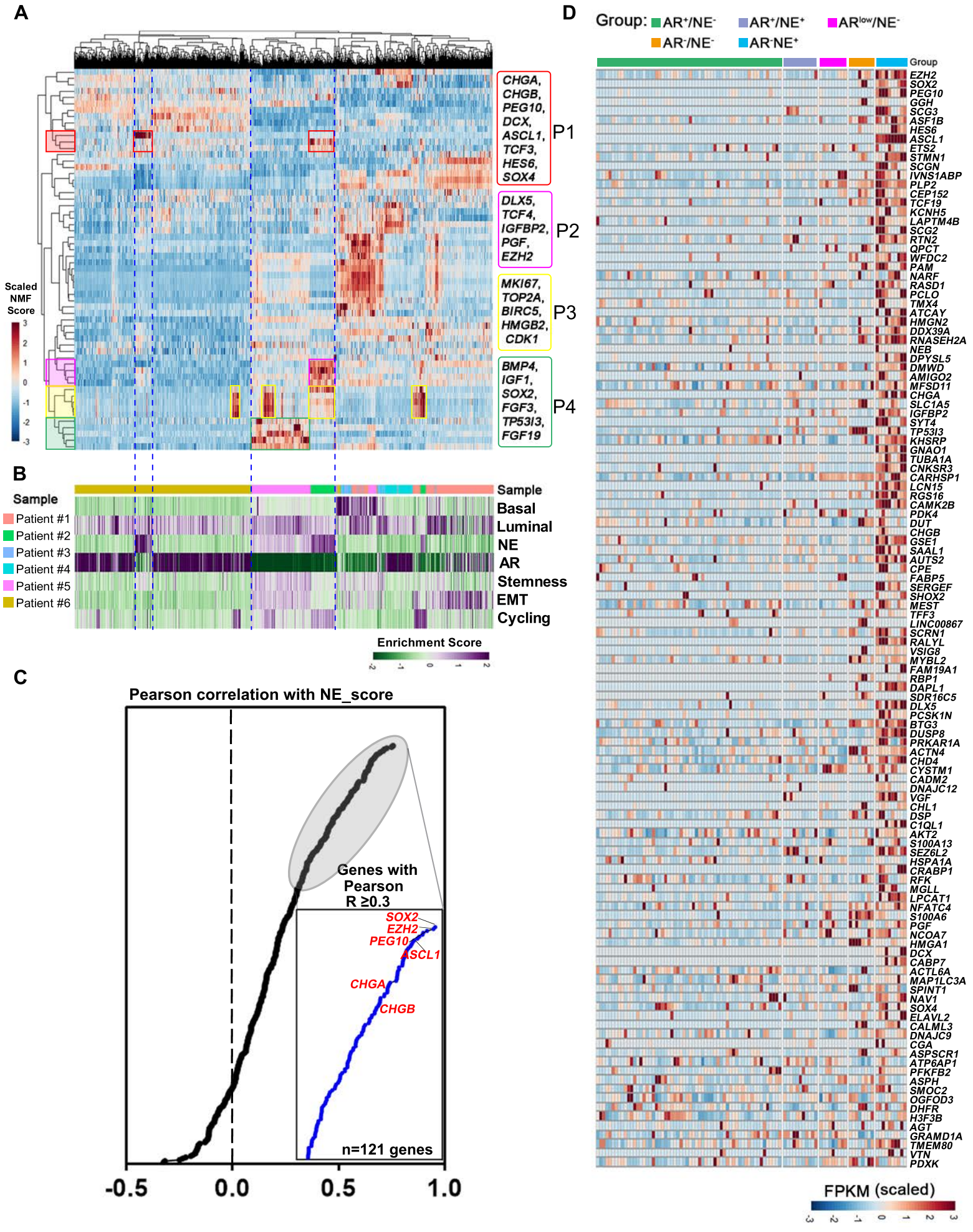
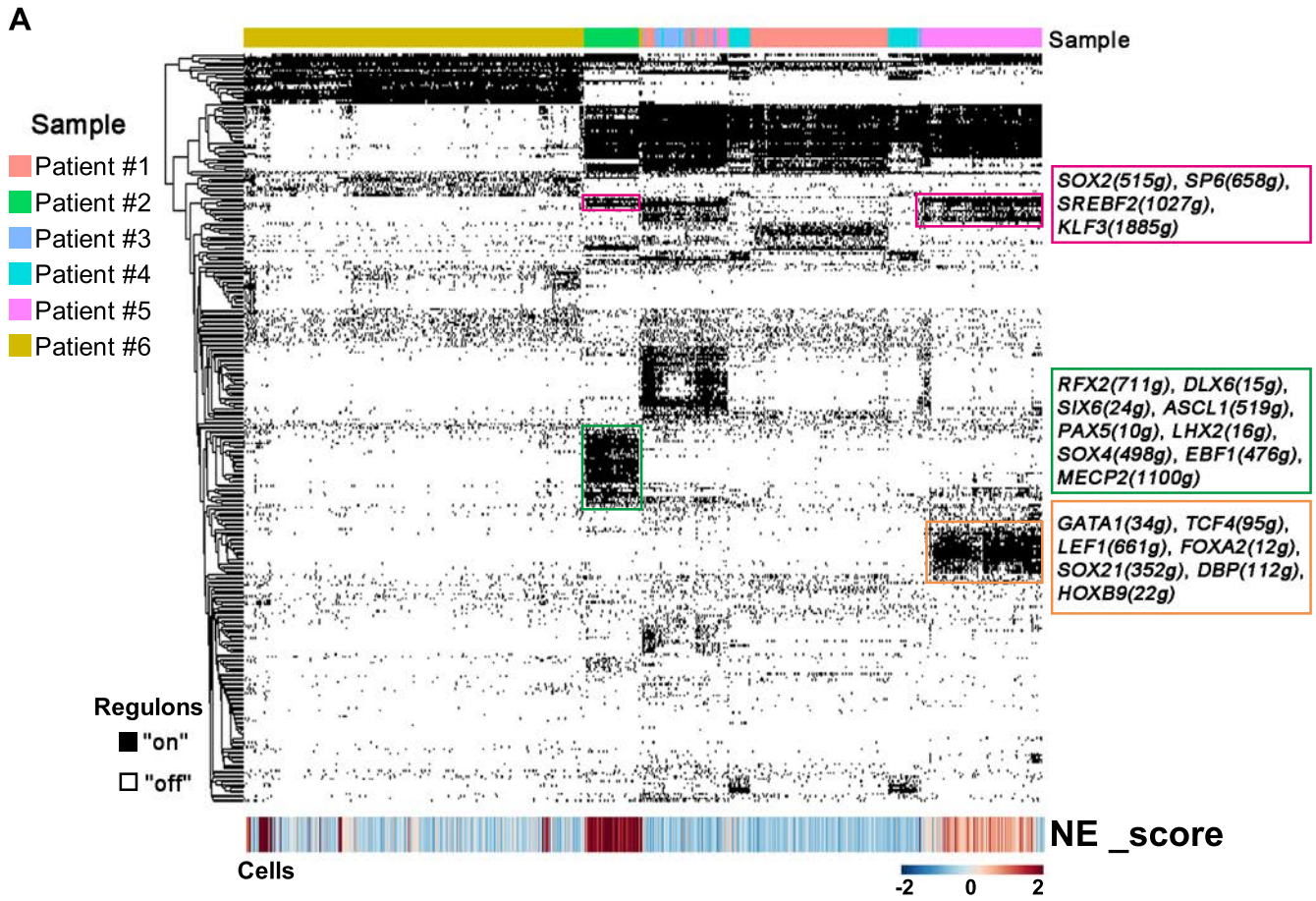
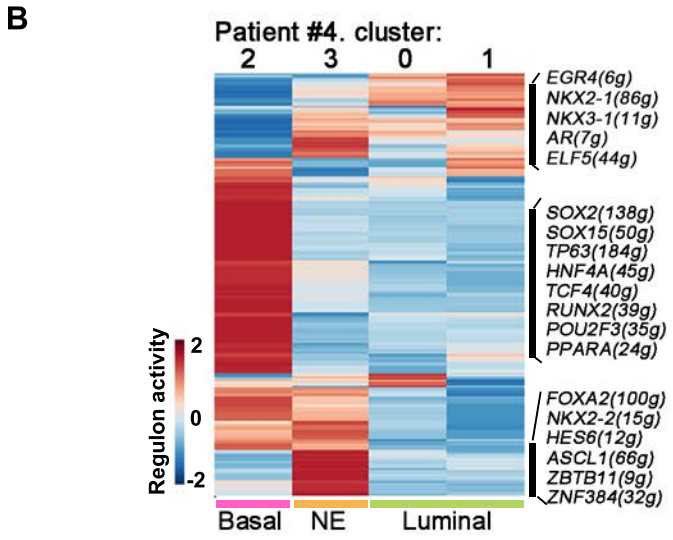
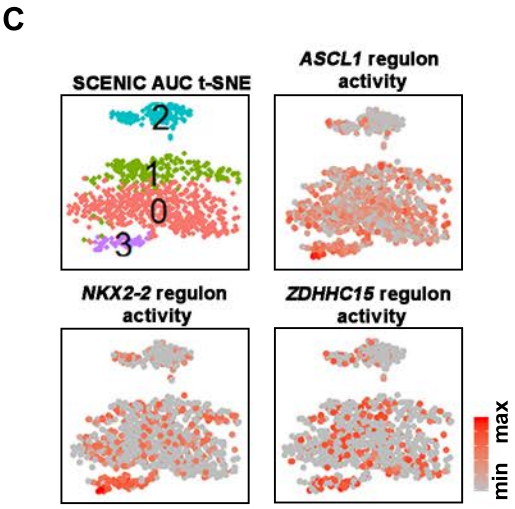
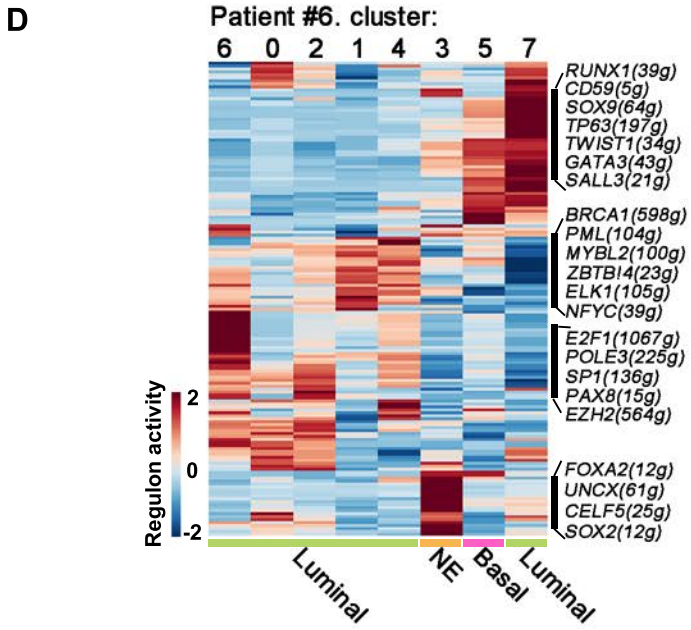


Fig. 1









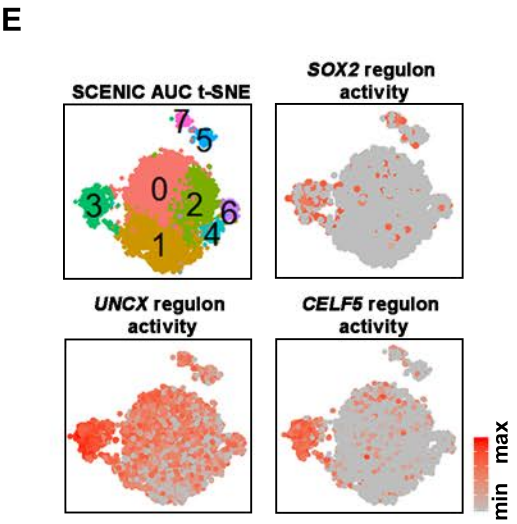
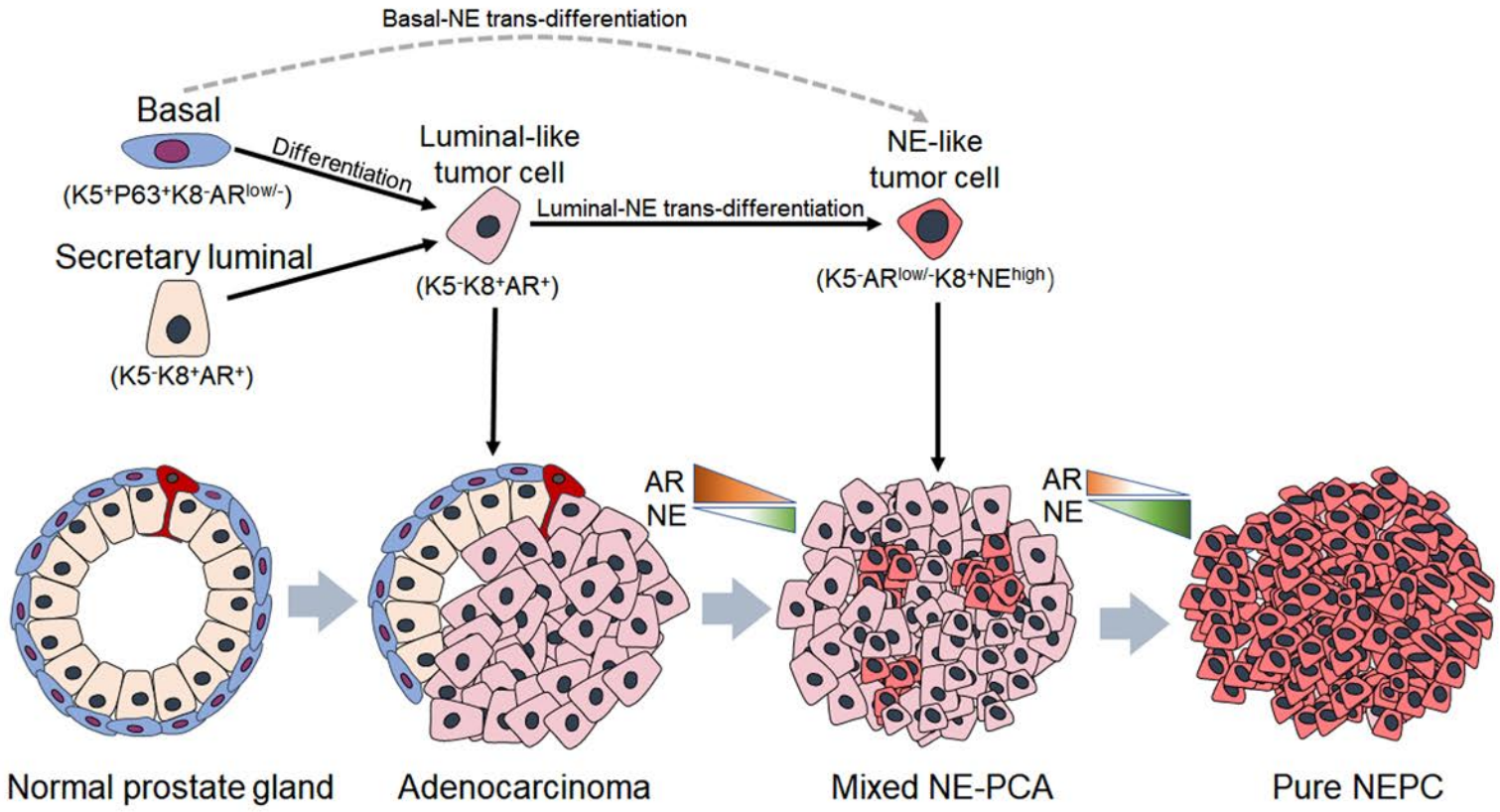
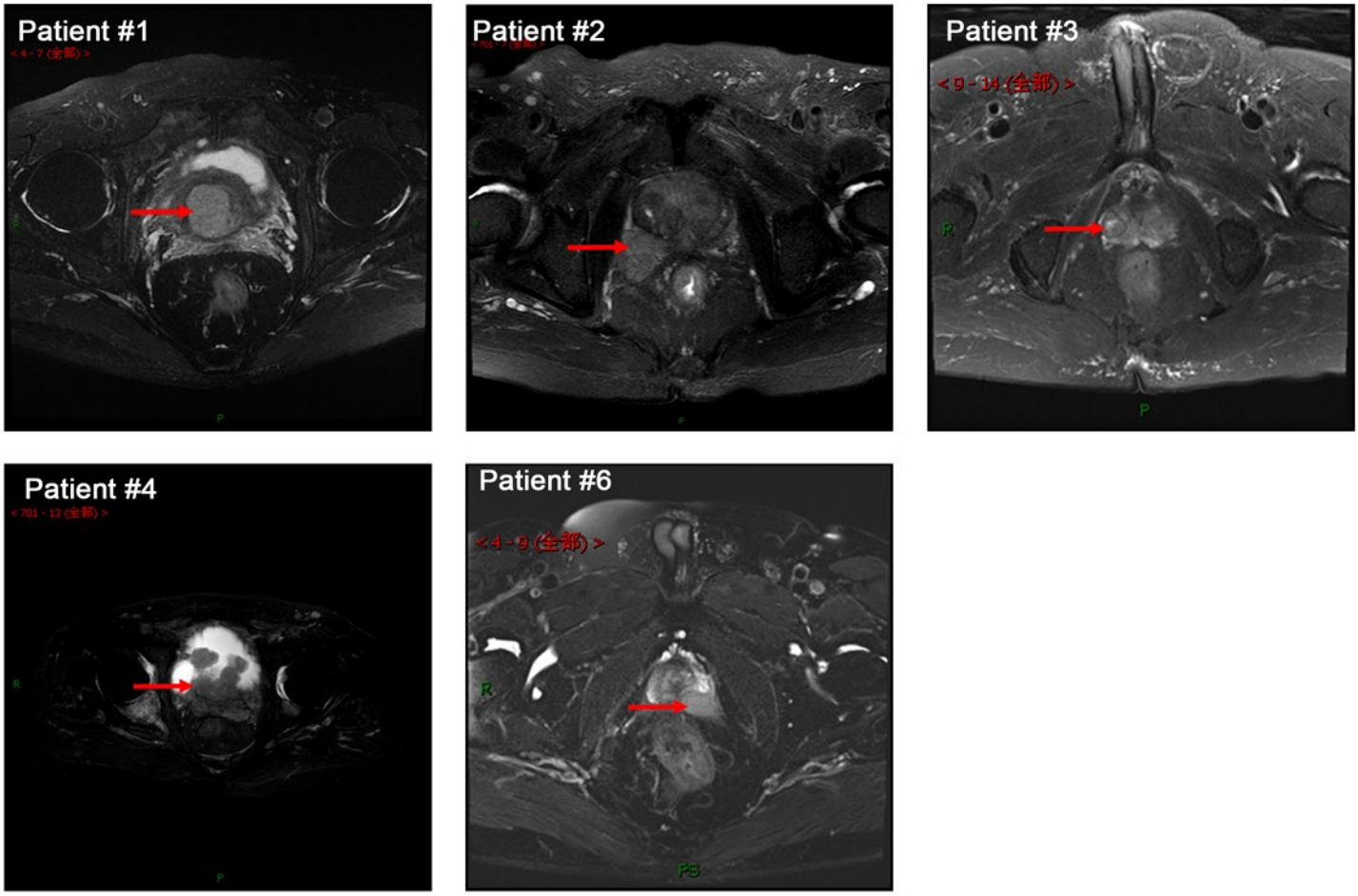


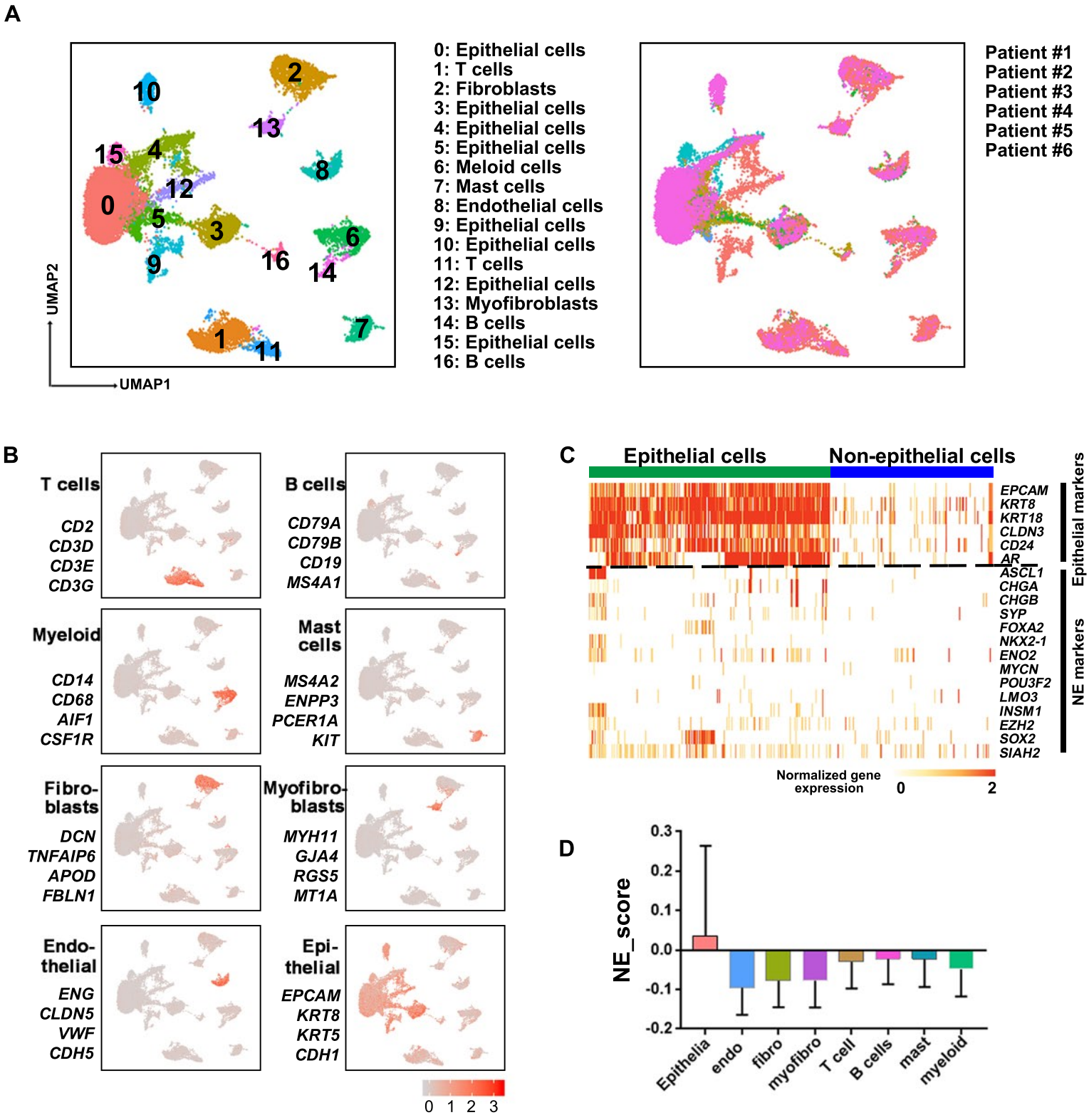
Fig. 8



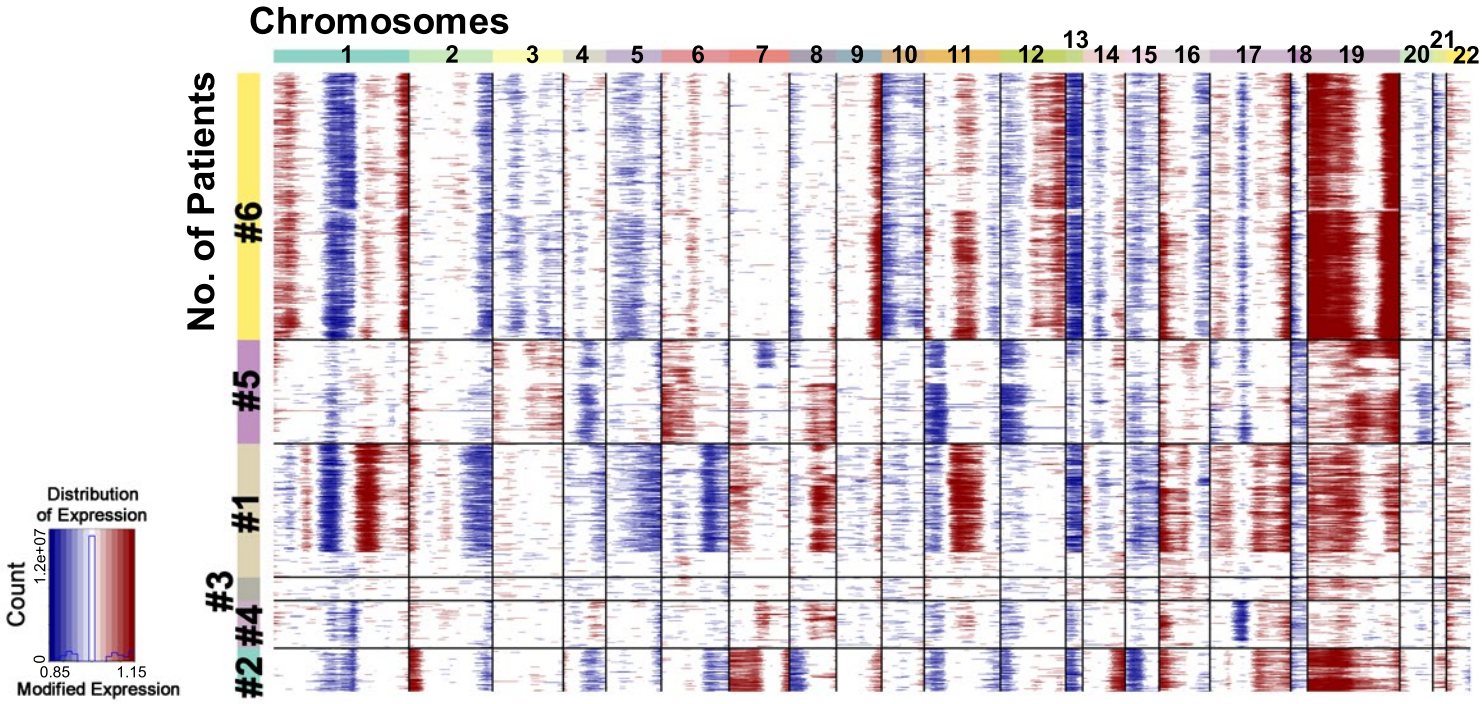
Supplementary fig. 1



Supplementary fig. 2



Supplementary fig. 3



Supplementary fig. 4