

**Stem Cell Reports**

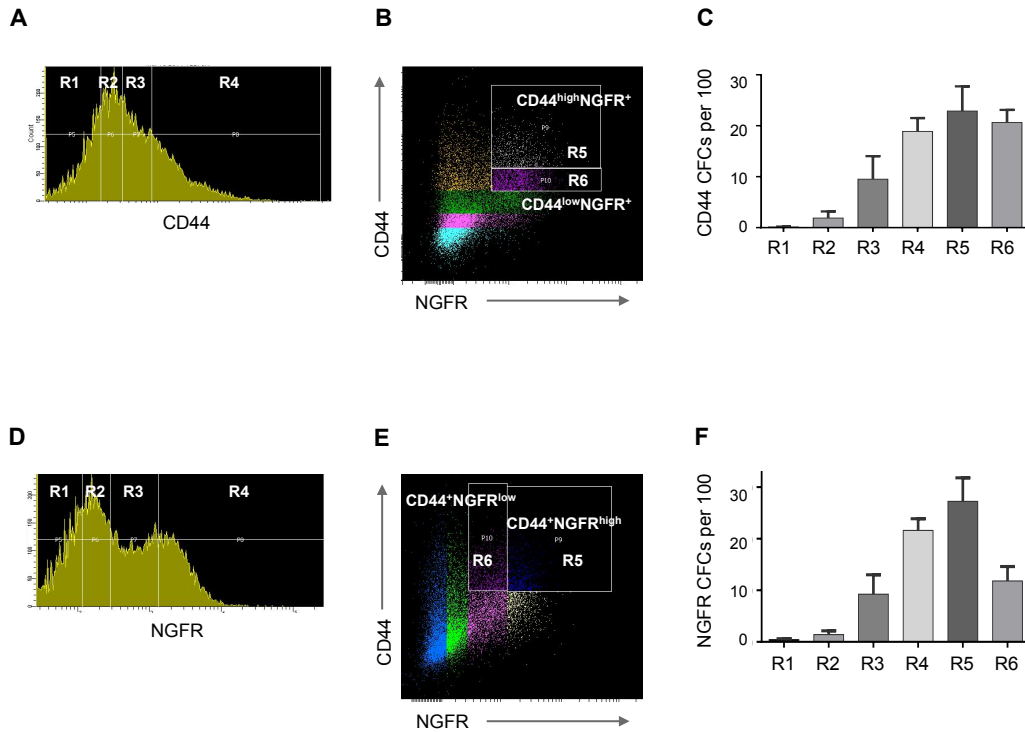
**Supplemental Information**

**Characterization of Epithelial Progenitors in Normal Human  
Palatine Tonsils and Their HPV16 E6/E7-Induced  
Perturbation**

**Sung Yoon Catherine Kang, Nagarajan Kannan, Lewei Zhang, Victor Martinez, Miriam P.  
Rosin, and Connie J. Eaves**

## Supplemental Figures

## Figure S1



### Figure S1. Functional comparison of subsets defined by expression of CD44 or NGFR, Related to Figure 4

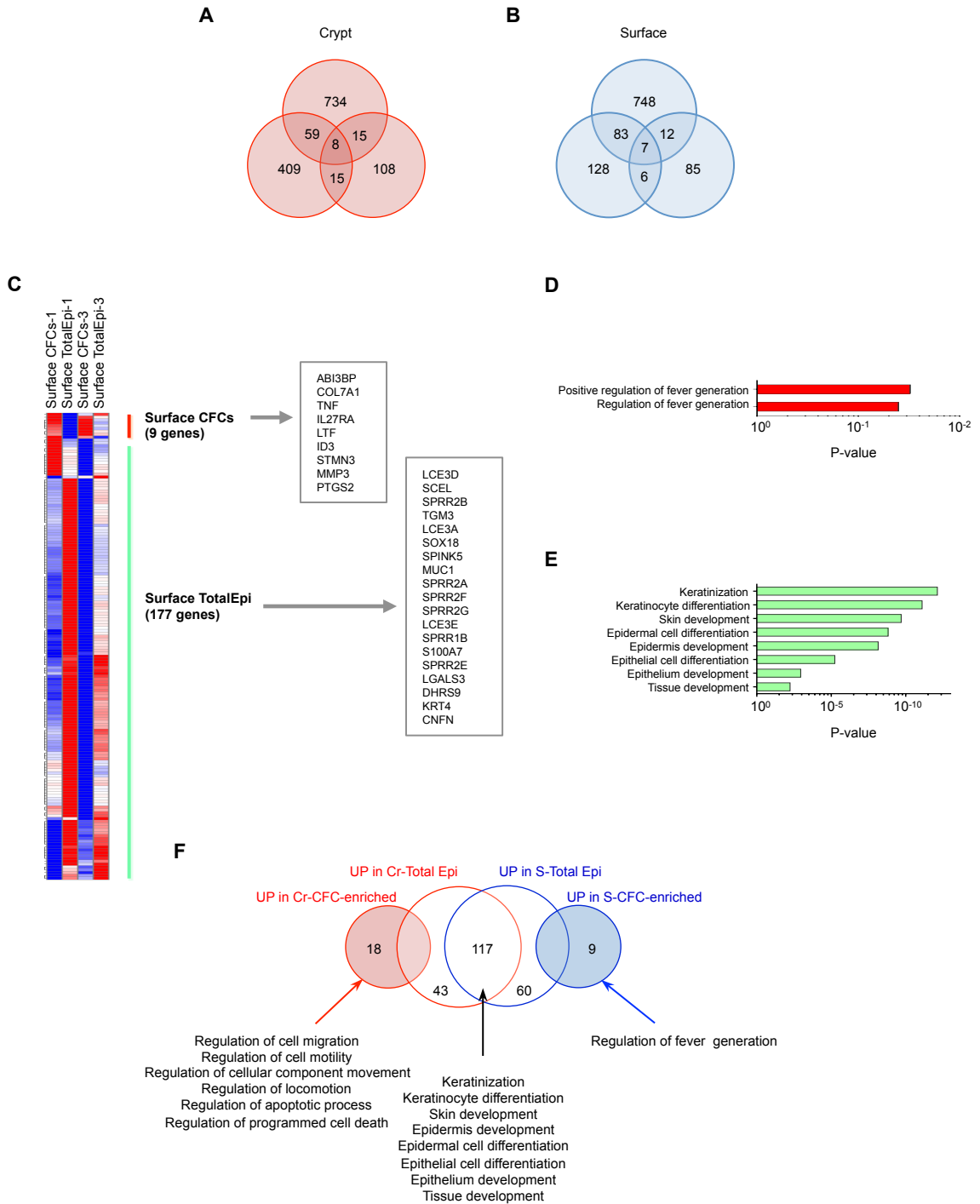
(A-B) FACS plots showing the gating strategies used to separate 6 subsets defined by CD44 expression within the CD45<sup>+</sup>CD31<sup>-</sup> population.

(C) CFC frequencies in each of the 6 subpopulations shown in A and B (n = 4).

(D-E) FACS plots showing the gating strategies used to separate 6 subsets defined by NGFR expression within the CD45<sup>+</sup>CD31<sup>-</sup> population.

(F) CFC frequencies in each of the 6 subpopulations shown in D and E (n = 4).

**Figure S2**



**Figure S2. Transcriptome analysis of different subsets of human tonsillar cells, Related to Figure 6**

(A) Number of genes overexpressed by  $\geq 1.5$ -fold in crypt CD44<sup>+</sup>NGFR<sup>+</sup> cells compared to their surface counterparts, with overlap shown for 3 donors.

(B) Number of genes overexpressed by  $\geq 1.5$ -fold in surface CD44<sup>+</sup>NGFR<sup>+</sup> cells compared to their crypt counterparts, with overlap shown for 3 patients.

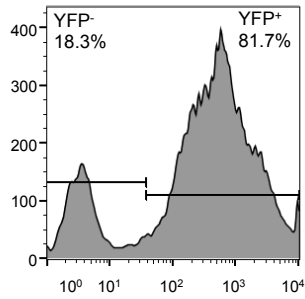
(C) Hierarchical cluster analysis of surface CD44<sup>+</sup>NGFR<sup>+</sup> cells and their respective TotalEpi subsets for genes that displayed 2-fold or higher differential expression ( $n = 2$ ; red, genes commonly overexpressed in the CD44<sup>+</sup>NGFR<sup>+</sup> cells; green, genes commonly overexpressed in TotalEpi cells;  $p \leq 0.05$ ).

(D) The top biological GO categories found to be upregulated in the CD44<sup>+</sup>NGFR<sup>+</sup> surface cells as compared to the TotalEpi cells.

(E) The top biological GO categories upregulated in the surface epithelial-enriched subsets cells as compared to CD44<sup>+</sup>NGFR<sup>+</sup> cells.

(F) A summary of Figures 6E-6G and Figures S3C-S3E. The areas indicated by the red or blue arrows represent the transcripts that are more highly expressed in the CD44<sup>+</sup>NGFR<sup>+</sup> crypt (Cr) or surface (S) cells, respectively, by  $>2$ -fold (compared to the TotalEpi cells). The region indicated by a black arrow shows the number of transcripts that are highly expressed in the TotalEpi subsets (compared to the CD44<sup>+</sup>NGFR<sup>+</sup> cells) that are common to both the TotalEpi crypt and surface cells.

**Figure S3**



**Figure S3. FACS analysis of cells infected with lentiviral E6/E7-YFP vectors and analyzed on day 4 post-infection, Related to Figure 7**

## Supplemental Tables

Marker	CL	Surface epithelium								Crypt epithelium					
		Basal layer		Parabasal layer		Mid layer		Upper layer		Basal layer		Mid layer		Upper layer	
		% Pos	intensity	% Pos	intensity	% Pos	intensity	% Pos	intensity	% Pos	intensity	% Pos	intensity	% Pos	intensity
CK5	C	4	++/+++	4	++/+++	3	+/++	0	-	4	++/+++	4	++/+++	3	++/+++
CK14	C	4	+++	4	+++	4	+++	3	+++	4	+++	4	+++	4	+++
CK19	C	4	++	1	±	1	±	0	-	3	++	4	+++	2	++
CD44	M	4	+++	4	+++	3	++	0	-	4	+++	4	++	0/1	+/-
NGFR	M	4	+++	0	-	0	-	0	-	3	++	2	+	0	-
P63	N	4	+++	4	+++	3	++	0	-	4	+++	3	+++	2	+++
CK8/18	C	4	±	4	+	3	+	0	-	2	++	4	+++	3	+++
CK13	C	4	±	4	++	4	++	4	+++	4	±	4	++/+++	4	+++
CK4	C	0	-	1	+	4	++	3	+	0	-	4	+++	4	++/+++
Involucrin	C	0	-	3	+++	4	+++	4	+++	4	++	4	+++	4	+++
MUC1	M	1	±	3	+	4	++	4	+++	4	++	4	+++	4	+++
Ki67	N	0	-	4	+++	0	-	0	-	4	+++	3	+++	1	++

**Table S1. Immunohistological examination of human tonsillar epithelium, Related to Figure 1**

Cellular locations, percent positivity and staining intensity of each marker tested by IHC. CK5, cytokeratin-5; CK14, cytokeratin-14; CK19, cytokeratin-19; NGFR, nerve growth factor receptor; CK8/18, cytokeratin-8/18; CK13, cytokeratin-13; CK4, cytokeratin-4; MUC1, Mucin 1; CL, cytological location: M, membranous; C, cytoplasmic; N, nuclear. % Pos, percentage of stained cells: 0, no staining; 1, ≤30% cells stained; 2, 30 to 50% cells stained; 3, 50 to 70%; 4: 70 to 100%. Intensity: -, negative; ±, equivocal; +, weak; ++, moderate; +++, strong.

	CD44 <sup>-</sup> NGFR <sup>-</sup>			CD44 <sup>+</sup> NGFR <sup>-</sup>			CD44 <sup>+</sup> NGFR <sup>+</sup>			CD44 <sup>-</sup> NGFR <sup>+</sup>		
	U (%)	S (%)	Cr (%)	U (%)	S (%)	Cr (%)	U (%)	S (%)	Cr (%)	U (%)	S (%)	Cr (%)
CK19	ND	17, 40	19	52	40	56	72	74	95	ND	32, 94	89, 92
CK5	ND	5, 6	8	ND	10	20, 23	69	70	85	ND	25, 20	28
p63	49	10	90	60, 84	ND	80	94	90	100	35	36, 51	ND
Involucrin	21	ND	21, 17	14	ND	12, 20	96	20	69	ND	8	14, 17
CK4	27	22, 42	ND	ND	14, 19	20	26	28	86	74	41	51

**Table S2. IHC analysis of freshly sorted and cytopun CD45<sup>+</sup>CD31<sup>-</sup> subpopulations defined by CD44 and NGFR, Related to Figure 2**

The proportion of cells positive for each marker tested with the site of origin specified. U, unseparated; S, surface; Cr, crypt; ND, not done.