Transition of Wnt signaling microenvironment delineates the squamo-columnar junction and emergence of squamous metaplasia of the cervix

Cindrilla Chumduri[1,#,*], Rajendra Kumar Gurumurthy[1,#], Hilmar Berger[1], Stefanie Koster[1], Volker Brinkmann[1], Uwe Klemm[1], Hans-Joachim Mollenkopf[1], Hermann Herbst[2], Mandy Mangler[3,4], Thomas F Meyer[1,*]

[1]Department of Molecular Biology, Max Planck Institute for Infection Biology, Berlin, Germany

[2]Institute of Pathology, Vivantes Klinikum Berlin, Rudower Straße 48, 12351 Berlin, Germany

[3]Charité University Medicine, Department of Gynecology, Berlin, Germany

Current address: [4]Auguste-Viktoria-Klinikum, Vivantes, Klinik für Gynäkologie und Geburtsmedizin 12157 Berlin, Germany

[#]These authors contributed equally

[*]Corresponding authors:

Emails:

tfm@mpiib-berlin.mpg.de

chumduri@mpiib-berlin.mpg.de

Phone: +49-30-28-460-400

Abstract

The transition zones (TZ) between squamous and columnar epithelium constitute hotspots for the emergence of cancers. Carcinogenesis at these sites is often preceded by the development of metaplasia, where one epithelial type invades the neighboring one. It remains unclear how these niches are restrained at the boundary between the two epithelial types and what factors contribute to metaplasia. Here we show that the cervical squamo-columnar junction derives from two distinct stem cell lineages that meet at the TZ. In contrast to the prevailing notion, our analysis of cervical tissue showed that the TZ is devoid of any locally restricted, specialized stem cell population, which has been implicated as precursor of both cervical squamous cell carcinoma and adenocarcinoma. Instead, we reveal that these cancers originate from two separate stem cell lineages. We show that the switch in the underlying Wnt signaling milieu of the stroma is a key determinant of proliferation or quiescence of epithelial stem cell lineages at the TZ. Strikingly, while the columnar lineage of the endocervix is driven by Wnt signaling, the maintenance of squamous stratified epithelium of the ectocervix and emergence of squamous metaplasia requires inhibition of Wnt signaling via expression of Dickkopf2 (Dkk2) in the underlying stroma. Moreover, Notch signaling is required for squamous cell stratification. Thus, our results indicate that homeostasis at the TZ is not maintained by a transition from one epithelial type to another but rather results from alternative signals from the stromal compartment driving the differential proliferation of the respective cell lineages at the squamo-columnar junction.

Introduction

Recent years have seen rapid progress in our understanding of how adult epithelial tissues are maintained by dedicated stem cell niches. In many cases, Wnt signaling initiated from the underlying stroma plays decisive roles. What is entirely unclear so far is how these niches are restrained at the boundary between two different types of epithelia, e.g. columnar and stratified ones. Such transition zones are found in several organs in the human body, e.g. the gastro-esophageal junction, as well as the cervix. These sites almost invariably show a predisposition towards transformation, which is preceded by metaplasia, where one epithelium type invades another.

Cervical cancer predominantly occurs at the TZ, where the stratified squamous epithelium of the ectocervix and the simple columnar epithelium of the endocervix meet [1,2]. It is one of the most common and deadly cancers in women and occurs as two histologically distinct types: adenocarcinomas (ADC) and squamous cell carcinomas (SCC) [3]. Both ADC and SCC are proposed to originate from a common precursor stem cell population with a unique immunophenotype that is localized exclusively at the TZ. This precursor cell population is proposed to consist either of residual embryonic cytokeratin 7 (KRT7) positive cells [4], p63+/KRT7+/KRT5+ transitional basal cells [5], or reserve cells [6]. As a result, prophylactic ablation of residual embryonic cytokeratin 7 (KRT7) positive cells has been proposed as a method to prevent cervical cancers [7,8]. However, how this population gives rise to two epithelial types and the molecular mechanisms that govern maintenance of the TZ and define the distinct epithelial compartments has not been studied. In addition, the changes that contribute to metaplasia at the TZ of the cervix (or indeed other tissues) also remain unclear.

Here, using *in vivo* lineage tracing, cervical organoid models, single molecule RNA in situ hybridization (RNA-ISH) and a mouse model of squamous metaplasia, we show that endo- and ectocervical epithelia are derived from two separate lineage-specific stem cell populations that meet at the TZ. Using bioinformatic analysis as well as immunohistochemistry of cancer tissue, we find that the transcriptional signatures of these lineages correspond to those of SCC and ADC, indicating that these two histologically distinct cancer types also arise from the two different lineages rather than from a common precursor. We show that two distinct active or inhibitory Wnt signaling milieus are established in the endo- and ectocervical region by Wnt-pathway agonists and antagonists, respectively, which are differentially expressed in the epithelium and sub-epithelial stroma on either side of the TZ *in vivo*. Organoids derived from these two distinct stem cell lineages show a strong divergence in their requirement for Wnt signaling. Strikingly, we demonstrate that endocervix also harbors squamous stem cells, which are kept in a quiescent state by the Wnt signaling microenvironment that promotes proliferation of the columnar lineage. *In vitro*, these quiescent stem cells can give rise to squamous epithelium in the absence of Wnt. This is the first time that the presence of two stem cell lineages with opposing signaling requirements within the same niche has been observed. In addition, using an *in vivo* mouse model of cervical metaplasia, we show that loss of Wnt signals by induction of Wnt inhibitory signaling molecules in the endocervical region leads to the outgrowth of the pre-existing ectocervical stem cells within the transition zone and the endocervix. These findings provide the first mechanistic underpinning of how homeostasis is maintained at the transition zone and how development of metaplasia is initiated.

Results

Distinct cellular origins of squamous and columnar epithelium

To obtain deeper insight into the cellular composition and molecular determinants of TZ maintenance, we carried out a detailed analysis of marker profiles in human and mouse cervical epithelium. Strikingly, our comprehensive, unbiased analysis including the entire endo- and ectocervix regions failed to detect any specific cell type that was restricted exclusively to the TZ - in contrast to the prevailing concept. Instead, we observed two distinct epithelial lineages, with KRT5 expressed throughout the squamous stratified epithelium and KRT7/KRT8 expressed throughout the columnar epithelium. At the TZ there is an overlap of both lineages, where KRT5+ basal cells appear to displace overlying KRT7+/KRT8+ columnar cells to form squamous stratified epithelium (Fig. 1 A-D and Fig. S1 A-D). RNA-ISH confirmed that KRT8 and KRT5 expression was restricted to the columnar epithelium and basal/parabasal cells of squamous epithelium, respectively (Fig. 1 E-F, and Fig. S2 A-B). In contrast to previous reports describing a discrete KRT7 population, restricted to the TZ [4,5], we observed KRT7 expression at high levels throughout the endocervical epithelium and sparse expression in the ectocervical epithelium (Fig. 1 G and Fig. S2 C). We also observed that patches of subcolumnar KRT5+ cells occurred sporadically beneath KRT7+/KRT8+ cells within the endocervix (Fig. S1A). These islands of KRT5+ cells may correlate with foci of squamous metaplasia that are frequently observed within the endocervix and account for 10% of premalignant squamous intraepithelial lesions (SIL) [9,10].

Further, using KRT5CreErt2/Rosa26-tdTomato and KRT8CreErt2/Rosa26-tdTomato mice for genetic lineage tracing we confirmed the presence of two distinct epithelial lineages in the cervix (Fig. 1 H-I and J). In these mice, Cre was induced by tamoxifen injection 4 weeks after birth. At 16

114 weeks of age, KRT5+ cells exclusively populated the stratified epithelium, including all
115 differentiated cells, while KRT8+ cells exclusively generated endocervical epithelium.

116 **Wnt agonists and antagonists in epithelia and stroma orchestrate TZ**

117 To gain insight into the factors that regulate the two lineages and maintain the TZ, we established
118 and defined conditions that facilitate long-term *in vitro* propagation of ecto- and endocervical
119 epithelial stem cells as 3D organoids. Wnt signaling was described to be essential for generation
120 and long-term maintenance of adult epithelial stem cell-derived organoids from the various
121 tissues so far described, as shown by the requirement of Wnt3a and R-spondin 1 in the tissue-
122 specific organoid culture media [11]. In contrast to this, the presence of Wnt3a and R-spondin 1
123 (RSPO1) in the culture medium was detrimental for the formation and expansion of human and
124 mouse squamous stratified organoids derived from single ectocervical stem cells (Fig. 2A and Fig.
125 S3A-C). The presence of epidermal growth factor (EGF), fibroblast growth factor 10 (FGF-10), and
126 the inhibition of transforming growth factor beta (TGF-β) and bone morphogenetic protein (BMP)
127 signaling, on the other hand, was essential for long-term maintenance of these organoids.
128 Squamous stratified organoid growth was further increased in the presence of the cAMP pathway
129 agonist forskolin (FSK) (Fig. 2A and Fig. S3D-E). These organoids are KRT5+/KRT7- and fully
130 recapitulate the *in vivo* tissue architecture, with stratified epithelial layers decorated with the
131 adhesion molecule E-cadherin (CDH1) (Fig. S3F and H). The outer layer consists of p63+ basal
132 cells, which differentiate into parabasal cells with p63 staining fading out towards the lumen.
133 Also, typical of ectocervical tissue, only basal cells express the proliferation marker Ki67 (Fig.
134 S3G).

In contrast to the ectocervix, stem cells derived from both proximal and distal endocervix give rise to organoids consisting of a simple columnar epithelial layer when cultured in the presence of Wnt proficient medium containing Wnt3a and RSPO1 (Fig. 2B-C). These organoids faithfully resemble the *in vivo* endocervical epithelium, are KRT7+/KRT5- and exhibit sporadic Ki67 staining (Fig. S4A and S3I). Their self-renewal capacity in culture can be maintained for more than seven months (Fig. S4B). Further, transcriptional profiling of organoids derived from human ecto- and endocervix revealed distinct keratin expression patterns (Fig. S3J).

Strikingly, if cells derived from endocervix tissue were cultured in Wnt deficient medium (FSK+ medium without Wnt3a or RSPO1), they gave rise to p63+ stratified organoids, resembling those derived from ectocervix (Fig. 2B-C, S4C). Since the formation of columnar rather than squamous organoids from endocervical stem cells was dependent on supplementation with Wnt agonists, we investigated the source of Wnt signaling in the cervix. Microarray analysis of organoids and RNA-ISH showed that the transcriptional regulation of Wnt in the endocervix diverges from that in the ectocervix: Wnt agonists are upregulated in columnar epithelium, while the Wnt antagonists Dickkopf WNT signaling pathway inhibitor 3 (DKK3), DKK1 and KREMEN1 are upregulated in squamous epithelium (Fig. 2D, E, G).

Further, we observed that the spatial distribution of extrinsic Wnt agonists and antagonists in the underlying stroma defines the borders between the two epithelial types. Both the Wnt agonist RSPO1 and its downstream target Axin2 are highly expressed in the lamina propria (stroma) beneath the columnar epithelium and RSPO3 in the muscularis of the endocervix (Fig. 2E, S4D-F). Notably, the Wnt antagonist DKK2 is specifically expressed in stroma proximal to the basal cells of the ectocervical squamous epithelium, which express high levels of DKK3 (Fig. 2F-H, S4G-

H). In contrast to the ectocervix, high levels of DKK3 expression were observed in the endocervical stroma, while expression of DKK1 was negligible in either region of the cervix (Fig. S5A). Expression levels of DKK4, RSPO2 and RSPO4 also did not show notable regional variation (Fig. S5B-D). Thus, the epithelium of the cervix is maintained by two distinct stem cell populations whose fate is determined by opposing Wnt signaling microenvironments, which are established through the interplay of the epithelial and stromal compartments of the endo- and ectocervix respectively, with a defined switch at the TZ.

Wnt antagonists, Notch and EGFR signaling maintain ectocervical stemness and differentiation

Next, we sought to identify the cellular pathways that control self-renewal and differentiation in human ectocervical tissue. Microarray analysis showed that squamous ectocervical organoids have a higher expression of Notch-related genes than organoids derived from the endocervical columnar epithelium (Fig 3A). We thus carried out a comparative analysis of 2D cells (2D-ecto), three-day-old early organoids (EO-ecto), and two-week-old, mature differentiated organoids (DO-ecto). 2D cultures were enriched for CDH1+ and p63+ cells, with >60% and >30% of cells showing organoid-forming potential at passage 1 and 8, respectively (Fig. 3B, Fig. S6A). Early organoids consist of 8-16 cells that are undifferentiated and positive for Ki67 and p63 (Fig. 3C and Fig. S6B). Mature organoids consist of several stratified differentiated layers with more than two-thirds of cells differentiated and less than one-third of cells proliferating (Fig. 3D and Fig. S6B). Gene expression patterns of cells from 2D cultures and early organoids show high similarity and display a distinct set of differentially expressed genes compared to mature organoids (Fig. 3E, Table S1-2). Recent studies reported that stem cells from diverse tissue types show similar transcriptional signatures compared to the large divergence observed in the ensuing

differentiated tissues [12]. Comparative analysis of the ectocervical cells (either 2D or EO) and differentiated cell expression profiles to that of frequently upregulated genes in stem cells from diverse tissue types confirmed a high similarity shared with the ectocervical 2D cells and EO in contrast to DO-ecto cells (Fig. 3F, Table S1). This is further supported by the expression profile of genes that are concordantly up- or downregulated in ectocervical 2D-ecto and EO-ecto vs. DO-ecto cells with those of ground state stem cells derived from different tissue types vs. their respective differentiated cells (Fig. S6C, Table S1). Thus 2D-ecto and EO-ecto define characteristics of ectocervical stem cells.

A survey of genes that are upregulated in ectocervical stem cells compared to differentiated cells revealed high expression of the Notch ligands Delta-Like Ligand 3 (DLL3) and Manic Fringe (MFNG), the latter facilitating binding of DLL to the Notch receptor (Fig. 3G). In contrast, differentiated cells expressed higher levels of Notch 2 and Notch 3 receptors as well as their targets, including the transcription factor HES1 and Presenilin 1 (PSEN1), a core component of γ-secretase (Fig. 3G). Ectocervical stem cells also showed highly upregulated expression of the WNT antagonists DKK1, DKK3 and the DKK receptor KREMEN2 (Fig. 3G). Concordantly, inhibition of Notch activation using the γ-secretase inhibitor DBZ reduced organoid growth (Fig. 3H, Fig. S6D), as these organoids failed to differentiate and stratify (Fig. 3I). Thus the ectocervical stem cells act as Notch signal-sending cells, while the differentiated cells show the signature of Notch signal-receiving cells, leading to the trans-activating interaction that facilitates differentiation and ultimately epithelial stratification.

Further, gene set enrichment analysis (GSEA) revealed that genes regulated by several transcription factors downstream of Notch ligand and EGF receptor (EGFR)-RAS-MAPK signaling

were highly enriched among genes upregulated in ectocervical stem cells, including AP1 [13,14], CREB, ETS, new ETS-related factor (NERF), ELK1, E2F, SRF, MYC and YY1 [15-18] (Fig. 3J). The two pathways function together to regulate proliferation and differentiation, with the EGFR pathway promoting the expression of Notch DLL ligands [19]. On the other hand, genes belonging to the RAS antagonistic NF1 pathway [20] were enriched in genes highly expressed in differentiated cells. Together, these observations indicate that the Wnt antagonists together with EGFR and Notch-inducing pathways regulate ectocervical stemness and differentiation.

The emergence of squamous metaplasia from quiescent KRT5+ stem cells in the endocervix

We next performed *in vitro* and *in vivo* analysis to determine the cellular origin and mechanism of squamous metaplasia. Primary endocervix-derived cells showed a clear enrichment of KRT5+ and p63+ cells if cultured in 2D in Wnt deficient medium. After transfer to organoid culture conditions, these cells produced only organoids of the squamous type, even in the presence of Wnt3a /RSPO1 (Fig. 4A). However, if primary endocervix-derived cells were grown in 2D in a Wnt proficient medium such cultures contained only a few KRT5+ or p63+ cells and gave rise to columnar organoids in the presence of Wnt. Yet, the absence of Wnt favored the growth of squamous organoids, including the characteristic basal and parabasal p63+ cells (Fig. 4A). Importantly though, endocervical organoids derived from single cells remained columnar even when transferred to Wnt deficient medium, thus excluding the possibility that columnar cells transdifferentiate to the squamous lineage (Fig. 4B). In contrast, primary ectocervical cells grown in 2D with either Wnt proficient or deficient medium give rise only to stratified organoids in Wnt deficient medium (Fig. 4C).

Although the expression of HOX genes, a family of decisive regulators during embryonic development, is largely unknown for the cervix, HOXA11 has previously been associated with cervix development [21] and deregulation of HOXB2, HOXB4, and HOXB13 have been implicated in cervical carcinogenesis [22]. Here we analyzed the pattern of HOX gene expression in the ecto- vs. endocervical organoids (Fig. S7A). Strikingly, we observed substantial differences between the two cultures, supporting the notion that the two tissue types represent different biological lineages in the cervix.

To further consolidate the lineage properties of stratified and columnar epithelial cells and spatial changes in the microenvironment, we performed lineage tracing, single cell RNA-ISH and IHC in a mouse model of squamous metaplasia induced by retinoid depletion [23]. These retinoid-depleted mice showed an upregulation of DKK2 gene expression in the stroma of the endocervix and uterine horns (here also referred to as endocervix), and the emergence of subcolumnar quiescent KRT5+ cells that eventually developed into metaplastic squamous stratified epithelium (Fig. 4D-F, S7B, compare to Fig. 2F and S4H). However, production of the Wnt target Axin2, which is normally expressed in the endocervix, remained unaltered in these mice, while KRT8 and KRT7 expression were restricted to the columnar epithelium (Fig. 4E, Fig. S7C-E). Further, by performing lineage tracing analysis in retinoid-depleted KRT5CreErt2/Rosa26-tdTomato and KRT8CreErt2/Rosa26-tdTomato mice, we confirmed that KRT8+ cells give rise to columnar epithelium while KRT5+ cells give rise to squamous metaplasia in the endocervix (Fig. 4G-H). Together, these data demonstrate that the endocervix harbors two distinct, unipotent stem cell populations with the potential to develop columnar or stratified lineages, respectively. Which one is activated thus appears to depend on the microenvironment and the opposing Wnt-related

245 signals in particular. While Wnt agonists support the formation of columnar epithelium, the local
246 upregulation of Wnt antagonist in the stroma drives the proliferation of quiescent KRT5+ reserve
247 cells to cause squamous metaplasia.

248

249 **Cellular origins of cervical squamous and adenocarcinomas**

250 A number of studies have shown that adult stem cells are susceptible to transformation and often
251 constitute the cells of origin for a variety of cancers [24]. The origin of ADC and SCC is controversial
252 and uncertain. Here we assessed the expression signatures of squamous and columnar cervical
253 organoids to determine the cells of origin of cervical cancers. We retrieved publically available
254 mRNA expression data for 302 cervical cancers from The Cancer Genome Atlas (TCGA,
255 http://cancergenome.nih.gov/). We used the similarity of differential gene expression profiles
256 between ectocervical squamous and endocervical columnar organoids to those between cancer
257 samples to classify the latter into squamous-like, columnar-like or undetermined cases (Fig.5A,
258 Table S1 and S3, Methods). We found that cancers classified as squamous-like matched a
259 histological diagnosis of SCC in all cases (n=111) while for those classified as columnar-like we
260 found 48/77 matching a histological diagnosis of ADC (Fig. 5B). For 111 cases histologically
261 diagnosed as SCC and 3 ADC, we could not determine a clear classification and 29 SCC were
262 assigned to the columnar-like group by our classifier. Importantly, cancer samples classified as
263 columnar-like were mainly $KRT5^{low}$, $KRT7^{high}$ and $p63^{low}$, while samples in the squamous-like and
264 undetermined group were mainly $KRT5^{high}$ and $p63^{high}$ with mixed KRT7 status (Fig. 5A),
265 suggesting that the undetermined group could consist of SCCs within or outgrown into columnar

266 endocervix, leading to the presence of contaminating endocervical columnar KRT7+ cells in the
267 samples.

268 To validate these results, we also obtained clustering results based on genome-wide methylation,
269 global mRNA and microRNA expression data for the same cancer samples from TCGA. The TCGA
270 cluster containing most ADCs in each of the clustering analyses from those three levels of cellular
271 regulation showed strong overlap with our columnar-like class. Using the majority vote among
272 mRNA, miRNA and DNA methylation clusters (Fig. 5B) we find that 69/77 samples from the
273 columnar-like group are also in the TCGA clusters enriched for ADC, while all other TCGA clusters
274 together contain mainly squamous-like and undetermined samples (228/231) (Fig. 5B and S8).
275 Interestingly, 21/29 cancers defined as columnar-like based on our classifier, but histologically
276 classified as SCC, showed strong similarity to ADC according to TCGA molecular profiles and might
277 therefore be misdiagnosed.

278 A recent study suggested that only a small population of cells located in the TZ (the so-called
279 squamo-columnar junction (SCJ) cells) express KRT7 and that these are the precursors of both
280 SCC and ADC [4]. We further investigated the mRNA expression levels in organoids and 302 cervical
281 cancer samples from TCGA with regard to SCJ markers proposed in that study [4], as well as KRT5.
282 In contrast to KRT5, expression of the proposed SCJ markers is significantly higher in healthy
283 endocervical organoids as compared to healthy ectocervical organoids and the same trend is seen
284 in ADC vs. SCC (Fig. 5D-E). This indicates that the reported SCJ cells are not distinct from the
285 endocervical columnar lineage and are not the cells of origin for SCC.

286 Our study also revealed a set of genes that are differentially expressed between squamous and
287 columnar organoids and show a strong correlation with columnar-like and squamous-like

cancers, including MUC5B, KRT5, CSTA, while the proposed SCJ markers KRT7, AGR2 and GDA specifically labeled ADC but not SCC sections (Figure 5C and S9). Thus, the majority of cervical cancers can be divided into two subgroups based on molecular signatures that correlate with signatures of KRT5+ stem cells for squamous or KRT7+/KRT8+ stem cells for columnar cervical epithelia. Together, these results indicate that the cervix harbors two distinct stem cell lineages, reflecting the cells of origin for SCC and ADC, respectively.

Discussion

The TZs of the mucosal epithelium constitute critical zones of enhanced disposition to infections and carcinogenesis [25-29]. Revealing the principles of cellular regulation and homeostasis of these tissue regions is key to understanding the impact of intrinsic and extrinsic disturbances, as well as for prospective and therapeutic disease prevention.

We show that distinct microenvironmental conditions and molecular signals from the epithelial and stromal tissues drive the dominance of specific epithelial lineages of the TZ. We reveal Wnt signaling as a key determinant in regulating the homeostasis at borders between two epithelial types. Wnt signaling has been shown to be indispensable for the maintenance and homeostasis of adult stem cells in several mammalian tissues [11,30]. However, here we show that in the cervix, Wnt signaling stimulated by the underlying stroma drives the columnar lineage while imposing quiescence of squamous lineage-specific stem cells that exist in the same milieu. With the transition to a Wnt repressive microenvironment, these quiescent squamous lineage stem cells are activated and replace the columnar epithelia at the TZ or as an island of metaplasia within

309 the endocervix. Further, the fact that ADC or SCC arise from two distinct stem cell lineages rather
310 than a common cellular origin has important clinical implications for choice of therapy and
311 suggests that preventive ablation of the SCJ alone may not fully eliminate potential cervical
312 cancer precursor cells [7,8].

313 Our data constitute a major conceptual progress in our understanding of how epithelial junctions
314 are maintained in our body. Accordingly, homeostasis at these sites is not maintained by the
315 transition from one epithelial type to another but rather that the adult tissue is composed of
316 different stem cell populations that are retrieved upon extrinsic signals to generate respective
317 cell lineages, forming the adult tissue. This novel concept of homeostasis of the mucosal TZs fits
318 well with other recent observations on the mucosal stem cell identity and may stimulate future
319 investigations with therapeutic relevance.

320

321 **Author contributions**

322 CC, RKG and TFM conceived the study; CC, RKG designed and CC, RKG and SK performed
323 experiments;HB conceived and performed the in silico analyses; VB, UK, and HM contributed
324 imaging, mouse breeding, microarray studies, respectively; MM and HH provided human
325 samples; CC and RKG analyzed the data, and wrote the manuscript; HB and TFM revised the
326 manuscript; TFM supervised the study.

327 **Competing interests statement**

328 The authors declare to have no competing interests

329 **Acknowledgements**

The authors would like to thank Marina Drabkina and Christiane Dimmler for excellent technical assistance, Kirstin Hofmann for excellent help with mouse experiments, Ina Wagner for the microarrays, Gesa Rausch and Michael Meyer for assistance with the animal application, Diane Schad for assistance with preparing the graphics and Rike Zietlow for editing the manuscript.

This work was supported by the BMBF through the Infect-ERA project CINOCA (FK031A409A to TFM). The funders had no influence on the study design or analysis of the data.

Data availability

Microarray data have been deposited in the Gene Expression Omnibus (GEO; www.ncbi.nlm.nih.gov/geo/) of the National Center for Biotechnology Information and can be accessed with the GEO accession number GSE87076.

References

1 Fu, Y. S. *Pathology of the Uterine Cervix, Vagina, and Vulva*. (Saunders, 2002).

2 Singer, A. The uterine cervix from adolescence to the menopause. *Br. J. Obstet. Gynaecol.* **82**, 81-99 (1975).

3 Ferlay, J. *et al.* GLOBOCAN 2012 v1. 0. *Cancer incidence and mortality worldwide: IARC CancerBase, http://globocan.iarc.fr, (28/09/2016).* (2013).

4 Herfs, M. *et al.* A novel blueprint for 'top down' differentiation defines the cervical squamocolumnar junction during development, reproductive life, and neoplasia. *J. Pathol.* **229**, 460-468, doi:10.1002/path.4110 (2013).

5 Jiang, M. *et al.* Transitional basal cells at the squamous-columnar junction generate Barrett's oesophagus. *Nature* **550**, 529-533, doi:10.1038/nature24269 (2017).

6 Smedts, F. *et al.* Keratin expression in cervical cancer. *Am. J. Pathol.* **141**, 497-511 (1992).

7 Franceschi, S. Past and future of prophylactic ablation of the cervical squamocolumnar junction. *Ecancermedicalscience* **9**, 527, doi:10.3332/ecancer.2015.527 (2015).

8 Herfs, M. *et al.* Unique recurrence patterns of cervical intraepithelial neoplasia after excision of the squamocolumnar junction. *Int. J. Cancer* **136**, 1043-1052, doi:10.1002/ijc.28978 (2015).

9 Abdul-Karim, F. W., Fu, Y. S., Reagan, J. W. & Wentz, W. B. Morphometric study of intraepithelial neoplasia of the uterine cervix. *Obstet. Gynecol.* **60**, 210-214 (1982).

10 Burghardt, E. in *Viral etiology of cervical cancer* (eds R Peto & H Zur Hausen) (Cold Spring Harbor Laboratory, 1986).

11 Clevers, H., Loh, K. M. & Nusse, R. Stem cell signaling. An integral program for tissue renewal and regeneration: Wnt signaling and stem cell control. *Science* **346**, 1248012, doi:10.1126/science.1248012 (2014).

12 Wang, X. *et al.* Cloning and variation of ground state intestinal stem cells. *Nature* **522**, 173-178, doi:10.1038/nature14484 (2015).

13 Ranganathan, P., Weaver, K. L. & Capobianco, A. J. Notch signalling in solid tumours: a little bit of everything but not all the time. *Nat. Rev. Cancer* **11**, 338-351, doi:10.1038/nrc3035 (2011).

14 LaVoie, M. J. & Selkoe, D. J. The Notch ligands, Jagged and Delta, are sequentially processed by alpha-secretase and presenilin/gamma-secretase and release signaling fragments. *J. Biol. Chem.* **278**, 34427-34437, doi:10.1074/jbc.M302659200 (2003).

15 Cargnello, M. & Roux, P. P. Activation and function of the MAPKs and their substrates, the MAPK-activated protein kinases. *Microbiol. Mol. Biol. Rev.* **75**, 50-83, doi:10.1128/mmbr.00031-10 (2011).

16 Yang, S. H., Sharrocks, A. D. & Whitmarsh, A. J. MAP kinase signalling cascades and transcriptional regulation. *Gene* **513**, 1-13, doi:10.1016/j.gene.2012.10.033 (2013).

17 Avraham, R. & Yarden, Y. Feedback regulation of EGFR signalling: decision making by early and delayed loops. *Nat. Rev. Mol. Cell Biol.* **12**, 104-117, doi:10.1038/nrm3048 (2011).

18 Li, M. *et al.* Guanylate binding protein 1 is a novel effector of EGFR-driven invasion in glioblastoma. *J. Exp. Med.* **208**, 2657-2673, doi:10.1084/jem.20111102 (2011).

19 Tsuda, L., Nagaraj, R., Zipursky, S. L. & Banerjee, U. An EGFR/Ebi/Sno pathway promotes delta expression by inactivating Su(H)/SMRTER repression during inductive notch signaling. *Cell* **110**, 625-637 (2002).

20 Yap, Y. S. *et al.* The NF1 gene revisited - from bench to bedside. *Oncotarget* **5**, 5873-5892, doi:10.18632/oncotarget.2194 (2014).

21 Taylor, H. S., Vanden Heuvel, G. B. & Igarashi, P. A conserved Hox axis in the mouse and human female reproductive system: late establishment and persistent adult expression of the Hoxa cluster genes. *Biol. Reprod.* **57**, 1338-1345 (1997).

22 Lopez, R. *et al.* HOXB homeobox gene expression in cervical carcinoma. *Int. J. Gynecol. Cancer* **16**, 329-335, doi:10.1111/j.1525-1438.2006.00350.x (2006).

23 Darwiche, N., Celli, G., Sly, L., Lancillotti, F. & De Luca, L. M. Retinoid status controls the appearance of reserve cells and keratin expression in mouse cervical epithelium. *Cancer Res.* **53**, 2287-2299 (1993).

24 White, A. C. & Lowry, W. E. Refining the role for adult stem cells as cancer cells of origin. *Trends Cell Biol.* **25**, 11-20, doi:10.1016/j.tcb.2014.08.008 (2015).

25 Golusiński, W., Leemans, C. R. & Dietz, A. *HPV Infection in Head and Neck Cancer*. (Springer, 2016).

26 Yang, E. J. *et al.* Microanatomy of the cervical and anorectal squamocolumnar junctions: a proposed model for anatomical differences in HPV-related cancer risk. *Mod. Pathol.* **28**, 994-1000, doi:10.1038/modpathol.2015.54 (2015).

27 Wang, X. *et al.* Residual embryonic cells as precursors of a Barrett's-like metaplasia. *Cell* **145**, 1023-1035, doi:10.1016/j.cell.2011.05.026 (2011).

28 Rajendra, S. *et al.* Transcriptionally active human papillomavirus is strongly associated with Barrett's dysplasia and esophageal adenocarcinoma. *Am. J. Gastroenterol.* **108**, 1082-1093, doi:10.1038/ajg.2013.94 (2013).

29 Morbini, P. *et al.* Markers of squamocolumnar junction cells in normal tonsils and oropharyngeal cancer with and without HPV infection. *Histol. Histopathol.* **30**, 833-839, doi:10.14670/hh-11-590 (2015).

30 Yamaguchi, T. P. Heads or tails: Wnts and anterior-posterior patterning. *Curr. Biol.* **11**, R713-R724, doi:10.1016/S0960-9822(01)00417-1 (2001).

31 Rock, J. R. *et al.* Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 12771-12775, doi:10.1073/pnas.0906850106 (2009).

32 Van Keymeulen, A. *et al.* Distinct stem cells contribute to mammary gland development and maintenance. *Nature* **479**, 189-193, doi:10.1038/nature10573 (2011).

33 Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133-140, doi:10.1038/nn.2467 (2010).

34 Willert, K. *et al.* Wnt proteins are lipid-modified and can act as stem cell growth factors. *Nature* **423**, 448-452, doi:10.1038/nature01611 (2003).

35 Farin, H. F., Van Es, J. H. & Clevers, H. Redundant sources of Wnt regulate intestinal stem cells and promote formation of Paneth cells. *Gastroenterology* **143**, 1518-1529.e1517, doi:10.1053/j.gastro.2012.08.031 (2012).

36 R-Core-Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* (2015).

37 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47, doi:10.1093/nar/gkv007 (2015).

38 Mootha, V. K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267-273, doi:10.1038/ng1180 (2003).

39 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

40 The Cancer Genome Atlas Research Network. Integrated genomic and molecular characterization of cervical cancer. *Nature* **543**, 378-384, doi:10.1038/nature21386 (2017).

41 Broad Institute TCGA Genome Data Analysis Center. Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (Primary solid tumor). *Aggregate Analysis Features. Broad Institute of MIT and Harvard. doi:10.7908/C14B30G5* (2015).

Figure legends

Figure 1. Cervix consists of two distinct KRT5+ stratified and KRT7+/8+ columnar epithelial lineages.

Transition zone (TZ) including stratified and columnar epithelium from human (A,C) and mouse (B, D) cervix tissue sections immunolabeled with antibodies against KRT5, KRT7, and KRT8; nuclei are shown in blue. (E-G) Single molecule RNA ISH brightfield images of mouse cervix TZ for KRT8, KRT5 and KRT7; nuclei are shown in blue. (H-I) Tiled images of tissue sections from the genital system of 16 week-old KRT5CreErt2/Rosa26-tdTomato and KRT8CreErt2/Rosa26-tdTomato mice after tamoxifen induction at the age of 4 weeks. (J) Schematic of the stratified and columnar lineages and the TZ of the cervix. Tiled images were acquired with AxioScan imager and are representative of *n* = 3 biological replicates. Arrows indicate squamous epithelium (Sq) and columnar epithelium (Co).

Figure 2. Wnt signaling pathway agonists and antagonists play a key role in ecto and enodcervical development.

(**A**) Bright-field images of human ectocervical organoids. Efficient organoid formation depends on absence of Wnt3a and RSPO1 and presence of FSK. (**B**) Cells isolated from endocervical tissue grown in Matrigel with different factors. Wnt signaling is essential for columnar organoid formation, while absence of Wnt drives formation of squamous stratified organoids. (**C**) Columnar and stratified organoids derived from endocervix, containing $p63^-$ (columnar) and $p63^+$ (stratified) cells. Both express the epithelial marker E-cadherin (CDH1). *n* = 5 biological replicates.

(**D**) Expression analysis of differentially regulated genes in human ecto- vs endocervical organoids. Wnt-related genes are expressed at higher levels in endocervical, Wnt inhibitors at higher levels in ectocervical organoids. Columns = biological replicates. (**E-G**) Single molecule RNA ISH of mouse TZ for (**E**) AXIN2, (**F**) DKK2, (**G**) DKK3, nuclei are shown in blue. Tiled images were acquired with AxioScan imager and are representative of *n* = 3 biological replicates. (H) A schematic representation of the distinct epithelial lineages and the underlying tissue microenvironment at the TZ. Arrows indicate squamous (Sq) and columnar (Co) epithelium.

Figure 3. Stemness and differentiation of ectocervix depend on Wnt antagonist, Notch and EGFR signaling.

(**A**) Expression analysis of differentially regulated genes in human ecto- vs endocervical organoids. Notch-related genes are expressed at higher levels in ectocervix. Columns = biological replicates. (**B**) Confocal images of 2D human ectocervical stem cell cultures immunolabeled for progenitor cell marker p63 and epithelial cadherin (CDH1). (**C, D**) 3D reconstruction of whole-mount confocal images of 3 day-old early ectocervical organoids labeled for p63 and Ki67 (**C**) and a two-week-old differentiated ectocervical organoid labeled for Ki67 and actin (phalloidin) (**D**). (**E, F**) Heatmaps of differentially regulated genes in 2D as well as early and corresponding differentiated ectocervical organoid cultures (**E**) and genes frequently upregulated in stem cells (**F**) (details see Methods section). (**G**) Heatmap of selected differentially expressed genes showing increased Wnt inhibitors and Notch inducers in 2D cultures and early organoids from ectocervix, in contrast to Notch activation-associated genes in differentiated organoids; columns = biological replicates.(**H**) Quantification of the area of human ectocervical organoids grown in the presence

or absence of γ-secretase inhibitor (DBZ), *n* = number of organoids, representative of 3 biological replicates; data represent mean ± s.e.m. (**I**) Confocal images of human ectocervical organoids immunolabeled for CDH1, Ki67 or p63. Inhibition of Notch activation by DBZ prevents differentiation and reduces proliferation. *n* = 3 biological replicates. (**J**) Heatmap showing GSEA enrichment –log10(p-value) of GSEA revealing enrichment of genes upregulated in stem cells regulated by transcription factors downstream of Notch, EGFR-RAS-MAPK target genes signaling among genes upregulated in 2D and EO ectocervical organoids, while the RAS antagonistic NF1 pathway is enriched among genes highly expressed in differentiated ectocervical organoids.

Figure 4. Two distinct stem cells from the endocervix give rise to columnar or squamous stratified lineages depending on the microenvironment.

(**A**) Wnt deficient medium enriches for $p63^+/KRT5^+$ endocervical stem cells that only give rise to stratified organoids, while Wnt proficient medium supports both $KRT7^+$ and $p63^+/KRT5^+$ cells, which can give rise to columnar or stratified organoids, depending on culture conditions (**B**) Endocervical stem cells that give rise to columnar epithelium are unipotent and fail to transdifferentiate into stratified organoids. Single endocervical organoids were grown in Wnt proficient medium, dissociated into single cells and transferred to Wnt proficient or deficient medium. (**C**) Confocal images of ectocervical epithelial cells grown in 2D. $p63^+$ cells are present in Wnt proficient and Wnt deficient media but organoids are formed only in Wnt deficient medium. (**D**) Treatment scheme of Vitamin A deficient diet study in WT and lineage tracing mice. (**E**) Tissue sections from the genital system of C57BL6 mice fed a vitamin A deficient diet for 15 weeks were labeled with antibodies against KRT7 and KRT5. Zoom: outgrowth of subcolumnar

KRT5+ stem cells that give rise to squamous metaplastic epithelium in the endocervix. (**F**) Single molecule RNA ISH of tissue from a mouse fed with a vitamin A deficient diet. Expression of DKK2 is enhanced in endocervical stroma. Boxed areas in panel **F** are magnified on right. (**G, H**) Lineage tracing in KRT8CreErt2/Rosa26-tdTomato (**G**) and KRT5CreErt2/Rosa26-tdTomato (**H**) mice fed a vitamin A deficient diet reveals that squamous metaplasia arising in the endocervix is negative for KRT8-tdTomato (**G**) and positive for KRT5-TdTomato (**H**) lineage markers. Fluorescent and brightfield tiled images were acquired with AxioScan imager. Data representative of *n* = 3 biological replicates.

Figure 5. Cervical squamous carcinomas originate from KRT5^{+} and adenocarcinomas from KRT7^{+} stem cells.

(**A**) Expression profiles of SCC and ADC correlate well with genes differentially expressed between ecto- and endocervical organoids. (B) Classification of cancer samples based on majority voting from hierarchical mRNA and miRNA or methylation status clustering suggests that 29 samples are histologically incorrectly diagnosed as squamous carcinoma. (**C**) Heatmap showing the mean-substracted expression for selected bimodal genes in cancer samples that are differentially expressed in squamous and columnar organoids. Colour denotes fold-change from mean gene expression across all samples. (**D, E**) Expression profiles of proposed squamocolumnar junction markers together with KRT5 in 302 cervical cancer samples (**D**) and in cervical organoids (**E**). Expression of these markers is higher in endocervical organoids (*n*=6) and ADCs (*n*=51) compared to ectocervical organoids (*n*=10) and SCCs (*n*=251), in contrast to KRT5 expression; * = p<0.05. (**F**)

523 Model depicting the KRT5^{+} and KRT7^{+} stem cell organization and Wnt/Notch microenvironment

524 in TZ and during squamous metaplasia.

525

Materials and Methods

Antibodies and Chemicals

The following antibodies and chemicals were used: mouse-anti-p63 (Abcam, # ab375), rabbit-anti-p63 (Abcam, # ab53039), mouse-anti-E-Cadherin (BD Biosciences, # 610181), rabbit-anti-Ki67 (Abcam, # ab16667), mouse/rat-anti-Ki67-FITC (eBioscience, # 11-5698), mouse-anti-KRT5 (Sigma, # C-7785), rabbit-anti-KRT5 (Abcam, # ab52635), rabbit-anti-cytokeratin 5-Alexa488 (Abcam, # ab193894), mouse-anti-KRT7 (Santa Cruz, # sc-23876), rabbit-anti-cytokeratin 7 (Abcam, # ab181598), rabbit-anti-cytokeratin 7-Alexa555 (Abcam, # ab209601), rabbit-anti-CSTA (Cystatin A) (Sigma # HPA001031), rabbit-anti-AGR2 (Proteintech, # 12275-1-AP), mouse-anti-MUC5B (Abcam, # ab77995), rabbit-anti-GDA (Sigma # HPA019352), Hoechst (Sigma, #B2261), Draq5 (Cell Signaling, #4085), γ-secretase inhibitor XX (DBZ) (Calbiochem # 565789) and p38 inhibitor SB202190 (Sigma, # S7067). Secondary antibodies labeled with the fluorochromes Cy2, Cy3 or Cy5 were obtained from Jackson ImmunoResearch Laboratories.

Mouse experiments

All procedures involving animals were approved by the national legal as well as institutional and local authorities at the Max Planck Institute for Infection Biology. Wild-type C56BL6, KRT5CreErt2 [31] and KRT8CreErt2 [32] mice were obtained from the Jackson Laboratory. These strains were bred to Rosa-tdTomato [33] mice in order to generate mice expressing a fluorophore in Cre-expressing cells. For lineage analysis for the cell of origin of Krt5+ or KRT8+ cells, Cre recombinase was induced in female mice by administering tamoxifen (Sigma) intraperitoneally at 0.25 mg g^{-1} body

546 weight in 50 µl corn oil at week 4 on three consecutive days. Mice were euthanized at 14-20
547 weeks and the genital tracts removed for further analysis.

548

549 **Depletion of retinoid signaling in mice using Vitamin A-deficient diet**

550 At birth experimental mice and their mothers were placed on a vitamin A-deficient test diet
551 (SAFE, U8978P-0074) or control diet with added Vitamin A at physiological levels of 6 IU/g (SAFE,
552 U8978P-0075) following a protocol developed for BALB/c mice [23]. Littermates were weaned at 3
553 weeks of age and maintained on the deficient or control diet for a period of 14-20 weeks before
554 being euthanized for further analysis.

555 **Mouse cervical medium**

556 Cervical cells were cultured in ADF medium (Invitrogen, # 12634) supplemented with 12 mM
557 HEPES (Invitrogen, # 15630-056), 1% GlutaMax (Invitrogen, # 35050-038), 1% B27 (Invitrogen, #
558 17504-044), 1% N2 (Invitrogen, # 17502048), 50 ng/ml mouse epidermal growth factor (EGF)
559 (Invitrogen, # PMG8043), 100 ng/ml mouse noggin (Peprotech, # 250-38-100), 100 ng/ml human
560 fibroblast growth factor (FGF)-10 (Peprotech, # 100-26-25), 1.25 mM N-acetyl-L-cysteine (Sigma,
561 # A9165-5G), 10 mM nicotinamide, (Sigma, # N0636), 2 µM TGF-β R Kinase Inhibitor IV
562 (Calbiochem, # 616454), 10 µM ROCK inhibitor (Y-27632) (Sigma, # Y0503), 1%
563 penicillin/streptomycin (Gibco, # 15140-122) with or without 25% Wnt3A- and 25% R-spondin1-
564 conditioned medium, as described in Willert et al [34] and Farin et al [35].

565

566 **Human ectocervical (Wnt deficient) medium**

Consisted of ADF, 12 mM HEPES and 1% GlutaMax, supplemented with 1% B27, 1% N2, 0.5 µg/ml hydrocortisone (Sigma, # H0888-1G), 10 ng/ml human EGF (Invitrogen, # PHG0311), 100 ng/ml human noggin (Peprotech; # 120-10C), 100 ng/ml human FGF-10 (Peprotech, # 100-26-25), 1.25 mM N-acetyl-L-cysteine, 10 mM nicotinamide, 2 µM TGF-β R kinase Inhibitor IV, 10 µM ROCK inhibitor (Y-27632), 10 µM forskolin (Sigma, F6886) and 1% penicillin/streptomycin.

Human endocervical (Wnt proficient) medium

Consisted of ADF, 12 mM HEPES, 1% GlutaMax, supplemented with 1% B27, 1% N2, 10 ng/ml human EGF, 100 ng/ml human noggin, 100 ng/ml human FGF-10, 1.25 mM N-acetyl-L-cystein, 10 mM nicotinamide, 2 mM TGF-β R kinase Inhibitor IV and 10 µM ROCK inhibitor (Y-27632) with 25% Wnt3A- and 25% R-spondin1 conditioned medium.

Epithelial stem cell isolation from human and mouse cervix

Human ecto- and endocervix samples were provided by the Department of Gynecology, Charité University Hospital, Berlin, Germany. Scientific usage of the samples was approved by the ethics committee of the Charité University Hospital, Berlin (EA1/059/15); informed consent to use their tissue for scientific research was obtained from all subjects. Only anatomically normal tissues were used, within 2-3 h after of removal. Mouse cervix was removed from euthanized 4-8 week old healthy female wild type BALB/c mice (from Charles River) immediately preceding the isolation of the cells. Tissue samples were washed thoroughly in sterile PBS (Gibco, # 14190-094) and minced with surgical scissors. Minced tissue was incubated in 0.5 mg/ml collagenase type II

(Calbiochem, # 234155) for 2.5 h at 37 °C in a shaker incubator. Tissue and dissociated cells were pelleted by centrifugation (5 min at 1000 g, 4 °C), supernatant discarded, cells resuspended in TrypLE express (Gibco, # 12604021) and incubated for 15 min at 37 °C in a shaker incubator. After dissociation, the cell and tissue pellet was resuspended in ADF (Invitrogen) medium and passed through a 40 µM cell strainer (BD Falc, # 352340) to separate the single dissociated cells from tissue pieces. Cells were pelleted by centrifugation (5 min at 1000xg, 40 °C), resuspended in either human ecto- or endocervical or mouse cervical medium and cultured either directly as organoids or in 2D.

Human epithelial stem cell culture and maintenance in 2D

Human epithelial stem cells isolated from the tissue were resuspended in either ecto- or endocervical medium and plated in collagen-coated tissue culture flasks. Cells were incubated at 37 °C, 5% CO2 in a humidified incubator. Once they reached 70-80% confluence, cells were detached using TrypLE Express, and centrifuged at 1000xg for 5 mins at 40 °C. The cells were then used for culturing organoids or for maintenance of 2D stem cells. 2D stem cells were maintained by seeding the 2D cells from P1 into tissue culture flasks containing lethally irradiated J2-3T3 fibroblast feeder cells in ecto- or endocervical medium. Medium was replaced and irradiated fibroblasts added every 4 days until the colonies reached a confluence of 60-70%, at which stage they were detached and reseeded onto freshly irradiated feeders at a 1:5 ratio or cryopreserved for later use.

609 **Organoid culture and maintenance**

610 Cells isolated from tissue or the stem cells grown in 2D culture were mixed with 50 µl of ice-cold
611 Matrigel (BD, # 356231) at a density of 20,000 cells, and the Matrigel droplet was placed in a pre-
612 warmed 24-well plate and allowed to polymerize for 10 min at 37 °C. The Matrigel droplet was
613 then overlaid with 500 µl of pre-warmed human ecto- or endocervical medium. Cultures were
614 kept at 37 °C, 5% CO2 in a humidified incubator for 2-3 weeks and medium replaced every four
615 days. For passaging the organoids, Matrigel was first dissolved by adding 1 ml of ice-cold ADF and
616 pipetting up and down 5 times. Organoids were collected in a 15 ml Falcon tube and a further 4
617 ml of ice-cold ADF medium was added and organoids resuspended well to completely dissolve
618 the Matrigel, followed by centrifugation at 300xg for 5 mins at 4°C. Medium was discarded and
619 the ectocervical and mouse organoids were incubated with 1 ml of TrypLE Express for 30 min at
620 37 °C followed by mechanical fragmentation using a fire-polished glass Pasteur pipette by
621 vigorous pipetting (8–10 times) to generate single cells. The single cells were then seeded at a
622 1:10 ratio back into Matrigel for expanding and culturing. For the endocervical organoids, after
623 centrifugation organoids were subjected to mechanical fragmentation as described above to
624 generate fragments that were seeded back into Matrigel at 1:5 ratio. Matrigel was allowed to
625 polymerize for 10 min at 37°C, overlaid with pre-warmed medium and cultured as described
626 above.

627

628 **Organoid forming ability**

629 Stem cells were counted and a defined number resuspended in 50 µl Matrigel to generate
630 organoids as described above. Between 2-3 weeks after plating images were taken of the whole
631 well and the number and area of organoids formed were determined using ImageJ to calculate
632 organoid forming efficiency.

633

634 **Immunofluorescent histochemistry**

635 Organoids were washed five times with cold PBS to remove Matrigel before fixing with 4%
636 paraformaldehyde for 1 h at room temperature (RT) followed by washing with PBS twice.
637 Organoids were then subjected to dehydration in an ascending ethanol series followed by
638 isopropanol and acetone for 20 min each. The dehydrated organoids were paraffin-embedded
639 and 5 µM sections cut on a Microm HM 315 microtome. Mouse and human tissues were
640 extensively washed with PBS and fixed using 4% PFA overnight at RT. Samples were subjected to
641 dehydration in an ascending ethanol series followed by isopropanol and xylene (60 min each)
642 followed by paraffinization using a Leica TP1020 tissue processor. The tissue was embedded and
643 5 µM sections cut on a microtome. For immunostaining, paraffin sections were deparaffinized
644 and rehydrated, followed by treatment with antigen retrieval solution (Dako, # S1699). Sections
645 were blocked using blocking buffer (1% BSA and 2% FCS in PBS) for 1 h at RT. Primary antibodies
646 were diluted in blocking buffer and incubated for 90 mins at RT followed by five PSB washes
647 before 1 h incubation with secondary antibodies diluted in blocking buffer along with Hoechst or
648 Draq5. Sections were washed with PBS five times and mounted using Mowiol. Images were
649 acquired with a Leica TCS SP8 confocal microscope.

Fresh epithelial isolates were grown on collagen-coated coverslips in 2D and fixed with 4% paraformaldehyde for 30 min at RT. Cells were permeabilized and blocked with 0.5% Triton X-100 and 1% BSA in PBS. Primary antibodies were diluted in 1% BSA in PBS and incubated for 1 h at RT followed by three washes in PSB-T (0.1% Tween 20 in PBS), followed by 1 h incubation with secondary antibodies diluted in 1% BSA in PBS along with Hoechst or Draq5. Coverslips were washed three times with PBS-T and once with PBS and mounted using Mowiol. Images were acquired on a Leica TCS SP8 confocal microscope. Images were processed with Adobe Photoshop; 3D reconstruction was done with the Volocity 6.3 software package (Perkin Elmer).

Whole mount staining

Matrigel was removed from the organoids by extensive washing with ice-cold PBS prior to fixation (4x 45 min) and allowed to settle by gravity to maintain the 3D structure. Organoids were then fixed using pre-warmed (37 °C) 3.7% PFA for 1 h at RT followed by three PBST washes. Permeabilization and blocking was performed overnight at 40 °C using 5% donkey serum, 1% FCS, 0.05% Tween20, 2% Triton X- 100, 0.02% sodium azide in PBS. Organoids were incubated with primary antibodies diluted in blocking buffer (5% donkey serum, 1% FCS, 0.25% Triton X- 100, 0.02% sodium azide in PBS) at 4°C for 3-5 days followed by three PBST washes for 45 min each at RT. Next, organoids were incubated with secondary antibodies diluted in blocking buffer for two days at 4°C followed by one PBST wash for 45 min and three washes with PBS containing 5% glycerol for 45 min each. Organoids were then carefully transferred to an ibidi µ-slide (# 81822) together with some PBS and glycerol solution and Z stack images were acquired with a confocal microscope and image processing and 3D reconstructions were done using Volocity 6.3 software.

672

673 **Single-molecule RNA in situ hybridization (RNA-ISH).**

674 For single molecule RNA in situ labelling, paraffin embedded 10 µM tissue sections were used
675 with RNAscope 2.5 HD Red Reagent kit (Advanced Cell Diagnostics). Hybridizations were
676 performed according to the manufacturer's protocol. In each experiment, positive (PPIB) and
677 negative (DapB) control probes were used as per the manufacturer's guidelines. Tiled bright field
678 images were obtained with Axio Scan.Z1 tissue imager (Zeiss). Images were further processed
679 with Zen 2.3 (Blue edition) image analysis software and further compiled using Adobe illustrator.

680

681 **RNA isolation and quality control**

682 Microarrays were hybridized for human ectocervical cells cultured in 2D in Wnt-deficient medium
683 (n=3 biological replicates from 2 human donors) or as organoids (EO: n=3 biological replicates
684 from 3 human donors, DO: n=4 biological replicates for 4 human donors), human endocervical
685 cells cultured in 2D Wnt-proficient medium (n=3 biological replicates from 3 human donors) or
686 as DO organoids (n=3 biological replicates from 3 human donors), as well as mouse cervical EO
687 and DO organoids cultured in Wnt-proficient or -deficient medium, respectively (n=2 biological
688 replicates per condition). In the absence of any pre-existing knowledge on expected effect sizes
689 sample sizes were selected based on available samples. Cells and organoids were pelleted and
690 resuspended in 1 ml Trizol (Life Technologies) and RNA was isolated according to the
691 manufacturer's protocol. Quantity of RNA was measured using a NanoDrop 1000 UV-Vis

692 spectrophotometer (Kisker) and quality was assessed by Agilent 2100 Bioanalyzer with an RNA
693 Nano 6000 microfluidics kit (Agilent Technologies).

694

695 **Microarray expression profiling and Data analysis**

696 Microarray experiments were performed as single-color hybridizations on custom whole genome
697 human 8x60k Agilent arrays (Design ID 048908) and Agilent Feature Extraction software was used
698 to obtain probe intensities. The extracted single-color raw data files were background corrected,
699 quantile normalized and further analyzed for differential gene expression using R [36] and the
700 associated BioConductor package LIMMA [37] (Table S2). Microarray gene expression comparisons
701 between groups were performed using unpaired tests for all human comparisons. R was also
702 used for all statistical analyses unless stated otherwise. Mann-Whitney-U test was used for
703 comparisons of gene expression in SCJ marker genes with a threshold of $p<0.05$. Microarray data
704 have been deposited in the Gene Expression Omnibus (GEO; www.ncbi.nlm.nih.gov/geo/) of the
705 National Center for Biotechnology Information and can be accessed with the GEO accession
706 number GSE87076.

707 The signature of differentially expressed genes between ectocervical 2D/EO vs DO organoids was
708 selected from all genes with a false discovery rate (FDR) < 0.05 and log2 fold change < -1.5 or >
709 1.5 in any of the two comparisons (2D vs. DO or EO vs. DO) and the largest absolute fold change
710 from both comparisons and possible replicate probes was taken for each gene.

711

712 **Analysis of stem cell related genes**

Raw data from different microarray data sets obtained from adult tissue stem cells (SC) cultured on feeder cells and corresponding differentiated cells from air-liquid interface (ALI), Matrigel or self-assembly sphere (SAS) were downloaded from GEO (GSE57584, GSE66115, GSE69453, GSE65013, GSE32606, GSE69429, GSE49292) and normalized together using method 'RMA-sketch' with Affymetrix Power Tools. We assessed differentially expressed genes between SC and corresponding differentiated cell cultures for normal esophagus, Barrett's esophagus, gastric cardia, duodenum, jejunum, ileum, colon ascendens, colon transversum, colon descendens, KRT5+ and KRT7+ fetal esophageal cells, fallopian tube, nasal turbinated epithelium, tracheobronchial epithelium and distal airway epithelium. We selected stem cell-related genes as those genes with significant (adjusted p-value < 0.05) up- or down-regulation (abs(logFC) > 1) in at least 5 out of 18 comparisons (Table S1).

Gene Set Enrichment Analysis (GSEA)

We performed a pre-ranked GSEA analysis using GSEA software v2.1.0 [38,39] obtained from http://software.broadinstitute.org/gsea. The t-statistics from comparisons of ectocervical organoids (2D vs. Differentiated organoids or Early organoids vs. Differentiated organoids) were used to rank probes and enrichment of MSigDB Motif gene sets [http://software.broadinstitute.org/gsea/msigdb] (c3.all.v5.1.symbols.gmt) was computed using standard settings, collapsing probe sets within genes using the Max_probe method and using 1000 permutations. For further analysis we kept only motif gene sets that were significant in at least one of the up or down regulated genes in the two comparisons mentioned above at FDR <

734 5%. For the heatmap visualization, we chose the smallest p-value for motif gene sets referring to
735 the same transcription factor use the negative log10 of this value for visualization.

736

737 **Cervical cancer data**

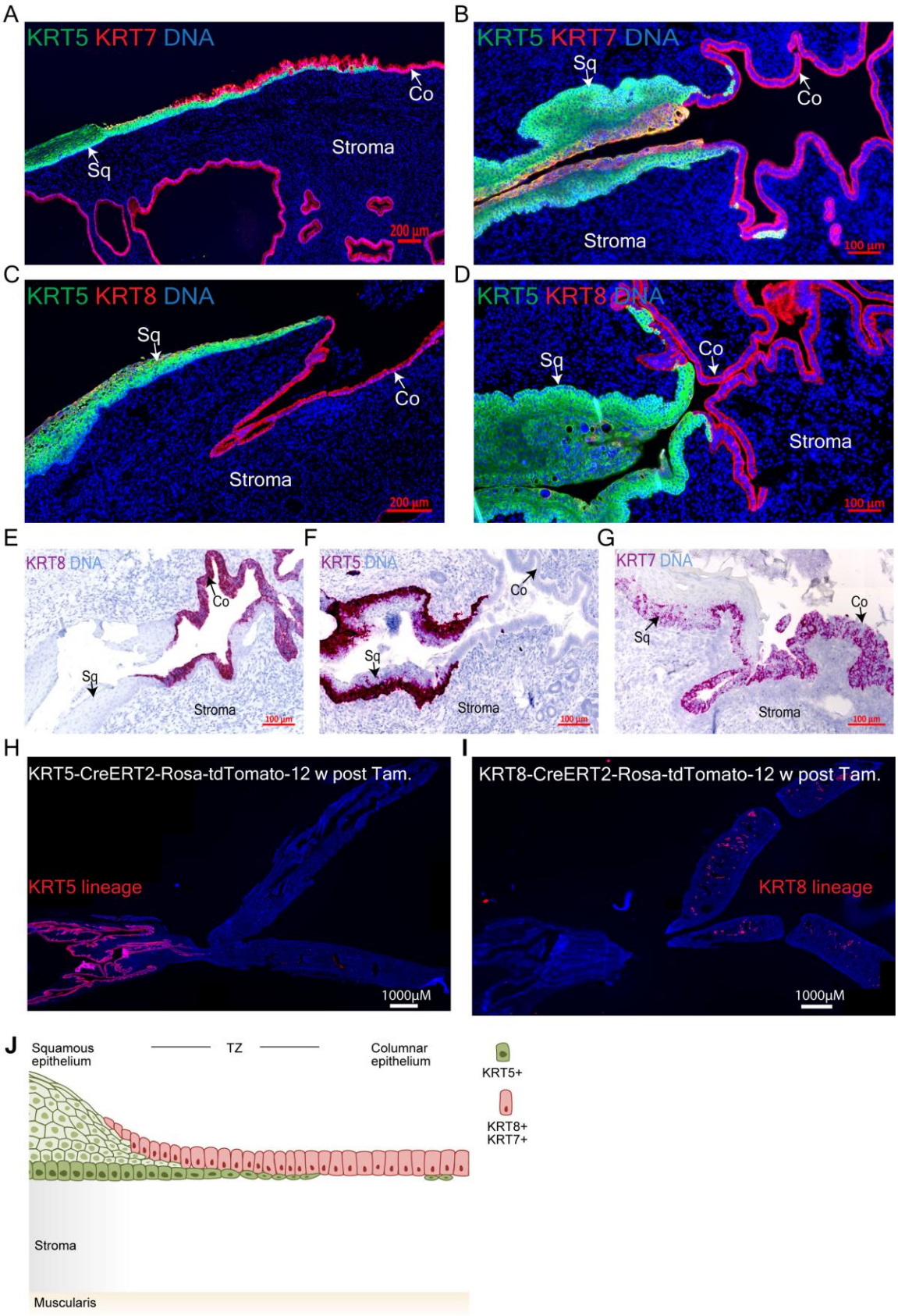
738 Expression data (Level 3 processed RNASeq_v2) was obtained for 302 unique samples with
739 available histological diagnosis from The Cancer Genome Atlas (TCGA) data portal (https://gdc-
740 portal.nci.nih.gov/). This data was generated within the Cervical Squamous Cell Carcinoma and
741 Endocervical Adenocarcinoma project (TCGA-CESC) and is a superset of the published cohort [40].
742 Per gene expression levels were extracted from "*.rsem.genes.normalized_results" files using
743 custom scripts. Public clinical sample annotations for t0hose samples were also obtained from
744 the same source. Aggregated features including clustering results based on DNA methylation,
745 mRNA and microRNA expression was obtained from the Cervical and Endocervical Cancer (CESC)
746 project Firehose site of TCGA [41]. For the details on the majority vote see Fig. S8. To classify
747 samples into squamous-like and columnar-like classes, the gene expression levels were log2
748 transformed and Z-score was applied to make genes comparable. A squamous vs columnar
749 organoid signature was defined based on the fold changes between ectocervical squamous and
750 endocervical columnar differentiated organoids for 2,834 genes with FDR < 0.05 and absolute
751 log2 fold change > 1, selecting the probe with the lowest p-value for each gene. Spearman
752 correlation coefficients (referred to as Co-Sq Score) were computed between Z-scored gene
753 expression values from each cancer sample and the corresponding fold change for the same gene
754 from the squamous vs columnar organoid signature. We defined samples with Co-Sq Score > 0.2
755 as squamous-like, those with < -0.2 as columnar like and all other as ‘undetermined’ (Fig S8).

Applying the same procedure to 1,000 random sets of genes of the same size with the same fold changes produced sample correlation coefficients generally lower than |0.06|. Thresholds for classification of samples into KRT5-high/low and KRT7-high/low as well as TP63 high/low classes were selected manually to separate the highest cluster from all other samples (Fig. S8). For simplicity, we combined all diagnoses with an adenoma component (Endocervical Adenocarcinoma, Endometrioid Adenocarcinoma, Mucinous Adenocarcinoma and Adenosquamous Carcinoma) into Cervical Adenocarcinoma (Tables S1 and S3).

Code Availability

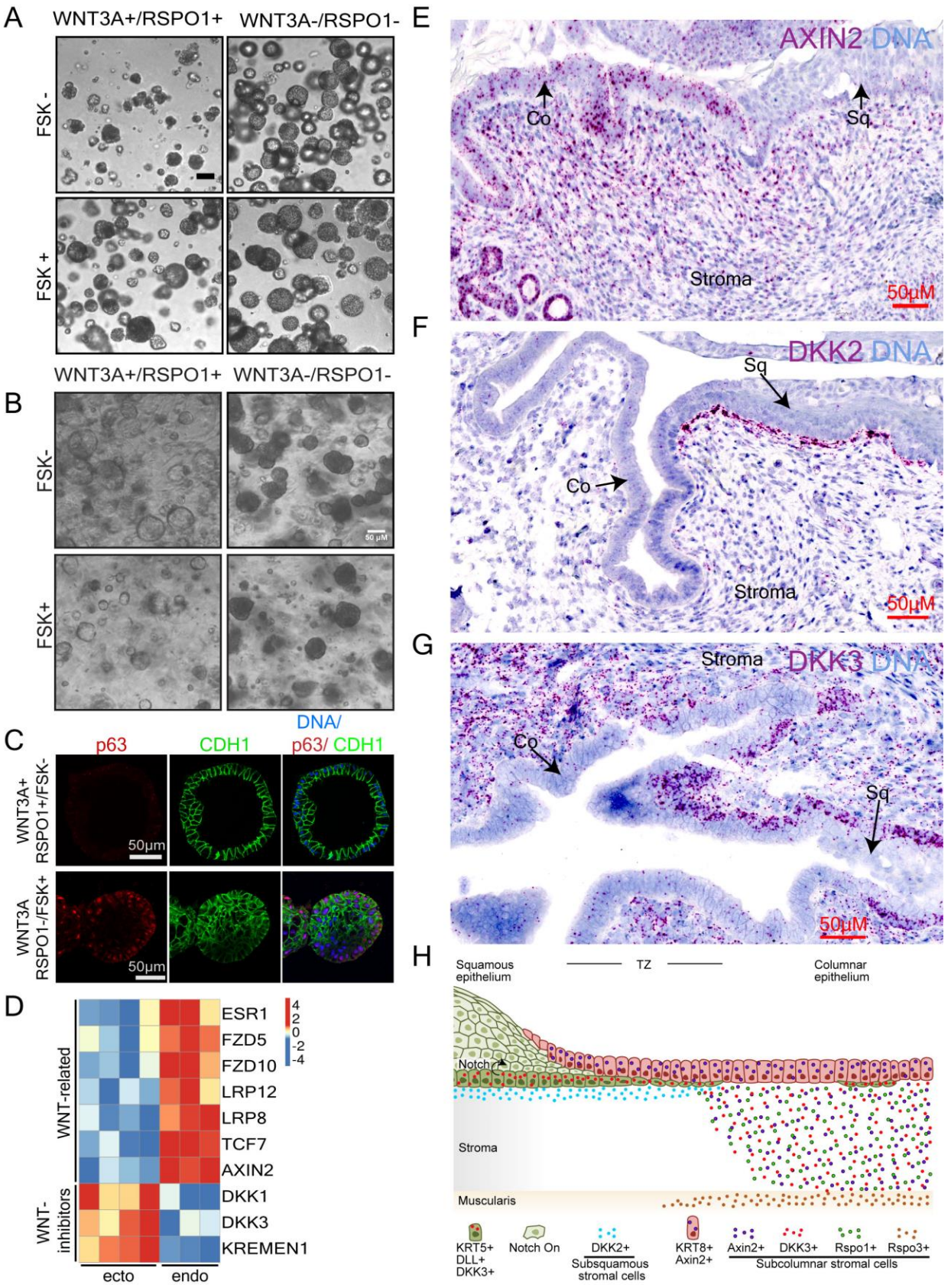
All R code used for generating analyses used in this publication is available from the authors on request.

Figure 1



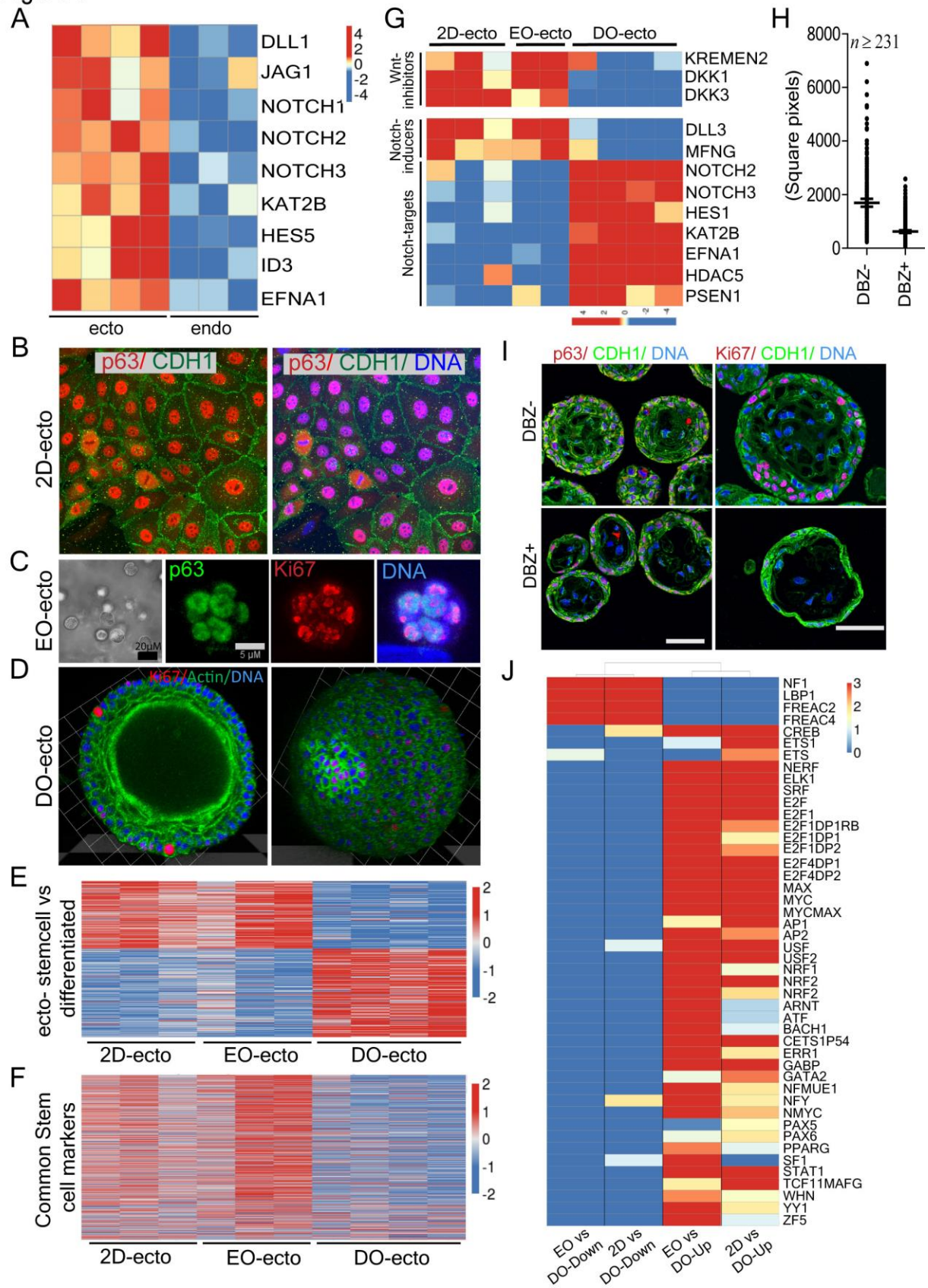
769

Figure 2

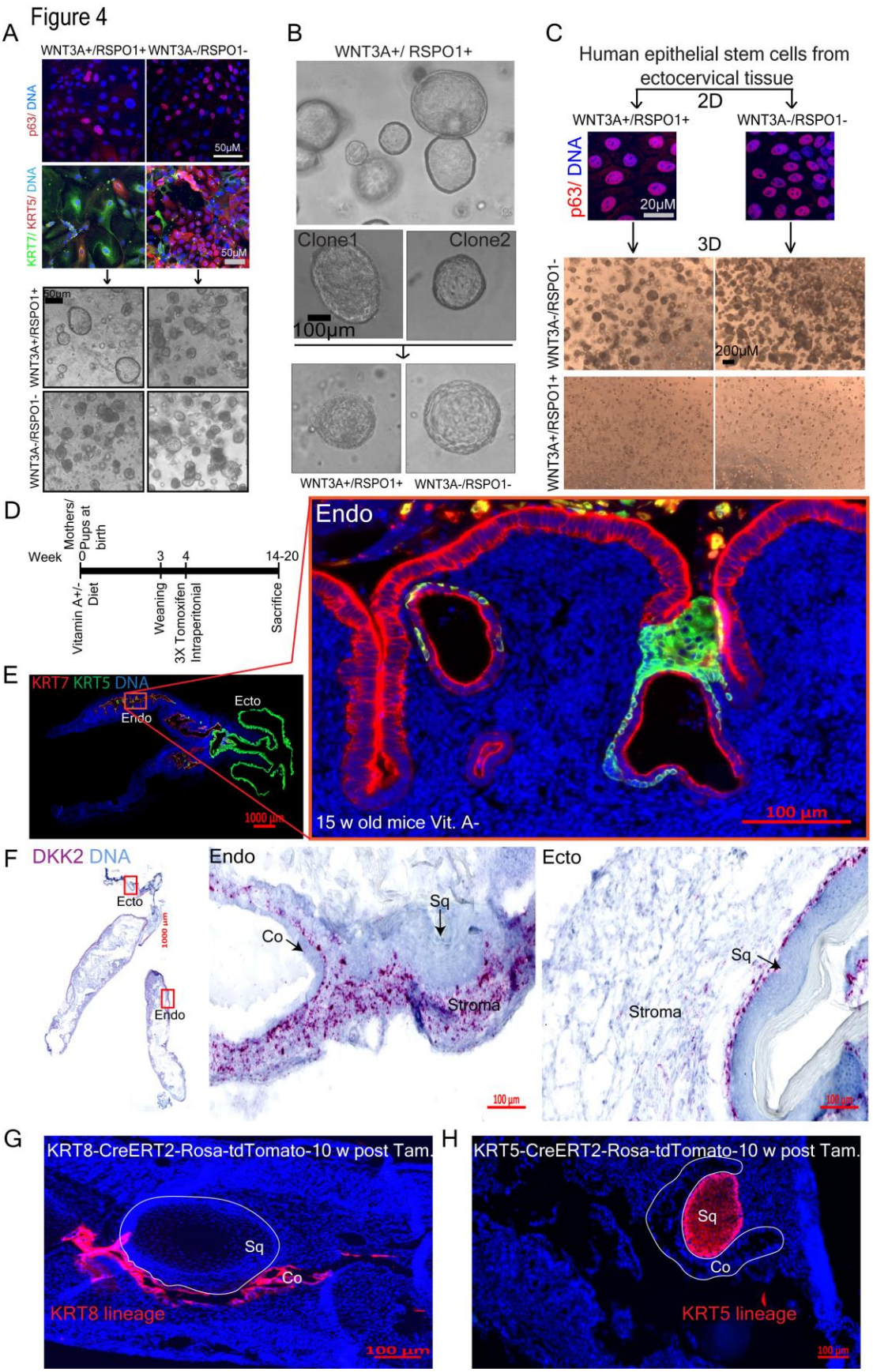


770

Figure 3

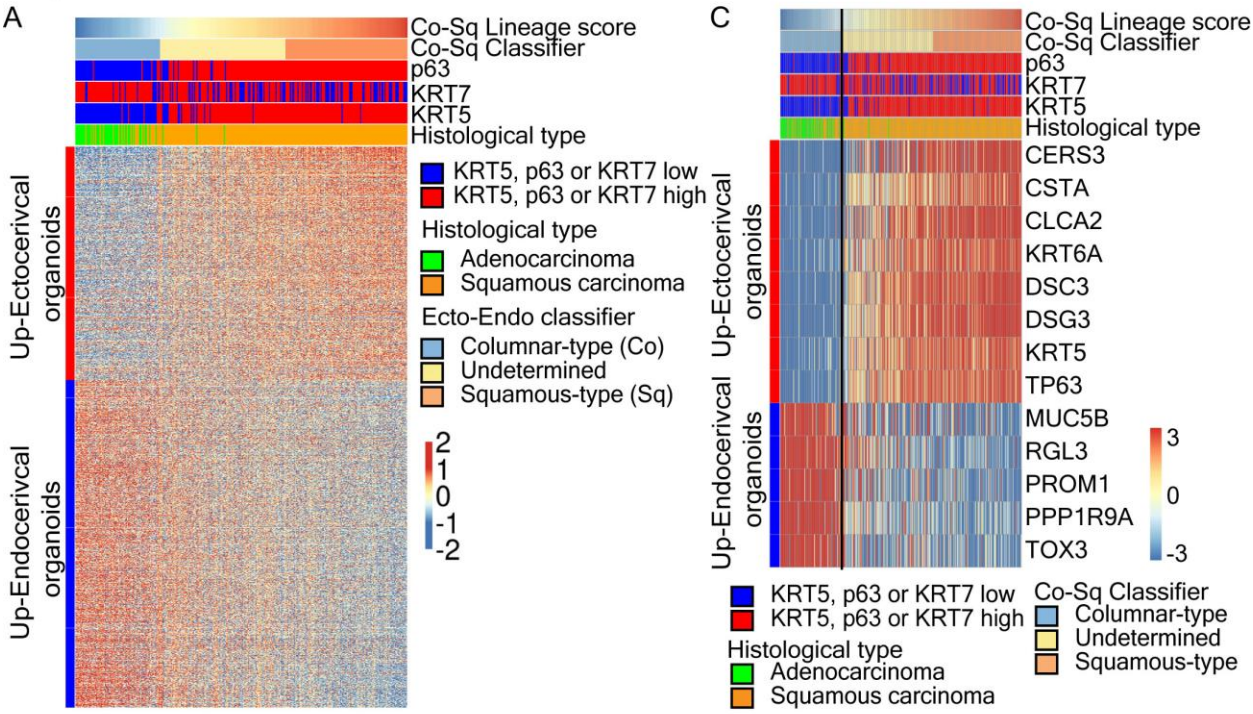


771



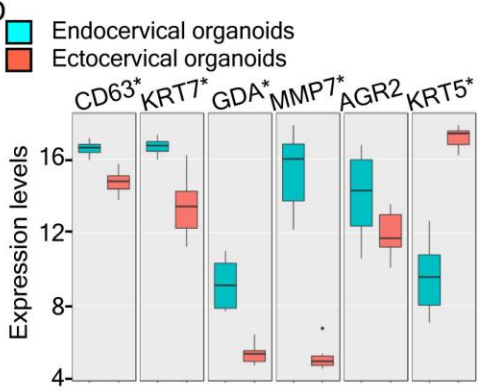
772

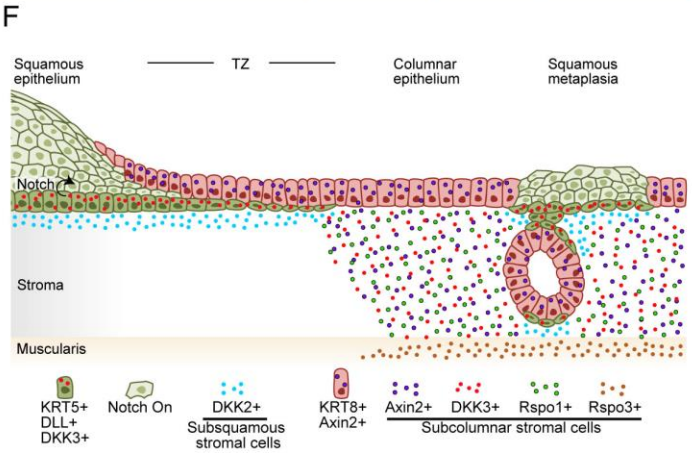
Figure 5

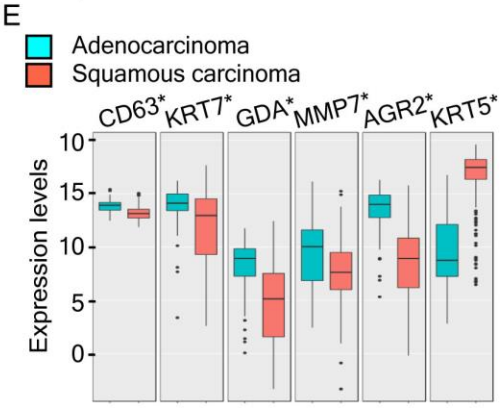


B

	TCGA classification	*Squamous/Columnar cervical lineage classifier*		
Histology	Majority vote from Methylation / miRNA/ mRNA cluster	Columnar	Squamous	undetermined
Cervical Adeno-carcinoma	Adenocarcinoma enriched cluster	**48**	0	2
	Other clusters	0	0	1
Cervical Squamous-Carcinoma	Adenocarcinoma enriched cluster	**21**	0	3
	Other clusters	8	111	108







773