Characterization of Epithelial Progenitors in Normal Human Palatine Tonsils

by

Sung Yoon Catherine Kang

M.Sc., Simon Fraser University, 2010

Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Biomedical Physiology and Kinesiology
Faculty of Science

SIMON FRASER UNIVERSITY

Summer 2016

Approval

Name:	**Sung Yoon Catherine Kang**
Degree:	**Doctor of Philosophy**
Title:	***Characterization of epithelial progenitors in normal human palatine tonsils***
Examining Committee:	**Chair:** Will Cupples Professor

Miriam P. Rosin
Senior Supervisor
Professor

Lewei Zhang
Supervisor
Professor
Department of Oral Biological and Medical Sciences
University of British Columbia

Connie J. Eaves
Supervisor
Professor
Department of Medical Genetics
University of British Columbia

Sharon Gorski
Internal Examiner
Associate Professor
Department of Molecular Biology and Biochemistry

Rama Khokha
External Examiner
Professor
Department of Medical Biophysics
University of Toronto

Date Defended/Approved: May 17, 2016

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

The palatine tonsils are a collection of lymph nodes overlaid by stratified non-keratinized epithelium that invaginates deep into the tissue, forming tonsillar “crypts” where ingested and inhaled pathogens are collected and initiate an immune response followed by transepithelial lymphocyte infiltration. The dynamic nature of this site suggests the existence of primitive cells responsible for the constant tissue repair and regeneration; however, such cells in the tonsils have not been characterized. Human Papillomavirus (HPV)-associated cancer of oropharynx is a global health concern that is on the rise, with HPV16 oncoproteins E6/E7 frequently detected in the epithelium of tonsillar crypts. It is hypothesized that the long-term self-renewing progenitor cells are the target of HPV-induced malignancy, but the lack of a method to specifically study these critical cells has been a barrier to further investigation. In this study, I have developed and optimized the methodology to identify, isolate and quantitate epithelial progenitors from human palatine tonsil. I show that tonsillar progenitors that form colonies *in vitro* in 2D colony assays and differentiate into multilayered epithelial tissues in a 3D culture system are $CD44^{+}NGFR^{+}$ and present in both surface and crypt regions. Transcriptome analysis indicates a high similarity between $CD44^{+}NGFR^{+}$ cells in both regions, although those isolated from the crypt contained a higher proportion of the most primitive (holo)clonogenic cells. The method was then applied to study effects of HPV infection on purified $CD44^{+}NGFR^{+}$ cells from both regions. Lentiviral transduction of $CD44^{+}NGFR^{+}$ cells with HPV16 E6/E7 oncogenes prolonged their growth in 2D cultures and caused aberrant differentiation in 3D cultures. Interestingly, in the presence of the normal cells, the E6/E7-transduced cells proliferated more slowly in 2D cultures and formed uniquely heterogeneous epithelial structures displaying varying degrees of perturbation in 3D cultures suggesting possible inhibitory effects of the cocultured normal cells. The system developed and presented in this thesis allows for a targeted approach to study a specific subset of epithelial cells purified from the tonsillar crypts and their response to E6/E7 infection, setting the stage for addressing many unanswered questions pertaining to the early stages of tonsillar oncogenesis.

Keywords: tonsil; epithelial progenitors; oropharyngeal cancer; HPV

To my parents

Acknowledgements

This thesis would not have come into existence without the help of a number of people I have been fortunate to work with. First, I would like to thank my supervisor, Dr. Miriam Rosin, for giving me the opportunity to take on this unique project and for her continuous support, thoughtfulness and encouragement throughout this journey. I was always inspired by her unceasing curiosity and her passion to make a difference.

I also would like to thank my thesis committee members, Dr. Connie Eaves and Dr. Lewei Zhang, for their expertise and insightful comments that helped guide this research project, but also for the hard questions that taught me how to think and critique.

I am truly indebted to my colleague, Dr. Nagarajan Kannan, for his resourcefulness, patience and support, scientifically and morally, from the very start of my doctoral training. I would not have made it this far if it had not been for all those hours of discussions and his timely, helpful replies to numerous emails. Thank you, Raj.

I would like to acknowledge the excellent support of Ivan Sun with immunohistochemistry work and tonsil sample acquisition. I would like to thank the rest of the Rosin lab, Dr. Denise Laronde, Dr. Tarinee Lubpairee, Dr. Ajit Auluck, Leigha Rock, Jade Lavallee, Salima Walji, Marco Wu and anyone else that I may have missed, for their support, input and encouragement at every milestone. I thank Helen Chiu, Anita Fang, Huijun Jiang, Emily Morgan, Sarah Gustin, and Arla Yost for the administrative support.

I would like to thank the surgeons, pathologists, and technologists for their assistance with obtaining tonsillectomy samples: Dr. Scott Durham, Dr. Ken Berean, Dr. Robert O'Connor, Dr. Matthew Dickson, Dr. Patricia Lee, Dr. Paul Mick, Dr. Donald Mintz, Shabniz Ramji and Margaret Luk. I thank Dr. Victor Martinez for his help with transcriptome analysis, and the staff of the FACS facility at the BC Cancer Research Centre, Gayle Thornbury, Wenbo Xu and David Ko, for their assistance with cell sorting and analysis.

I also gratefully acknowledge the generous advice and assistance I received from the staff at the Eaves lab: Darcy Wilkinson for sharing her experience in tissue processing with me when I first started, Glenn Edin for preparing the lentivirus, and Margaret Hale for her help with some of the molecular biology experiments. I would also like to express my appreciation to the Mammary team in the Eaves lab, past and present, in particular, Dr. Maisam Makarem, Dr. Long Nguyen, Dr. Peter Eirew, Dr. Davide Pellacani, Dr. Sylvain Lefort and Sneha Balani, for the discussions, invaluable suggestions and input.

Finally, my dearest gratitude goes to my parents, for all their sacrifice, support and prayers, especially during the last stage of my graduate studies when I had to face a major medical decision. I am deeply grateful for the health I have regained and for the successful completion of this dissertation in spite of the unexpected challenges, and I owe this to their love and support, physically and emotionally.

Last but foremost, I thank God for giving me the strength to carry on even when it seemed impossible.

Table of Contents

List of Tables

List of Figures

List of Acronyms

2D	2-dimensional
3D	3-dimensional
APC	allophycocyanin
CD	cluster of differentiation
CFC	colony-forming cell
CK	cytokeratin
DAB	3,3'-diaminobenzidine
DAPI	4',6-diamidino-2-phenylindole
DNA	deoxyribonucleic acid
ECM	extracellular matrix
EDTA	ethylenediaminetetraacetic acid
FACS	fluorescence activated cell sorting
FBS	fetal bovine serum
FFPE	formalin-fixed paraffin-embedded
FITC	fluorescein isothiocyanate
GO	Gene Ontology
H & E	hematoxylin and eosin
HPV	human papillomavirus
HRP	horseradish peroxidase
IHC	immunohistochemistry
MUC1	mucin-1
NGFR	nerve growth factor receptor
OPSCC	oropharyngeal squamous cell carcinoma
PCR	polymerase chain reaction
PE	phycoerythrin
PI	propidium iodide
PF	Phosphate-buffered saline supplemented with FBS
RNA	ribonucleic acid
ROCK	rho-associated kinase
ROCKi	rho-associated kinase inhibitor

SEM	standard error of the mean
SCC	squamous cell carcinoma
SD	standard deviation
YFP	yellow fluorescence protein

Chapter 1. Introduction

1.1. Palatine tonsils

1.1.1. Anatomy and physiology of the palatine tonsils

A human body is constantly challenged by inhaled or ingested microorganisms throughout its lifetime. To protect it from potentially hazardous foreign antigens, the body utilizes various physiological mechanisms as a barrier. Such mechanisms are largely categorized into two types: skin-associated lymphoid tissue (SALT) and mucosa-associated lymphoid tissue (MALT). The paired palatine tonsils, along with other lymphoid tissues situated in the oropharynx, are major components of MALT[1].

The oropharynx, an anatomic location that refers to the midportion of the throat where the oral cavity joins with the pharyngeal cavity, situates different types of tonsils collectively called "Waldeyer's ring"[2,3]. Waldeyer's ring consists of submucosal and subepithelial lymphatic tissues and includes the nasopharyngeal tonsil (also known as adenoid), tubal tonsil, lingual tonsils and palatine tonsils (**Figure 1-1A**). The paired palatine tonsils, commonly known (and will hereafter be referred to) as "the tonsils", are located in the palatine fossa, partially covered by the semilunar fold[1] (**Figure 1-1B**).

The tonsils are aggregates of lymphoid follicles (also known as lymphoid nodules) composed of the central germinal centres (B cell areas) surrounded by an extrafollicular zone (T cell areas). Antigens that enter the crypt are taken up by antigen handling dendritic cells, macrophages and M cells (specialized epithelial cells of the mucosa-associated lymphoid tissues[4]). Dendritic cells transport antigens to the extrafollicular (EF) zone, process them and present immunogenic peptides to B cells, thus promoting clonal affinity maturation by selection for B-cell survival. The lymphoid follicles are encapsulated by the epithelium overlying 10-30 crypts and invaginate deep

within the lymphoid tissues, thereby extending the surface area of the tonsil up to 300 cm^2, facilitating both protective and antigen sampling roles[5,6]. The stratified epithelium and multiple crypts are key features that distinguish the palatine tonsils from other tonsils. The adenoid and tubal tonsils are overlaid by ciliated respiratory epithelium and do not exhibit crypts; the lingual tonsil is covered with stratified epithelium but only exhibits a single crypt along with numerous salivary glands and striated skeletal muscle fibres that are not found in the palatine tonsils. The existence of multiple crypts in the palatine tonsils highlights their unique capacity for antigen sampling, serving as the first line of defense against many potentially pathogenic microbial organisms[7]. The palatine tonsils are absent in murine species[8] and might thus be considered as relatively modern organs specific to mammals. Even amongst the mammals, varying numbers of tonsillar crypts have been noted: the tonsils of the pig, rabbit and cat having a single crypt; sheep and camel tonsils having a few short crypts; the cow tonsils, a few large crypts; the dog tonsil having multiple crypts and focal infiltrations of lymphocytes, and the human tonsil with up to 30 crypts and marked lymphocytic infiltrations[9,10]. Salivary glands are only found in the palatine tonsils of animals and are absent in human palatine tonsils[11]. Although the reason for such inter-species variations is not clear, it is speculated to reflect an evolutionary advantage[10].

1.1.2. Development of the oral cavity

The oral cavity and the pharynx regions are the anatomic space bounded by hard and soft tissue structure, of which shape constantly changes with normal physiologic functions and movements such as speech, swallowing and respiration. The oral cavity is enclosed by lips (the anterior end), the cheeks (the lateral sides), the hard palate (the superior side), as well as the tongue and the sheet of muscles overlying the inner side of the mandible (the inferior side).

Embryologically, the epithelium overlying the inner surface of the lips, floor of the mouth, gingiva, cheeks and hard palate is derived from ectoderm, whereas the epithelium covering the tongue is derived from both endoderm and ectoderm[12,13]. The majority of the connective tissue components are derived from neural crest cells[14,15].

1.1.3. Development of the tonsils

The tonsillar epithelium is derived from the inner endodermal layers of the second pharyngeal (or branchial) pouch, which “outpouches” towards the ectoderm, forming pharyngeal pouches. The inner endodermal layer lining the second pharyngeal pouch dips into underlying mesenchymal cells and forms tonsillar crypts. Mesodermal cells form pharyngeal musculature and endothelial cells that surround the arteries, connective tissue stroma of the tonsil. In the 9th week of gestation, the first branched and solid crypt develops from the lining epithelium of the second pharyngeal pouches. In the 13th week, this formation stimulates the lymphocytic and mesenchymal infiltration of the connective tissue around the crypt branches, followed by a progressive development of complex epithelial invaginations until the 28th week (**Figure 1-2A**). In the 29th-35th week, populations of lymphocytes in the primary follicles (B-zone) and parafollicular area (T-zone) increase (**Figure 1-2B**); however, germinal centres are not observed in embryonic life[16–18]. Antigenic stimulations shortly after birth result in the development of secondary follicles with active germinal centres (**Figure 1-2C**). The subsequent proliferation of germinal centres results in the rapid growth of the tonsil in the first decade of life (**Figure 1-2D**). As a person ages, the overall size of the tonsils is much reduced as the lymphoid elements undergo involution, accompanied by the proliferation of fibrous tissue in the capsule and trabeculae[19].

1.2. Oral and tonsillar epithelium

1.2.1. Oral epithelium

The oral mucosa is characterized functionally as masticatory (hard palate and gingival), specialized (dorsal surface of the tongue) or lining (buccal mucosa, ventral surface of the tongue, soft palate, intra-oral surfaces of the lips and alveolar mucosa)[12,20].

Histologically, finger-like projections called rete ridges (undulation of epithelium) are protruded downwards into the connective tissues, in turn creating upward projections of the connective tissues into epithelium called dermal papillae. This structure facilitates

tight contact between the oral epithelium and the underlying lamina propria during mastication[21].

The total surface of the oral mucosa lining can be divided into 3 groups: Keratinized lining resembling the skin epidermis that is designed to withstand external mechanical forces (~25%), non-keratinized lining overlying the floor of mouth, buccal regions, etc. (~60%), and the specialized mucosa on the dorsal tongue (~15%)[22]. The keratinized and non-keratinized epithelia are stratified squamous epithelium comprised of various layers. Basal, spinous, granular and corneal layers are the components of the keratinized epithelium; granular and cornified layers are not present in the non-keratinized epithelium. Non-keratinized epithelium consists of basal, spinous (or parabasal), intermediate (or middle) and superficial layers. The stratified squamous epithelium is in direct contact with an underlying dense connective tissue called lamina propria, which contains blood vessels, fibroblasts, salivary glands, fibers and other cell types[12,13,15].

1.2.2. Tonsillar epithelium

Unlike the majority of the oral epithelium in which structure is consistent and predictable throughout the entire mucosal linings, the tonsillar epithelium exhibits an irregular and dynamic structure. The epithelium overlying the palatine tonsils is commonly classified into two types: (1) the stratified squamous epithelium lining the tonsillar surface and the upper part of crypt epithelium, and (2) the crypt epithelium that plunges deeply into the interfollicular areas running full-thickness through the tonsil. The latter is rarely stratified and mostly disrupted by infiltrating lymphocytes and other non-epithelial cells, hence referred to as lymphoepithelium[3].

The surface epithelium, similar to skin or oral mucosa, is underlined by connective tissue with intact basement membrane in between, with clear stratification displaying basal, spinous (or parabasal), intermediate and superficial layers[23]. The most primitive cells – those that can give rise to daughter cells that can self-renew and differentiate into mature cells – are believed to reside in the basal layer, supported by

the underlying stroma[24]. The differentiated cells then migrate upward and eventually are sloughed off upon terminal differentiation.

In contrast, the crypt epithelium usually either lacks collagenous connective tissue or has only a very few fibres associated with the basement membrane. Its structure is affected by the close proximity to lymphoid follicles that are actively producing immune cells[25]. The epithelium is characterized as a porous, sponge-like, “reticular” (net-like) epithelium[26], that contains non-epithelial cells including lymphocytes, macrophages and dendritic cells[27]. The thickness of the crypt epithelium varies depending on the degree of such infiltration. In some regions it has 3 compartments: the basal, the intermediate or prickle cell layer, and the superficial layer. However, in areas of extensive lymphocyte infiltration it may be comprised of a single layer[6,9]. Interestingly, despite the ensuing loose distribution of epithelial cells in the epithelium, with little or no cell-cell contact and constant environmental insults, the epithelial cells in the crypt continue to maintain protective layers, suggesting that primitive cells do exist in these sites and are able to replenish rapidly differentiating cells.

1.3. Stem cells

1.3.1. Adult stem cells

Stem cells can be classified as embryonic stem cells (ESCs) or adult stem cells, according to their developmental status. ESCs are derived from cells in the inner mass of the blastocyst, and are totipotent; this refers to the fact that these cells can give rise to any of the three germlines, endoderm, mesoderm and ectoderm, which collectively form all organs in the body.

Adult stem cells, in contrast, are tissue-specific, in that they are specialized to produce some or all of the mature cell types found within particular tissues or organs. Adult stem cells are defined as undifferentiated cells that (1) can self-renew for the long term, (2) have the ability to give rise to transit-amplifying (TA) and differentiated cells, and (3) reside in adult tissues[28,29].

Different types of adult stem cells can be identified based on their cycling patterns and cell output capacity. For example, (1) those constantly cycling with fast tissue turnover, such as intestinal stem cells, and short- or long-term cycling haematopoietic stem cells[30–33], (2) those of which proliferation can be strongly induced by injury, such as lung[34] or muscle[35,36], and (3) those that remain quiescent or transiently activated, such as hair follicle stem cells[37]. Although seemingly distinct, all adult stem cells exhibit long-term proliferation capacity that is "plastic", and hence the rate of turnover can be altered in response to external cues or tissue damage/dead cells[34,38–41]. Adult stem cells can also be found in less-regenerative organs such as brain[42].

While the term "embryonic stem cells" is generally well-defined and accepted, the definition of "stem cells" in general still remains ambiguous as scientists from different disciplines including developmental biologists, immunologists and many others use the term in different ways[43]. Other relevant terms such as "progenitors" and "precursors" are also often used interchangeably with stem cells[43]. Nevertheless, a certain level of tolerance is necessary in order not to be distracted by semantic arguments but to focus on more important research questions. For clarity, I will define stem cells as cells that can give rise to the entire tissue, and progenitors as cells that have more limited self-renewal potential and typically differentiate into mature cells more rapidly, maintaining homeostasis of the tissue in which they reside.

1.3.2. Stem cell niche

Regulation of stem/progenitor cells, regardless of their intrinsic potential, is influenced by inputs from their microenvironment or the "niche". The stem cell niche hypothesis was first proposed by Schofield, stating that stem cells reside within fixed compartments that provide signals via secreted or surface molecules[44]. The surrounding supporting cells in the niche (1) release signals to regulate stem cell self-renewal, survival and maintenance[45,46], (2) facilitate maintenance of the correct stem cell polarization with respect to the niche, allowing for asymmetric cell division[47], and (3) attach to stem cells to maintain close proximity of stem cells to self-renewal and survival signals emanating from their microenvironment[48]. In mammals, epithelial stem cells orient in such a way that, the basal, and not the apical, plasma membrane of

stem/progenitor cells is in contact with the basement membrane. Basement membranes are typically rich in laminins and type IV collagen. The basal cells make contacts with these extracellular matrix components via cell surface receptors such as integrins[49], facilitating the correct polarity within stem cells and providing a scaffold and an optimal stiffness. The cell-matrix interaction and reception of various signals produced by niche components are key to maintenance of the stem cell pool and determination of their differentiation fates[50].

1.3.3. Stem cell hierarchy in other tissues

Hematopoiesis is one of the best-characterized examples of a hierarchical differentiation process originating in a hematopoietic stem cells (HSC) compartment that is predominantly localized in the bone marrow throughout normal adult life[51]. Individual HSCs can differentiate into multiple cell lineages of the blood and immune systems, including erythrocytes, platelets, basophils, eosinophils, neutrophils, macrophages, T cells and B cells. HSCs comprise subsets that have variable durabilities of mature cell outputs ranging from a few weeks to the lifetime of the organism[33]. Those with shorter and longer term output capacity have different phenotypes which allow their differential prospective isolation using fluorescence-activated cell sorting (FACS) technology.

Unlike hematopoietic cells, epithelial cells are organized in layer(s) with close cell-cell interactions confined in the specific tissue, overlying most of the organs of the body, such as skin, glandular tissues, digestive respiratory tracts, and the urogenital system[52]. These epithelial sheets not only protect the organs from the external environment, thereby maintaining tissue integrity, but also facilitate water retention and nutrient absorption[52], and in some cases, trans-epithelial transport of non-epithelial cell[53]. The number of epithelial layers varies depending on anatomic locations and functions. For instance, skin epidermis is made up of a stratified multilayer; Glandular tissues such as mammary and prostate epithelium are organized as bilayers; Intestine and hair follicles are lined by a single layer[54]. Regardless of the organ of interest, most of these epithelia are maintained by slow-cycling stem/progenitor cells that are long-lived and able to generate short-term proliferative progeny often referred to as transit amplifying (TA) cells[52,55]. TA cells then actively produce daughter cells through a limited number of

divisions to replace terminally differentiated dead cells that are constantly being removed. This process of constant cell replacement and organ renewal is called tissue homeostasis[52].

Multipotency and hierarchical relationships within the adult epithelial cell populations have been extensively characterized in the skin epidermis using transgenic mouse models. These have demonstrated that multipotent stem cells within the bulge of the hair follicle can differentiate into all three lineages, hair, sebaceous glands and interfollicular epidermis[56–59]. Interfollicular epidermis is organized in stacks of distinct cell layers or "stratified" squamous epithelium that consists of the innermost basal layer where the most primitive cells reside, suprabasal layers (spinous and granular layers) composed of more differentiated cells migrating outwards, and the outermost stratum corneum layers that function as an impermeable barrier to the external environment[60]. The interfollicular epidermis and sebaceous glands contain lineage-restricted unipotent progenitors that differentiate into mature epithelial cells that replace dead cells[60]. The notion of "epidermal proliferative unit" was proposed by Potten in 1974[61], stating that the epidermis is organized in stacks of cells with ~10 basal cells (1 unit), and that each unit contains one stem cell. This notion has been experimentally demonstrated by engrafting cells onto Nude mice[62–64] or lineage tracing analysis[65] (discussed in Section 1.4.2.).

1.4. Detection, isolation and quantification of epithelial stem/progenitor cells

1.4.1. Qualitative methods

Label-retaining cells

Identification of epithelial stem cells has been inferred from various methods. One of the earliest endpoints used was label-retention *in vivo* based on the hypothesized slow-cycling activity of normal adult stem cells[66–69]. To identify quiescence cells, pulse chase experiments are carried out using DNA analogues such as tritiated-thymidine (3H-TdR) and more recently, 5-bromo-29-deoxyuridine (BrdU), both of which are incorporated into proliferating cells (S-phase). After long-term exposure to 3H-TdR or

BrdU, actively cycling and proliferating cells lose their label shortly after the exposure has discontinued, whereas a minor subset of cells may retain labeled DNA for a longer period of time. These “label-retaining cells” (LRCs) are therefore considered to represent slow-cycling or quiescent cells. In the oral cavity, BrdU was used to identify LRCs in mouse gingiva[70] as well as rat buccal mucosa, tongue and hard palate[71], which were predominantly but randomly distributed in the basal layer[72]. However, not all stem cells are slowly cycling, and a slowly-cycling property does not indicate their functional potential[28], which cannot be experimentally assessed as visualization of LRCs requires cell fixation. Additional qualitative and/or quantitative methods should therefore be employed in order to circumvent this caveat and strengthen the claims put forward.

Lineage tracing

Recent advances in the development of *in vivo* lineage tracing methods to track the cell output activity of different stem cell types based on their expression of fluorescently labeled (or substituted) transcription factors has provided insights into differentiation hierarchies and stem cell lineage outputs within epithelial tissues. Prime examples include the skin epidermis[60] and the gut[73]. In this technique, a particular stem cell type is genetically labeled with a fluorescence marker that is then transmitted to its progeny, thereby allowing these daughter cells to be identified histologically as they migrate away from the founder cell[74].

In their pioneering work, Barker et al.[73] performed lineage tracing and proposed Lgr5 as a marker of stem cells after having shown that the villi cells that were initially labeled with multiple colours, within 8 weeks, were replaced with a single colour that originally only marked $Lgr5^+$ cells near the base of the crypt[73]. Subsequently, $Lgr5^+$ cells have also been shown to comprise an actively proliferating and multipotent cell population in mouse hair follicle[75], and to repopulate hair follicles for more than 14 months, supporting the existence of long-lived hair follicle stem cells.

In vivo lineage tracing has also been used to study oral epithelial progenitor cells. One study used a tamoxifen-inducible Cre recombination-based cell lineage labeling system (K14-CreER; Rosa26-LSL-LacZ mouse model) to trace the origin of taste bud cells of postnatal and adult mice. The study showed that both the mature keratinocytes

and the taste bud receptors have the common origin defined by $K14^+K5^+Trp63^+Sox2^+$ immediately outside the taste bud. This was inferred from the detection of labeled cells on both the dorsal tongue and in the buccal mucosa after a one-month chase, suggesting bipotency of $K14^+K5^+Trp63^+Sox2^+$ cells[76]. In another study, a Sox2-Cre-ER; Rosa26-LSL-EYFP mouse model and transplantation experiments were used, demonstrating that Sox2-expressing cells in the basal layers can give rise to multiple mature cell types within the tissue and these could be detected for at least 10 months after labeling the dorsal tongue[77].

X-chromosome inactivation

Although *in vivo* lineage tracing is a powerful technique, not all tissues can be readily examined utilizing this strategy. X-chromosome inactivation in females is another approach that has provided useful information in this regard. One of the X-chromosomes in each cell becomes transcriptionally inactive during early embryogenesis and this epigenetic condition is then maintained in all subsequent daughter cells thereby serving as a clonal indicator of cells that remain relatively immobile throughout the lifetime of the tissue. For example, in the human breast, an indication of how stem cells organize was provided by analyses of the pattern of X-chromosome inactivation, where multiple discrete regions of breast epithelium were found to share the same inactive X chromosome[78]. In head and neck, little has been reported on utilizing this method to study progenitor cells. A few studies, however, used the method to examine synchronous tumors in the upper aero-digestive tract, showing an inactivation of a single allele indicative of monoclonal proliferations[79,80].

Mitochondrial mutations

Recent studies have shown the considerable potential also of mitochondrial mutations for lineage tracing of human stem cell activity especially in epithelial tissues with a discrete architecture. In normal human tissues, it is common that cells and their progeny contain nonpathogenic mutations in their mitochondrial DNA including mutations in the cytochrome c oxidase (COX)[81,82] due to a lack of protective histones and poor DNA repair mechanisms. These mutations, however, can expand stochastically and affect all the mitochondria in the cell (homoplasmic) or a certain proportion (heteroplasmic). This expansion process takes up to years to be seen either

in the form of homoplasmy or heteroplasmy. Thus stem cells or long-lived progenitors are thought to be the most likely cell types that accumulate enough of these mitochondrial mutations for them to become biochemically detectable. The technique involves two-colour enzyme histochemistry to detect activity of the mtDNA-encoded COX and DNA-encoded SDH (a component of complex IV and II of the respiratory chain, respectively), where blue indicates a lack of COX activity in cells and brown indicates the presence of active COX. This technique is then combined with polymerase chain reaction (PCR) and mtDNA sequencing to determine whether the mitochondrial mutations were derived from the same clone. This method has been used as a reliable marker of clonal expansion in the intestinal crypt and stomach[81,83,84] as well as liver, pancreas and skin[85].

Fluorescence-activated cell sorting

Fluorescence-activated cell sorting (FACS) allows the isolation of single cells of based on their unique, cell-type specific protein expression. This method involves labeling antibodies to specific proteins with different fluorochomes of distinct emission wavelengths. As the labeled cells pass through the flow cytometer, cells labeled with a specific set of fluorochromes are then individually detected in droplets that pass in front of lasers that activate the fluorochromes to allow the cell to be subsequently analyzed and isolated as desired[86].

Originally developed in the immune and hematopoietic fields, this method enabled the identification of stem cell markers that can be used to prospectively enrich for subpopulations of viable cells that display different functional properties[87–91]. This effort was facilitated by the fact that these cells often exist *in situ* as isolated cells and do not depend on chemical linkages to other cells or tissue components for their viability. Adaptation of this methodology to other tissues has been slower, because of the need to develop methods for obtaining viable single cell suspensions from them. However, a number of markers, including markers of embryonic stem cells, neural, hematopoietic, mesenchymal, epithelial and limbal epithelial stem cells, endothelial progenitor cells, and cancer stem cells have now been identified[86]. But, marker expression patterns are often overlapping on multiple cell types, necessitating the use of multi-color (multiparameter)

detection approaches to enhance the purity and specificity of prospectively isolated subpopulations. **Table 1-1** summarizes markers of epithelial stem/progenitor cells.

Many of the proteins that have been shown to mark primitive cells in the oral epithelium were first identified in skin stem cell studies of the hair follicle and interfollicular epidermis. These markers include the integrin pair α6 (CD49f) and β1 (CD29)[92–94], transmembrane proteins such as melanoma chondroitin sulphate proteoglycan[95], intracellular proteins such as cytokeratins (CK) 15 and 19[96,97]; and a nuclear protein p63[98,99]. Although immensely useful, no single or combination of phenotypic markers has yet been found to be an exclusive indicator of a stem cell[100], and it remains critical to confirm the identity of FACS-purified viable subpopulations using robust functional assays that enable stem cells to be specifically quantified and characterized.

1.4.2. Quantitative methods

In vivo transplantation assay

In vivo transplantation is currently the 'gold-standard' assay for evaluating stem cell activity. In the mammary field, the assay for mouse mammary stem cells involves injecting dissociated mammary cells into the cleared (or de-epithelialized) fat pad of a pubertal female mouse. Stem cell growth and differentiation properties are inferred from their ability to generate an entire mammary gland composed of branching ducts and sac-like lobules in the fat pad[101]. Such cells are operationally referred to as mammary repopulating units (MRUs), of which reported frequencies ranges from 1/30,000 to 1/1,400 according to limiting dilution analysis[101,102]. Coupling with cell sorting strategy to fractionate $CD29^{hi}$(or $CD49f^{hi}$)$CD24^{+}$ cells has enabled an increase in the MRU purity up to ~1-2% (reviewed in ref. 103). A similar method has also been used to detect and quantitate human mammary stem cells by combining cells dissociated from human reduction mammoplasty samples with irradiated fibroblasts in collagen, and transplanting them under kidney capsules of immunodeficient mice[88]. Human MRU activity is then assessed based on whether the gels contain cells that can form colonies of mammary cells *in vitro* when subsequently dissociated and plated in two-dimensional cultures[88]. The human MRU was detected at a frequency of 1/1,000 to 1/10,000 cells, had a distinct

phenotype ($CD49f^{+}EpCAM^{-}$) and was serially transplantable, demonstrating self-renewal and differentiation activities[88]. A similar approach has been used to study epidermal stem cells, where keratinocytes dissociated from human neonatal foreskin tissue were inoculated into cleared rat tracheas and transplanted subcutaneously into immunnodeficent mice[104]. Complete epithelialization of the tracheal lumen required transplantation of at least 5×10^5 unsorted keratinocytes, whereas 3×10^4 cells was sufficient for $\alpha6^{bri}CD71^{dim}$ cells (putative stem cell fraction) to regenerate a complete skin epidermis.

Although *in vivo* assays have clear merits, the fact that they are costly and time consuming has led to the development of various *in vitro* alternatives to study stem/progenitor cells.

In vitro sphere-forming cell assay

The sphere-forming cell assay was first introduced in 1992 in a report by Reynolds and Weiss, that a small subpopulation of neuronal cells from adult brain was able to grow as free-floating spheres (termed “neurospheres”) in non-adherent culture conditions in the presence of epidermal growth factor (EGF) and fibroblast growth factor (FGF) without serum[105]. The “stemness” of the neuropheres was demonstrated by (1) the expression neuroepithelial stem cell markers within these neuropheres, (2) their ability to self-renew evidenced by the formation of secondary neuropheres when dissociated and cultured in non-adherent conditions, and (3) their ability to differentiate into both neurons and glial cells when dissociated and plated on an adherent culture plate. The sphere-forming assay has since been widely used to study many different epithelial tissues, including breast[88,106], cornea[107], skin dermis[108], pancreas[109] and lung[110]. The assay has also become a popular method of choice to study stem cell properties in several tumour types[111–114] including oral squamous cell carcinomas[115].

Generally, activity in the sphere-forming assay is reported as both the size and number of spheres formed in given conditions, often with a wide range of seeding densities. However, it is important to be mindful of the caveats of this system and not assume that each sphere is always derived from a single cell and is therefore a clone[116]. Although seeding densities anywhere between 0.2 and 20 cells per µl is considered a

'clonal density'[117–119], cell aggregation is a common property of many cell types and occurs even with low densities of cells[120,121]. In addition to physical aggregation, inaccurate estimation of sphere-forming cells can result from paracrine signals released by neighbouring cells, which can greatly affect the individual sphere-forming efficiencies[116].

Despite these limitations, generation of spheres can serve as an informative tool to detect cell survival or proliferation rather than quantifying input progenitor numbers[116]. Recently, attempts have been made to circumvent the issue of cell aggregation by culturing spheres in semisolid medium such as collagen[122] or methylcellulose[123–125]. If the experiment is designed to characterize stem cell activities (self-renewal and differentiation), however, a better practice is still to plate cells at a density of a single cell per well so that true clonality can be guaranteed[116].

In vitro two-dimensional (2D) colony-forming cell assay

In studies of human skin epidermis, cells with different proliferative capacities have been identified and categorized into three types: holoclones, paraclones and meroclone[126,127]. The holoclone is defined as a tightly clustered large colony with smooth perimeter containing less than 5% of terminally differentiated cells; it is derived from cells with the greatest proliferative capacity under the culture conditions in which these colonies are generated (120-160 cell divisions)[126,128,129]. Paraclones are colonies obtained under the same culture conditions but are small, highly irregular in shape because they contain only terminally differentiated cells and are derived from cells with the lowest replicative potential (no more than 15 generations)[126]. Meroclones are colonies with intermediate properties; i.e. they achieve an intermediate size and containing a heterogeneous mixture of cells with varying proliferative capacity[126]. Holoclones, when serially passaged in liquid culture, give rise to meroclones and/or paraclones, whereas paraclones soon abort and terminally differentiate, consistent with the concept that holoclones derive from the most primitive cells that maintain a unidirectional differentiation process comparable to that responsible for sustaining the tissue architecture *in vivo,* where stem cells in the basal layers continuously generate transit amplifying cells that are destined to differentiate and renew the tissue[130,131]. The method used to test clonogenic efficiency is the colony forming cell (CFC) assay, where

the number of colonies formed in a culture dish after culturing a given number of cells is used to determine their proliferative ability.

Colony morphology and associated proliferative potential have been widely used as a measure of functional activity of stem/progenitor cells from skin epidermis[132], follicles[133], limbal tissues[134] and mammary epithelium[135,136]. In oral mucosa, a recent report suggest that holoclonal morphology and their high clonogenicity may be the characteristics of progenitor cells isolated from oral epithelium, based on the finding that holoclones intensely expressed a basal cell marker p75 (NGFR) on both the mRNA and protein level[137].

Several factors have been shown to affect colony-forming efficiencies. One of the most widely used additions is the use of irradiated murine fibroblasts (NIH 3T3 cells) as feeder cells to enhance colony-forming capacity. One of the earliest demonstrations of this ability appeared in the study of human keratinocytes by Rheinward and Green, who showed that co-culture of epidermal keratinocytes obtained from skin biopsies (10^3-10^5 cells) with lethally irradiated 3T3 fibroblasts markedly improved both colony-forming efficiencies and their capacity for serial passage[138]. They concluded that the limited proliferative capacity which had been historically observed was not cell-intrinsic, but rather due to the suboptimal culture environment where their association to fibroblasts was lost[138]. Since then, feeder cells have been widely used in many *in vitro* epithelial culture systems to improve the detection of colony-forming cells which requires that these be plated at very low cell densities[89,135,139–141].

An additional factor that has been shown to influence plating efficiencies of epithelial cells is the inhibitor of Rho-associated protein kinases (ROCKs), the downstream effector proteins of RhoA GTPase that regulate the actin cytoskeleton and cell migration and proliferation[142,143]. Initially discovered as an agent that promoted the survival and cloning efficiency of human embryonic stem cells, ROCK inhibitor (Y27632) is now widely used during *in vitro* manipulations of tissue cells[144–147]. ROCK inhibition, when combined with feeder cells, enhances proliferation and survival of keratinocytes[148,149]. Recently, this method has been demonstrated to induce an indefinite proliferation and to rapidly generate both normal and tumor epithelial cells from many

tissues obtained from small biopsy specimens[150]. ROCK inhibition has been incorporated into the *in vitro* murine prostate colony assay, showing an increase in cloning efficiency by 8-fold[151]. An enhancing effect has also been noted in mammary colony-forming cell assays[152].

In vitro three-dimensional (3D) organotypic culture system

The 3D organotypic culture system is used to study keratinocyte differentiation and maintenance *in vitro*[153]. Epithelial keratinocytes are seeded on top of an extracellular layer that consists of collagen and fibroblasts and allowed to form a monolayer. Culture medium is then removed from the keratinocyte cell surface[154]. When undifferentiated confluent monolayers with underlying extracellular matrix are exposure to air, the undifferentiated cells are polarized and form tight junctions. The basal layer that contains stem/progenitor cells maintains its proliferative potential, while daughter cells lose their cell division activity and differentiate into mature cells altering gene expression as they leave the basal layer and migrate upward, leading to a regeneration of full-thickness human skin[155]. The different stages of maturation in this regenerated tissue have been characterized via expression of keratins and other associated proteins known to be associated with normal differentiation *in vivo*. Similar methods have been used to reconstitute realistic model of skin epidermis as well as oral, esophageal and cervical epithelium[137,154,156–159].

In vitro 3D organotypic culture systems have a wide range of applications including examinations of normal development and pathogenic processes such as oncogenic transformation, viral growth and life cycle and virus-host interaction[160] and toxicological studies[161]. In the later part of this dissertation, I explored the utility of this culture system to examine the interaction between normal tonsillar epithelial cells and those expressing HPV16 E6/E7 oncoproteins and impact of such co-cultures on tonsillar epithelial differentiation.

1.5. Human papillomavirus (HPV) in the oropharynx

1.5.1. HPV in oropharyngeal carcinoma

Head and neck cancers comprise a group of cancers of the upper aerodigestive tract. These include cancers derived from cells of the oral cavity (including lips), pharynx (nasopharynx, oropharynx, and hypopharynx) and larynx[162]. Head and neck cancer is the sixth most common type of cancer in the world, with approximately 690,000 new cases worldwide and 375,000 deaths in the year 2012, representing 4.9% of total cancer incidence and 4.6% of total cancer mortality[163].

Cancer of the oropharynx is a subset of head and neck cancer, with the palatine tonsils being the most common site of malignancy in the oropharynx. The majority (85%) of such cases are oropharyngeal squamous cell carcinomas (OPSCCs)[164–166], followed by malignant lymphoproliferative disease including large B-cell lymphoma[167].

Until several decades ago, the primary risk factors for oropharyngeal carcinoma were tobacco and alcohol[168]. Tobacco cessation during that time span has led to a decrease in the incidence of smoking-related OPSCCs, However, the incidence of HPV-related OPSCCs - especially in the tonsil and base of tongue - has been steadily increasing in many developed countries where tobacco use has declined, resulting in a continual increase in the overall incidence of OPSCC[169,170]. The proportion of HPV-positive oropharyngeal cancers has risen to over 70% in North America and over 90% in Europe, with a striking increase in Sweden from 23% in the 1970s to 57% in the 1990s and to 93% in 2007[171]. These data suggest that HPV is now the primary cause of tonsillar malignancy in North America and Europe[171,172]. A pronounced increase is seen among men, younger individuals, and those with a high lifetime number of vaginal or oral-sex partners[173,174].

Although cervical cancer is the best-known HPV-associated cancer, with over 99% of total cases at this tissue site attributed to HPV infection[175], there has also been a dramatic decrease (over 80%) in the incidence of these cancers in the United States, which is attributed to the success of routine cervical screening programs[175]. The rise of the incidence of HPV-associated oropharyngeal cancer in men in parallel with the

marked decrease in the incidence of cervical cancer in women has transformed the epidemiological landscape of HPV-associated disease; it was estimated that, by 2010, the annual number of oropharyngeal cancer in men in the USA will have surpassed that of cervical cancer in women[173] (**Figure 1-4**).

HPVs are a large family of viruses, classified into different types based on percentages of identity at the nucleotide level in the L1 gene[176]. Over 120 types of HPV have been identified, of which at least 13 types have been categorized as high risk and considered causative agents of cancer and have been classified as carcinogenic (Group 1) by the International Agency for Research on Cancer (IARC)[177,178]. Of these high-risk types, HPV16 and HPV18 are responsible for approximately 70% of cancers of the cervix, vigina and anus[179,180]. In HPV-associated head and neck squamous cell carcinoma (including oral cavity, oropharynx and larynx), HPV16 is the most prevalent type, accounting for more than 80% of HPV positive cases[181] and 90% of HPV positive oropharyngeal cancers[181–184].

In a case-control study of 100 OPSCC patients and 200 control patients in the United States, D'Souza et al. reported that the presence of HPV16 in oral exfoliated cells was associated with a more than 13-fold increase in risk of oropharyngeal cancer[183]. In a hospital-based case-control study in Canada, HPV was detected in 43% of oral exfoliated samples obtained from oropharyngeal cancer patients, and those that harboured HPV16 were 31.5 times more likely to be cancer patients[185]. The HPV involvement was less clear for non-tonsillar oral cancers[185]. Interestingly, a recently published large-scale cohort study showed that HPV16-encoded E6 antigens were present in prediagnostic sera of approximately one-third of individuals who were later diagnosed with oropharyngeal cancer, suggesting a strong association between HPV infection and development of OPSCC[186].

1.5.2. Biology of HPV infection

HPVs are small DNA viruses (approximately 8kb in size) that encode 8 viral proteins: E1, E2, E4, E5, E6 and E7 (early genes), and L1 and L2 (late genes). HPV infections specifically arise in stratified squamous epithelia such as the skin, the

anogenital tract, cervical epithelium, and oral epithelium. The life cycle of HPV begins with HPV infection in undifferentiated proliferating cells (or “stem cells”) in the basal layer. This is thought to be facilitated through microwounds at several sites[187]. Once the HPV genome is integrated into the host genome of the infected basal cell, the host DNA replication machinery mediates synthesis of HPV DNA. The replication cycle can be divided into two parts. (1) First, the viral genome replicates to a low copy number (~100) in the infected proliferating cells and this state can be maintained for varying periods of time with production of E1 and E2 viral proteins that are essential for their initial replication. (2) Later, the synthesis of other viral proteins (E4, E6 and E7) occurs as the cells leave the basal layer and migrate upwards. Since the cells expressing human papillomaviral proteins can remain active in the suprabasal layers as they differentiate a subset of cells that reached the top layer of the epithelium can then re-enter S phase to replicate the HPV genome, followed by production of capsid proteins. The HPV genome is then subsequently re-assembled into a complete virus in episomal form which then can re-initiate the entire viral cycle. The HPV life cycle therefore strictly depends on the differentiation of infected cells[175,188]. Although the majority of infections are cleared by the immune system or kept in the initial replication stage, a few manage to evade the immune surveillance and persist, thereby promoting cancer development even after decades[189–191]. E6 and E7 oncoproteins are the major factors responsible for HPV-induced malignant transformation[175] and E5 cooperates with E6/E7 to promote hyperproliferation of infected cells, likely facilitating malignant progression[192]. Most research has been done with E6 and E7 oncoproteins that are strong drivers of progression.

E6 and E7 oncoproteins function mainly by altering the mechanisms responsible for cell cycle progression and for DNA damage recognition and repair. E7 targets members of the retinoblastoma (Rb) family. The binding of E7 to Rb releases E2F transcription factors allowing these factors to bind to regulatory regions of genes that code for proteins that are necessary for cell cycle entry. In normal cells, aberrant, uncontrolled cell cycle progression is interrupted by tumor suppressor p53-dependent apoptosis. However, in the presence of E6 oncoproteins, this p53-mediated process is abrogated as E6 binds to the p53 protein and promotes its degradation. In addition to RB destabilization, E7 contributes to immortalization through interaction with other key cell

cycle regulators including the cyclin-dependent kinase (CDK)-inhibitors p21 and p27 which play an important role in growth arrest during epithelial cell differentiation[193]. Activation of E2F transcription factors is also achieved by cyclinD/cdk that phosphorylates Rb to release E2F and subsequently synthesize E2F-activated genes (MCM, Ki-67, PCNA) that drive cell cycle progression. This process normally can be inhibited by the CDK-inhibitor p16Ink4a. However, in the presence of active E7, destabilization of Rb is not dependent on cyclinD/cdk. p16 is, therefore, no long able to control cell cycle progression as the negative feedback loop by p16 has been abrogated, and this results in accumulation of intracellular p16 protein. p16 positivity *in situ* is thus considered indicative of active E7 and frequently used as a biomarker of HPV infection[194]. The E6 and E7 proteins have also been shown to activate telomerase[195] and alter a numerous array of pathways involved in malignant transformation[196].

Although functional E6 and E7 are the primary requirement for transforming activity[197], neither of these oncoproteins by itself is able to efficiently achieve immortalization. E6 alone has no immortalization activity, and E7 alone immortalizes cells at very low frequency[198]. However, in combination, they act cooperatively and immortalize most primary cells including oral keratinocytes at very high efficiency, thereby contributing to the development of HPV-induced malignant progression[199,200]. When subjected to 3D organotypic (raft) culture system, keratinocytes expressing both E6 and E7 stratify and differentiate in a way that mimics the histological morphology observed in dysplastic lesions[201]. With respect to their tumorigenic potential, experiments using nude mice have shown that E6 and E7 in combination are still insufficient to form tumors from infected cells, and that tumorigenesis requires further insults, for instance, exposure to tobacco carcinogens[202–204]. Also, in oral keratinocytes as in cervical keratinocytes, HPV-induced tumor formation requires collaboration with other oncogenes, further genomic instabilities, additional genetic mutations and/or chromosomal rearrangement, which is in alignment with the fact that most people infected with HPV do not develop cancers and if they do, there is a long latency[205,206].

1.5.3. HPV in oropharynx of healthy individuals

Although HPV is now well-recognized as a major risk factor for head and neck cancer, the frequency of an HPV infection in the oral cavity in individuals in the general population and its association with cancer risk is still poorly understood. Even less is known about the prevalence of the infection in the tonsillar region, which is much more difficult to sample.

Previous studies reported that HPV was detected in oral rinse specimens of 3-5% of adolescents[207–209] and 4.5-10% of adults[210–212]. Among adults, a recent large scale study of 5579 participants (aged 14 to 69) identified a bimodal pattern, showing peaks in the prevalence plot with two age groups, individuals aged 30-34 years (7.3%) and people aged 60-64 years (11.4%)[213]. With respect to HPV types, a comprehensive systemic review of 18 studies showed that 4.5% of 4070 healthy subjects had any type of HPV, 3.5% of 4441 subjects had carcinogenic HPV, and 1.3% of 3977 subjects had HPV16[214]. However, it should be noted that only two of these 18 studies used specimens from the palatine tonsils as opposed to oral scrapes or rinses[215,216]; Even then, the data were based on the analysis of brushing samples collected from the outer surface of the tonsils, which did not provide information on the actual frequency in the crypt region. To address this issue, Franceschi et al. examined cells obtained from deep brushing of the resected tonsils with the crypts exposed[217]. Of the total 200 samples examined, only 3 were tested positive for HPV (2 for HPV16 and 1 for HPV39). This study, however, encountered a major challenge with sample evaluation due to low number of squamous epithelial cells or a large number of lymphocytes masking epithelial cells[217].

The relative rarity of HPV prevalence in the tonsils of healthy (cancer-free) individuals is contrary to the findings from the cervix, where HPV is frequently detected and often persists for many years[218,219]. Such obscurity as well as other issues to be discussed in the following section raises a question of whether HPV positive tonsillar malignancy follows the same course of oncogenic progression as seen in cervical cancers.

1.5.4. Natural history of HPV-associated tonsillar carcinoma

The natural history of HPV-associated oropharyngeal cancer is poorly understood. This ambiguity is attributed to several factors; First, HPV-positive oropharyngeal malignancy is believed to originate from deeply invaginated tonsillar crypt, evidenced by its tendency to display a growth pattern that emanates from the crypts and moves outward, sometimes spreading to the surface epithelium[220]. Tissue examination of sections from the inner tonsil using *in situ* hybridization has shown that HPV DNA is primarily present in the crypt region. A similar location has been noted for the overexpression of p16 protein, an alteration that is a commonly accepted biomarker for the presence of active HPV oncoproteins[194,220]. This location means that it is not possible to use a brushing technique such as that used in the Papanicolaou (PAP) test to screen for potential premalignant and malignant changes in the tonsil. Second, there is currently no way to identify tonsils harbouring premalignant lesions clinically in a noninvasive fashion, to allow for early diagnosis of the disease, and there is currently no visualization modality to facilitate such detection[221]. Third, from a histological standpoint, very few dysplastic lesions have been observed in archival samples of tonsils; the few that have been reported have mainly been present in tissue adjacent to tonsils. Whether this is due to the fact that tonsillar tissues are permeated by lymphocytes, making it difficult to study the tissue, or whether the histological changes associated with early disease are different from classic dysplasia features, is unknown. Tonsil cancers often appear non-keratinized with a "basaloid" morphology[222], further complicating pathological determination of a HPV-associated tonsillar malignancy, as the reticulated epithelium of the "normal" tonsillar crypts is already highly infiltrated with lymphocytes and appears poorly differentiated with lack of keratinization[194,222]. In addition, the term "basaloid" inaccurately reflects the true disease state, as basaloid SCC is known for its aggressive behaviour and poor prognosis, when in fact, HPV-associated tonsillar SCC has a better prognosis compared to HPV-negative oropharyngeal cancers[223]. Together, these challenges remain barriers to understanding the natural history of HPV-associated tonsillar cancers.

1.6. Association between HPV and stem/progenitor cells

Cells with long-term tissue-specific regenerative potential would seem ideal target cells of HPV infections, as they may provide the opportunity of a longer-lived cellular reservoir for viral replication. Experimental evidence of the persistence of papillomaviral infection in basal cells came from a study by Malglennon et al. of rabbit oral papillomavirus (ROPV)[224]. In this study, the authors used tattoo-marking technique by pin-pricking the rabbit tongue with a mixture of ROPV virions and black tattoo ink in order to permanently mark the experimentally infected sites, with an aim to elucidate the fate of viral genomes following the immune regression of papillomas. Papillomas formed at 75% of the marked sites, continued to grow in size for 4-6 weeks, and started to regress until they were completely resolved by 8 weeks by the immune system. Tissues were taken from the tattoo-marked sites at different time points (from 10-, 15-, 19- and 22-week rabbits) to perform laser microdissection of basal layers followed by real-time PCR analysis. ROPV DNA was detected in the basal cells for up to 22 weeks post-infection (even after immune-mediated regression of induced papillomas), but not in the basal cells from uninfected adjacent sites. The analysis showed that less than one copy of ROPV per cell was detected in the majority of basal cell samples analyzed, meaning not all basal cells that resided in the infected sites were subjected to latent ROPV infection. This has led the authors to hypothesize that there is a small subset of "special" cells that are putative stem cells[224,225]. Another study used transgenic K14-HPV16E6 and K14-HPV16E7 mice expressing E6 or E7 of HPV16 (or both) directed by Keratin 14 promoter in the basal cells. Analysis of these mice showed that expression of the E6 and E7 oncoproteins (either one of the two or both) caused aberrant migration of hair follicle bulge stem cells evidenced by the significant reduction in the number of label-retaining cells and a parallel increase in the number of proliferating cells[226]. By utilizing a different hair follicle stem cell marker, Lgr5, in another transgenic mouse model, da Silva-Diz et al. also demonstrated abnormal migration and expansion of stem cells expressing $Lgr5^{+}$ and basal cytokeratin K15 to the upper region of hair follicles upon expression of E6 and E7, as well as accumulation of these cells at pre-neoplastic lesions and epidermal tumors[227].

There are also data supporting the notion that there is a subset of cells that are the most prone to HPV infection and persistence. For example, in the cervix, the majority of tumors develop in the transformation zone, particularly at the squamo-columnar junction. This is the area where a single columnar layer (overlying endocervix) meets the multilayered squamous cell epithelium (overlying ectocervix)[228]. This observation has led to the hypothesis that areas of transformation from one tissue type to another in the cervix may be preferential targets for malignant transformation by HPV. Recently, Herfs et al. have reported that the reserve cells at this squamo-columnar junction exhibit unique gene expression signatures and protein expression patterns that overlap with high-grade cervical lesions but not with low-grade lesions or adjacent normal cervical epithelium. The authors then proposed that the discreet population of cells at this junction may be the cells of origin of cervical cancer[229]. Considering the observation that tonsillar crypt region is more susceptible to HPV-infection than other regions of the oral cavity, it is conceivable that there may also be a discreet population of cells in the crypt epithelium that are intrinsically more prone to HPV infection that can subsequently serve as tumour-initiating cells.

1.7. Overall goal and specific aims

The structure of epithelium overlying the surface of the tonsils appears similar to oral mucosa where multiple layers are uniformly stratified with a clear demarcation between stroma and epithelium. On the other hand, the crypt epithelium is reticulated and heavily infiltrated with lymphocytes where epithelial cells intermingle with non-epithelial cells. Despite the constant insults from the external environment and accompanying trans-epithelial immunological activities in the tonsillar crypts, the tonsillar epithelium is continuously repaired and regenerated, maintaining the proper architecture and functions. The progenitor cells responsible for this active renewal, however, have not been characterized. Moreover, the continued increase in the incidence of HPV-associated tonsillar cancers, evidence that HPV-induced malignancy begins in the tonsillar crypt, as well as the widely accepted notion that HPV targets stem/progenitor

cells highlights an immediate need to better understand the biology of tonsillar epithelial progenitors, especially those that reside in the crypt epithelium.

A study of a specific subset of cells requires a robust method to purify them for further characterization, currently not available for the tonsil. In this thesis, I sought to develop a method to identify and isolate primitive normal epithelial cells in human tonsils by taking both quantitative and qualitative approaches. This method also enabled us to create a 3D epithelial structure in which to investigate functional and phenotypic deregulation that occurs in such cells as a result of HPV infection.

The specific aims of this dissertation were as follows:

(1) To examine *in situ* immunophenotype of tonsillar surface and crypt epithelial cells by using a panel of selected antibodies to delineate spatial distribution of candidate markers (Chapter 3.1.)

(2) To develop and optimize a quantitative assay to assess human tonsillar epithelial progenitor cell activities (Chapter 3.2.)

(3) To identify the markers that can be used to prospectively isolate cell subpopulations enriched for tonsillar epithelial progenitors (Chapter 3.3.)

(4) To compare the transcriptome profiles of human tonsillar epithelial subsets isolated from tonsillar surface and crypt epithelium (Chapter 3.4)

(5) To examine the effects of HPV16 E6/E7 oncoproteins on the growth of tonsillar epithelial progenitors (Chapter 3.5.)

The following publication arose from this thesis work and is incorporated into the material presented in this thesis.

Kang, S. Y. C., Kannan, N., Zhang, L., Martinez, V., Rosin, M. P. & Eaves, C. J. Characterization of epithelial progenitors in normal human palatine tonsils and their HPV16 E6/E7-induced perturbation. *Stem Cell Reports* **7**, (2015).

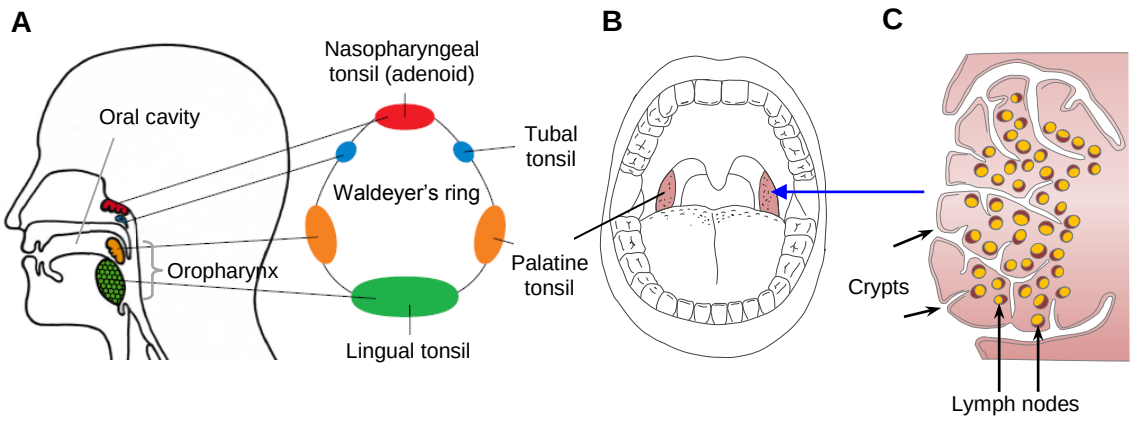


Figure 1-1. Waldeyer's ring and paired palatine tonsils in the oropharynx

(A) Midline sagittal view of the oral cavity and oropharynx, displaying the Waldeyer's ring components including the adenoid on the roof of the nasopharynx, tubal tonsils at the pharyngeal openings to auditory tubes, lingual tonsils and palatine tonsils at the base of tongue (the oropharynx). Modified from "Immunology of the tonsils, Perry, M. and Whyte, A., Immunology Today, 19 (9), 414-421.", Copyright (1998), with permission from Elsevier. (B) Posterior view of the oral cavity and oropharynx with the paired palatine tonsils on either side of the oropharynx. (C) Cross-sectional view of a palatine tonsil with multiple crypts and lymph nodes.

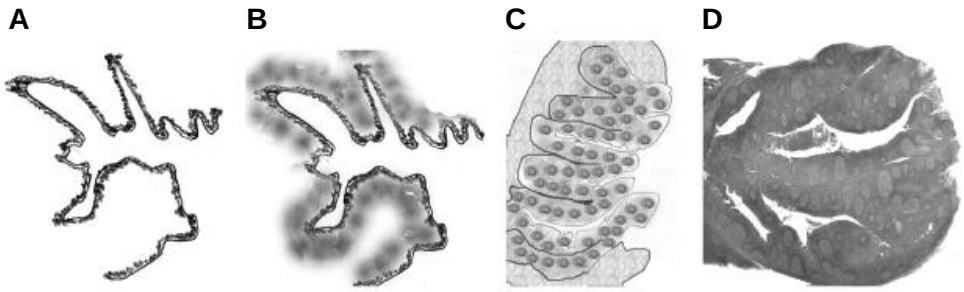


Figure 1-2. Development of the tonsil

(A) Invagination of solid epithelial cores into the surrounding mesenchymal tissue that begins in the 13th embryonic week. (B) Lymphocytic and mesenhchymal infiltration of the connective tissue around the crypt branches. (C) Development of primary germinal centers and secondary follicles upon antigenic stimulations shortly after birth. (D) Hyperplastic tonsil of a child resulting from the subsequent proliferation of germinal centres followed by rapid growth of the tonsil. Reprinted from “Developmental anatomy of the tonsil and its implications for intracapsular tonsillectomy, Isaacson, G. and Parikh, T., International Journal of Pediatric Otorhinolaryngology, 72 (1), 89-96.”, Copyright (2008), with permission from Elsevier.

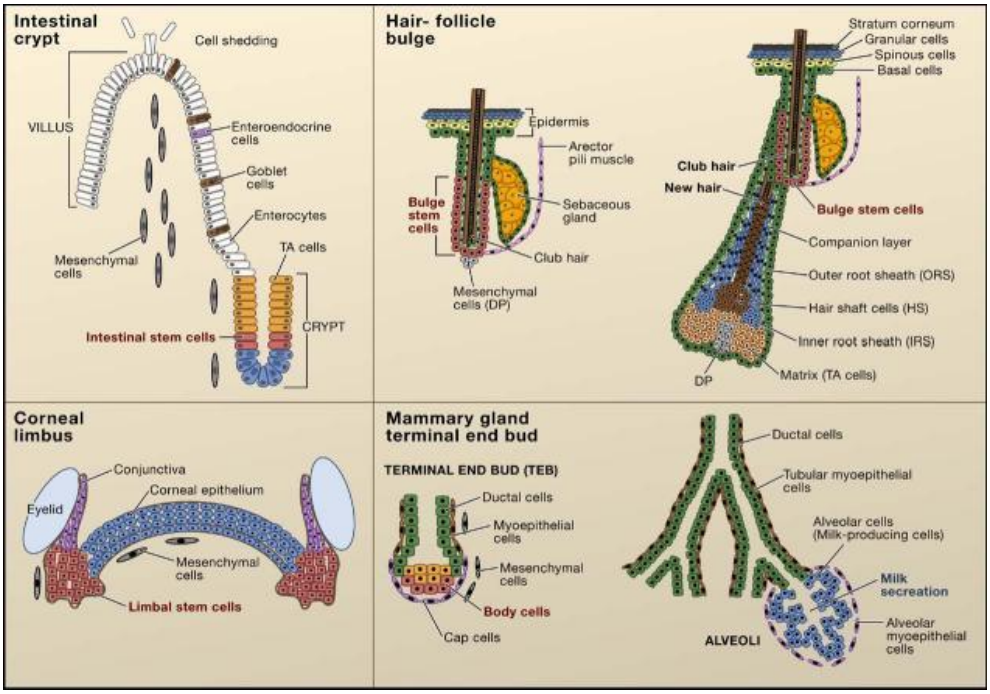


Figure 1-3. Adult epithelial stem cells

Examples of putative stem cells residing in their specialized niches. Reprinted from “Epithelial stem cells: Turning over new leaves, Blanpain, C. et al., Cell, 128 (3), 445-458.”, Copyright (2007), with permission from Elsevier.

Table 1-1. Markers used to identify and purify stem/progenitor cells from other human epithelial tissues

Organ	Cell type	Phenotype	References
Skin	Keratinocyte stem cells	$\alpha6^{bri}CD71^{dim}$	90
	Transit amplifying (TA) cells	$\alpha6^{bri}CD71^{bri}$	90,230
Prostate	Basal cell	$Trop2^{+}CD49f^{hi}$ $CD24^{lo/+}$	231,232
Intestine	Intestinal stem cell	$CD44^{+}CD24^{-/lo}CD166^{+}$	233
Breast	Mammary stem cell	$CD49f^{hi}EpCAM^{-}$	234,235
Lung	Bronchioalveolar stem cell	$Sca\text{-}1^{+}CD34^{+}$	91
Oral mucosa	Oral keratinocyte stem/progenitor cell	$p75NTR^{+}$	137,236

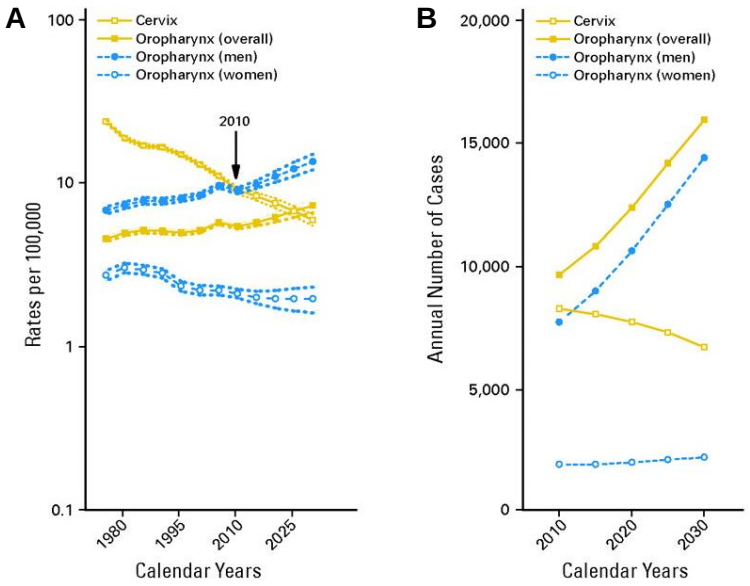


Figure 1-4. Observed and projected incidence and annual number of oropharyngeal and cervical cancer cases in the United States

(A) The incidence of oropharyngeal cancers has continued to increase for the past decades and was expected to have surpassed the incidence of cervical cancers by 2010. (B) The annual number of oropharyngeal carcinoma cases is expected to exceed those of cervical carcinomas by 2020. Modified from Chaturvedi, A. K. et al: Human papillomavirus and rising oropharyngeal cancer incidence in the United States, J Clin Oncol, 29 (32), 2011: 4294-301. “Reprinted with permission. © (2011) American Society of Clinical Oncology. All rights reserved.”

Chapter 2. Methods

2.1. Sample collection

This thesis used tonsil specimens collected at surgery from patients undergoing routine tonsillectomy for recurrent/chronic tonsillitis (n = 45, ages 17-45), and children undergoing tonsillectomy (n = 15, ages 4-13). The study was first introduced to the patients by the surgeon during an office visit prior to day of the surgery, and later described in detail by a study staff prior to consenting. Surgeries occurred at three hospitals (UBC Hospital, Vancouver General Hospital and Surrey Memorial Hospital). Basic demographical information (name, date of birth and gender) and reason for surgery were obtained. This study was approved by the UBC BC Cancer Agency Research Ethics Board.

2.2. Tissue processing

Immediately at surgery, freshly resected tonsil specimens were wrapped in sterile gauze moistened with cold saline and sealed in a sterile plastic container for transport on ice. Samples from patients undergoing surgery at either the UBC Hospital or Vancouver General Hospital delivered first to the Department of Anatomic Pathology at Vancouver General Hospital for removal of tissue for histological evaluation; excess tissue was then transported to the BC Cancer Research Centre (BCCRC) for research. Tonsillectomies that occurred at the Surrey Memorial Hospital (SMH) were from younger patients with no hospital requirement for pathological evaluation. All tissue from that site was directly transported to the BCCRC laboratory where a section was taken by the research laboratory technician for processing for histological evaluation. Healthy tonsils are sites of permanent activation of lymphoid cells; therefore, any non-malignant tonsils

determined by the absence of active pathological lesions were considered “normal” for the purposes of this study.

Processing of residual tissue for research began with a transfer of the tonsil samples into cold complete dissociation medium (CDM) containing DMEM (Life Technologies) supplemented with 5% fetal bovine serum (FBS), 10 µg/ml ciprofloxacin (Life Technologies) and 1:50 dilution of Antibiotic-Antimycotic (100x, Life Technologies). Specimens were thoroughly washed with CDM (more than 3 times) followed by mechanical separation into crypt and surface regions using a scalpel (**Figure 3-3**). Since the surface of the tonsil is smoother in texture than crypt regions, the two regions are easily distinguishable and separable. Tissue from the two regions was then finely minced with scalpels in glass petri dishes followed by enzymatic dissociation in CDM supplemented with 1 mg/ml collagenase (Life Technologies) and 500 µg/ml dispase (STEMCELL Technologies) for 7-12 hours on a rotary shaker at 37ºC. The resultant cell suspension was washed with CDM twice using centrifugation at 1200 rpm for 4 minutes. The pellet was suspended in 1x RBC lysis buffer (BioLegend) and left for 5 minutes on ice in order to remove contaminating RBCs. After centrifugation of the cells again at 1500 rpm for 4 minutes, the pellet was enzymatically dissociated into single cells in 0.05% Trypsin/EDTA and DNase I (Roche) for 4-6 minutes, passed through a 100 µm filter (BD Bioscience) to remove debris, washed with PBS (Life Technologies) containing 2% FBS (2% PF), and centrifuged at 1500 rpm for 4 minutes. This final pellet was resuspended in 2% PF and passed through a 40 µm filter to remove remaining cell aggregates and debris. In some experiments, red blood cells and granulocytes were removed by density centrifugation (Section 2.2.2).

2.3. Cell separation

2.3.1. Magnetic bead-based cell sorting

The cell yield obtained by the dissociation protocol described above varied, ranging from 4 x 10^7 to 4 x 10^8 per pair of palatine tonsils, with 80-90% cell viability based on the dye-exclusion test using trypan blue staining (to distinguish live cells that did not take up the dye from dead cells that did). Since the bulk of freshly dissociated

tonsillar single cell suspensions were heavily contaminated with lymphocytes along with some endothelial cells (Section 3.2.1.), I used a magnetic-assisted cell sorting method (EasySep, STEMCELL Technologies) to remove these cells. The cell concentration was adjusted to 10^8 cells/ml and then mixed with biotinylated antibodies against CD45 at 1:25 and CD31 at 1:50 and incubated for 15 minutes at room temperature (Antibody details are shown in **Table 2-1**). EasySep® Biotin Selection Cocktail was then added at 100 µl/ml cells and samples were incubated for another 15 minutes at room temperature, followed by addition of EasySep® Magnetic Nanoparticles (STEMCELL Technologies) at 50 µg/ml and a further incubation for 10 minutes at room temperature. I then brought the total volume to 2.5 ml by adding 2% PF, gently mixed the cells with a pipette, and let the suspension stand in the magnet at room temperature for 5 minutes to allow the magnetized $CD45^+$ cells and $CD31^+$ cells to bind to the wall of the tube. The unbound cell suspension was then poured off into a new tube. This incubation/pour-off step was repeated twice.

2.3.2. Density centrifugation

An alternate epithelial cell enrichment procedure employed density centrifugation. In this case, enzymatically dissociated tonsil cells were resuspended in 10 ml of 2% PF and slowly layered on top of 15 ml of Lymphoprep (STEMCELL Technologies) prepared in a 50 ml conical tube. The tube was centrifuged at 2000 rpm (Thermo Scientific Sorvall Legend™ RT+ Centrifuge) for 25 minutes at room temperature with the centrifuge brake turned off so as not to disturb the layers formed. Three layers were visible at the end of the centrifugation; the top layer was composed of epithelial cells, the interface of the Lymphoprep solution was composed of a mixture of epithelial cells and lymphocytes, and the bottom layer was composed of a mixture of lymphocytes and red blood cells. The top two layers were carefully transferred to a new 50 ml conical tube using a 10 ml disposable pipette, diluted by adding 20 ml of 2% PF and mixed thoroughly. The cell suspensions were then centrifuged at 1500 rpm for 15 minutes. The cell pellet obtained by using this method allowed for removal of red blood cells and also the depletion of $CD45/CD31^+$ cells approximately by half.

2.3.3. Fluorescence-activated cell sorting (FACS)

Debulked cell suspensions, either by the immunomagnetic depletion or density centifugation method were further processed to remove the remaining endothelial and hematopoietic cells from the analyses or final sorts by flow cytometry (Antibody details are shown in Table 2-1); When immunomagnetic depletion was used, the debulked cells were re-stained with streptavidin-allophycocyanin (SA-APC) or SA-fluorescein isothiocyanate (SA-FITC) to bind all cells labelled with biotinylated antibodies ($CD45^+$ and $CD31^+$). When density centifugation was used, the cell suspensions obtained were labelled with Pacific Blue-conjugated anti-CD45 and anti-CD31. Depending on the experiments, cells were then stained or co-stained with antibodies against individual candidate markers for in order to subfractionate $CD45^-CD31^-$ cells. Cells were then sorted using an Influx II, FACSAria II or FACSAria III (BD Biosciences), and data were analyzed using FlowJo software (FlowJo).

2.4. Immunohistochemical (IHC) analysis

2.4.1. Tonsil tissue macroarray construction

A tissue array was constructed from formalin-fixed paraffin-embedded (FFPE) blocks containing 10 normal tonsillectomy samples. The sample details relating to these tonsils are shown in **Table 2-2**. The process involved the microscopic examination of hematoxylin and eosin H & E) slides of tissue cut from blocks to identify the region that contained both tonsillar surface epithelium and crypts. The selected areas (typically 0.7-0.9 cm in diameter) were marked on each block. The block was then incubated in a 65°C oven to soften the paraffin, followed by excision of the selected areas using scalpels. The excised tissue was re-embedded to create macroarray blocks of 5 cases (hereby referred to as macroarray).

2.4.2. IHC staining

Immunohistochemical (IHC) analysis was performed on 4-5 µm sections cut from the macroarray blocks. These sections were placed on glass slides and the tissue was

deparaffinized by heating slides in a 65°C oven for 2-4 hours followed by immersion in xylene for 5 minutes twice at room temperature. The tissue was then serially hydrated in alcohol solutions at room temperature; 2 rounds of 100% ethanol for 5 minutes each, 2 rounds of 95% ethanol for 2 minutes each, 70% ethanol for 2 minutes, followed by 2 rounds of washing in deionized water for 2 minutes each.

Antigens were retrieved using a heat-induced epitope retrieval method in which the tissue was immersed in Dako Target Retrieval Solution (Dako Cytomation, Burlington, ON) for 20 minutes in a steamer. The slides were then removed from the steamer and cooled down at room temperature for 20 minutes; subsequently, 200 µl of 1X PBS added to the tissue on the slides and allowed to incubate for 3 minutes. This step was repeated twice. Endogenous peroxidase was blocked by using Dako Blocking Buffer for 10 minutes, followed by incubation in 1X PBS for 3 minutes 3 times. Non-specific binding of antibodies was prevented by a further incubation in PBS containing 10% serum (PF) for 15 minutes. Primary antibodies were diluted in Dako Antibody Diluent and tissue was incubated at room temperature for 2-3 hours or at 4°C overnight. Primary antibodies were detected using horseradish peroxidase (HRP)-linked secondary antibody previously diluted and included in the Dako mouse-specific or rabbit-specific EnVision+ Kits (Dako Cytomation, Burlington, ON). Slides were washed 3 times for 5 minutes each in PBS between incubations. HRP activity was detected using Dako Liquid DAB$^+$ Substrate Chromagen System for 15 minutes. Tissue was then washed with deionized water twice for 2 minutes each, and incubated in hematoxylin (Surgipath) for 1.5-2 minutes to visualize nuclei. After washing twice with deionized water for 2 minutes each, tissue was serially dehydrated at room temperature in the following sequence: once in 70% ethanol, twice in 95% ethanol, and twice in 100% ethanol, all for 1 minute each. Tissue was then transferred to xylene and left for 2 minutes before mounting. The slide was coverslipped and stored at room temperature. The antibodies used for IHC are shown in **Table 2-3**.

All of the 10 samples (2 macroarray blocks of 5 individual samples) were examined at 300x magnification. Samples were evaluated for positive or negative staining, its intracellular localization (nuclear, cytoplasmic, membranous), intensity (equivocal, weak, moderate, and strong) and the percentage of positive cells (≤30% cells

stained, score 1; 30 to 50% cells stained, score 2; 50 to 70%, score 3; 70 to 100%, score 4).

2.4.3. Dual p16/Ki-67 IHC

p16/Ki-67 dual staining was performed on organotypic 3D cultures (described in Section 2.7) in order to evaluate HPV16 E6/E7 oncoprotein-induced perturbation..The staining procedure was carried out using the CINtec Plus Kit (Roche). The kit contains a ready-to-use primary antibody cocktail comprised of a mouse monoclonal antibody (clone E6H4) directed toward human p16INK4a protein and a rabbit monoclonal antibody (clone 274-11 AC3) directed against the human Ki-67 protein.

Formalin-fixed, paraffin-embedded blocks of these 3D cultures were cut into 4-5 µm sections, deparaffinized by incubating in a 60°C oven for 50 minutes and subsequently in xylene for 5 minutes 3 times at room temperature. Tissue was then serially hydrated in alcohol solutions at room temperature; 2 rounds of 100% ethanol for 5 minutes each, 2 rounds of 70% ethanol for 5 minutes each, followed by washing in deionized water for 5 minutes. Sections were subjected to a 20-minute heat-induced epitope retrieval in Dako Target Retrieval Solution (Dako Cytomation), followed by a cooldown at room temperature for 25 minutes. The tissue was then washed by incubating in 200 µl of 1X CINtec Wash Buffer for 3 minutes 2 times. Endogenous peroxidase was blocked by using Peroxidase Blocking Reagent included in the Kit for 5 minutes followed by incubation in 1X CINtec Wash Buffer for 5 minutes (also in the kit); the tissue sections were incubated with primary antibodies (anti-p16/Ki-67 cocktail included in the kit) at room temperature for 30 minutes, followed by incubating with 1x CINtec Wash Buffer for 5 minutes. p16/Ki-67 reactivity was detected using their horseradish peroxidase (HRP)-linked secondary antibody previously diluted and included in the CINtec PLUS Kit. Slides were washed 3 times for 5 minutes each in the wash buffer. HRP activity was detected using DAB Substrate Chromagen Solution for 10 minutes. Tissue was then washed once with deionized water and once with the wash buffer, 5 minutes each, and incubated in Fast Red system (Fast Red chromogen in Naphthol Phosphate Substrate solution) for 15 minutes twice with a change in between. Slides were washed twice with deionized water for 5 minutes each, and then incubated

in Mayer's hematoxylin (Ethanol–free) for 1 minute to visualize nuclei. After washing twice with deionized water for 2 minutes each, slides were incubated in 1.5% NaHCO3, followed by the final wash with deionized water for a minute; then a 2-step mounting procedure was applied, first by using an aqueous mounting medium provided with the kit to prevent alcohol-based fading of the Fast Red signal, followed by a permanent mounting step.

Ki-67 expression was used to quantify the proportion of proliferating cells in 3 different compartments of the *in vitro*-generated tonsillar epithelium: the single basal layer, just above the ECM; the two parabasal layers just above the basal layer; and the mid/superficial layers abovethe parabasal layers. The analysis was done on microscopic images taken at 300x magnification to capture the full thickness of the epithelium, typically showing ~40 basal cells per field. At least 5 representative fields per slide were examined.

Dual p16/Ki-67 expression was analyzed to specifically identify E6/E7-induced proliferation. Concomitant expression of p16 and Ki-67 has been proposed to distinguish E6/E7-transformed proliferating cells from uninfected proliferating cells[237] and has recently been shown to be a useful marker for discerning high-risk cervical lesions[238]. If nuclear Ki-67 (red) and cytoplasmic p16 overexpression (brown) were present within the same cell, the cell was designated $p16^{+}Ki\text{-}67^{+}$ (or co-expressing/co-localized). The entire length of the embedded and immunostained epithelium was examined at 300x magnification to identify cells co-expressing p16 and Ki-67. Of all microscopic fields examined, the number of fields that showed $p16^{+}Ki\text{-}67^{+}$ cells were noted and divided by the total number of fields to calculate the percentages of E6/E7-transformed regions in the entire epithelium.

2.5. Immunocytochemistry (ICC)

FACS-purified subpopulations were collected in FACS tubes containing 2% PF and immediately spun down at 1500 rpm for 5 minutes. Cells were resuspended in appropriate volume of 2% PF to adjust the concentration to ~5 x 10^5 cells/ml, of which 200 µl aliquots (1 x 10^5 cells) were individually loaded in each cuvette and spun onto

glass slides at 700 rpm for 3 minutes (Shandon Cytospin 3). The cytocentrifuged cells were dried at room temperature, and fixed in a 1:1 vol/vol mixture of methanol and acetone for 15 minutes at -20ºC. The fixed cells were either immediately stained for intracellular proteins as described in Section 2.1.1, or kept at -80°C until used later.

2.6. 2D colony-forming cell (CFC) assay

Test cells combined with 10^5 irradiated (50 Gy of 300kVp X-rays) mouse embryo fibroblast NIH 3T3 cells and seeded in each well of a 6-well plate containing Tonsil Epithelial Cell Medium (TEpiCM; ScienCell) supplemented with 5% FBS were incubated at 37ºC in 5% CO_2 and 20% O_2. 3,400 irradiated NIH 3T3 cells were added when 96-well plates were used to keep the cell density consistent. After 2 days, the medium was changed to serum-free TEpiCM. Optimization of the CFC assay culture conditions was one of the objectives of this thesis and is described in greater detail in Section 3.2. Thus, in the later stages of the study, we also supplemented the medium with 10 µM ROCK inhibitor (Y-27632, Cellagen Technology) and cultured cells in 5% O_2 instead of 20% After 7-10 days, the medium was removed and 2 ml of a 1:1 vol/vol mixture of methanol and acetone was added. Plates were stored in a -20ºC freezer for 15 minutes to allow the cells to be fixed. The fixed colonies were immediately stained with Giemsa or hematoxylin for scoring, or with various antibodies for IHC analysis as described above. Remaining colonies were kept in PBS at 4ºC for a maximum of 4 weeks until used for IHC.

As part of initial culture optimization study, effects of irradiated fibroblasts obtained from human tonsils on growth of tonsillar epithelial cells were tested in parallel with murine 3T3 fibroblasts. Since human tonsillar fibroblasts did not perform as well as the murine 3T3 cells (lower CFC plating efficiencies), murine fibroblasts were used in all subsequent experiments. Inhibitory effects of human fibroblasts have been reported previously[239].

2.7. 3D organotypic culture system

A 3D collagen matrix consisting of a bottom acellular layer and a top cellular layer was prepared by mixing a rat tail collagen I solution (BD Bioscience) with DMEM (Life Technologies) and reconstitution buffer, FBS and NaOH to give a 75% collagen solution in DMEM containing 5% FBS. To create the acellular layer, 1 ml of the collagen solution was added into each insert in a 6-well transwell plate (Corning) and placed in a 37°C incubator until the collagen solidified. To create the cellular layer, mouse fibroblasts were added to the remaining collagen solution to give a final concentration of 5,000 cells/ml, 3 ml of which were added to each well already containing a solidified acellular layer. These collagen matrices were allowed to contract for 5-7 days in DMEM supplemented by 10% FBS at 37°C, at which point freshly dissociated human tonsillar epithelial cells were seeded on top of the matrices. For the first 4 days following addition of test cells, cultures were fed with Epithelialization (Ep 1) medium containing DMEM:Ham's F12 (3:1) supplemented with 180 μM adenine, 5 μg/ml insulin, 5 μg/ml transferrin, 20 pM T3, 0.4 μg/ml hydrocortisone and 0.2% FBS. For the following 2-3 days, the medium was replaced with Ep 2 medium, which was identical to Ep 1 medium except that it contained 2% FBS instead of 0.2% FBS. The medium was then removed from the inserts to create an air-liquid interface and cultured in fresh Ep 2 medium for 10-12 days, after which cultures were harvested by fixing in 10% neutralized buffered formalin (Fisher Scientific) and paraffin-embedded.

2.8. Transcriptome analysis

Total RNA was extracted from freshly sorted tonsillar epithelial subsets using AllPrep DNA/RNA Micro Kit (Qiagen) according to the manufacturer's protocol and subsequently quantified by Nanodrop to obtain their concentrations. The details of the donors including their age, gender and clinical presentation are shown in Figure 3-17B.

RNA quality assessment, quality control and gene expression profiling were performed at The Centre for Applied Genomics (Toronto, ON) as follows; The RNA quality was assessed with the Agilent Bioanalyzer 2100 (Agilent Technologies) using the Agilent RNA 6000 Nanokit or Picokit (Agilent Technologies), and samples with an RNA

integrity number of 7.0 or higher were selected for gene expression profiling. 70 ng of RNA was labeled with TotalTM-96 RNA Amplification Kit (Ambion), and complementary RNA hybridized to Illumina HumanHT-12 V4 BeadChip (Illumina Inc., San Diego, CA, U.S.A.) according to the manufacturer's specifications for the “Whole-Genome Gene Expression Direct Hybridization Assay Workflow”. RNA was amplified using the Ambion Illumina TotalTM-96 RNA Amplification Kit and biotinylated using the Biotin-NTP Mix. Samples were incubated for 18 hours at 58°C for hybridization. Slides were subsequently washed and scanned using the Illumina BeadArray Reader in order to obtain raw image and data files. Raw signals were processed with GenomeStudio (Illumina Inc., version V20011.1) and further analyzed with Partek Genome Suite (Partek Inc., St. Louis, MO).

After quality control filtering and quantile normalization, 47,320 probes were included for further analyses. Unsupervised hierarchical clustering was performed on log2-transformed data using Gene-E (http://www.broadinstitute.org/cancer/software/GENE-E/index.html), while principal component analysis (PCA) was performed in log2-transformed data using R Statistical Software version 3.1.2, using the “prcomp” function[240]. Microarray data are deposited in the NCBI GEO repository (accession number: GSE69767).

2.9. Lentiviral preparations and transduction

Human papillomavirus E6 and E7 oncogenes were cloned from pLXSN-HPV16-E6/E7 (addgene plasmid ID: 52394; Ref. 198) into pLKO.1 YFP lentiviral vector (Lenti-HPV16-E6/E7-YFP) and sequence verified. Lentiviral particles were packaged using the human embryonic kidney 293T cell line, transiently transfected with vectors encoding VSV-G (envelope protein), POL (reverse transcriptase and integrase enzymes), and GAG (capsid protein) as previously described[241]. Following an initial media change at 24 hours post-transfection, the media containing active lentiviral particles was harvested at 48 and 72 hours post-transfection, filtered through a 0.45 µm mesh, and stored at -70°C.

Freshly sorted cells were directly plated onto a 96-well plate at a density of 1000 cells per well and cultured in TEpiCM media supplemented with 5% FBS, 10 µm ROCK

inhibitor and irradiated fibroblasts at 37°C, 20% CO_2 and 20% O_2 humidified chamber for 3-5 days until 70-80% confluency was reached. Lentviral transduction was performed in this liquid suspension cultures by adding lentiviral supernatant (either Lenti-HPV16-E6/E7-YFP or Lenti-mCherry) and incubating the cells at 37°C and 20% O_2 for 3 hours, followed by washing with serum-free TEpiCM. Subsequently, fresh serum-free TEpiCM supplemented with 10 µM ROCK inhibitor was added to the tissue culture plate and cells were incubated for 4-5 days in the same medium (until ~80% confluency has been reached). Stable expression of oncogenes in the transduced cells was verified by PCR using the following primers. Forward GAACCGGACAGAGCCCATTA; Reverse TCTGAGAACAGATGGGGCAC.

2.10. Statistical analysis

GraphPad Prism 6 software was used for generation of all graphs and statistical analyses. Numerical data are shown as mean ± SEM or ± SD. P values were generated using the parametric Student's t-test or one-way ANOVA with Bonferroni's multiple comparison test.

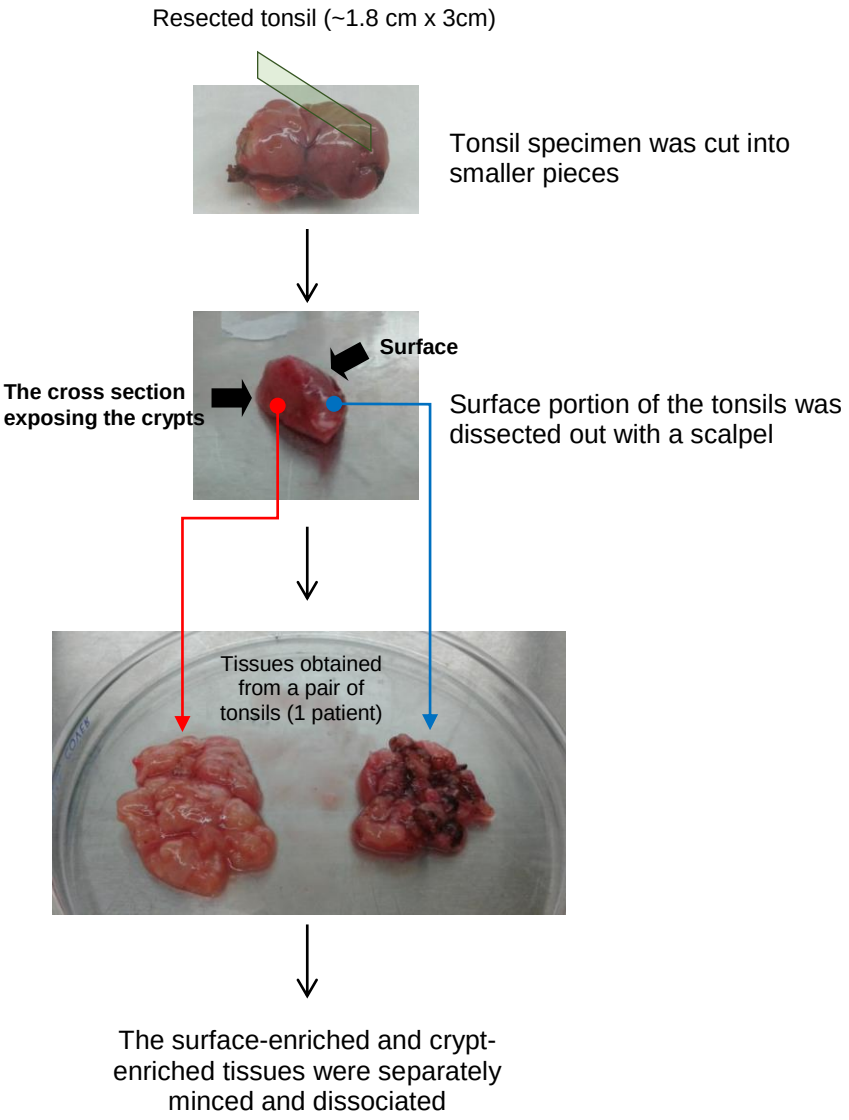


Figure 2-1. Separation of tonsillar crypt-enriched tissues from the surface epithelium

Table 2-1 Antibodies used for flow cytometry

Antibody target	Conjugate	Clone	Isotype	Supplier
CD45	Biotin	HI30	Mouse IgG1	BD Pharmingen
CD45	Pacific Blue	RA3-6B2	Rat IgG2a	BD Pharmingen
CD31	Biotin	WM59	Mouse IgG1	BD Pharmingen
CD31	Pacific Blue	WM59	Mouse IgG1	BioLegend
CD19	APC	HIB19	Mouse IgG1	eBiosciences
CD3	PE	SK7	Mouse IgG1	eBiosciences
CD44	PE	515	Mouse IgG1	BD Pharmingen
CD271(NGFR)	APC	ME20.4-1.H4	Mouse IgG1	Miltenyi Biotec
MUC1	Unconjugated	214D4	Mouse IgG1	STEMCELL
streptavidin	FITC	N/A	N/A	BD Pharmingen
streptavidin	APC	N/A	N/A	BD Pharmingen

Table 2-2. Samples used for tonsil macroarray IHC analysis

Sample ID	Age (years)	Sex[1]	Clinical presentation
12012	19	F	Recurrent & Chronic Tonsillitis
12013	23	M	Recurrent & Chronic Tonsillitis
12015	25	F	Chronic Tonsillitis
12016	41	M	Recurrent & Chronic Tonsillitis
12026	16	F	Nonspecific reactive lymphoid follicular hyperplasia
12032	30	F	Recurrent Tonsillitis
12034	26	F	Chronic Follicular Tonsillitis
12036	33	F	Recurrent & Chronic Tonsillitis
12037	17	M	Recurrent & Chronic Tonsillitis
12044	25	M	Recurrent Tonsillitis

[1] F, Female; M, Male

Table 2-3. Antibodies used for immunohistochemistry

Antibody target	Isotype	Clone	Supplier	Concentration[1]
CD44	Mouse IgG1	515	BD Pharmingen	1:100
NGFR	Mouse IgG1	ME20.4-1.H4	Miltenyi Biotec	1:25
MUC1	Mouse IgG1	HMPV	BD Pharmingen	1:100
CK4	Mouse IgG1	6B10	Santa Cruz	1:25
CK5	Mouse IgG1	3E2F1	Santa Cruz	1:200 (IHC)
CK5	Rabbit polyclonal	H-40	Santa Cruz	1:50 (ICC)
CK13	Rabbit IgG	EPR3671	Novus	1:100
CK14	Mouse IgG	SP53	Cell Marque	1:400
CK8/18	Mouse IgG1	4546	Cell Signaling	1:25
CK19	Mouse IgG	A53-B/A2-26	Cell Marque	1:50
CK19	Mouse IgG1	RCK108	Santa Cruz	1:40 (ICC)
Involucrin	Mouse IgG1	SPM259	Santa Cruz	1:50 (IHC) 1:20 (ICC)
Ki-67	Mouse IgG1	MIB-1	Dako	1:100
p63	Mouse IgG2a	4A4	Santa Cruz	1:50
Collagen IV	Mouse IgG1	CIV22	Dako	1:25

[1] IHC, Immunohistochemistry; ICC, Immunocytochemistry

Chapter 3. Results

3.1. *In situ* immunohistological phenotype of human tonsillar surface and crypt epithelial

The tonsillar morphology, structure, embryonic development and ontology, both in humans and in animals, have been thoroughly described in the past centuries[25,242–245]. However, a relatively fewer number of studies have examined and reported the immunohistochemical phenotypes of human palatine tonsillar epithelium. Therefore, the specific aims of this section were to identify the proteins that distinguish various types of epithelial cells in the human palatine tonsil based on their differentiation status and to compare the expression patterns seen in the surface with those in the crypt epithelium.

Candidate markers to examine were chosen to include many known to be expressed on different layers within epithelial tissues (i.e., markers of basal, parabasal, intermediate and superficial layers) others with different subcellular distributions (i.e. nuclear, cytoplasmic and membranous). The former were anticipated to be useful for describing cellular hierarchical changes within the tissue, with the basally located cells considered as the least differentiated and capable of most proliferative activity, and the superficially located cells the most differentiated. The subcellular localization was another feature considered as any marker protein to be used for prospective viable cell sorting must contain extracellular domains (membranous expression) that can be recognized by antibodies without having to penetrate the cell membrane. Three candidate markers, CD44, NGFR and MUC1 were selected for this purpose. CD44 is a well-established marker of basally located cells in many epithelia[246–249]. NGFR (also referred to as p75[NTR] or CD271) has been reported to characterize esophageal keratinocyte stem cells[250] and shows restricted expression in the basal layers of oral mucosal epithelium[236]. MUC1 is a membrane glycoprotein expressed on the apical surface of a variety of mucosal epithelia such as those of the gastrointestinal,

respiratory, reproductive, and urinary tracts as well as the surface of the eye[251], and plays a key role in lubrication and hydration of cells as a protection from microbial insults and digestive enzymes[252].

Evaluation of immunoreactivity was performed as described in Section 2.3.2 and the results are summarized in **Tables 3-1** (surface epithelium) and **3-2** (crypt epithelium).

3.1.1. Basement membrane

H & E stained images clearly showed contrasting morphologies between the highly stratified surface epithelium and irregularly structured crypt epithelia (**Figure 3-1A-1C**). IHC staining of different regions of tonsillar tissue in formalin-fixed paraffin-embedded samples revealed a continuous collagen IV-positive basement membrane separating the surface epithelium from the underlying stroma (**Figure 3-1D and 1E**). In contrast, in the crypt, collagen IV staining was detectable yet weak and discontinuously distributed (**Figure 3-1E and 1F**).

3.1.2. Cytokeratins

Cytokeratin-14 (CK14) was strongly stained in cells throughout the surface epithelial layer and on putative epithelial cells in the crypt (**Figure 3-2A**) suggesting its likely utility as a pan-tonsillar epithelial marker. CK19, CK4, CK13 and involucrin were all expressed but variably in the surface epithelium where their location could be associated with an increasingly differentiated state from the basal layer to the outer surface. In the surface epithelium, CK19 was mostly restricted to the basal layer and infrequently expressed in the mid layers (**Figure 3-2A**), however, in the crypt region, it was moderately or strongly expressed throughout the full thickness of the epithelium (**Figure 3-2A**). CK13 and CK4 were detected only in the parabasal, mid and upper layers and not in any basal cells regardless of the site (**Figure 3-2B**). The staining intensity, however, was overall much stronger in the crypt epithelium compared with their surface counterparts (**Table 3-2B**).

3.1.3. Cell surface antigens

In the epithelium overlying the tonsillar surface, CD44 was most strongly expressed in the basal and parabasal layers (lower 1/3), was moderately positive in the intermediate (mid) layers, and absent in the superficial (upper) layers (**Figure 3-1A**). NGFR, on the other hand, was exclusively expressed in a single basal layer adjacent to the basement membrane overlying the stromal layer (**Figure 3-1A**). MUC1 was either absent or very weakly positive in the basal layer, with a gradual increase in intensity from parabasal layer to the uppermost layers (**Figure 3-1B**). In the crypt epithelium, while their association with a decreasing differentiated state (CD44, NGFR) or an increasing differentiation state (MUC1) remained similar, positively stained cells were more widely distributed in the crypt epithelium (**Table 3-1 and 3-2**). A difference in subcellular localization of CD44 and NGFR was noted in the surface and crypt epithelia, where both CD44 and NGFR were strictly membranous in the surface epithelium but frequently expressed in the cytoplasm of the crypt epithelial cells (**Figure 3-2A**). MUC1 expression was both membranous and cytoplasmic in both sites (**Figure 3-2A**).

3.1.4. Other markers

The transcription factor, p63, a widely annotated marker of basal epithelial cells[98,99,110], was seen not only in the basal layers but also in the parabasal layers extending up to the mid-layers of the surface epithelium (**Figure 3-2A**). Cells expressing the proliferation marker, Ki-67, were also absent from the basal layer of the surface epithelium and found only in cells in the overlying 3 to 5 parabasal layers. In the crypt regions, all of the same markers were expressed in some cells but these appeared more randomly organized due to their dispersed distribution (**Figure 3-2B**).

In summary, the *in situ* IHC examination of tonsillar epithelium identified the spatial distribution of the candidate protein markers according to their differentiation status. The crypt epithelial cells displayed a much wider spectrum of expression profile compared with their surface counterparts, with multiple markers (both basal and differentiation markers) expressed by the majority of crypt epithelial cells. Despite this distinction, the intensity of CD44 and NGFR remained the strongest in the basal layer regardless of the site, suggesting that CD44 and NGFR might be useful to isolate basal

cells from both surface and crypt epithelia. This possibility has been further explored and discussed in detail in Section 3.3.

3.2. Development and optimization of a quantitative *in vitro* assay for human epithelial stem/progenitor cells

The aim of the experiments described in this section was to develop and optimize a functional assay to detect and quantify primitive cells in dissociated suspensions of human tonsillar epithelial cells. Because the goal was to develop an assay that can be reliably used to examine biological properties of individual cells, it was imperative that viable cells be obtained. The tonsil cell dissociation protocol used was adapted from protocols developed for other epithelial tissue models, with a few minor modifications, including the establishment of the protocol for mechanical separation of the surface portion of a resected tonsil specimen from the tonsillar crypt-enriched regions (Section 2.2).

The following sections describe the stepwise process used to optimize conditions for functional assay for this new tissue site, involving (1) a depletion of contaminating lymphocytes and endothelial cells by using a magnetic beads-based cell isolation method (EasySep; Section 3.2.1), and (2) the testing of exogenous factors to determine the best conditions for detecting and quantifying tonsillar epithelial cells that can give rise to a cluster of >50 cells (i.e. a colony) when seeded at clonal densities in 2D *in vitro* culture (Section 3.2.2). The exogenous factors tested include irradiated murine fibroblasts (NIH 3T3) as feeder cells, Rho-associated protein kinases (ROCK) inhibitor (Y27632), and different atmospheric oxygen levels (normal 20% O_2 versus reduced 5% O_2). These factors were tested independent of one another as well as in combination. Under these *in vitro* culture conditions, different types of tonsillar epithelial colonies were observed, displaying varying morphologies and immunophenotypes (Section 3.2.3)

3.2.1. Depletion of contaminating lymphocytes and endothelial cells by EasySep enriches for tonsillar CFCs

The minced tissues (either separated into surface-enriched and crypt-enriched portion or unseparated) were enzymatically dissociated in cell dissociation medium containing collagenase and dispase, followed by trypsin and DNase I as described in the Methods section (Section 2.2). Single-cell suspensions were then subjected to the a magnetic bead-based cell isolation method designed to remove human B-cells and analyzed by flow cytometry to determine the enrichment of epithelial cells achieved.

Flow cytometry analysis of freshly dissociated cells revealed that the majority of cells were $CD45^+$ (hematopoietic) and $CD31^+$ (endothelial) cells (~92%; **Table 3-3**). When these cells were plated in 2D culture at various densities, very few cells attached and formed colonies (~0.1% plating efficiencies; **Figure 3-3C**). Since such a low plating efficiency made it impractical to use bulk cells for the purpose of testing different culture conditions, I used magnetic bead-based sorting (EasySep biotin-negative selection, Section 2.2.1) to remove the majority of $CD45/CD31^+$ cells (**Figure 3-3A**). This procedure increased the purity of $CD45^-CD31^-$ cells up to ~95% (**Figure 3-3B**), and plating efficiencies by up to 13-fold (CFC frequencies ranged from 0.02% to 0.3% for bulk, and from 0.1% to 0.7% for the debulked; **Figure 3-3C**). The improvement in CFC frequencies was more pronounced in the crypt-derived single cells compared with those dissociated from the surface-enriched tissues (**Figure 3-3D**).

3.2.2. Identification of an optimized condition for detecting colony-forming tonsillar epithelial cells

Irradiated NIH 3T3 fibroblasts are widely used in epithelial culture systems for their enhancing effects on colony growth, and inhibition of rho-associated kinase (ROCK) by ROCK inhibitor (Y-27632) has been shown to improve colony-forming cell efficiencies of epithelial progenitor cells when seeded at clonal densities (discussed in Section 1.4.2). Therefore, I hypothesized that feeders and ROCK inhibitor might exert similar effects on the growth of primary human tonsil epithelial cells *in vitro*. Additionally, different oxygen conditions were tested as reduced oxygen tension has been reported to

improve efficiency of colony formation[152,253,254]. The schematic diagram of the experimental design is shown in **Figure 3-4**.

In 20% O_2, the addition of NIH 3T3 fibroblasts (i3T3) alone or ROCK inhibitor alone each enhanced plating efficiencies of tonsillar epithelial cells by ~4.4-fold, and ~4.8-fold, respectively (**Figure 3-5A**). In 5% O_2, each of the two factors contributed to a ~6.2-fold and a ~3.8-fold increase in the detection of tonsillar CFCs, respectively. When combined, i3T3 and ROCK inhibitor enhanced plating efficiencies by ~8.3-fold in a 20% O_2 and ~11-fold in a 5% O_2 conditions (P<0.0001) (**Figure 3-5A**) and a positive linear relationship was reproducibly obtained between the number of cells seeded per well and the number of colonies generated (**Figure 3-5C**), thus validating the use of this assay to quantify CFC frequencies. Of note, the use of reduced atmospheric oxygen did not have an enhancing effect on plating efficiencies in the absence of irradiated fibroblasts, irrespective of the presence of ROCK inhibitor (**Figure 3-5A**).

Overall, an enrichment of human tonsillar epithelial cells by magnetic bead-assisted depletion of CD45/CD31$^+$ cells followed by culturing in the optimized conditions (including irradiated NIH 3T3 fibroblasts, 10 µM ROCK inhibitor and 5% O_2) allowed for enhanced detection of clonogenic cells by over 14-fold when compared to the initial mean CFC frequency obtained from unprocessed cells (i.e., total bulk cells in suboptimal conditions, 0.11 ± 0.04%; debulked cells in suboptimal conditions, 0.37 ± 0.07%; debulked cells in the optimized condition 1.5 ± 0.3%) (**Figure 3-5B**).

3.2.3. Characterization of colonies generated in 2D cultures

When debulked tonsillar cells were seeded at low densities and then cultured for 7-9 days, a spectrum of colony sizes and morphologies were observed. These could be categorized into the 3 types previously described for epidermal colonies on the basis of colony shape and differentiation status of the cells in the colonies[126]. These are referred to as holoclones (tightly clustered large colonies with a smooth perimeter in which less than 5% of cells express terminally differentiated features), paraclones (highly irregularly shaped colonies exhibiting a dispersed cell phenotype with a terminally differentiated

appearance) and meroclones (colonies of intermediate size and mixed composition) (**Figure 3-6**).

The differentiation status of cells in these colonies was confirmed with ICC analysis using basal cell and differentiation markers. Most of the cells within holoclones were strongly positive for CK19 and p63, markers of basal/parabasal cells, with little evidence of the differentiation markers, involucrin, CK8/18 and CK13. In contrast, paraclones were negative for CK19 and p63, and generally positive for the differentiation markers. Meroclones showed a heterogeneous staining pattern with varying proportions of positive cells for all markers examined (**Figure 3-6**).

3.3. Functional and phenotypic characterization of the human tonsillar epithelial progenitor cells

The objective of the experiments described in this section was to evaluate potential markers of human tonsillar epithelial progenitors that could be used for their identification and isolation at sufficient purities as a basis for subsequent functional and molecular analyses. Prospective purification of heterogeneous cell subpopulations to homogeneity allows for more accurate interpretation of the biological and molecular profiles of specific cells of interest. FACS exploits the differential expression of antigens (also known as “markers”) on different subsets of cells so that different fluorochrome-labelled antibodies can be used to isolate these. When the cells are to be subsequently assessed in growth assays, only antibodies that detect markers on the surface of viable (unfixed) cells can be used. Second, the type of fluorochorome to couple to each antibody has to be carefully selected especially when co-labeling cells with two or more antibodies simultaneously to minimize spectral overlap and optimize fluorescence intensities.

A key question was whether subpopulations located in the surface epithelium of the tonsil would differ functionally and/or phenotypically from those located in the crypt epithelium. The *in situ* IHC analysis described in Section 3.1 revealed that while the surface epithelium displays a consistently uniform distribution of protein markers, the

same markers were distributed in a dispersed fashion across the full thickness of crypt epithelium.

Candidate markers that were tested (CD44, NGFR and MUC1) were selected based on the data obtained from the previous IHC analysis. The data showed that CD44 and NGFR were strongly expressed in basally located cells in the tonsillar surface epithelium with their expression patterns inversely correlated with that of MUC1, which was strongly expressed in the intermediate and superficial layers (**Figure 3-2A**). Prospectively isolated tonsillar subpopulations defined by the marker(s) were subjected to 2D CFC assays and 3D organotypic culture system to assess their progenitor activities, and also examined by IHC to characterize their intracellular immunophenotypes.

3.3.1. Crypt- and surface epithelial progenitors have dissimilar intracellular immunophenotypes

Resected tonsils were physically dissected to separate surface epithelium from the crypt regions and independently dissociated as described in Section 2.1. The majority of hematopoietic ($CD45^+$) and endothelial ($CD31^+$) cells were then removed immunomagnetically, and the negatively selected cells co-stained with fluorochrome-labeled antibodies to CD45 and CD31, as well as antibodies to CD44 and NGFR for multiparameter flow cytometry analysis (**Figure 3-7A**). Dead cells ($DAPI^+$), doublets, as well as any remaining $CD45^+/CD31^+$ cells were further depleted by FACS, and the resultant $CD45^-CD31^-$ cells were subfractionated into 4 subsets according to their levels of surface CD44 and NGFR expression (**Figure 3-7B**). Each of these 4 fractions was isolated at >90% purity by FACS and analyzed for expression of intracellular markers in fixed cytospin preparations (**Figure 3-7A**).

Analysis of the FACS profiles revealed that most of the $CD45^-CD31^-$ cells from both sites were distributed among 3 subsets: $CD44^-NGFR^-$, $CD44^+NGFR^-$ and $CD44^+NGFR^+$ cells, with very few $CD44^-NGFR^+$ cells detected (**Figure 3-7B**; Data pooled from 10 independent FACS profiles are shown in **Table 3-6** under '% of total cells'). The analysis showed that the $CD44^+NGFR^+$ subpopulation isolated from both sites was the most highly enriched for cells strongly expressing CK19, CK5 and p63

(markers of the basal and mid-layers of the surface epithelium) (**Figure 3-8**). In contrast, fewer $CD44^+NGFR^+$ cells from either site were positive for involucrin or CK4, both of which appear as differentiation markers in the surface epithelium (**Figure 3-8**). The proportion of involucrin-positive or $CK4^+$ cells varied significantly in a site-dependent manner; 20% and 28% of $CD44^+NGFR^+$ cells in the surface epithelium expressed involucrin and CK4, respectively (**Table 3-4**), with much higher proportions of these (69% and 86%, respectively) in the $CD44^+NGFR^+$ crypt cells (**Table 3-4**). $CD44^+NGFR^-$ and $CD44^-NGFR^+$ subpopulations expressed lower levels of CK5, CK19 and p63. In the $CD44^-NGFR^+$ cells, p63 expression was detected in the cytoplasm and appeared absent in the nucleus. $CD44^-NGFR^-$ cells showed low expression of the basal markers and strong expression of the markers of differentiated cells. These findings suggest that phenotypically similar subsets of epithelial cells are present on the surface and in the crypts of the human tonsil, with some variability in their distribution.

3.3.2. Human tonsillar epithelial progenitor cells are $CD44^+NGFR^+$

I then asked whether CFCs are preferentially enriched in a phenotypically distinct subset defined by their surface CD44 and NGFR expression. Subfractionation of the $CD45^-CD31^-$ cells based on their level of expression of CD44 alone yielded 4 subsets each comprising ~25% of the total $CD45^-CD31^-$ fraction (**Figure 3-9A**). Additional subsets (R5 and R6) were created by applying more stringent gates (by excluding $NGFR^-$ fractions from R4) (**Figure 3-9B**). To analyze the frequency and content of CFCs based on NGFR expression, I then subfractionated the same $CD45^-CD31^-$ subpopulations into 4 subsets based on the level of expression of NGFR alone, and created additional subsets (R5 and R6) by excluding $CD44^-$ cells from $NGFR^+$ subsets (**Figure 3-9D and E**). Separate aliquots from each of the 12 sorted fractions was seeded into CFC assays at seeding densities ranging from 200 to 1000 cells per well of a 6-well plate using optimized assay conditions (irradiated fibroblasts, ROCK inhibitor and 5% O_2). Colonies were enumerated at the end of 7-9 days.

The results showed that both CD44 and NGFR expression are positively correlated with the CFC frequencies (**Table 3-5, Figures 3-9C and F**). Of the 8 single-marker-defined fractions (4 CD44-defined and 4 NGFR-defined fractions), the $NGFR^{high}$

subset was most enriched in CFCs for both anatomic sites (22% ± 2%, n=4, surface and crypt data pooled). Exclusion of $CD44^-$ cells from the $NGFR^{high}$ subset further increased the CFC frequency (27% ± 2%, n=4, surface and crypt data pooled).

An analysis of the distributions (yield) of CFCs among the CD44- or NGFR-defined subsets revealed that more than 90% of the CFCs were confined to the $CD44^+$ subsets (R3 + R4) for both surface and crypt derived samples (**Figure 3-9G**) and over 95% of the total CFCs were contained in the $NGFR^+$ subsets, R3 and R4 (**Figure 3-9H**). Both surface- and crypt-derived subsets displayed similar enrichment and distributions of CFCs, with an exception to the $NGFR^{high}$ and $NGFR^{low}$ subsets; surface-derived CFCs appeared to be more highly concentrated in the $NGFR^{high}$ region than in the $NGFR^{low}$ region (67% and 27%, respectively, n=2, $P<0.05$), whereas the CFC distributions between crypt-derived $NGFH^{high}$ and $NGFR^{low}$ subsets were not significantly different (57% and 37%, respectively, n=2, $P>0.05$, **Figure 3-9H**).

These combined data demonstrate that CFCs in the human tonsillar epithelial cells share a $CD44^+NGFR^+$ phenotype. The CFC-enriched fraction ($CD45^-CD31^-CD44^+NGFR^+$) will be hereafter referred to as a tonsillar progenitor fraction.

3.3.3. Both crypt- and surface-derived $CD44^+NGFR^+$ cells regenerate stratified *in vitro*

To further compare tonsillar surface- and crypt-derived progenitors in terms of their ability not only to proliferate (colony formation) but also to differentiate into multilayered epithelium, I isolated 4 matched pairs of surface- and crypt-enriched cell samples in 4 independent experiments and subjected the sorted subsets to both CFC assays and the organotypic 3D culture system as illustrated in **Figure 3-10**.

Colony yields from the purified cells from both sites remained linearly related to the number of cells plated per well (**Figure 3-11A**), with the highest frequencies of CFCs measured in the $CD44^+NGFR^+$ fraction from both sites (14% ± 4% and 21% ± 4% CFCs for the surface and crypt, respectively; **Figures 3-11C, Table 3-6**). This subset from both sites also contained the majority of the CFCs in that region (64% ± 10% and 72% ± 3% for surface and crypt samples, respectively (**Figures 3-11D, Table 3-6**).

The growth and differentiation activities of FACS-purified $CD44^+NGFR^+$ cells were assessed in organotypic 3D cultures. In this 3D system, dissociated suspensions of cells are seeded on top of an extracellular matrix (ECM) layer containing collagen I and fibroblasts and incubated at an air-liquid interface (**Figure 3-10**). Unseparated $CD45^-$ $CD31^-$ cells from the crypt produced a patchy monolayer in such cultures (initiated with 1.25 x 10^5 cells/culture; **Figure 3-12**, left panel), whereas the same number of $CD44^+NGFR^+$ crypt cells formed a complete sheet of multilayered epithelium (**Figure 3-12,** right panel), with a similar result obtained in cultures initiated with purified $CD44^+NGFR^+$ surface epithelial cells. IHC analysis of the cells in the 3D structures produced by epithelial cells from either tonsillar region indicated that the expression pattern of CD44, NGFR, CK19, Ki67, MUC1, involucrin, CK8/18, and CK13 recapitulated the immunophenotype typical of the surface epithelium of the tonsil (**Figure 3-12**, right panel). These findings extend the results of the 2D assays, indicating that the majority of primitive tonsillar epithelial cells have a $CD44^+NGFR^+$ phenotype and that they are highly enriched in that subset.

3.3.4. MUC1 expression is not exclusive to differentiated cells

Since the cells in the basal layer strongly expressed CD44 but lacked MUC1 expression (**Figure 3-2**), I hypothesized that tonsillar progenitors are $MUC1^-$. To investigate this hypothesis, I carried out CFC assays of FACS-sorted subpopulations based on cell surface expression of CD44 and MUC1. Typically, for immunostaining for FACS analysis, the first step was to remove hematopoietic cells and endothelial cells by using EasySep to reduce the proportion of unwanted cells before staining them with antibodies. However, since a directly conjugated anti-MUC1 antibody was unavailable and a fluorochorme (APC)-conjugated secondary antibody was required, it was critical to complete MUC1 labeling (both primary and secondary) before introducing any other primary antibodies to cell suspensions for EasySep (biotinylated anti-CD45 and -CD31) and for labeling anti-CD44 in order to prevent non-specific binding of the secondary APC to other primary antibodies. After MUC1 staining was completed, cells were then magnetically debulked, stained with a viability dye (PI or DAPI), FITC-conjugated antibody to $CD45^+/CD31^+$ cells and PE-conjugated anti-CD44 to label CD44. CFC data were generated from 3 independent experiments (**Figure 3-13**).

Distinct subpopulations defined by CD44 and MUC1 expression were seen in all 4 quadrants of the FACS profile of CD45/CD31$^-$ fraction of either surface or crypt tissue (**Figure 3-13A**). CFC frequencies between CD44$^+$MUC1$^+$ and CD44$^+$MUC1$^-$ subsets were similar (11.1% ± 1.3% vs. 11.7% ± 1.0% for surface; 10.7% ± 5.4% vs. 9.9% ± 2.1% for crypt; **Figure 3-13C, Table 3-7**). The majority of CFCs were distributed in the CD44$^+$MUC1$^-$ subset (Surface, 73% ± 8%; Crypt, 50% ± 15%; **Figure 3-13D, Table 3-7**). Consistent with the previous analysis of CD44-defined subsets, those that did not express CD44 (CD44$^-$MUC1$^-$ and CD44$^-$MUC1$^+$) displayed much lower CFC frequencies (Surface, 2.0% ± 0.7% and 2.8% ± 2.2%, respectively; Crypt, 2.5% ± 0.6% and 5.2% ± 3.6%, respectively; **Figure 3-13C**).

In conclusion, although the majority of tonsillar CFCs appear to be MUC1$^-$, the similarity in CFC enrichment between CD44$^+$MUC1$^+$ and CD44$^+$MUC1$^-$ subsets suggest that MUC1 may not be a marker to use to discriminate differentiated cells for the purpose of enriching for tonsillar progenitors by FACS.

3.3.5. Crypt and surface regions are differentially enriched in the most primitive subsets of epithelial progenitors

Since CD44$^+$NGFR$^+$ subset was found to be most enriched for tonsillar progenitors regardless of their original sites, I sought to interrogate more deeply the functional equivalence of CD44$^+$NGFR$^+$ cells in the crypt and surface sites and compare the proliferative potential of CD44$^+$NGFR$^+$ cells by designing a more discriminating clonal experiment (**Figure 3-14**). Single CD45$^-$CD31$^-$CD44$^+$NGFR$^+$ cells were obtained at a purity of >97% from both tonsillar surface and crypt-enriched tissues and plated individually into each well of a 96-well plate. Cells were cultured in 5% O_2 with irradiated fibroblasts and 10 μM ROCK inhibitor for 12-14 days. Starting from Day 7, colonies were examined under the microscope, counted and categorized by colony types (holoclones, meroclones or paraclones as described in Section 3.2.3) every 2 days until the colonies were dissociated and replated. The repeated examination of the culture plates every 2 days suggested that individual CD44$^+$NGFR$^+$ cells were functionally heterogeneous, with a few being more slow-cycling or quiescent than others, evidenced by the delayed appearance of colonies in a few wells over the course of the 10-14-day culture period.

In all three single cell experiments, crypt $CD44^+NGFR^+$ cells displayed higher primary CFC frequencies than their surface counterparts (10.2% ± 2.4% versus 5.1% ± 2.0%; *P*=0.01; **Table 3-8**). A higher proportion of holoclones was also seen in the wells seeded with a $CD44^+NGFR^+$ cell from crypt compared with surface isolates (62% ± 2% versus 26% ± 4%, *P*=0.0011; **Figure 3-15A, Table 3-8**). Individual serial replating of their clonal progeny showed that, on average, those derived from crypt cells could be passaged longer than those derived from surface cells (Crypt, up to 4 passages; Surface, up to 2-3 passages; **Figure 3-15B, Table 3-8**), giving rise to markedly greater number of progeny in 2 of the 3 donors (~10-20-fold; **Table 3-8**). The analysis of the colony type distribution detected in each of the serial replating CFC assays showed that $CD44^+NGFR^+$ crypt CFCs were also more capable of generating daughter CFCs that produced holoclones in later passages than their surface counterpart (**Figure 3-15C and D, Table 3-8**).

The crypt-derived clones were also able to produce multilayered epithelial structures in 3D cultures at lower input cell doses (down to 1,400 cells obtained from five primary colonies; i.e., 100-fold fewer than the number of freshly isolated $CD44^+NGFR^+$ cells used). Differentiation of the cells generated in these 3D assays of cells produced *in vitro* was again evident from IHC staining, which showed the presence of cells expressing CK5, a marker of basal/parabasal cells, and CK13, a marker of differentiated cells (**Figure 3-15E**).

3.4. Transcriptome analysis of tonsillar subpopulations

The aim of this section was to compare gene expression profiles among different epithelial subsets obtained from the human tonsils, in order to identify the transcriptome signature unique to (1) specific subsets (total epithelial cells vs. progenitor-enriched subpopulations), and/or (2) specific anatomic site (crypt vs. surface).

Three individuals were studied, each with the following samples: Total epithelial-enriched $CD45^-CD31^-$ subpopulations, and epithelial progenitor-enriched $CD45^-CD31^-CD44^+NGFR^+$ cells from each of the two tonsillar sites separately (**Figure 3-16A and 16B**). RNAs were extracted directly from each of these freshly isolated samples and

their microarray data were generated by the Illumina Human HT-12 V4 BeadChip as detailed in Section 2.7. All the subsequent transcriptome analyses were performed in log2-transformed data.

3.4.1. CD44$^+$NGFR$^+$ tonsillar epithelial cells have a distinct site-independent transcriptional profile

The hierarchical clustering analysis revealed that transcript profiles were more tightly clustered by donor than by the region of the tonsil (crypt or surface) from which the epithelial cells had been isolated (**Figure 3-16C**). To identify the unique set of transcripts commonly expressed by CD44$^+$NGFR$^+$ cells regardless of their donor or specific tonsillar region of origin, I first selected the transcripts that displayed a <1.5-fold differential expression (Donor 1, 15068 genes; Donor 2, 15768 genes; Donor 3, 14447 genes; **Figure 3-17A**). Of these transcripts, 11,191 transcripts were shared by all 3 groups and were further narrowed down by excluding those with a $\log_2$-transformed crypt:surface ratio (for each individual) greater than 0.99 or less than 1.01. The most similarly expressed genes in all 6 CD44$^+$NGFR$^+$ subsets thus identified are hereafter referred to as a "site-independent" transcriptome signature of this tonsillar CFC-enriched (CD44$^+$NGFR$^+$) subset (the full list of the 227 genes is shown in **Table 3-9**). Gene ontology (GO) analysis of these 227 genes showed a significant association with biological processes such as protein targeting, localization and transport, and viral gene expression (P<0.05; **Figure 3-17B, Table 3-10**).

3.4.2. Transcripts associated with immunological processes are differentially upregulated in crypt-derived CD44$^+$NGFR$^+$ cells

The transcripts with >1.5-fold differential expression between crypt-CFCs and their surface counterparts were then combined by site to identify the genes that were commonly differentially expressed by 2 or more donors (**Figure 3-18A**). Of those more highly expressed by the crypt CFC-enriched subsets (Donor 1, 146 genes; Donor 2, 491 genes; Donor 3, 816 genes), 97 were commonly upregulated in two or more donors, with only 8 upregulated in the crypt CFC-enriched subset from all 3 donors (**Figure 3-18A**). Among the genes more highly expressed in the surface-progenitors compared with crypt-CFCs (Donor 1, 110 genes; Donor 2, 224 genes; Donor 3, 850 genes; **Figure 3-**

18A), 108 genes were shared by at least 2 donors and only 7 of them were commonly upregulated in the samples from all 3 donors (**Figure 3-18A**). GO analysis of those overlapping genes revealed that the transcripts enriched in crypt CFCs were significantly associated with immunological processes, whereas those more highly expressed in the surface progenitor subset were linked to regulation of cell proliferation and response to external stimuli (**Figures 3-18B and 18C**). I then applied more stringent criteria (≥2-fold differential expression) to identify the transcripts that are most highly differentially expressed between the 2 sites. Thirteen and 26 genes were found to be commonly overexpressed in 2 or more crypt-CFCs and surface-CFCs, respectively, with the donors 2 and 3 sharing more commonalities and donor 1 being the most dissimilar (**Tables 3-11 and 3-12**).

3.4.3. Comparison between unseparated tonsillar epithelial cells and their $CD44^+NGFR^+$ subpopulations

To identify the transcriptome signature that distinguishes the CFC-enriched subsets from the unsorted total epithelial cells for both the crypt and surface epithelia, I compared the transcripts in the $CD44^+NGFR^+$ subsets with those in their donor-matched $CD45^-CD31^-$ cells using a threshold difference in expression of ≥2-fold (**Figures 3-19 and 3-20**).

This analysis revealed that a group of genes that were more highly expressed in the crypt CFC-enriched cells from two donors (compared with their total crypt epithelial counterparts) and another gene set that was more highly expressed in the total crypt epithelial cell population **(Figure 3-19A**). Analysis of both of these sets of differentially expressed genes by hierarchical clustering and GO showed that the first set of genes that were more highly expressed in the crypt $CD44^+NGFR^+$ cells (**Table 3-13**) were associated with cell migration and motility (**Figures 3-19A and 19B**). Those more enriched in the surface-derived $CD44^+NGFR^+$ with respect to their unseparated TotalEpi cells (**Table 3-14**) were highly associated with organization of extracellular matrix and structure (**Figures 3-20A and 20B**), although these 2 also appeared in the list of top 10 biological processes enriched in the crypt CFCs (**Figure 3-19B**). The genes showing higher expression in both the surface and crypt total epithelial cells with respect to their

progenitor fractions were linked to epidermal development and keratinization (**Figures 3-19C and 20C; Table 3-15**). A summary of this analysis is graphically shown in **Figure 3-21**.

3.5. Manipulation of tonsillar progenitors with HPV16 E6/E7 oncogenes

The aim of this pilot study was to investigate the effects of HPV16 E6/E7 oncoproteins on tonsillar progenitors. Although oncogenic transformation induced by HPV16 E6 and E7 oncoproteins have been reported previously in several tissue models[197], very few have specifically studied their effects on a distinct subset(s) of epithelial cells. It is generally believed that HPV-induced malignant transformation occurs in the progenitor cells, however, whether the primitive cells are the only target of HPV has not been rigorously investigated due to the lack of a method to identify different subset of cells and isolate them with sufficiently high purity to enable them to be separately manipulated *in vitro*. Furthermore, the observation that tonsillar crypt epithelium, but not the surface epithelium, is the site of HPV-associated tonsil cancer[194,220] raises an interesting question as to whether this crypt-specificity is solely due to the architectural advantage (deep invaginations that function as a reservoir of pathogens) or whether it may be attributed to intrinsic differences between the surface and crypt progenitors. As the first step towards exploring these questions, I isolated the progenitor-enriched $CD44^+NGFR^+$ subset from both the surface and crypt regions of the tonsils and independently transduced them with lentivirus containing HPV16 E6/E7 oncogenes.

The procedure involved the dissociation of tonsil surface-enriched and crypt-enriched tissues into single cells followed by the depletion of the bulk of lymphocytes using the density centrifugation, followed by isolation of $CD44^+NGFR^+$ cells from each region by FACS (**Figure 3-22A**). The sorted $CD44^+NGFR^+$ cells were then transduced with either (1) lentivirus encoding HPV16 E6/E7 oncogenes and yellow fluorescent protein (YFP), or (2) a control lentivirus encoding a fluorescent mCherry reporter (Detailed procedure in Section 2.8). The cells were then cultured in fresh serum-free media for 4-7 days, trypsinized and FACS analyzed, which showed the transduction

efficiency to be over 80% for both YFP and mCherry control (**Figure 3-22B**). The transduced cells were then selected by sorting for YFP^+ (E6/E7) and $mCherry^+$ (control) and allowed to expand in 35 mm dishes over the period of 7 days to generate sufficient number of cells for the subsequent experiments. Expression of the E6/E7 cassette in FACS-purified YFP^+ (E6/E7-transduced) cells was confirmed by PCR amplification of the E7 sequence (**Figure 3-22C**). A portion of these cells was serially passaged to examine their proliferative capacities. The other portion was subjected to the 3D culture system to investigate whether E6/E7 oncoproteins would also affect the differentiation of tonsillar epithelial cells, and whether surface and crypt would differ from each other.

3.5.1. HPV16 E6/E7 oncoproteins perturb the *in vitro* growth of $CD44^+NGFR^+$ human tonsillar epithelial cells in 2D culture

Short-term 2D cultures were set up using 3 different types of input crypt cells; control transduced cells ($mCherry^+$) alone, E6/E7-transduced cells (YFP^+) cells alone and a 1:1 mixture of the 2 (for assessment of growth competition), all of which were seeded at a density of 10^5 cells per 10 cm^2 culture dish. By Day 5, dishes were ~30%, ~80% and ~70% confluent, respectively, and the latter 2 were subsequently split among multiple dishes to expand. After 4 days, cells were harvested and counted to quantify 9-day outputs. The results showed that E6/E7-transduced cells completed 6.6 population doublings (PD) in 9 days, more than 2-fold higher than control cells (3.1 PD), which resulted in a >10-fold greater output of cells (**Table 3-16**). In the competition assay, a portion of the 9-day output obtained from the mixed culture that initially seeded with the equal number of $mCherry^+$ and YFP^+ cells were analyzed by flow cytometry to determine the proportion of YFP^+ cells versus $mCherry^+$ cells. 12.5% of the total cell suspensions were $mCherry^+$ while 85.2% were YFP^+, representing 3.1 and 5.9 population doublings, respectively (**Table 3-16**).

In longer term cultures, both surface- and crypt-derived control cells maintained in 2D culture displayed characteristics of senescence within 4 passages *in vitro* displaying acquisition of a large 'fried egg' morphology and positive staining for senescence-associated β-galactosidase (SA β-gal) within 22-25 days (**Table 3-16, Figure 3-23**). In contrast, both surface and crypt E6/E7-transduced cells were able to

proliferate for many more passages for a period of >70 and >100 days, respectively without any apparent sign of senescence (**Table 3-16, Figure 3-23**).

3.5.2. HPV16 E6/E7 oncoproteins perturb normal differentiation of tonsillar epithelium in 3D organotypic culture

The 3D differentiation culture (either surface- or crypt-derived) was also set up using 3 different input cell compositions; E6/E7-expressing (YFP^+) cells alone, their controls ($mCherry^+$) alone, and 1:1 mixture of the two. 4×10^5 cells were seeded on top of each collagen matrix, and allowed to form monolayer for 5 days while submerged in the culture medium and to differentiate for 11 days in the air-liquid interphase.

Multilayered epithelial populations were generated from all cell sources tested, but the structures produced by the E6/E7-expressing cells (either surface- or crypt-derived) showed less differentiated morphology evidenced by an absence of keratinization, fewer cell layers, pan-epithelial expression of CK19 (normally a basal cell marker), and dispersed presence of Ki-67 positive cells (proliferation-associated antigen normally confined in the basal and parabasal layers) across the full thickness of the epithelium (**Figures 3-24, 3-25 and 3-27; Tables 3-17 and 3-18**). An expansion of CD44 positive cells (normally basal/parabasal cell-restricted) and a negative or equivocal expression of CK4 (a differentiation marker of the cells normally expressed very strongly in the suprabasal cell layers) were observed in the crypt-derived E6/E7 culture (**Figure 3-25**), but not in the cultures of E6/E7-transduced surface epithelial cells (**Figure 3-24**). Expression patterns of additional differentiation markers such as CK13, involucrin and MUC1 did not differ between E6/E7-transduced cultures and the control cells regardless of the anatomic site from which they were isolated (**Figures 3-24 and 3-25**).

The simultaneous expression of nuclear Ki-67 and cytoplasmic p16 (senescence-associated antigen) within the same cell is considered indicative of oncogenic molecular changes induced by a HPV infection[237] and was proposed as a new surrogate marker for predicting high-risk cervical lesions[255]. The utility of the marker has been recently validated in a prospective study[238]. Consistent with the hypothesis, concordant expression of cytoplasmic p16 and nuclear Ki67 was found throughout almost the entire length of the epithelium generated by E6/E7-transduced cells (Found in 72% and 93% of

the total fields scored, the surface- and crypt-derived cultures, respectively; **Table 3-18**), whereas markedly fewer $p16^{+}Ki\text{-}67^{+}$ cells were detected in the control cultures (8% and 32% of the total fields scored, the surface- and crypt-derived controls, respectively; **Table 3-18**).

3.5.3. Coculture of E6/E7-transduced and normal cells creates abnormal structures in 3D organotypic culture

The structures produced in 3D cultures of the mixed cell sources (1:1 mixture of control and E6/E7-expressing cells) showed unexpectedly heterogeneous morphology with a few distinctions noted between the cultures of surface- and crypt-derived cells. In the former, the epithelium produced by the mixed cells appeared almost identical to the E6/E7 culture that displayed thinner epithelium and pan-epithelial expression of CK19 (**Figure 3-24**). In the latter, 3 different epithelial morphologies were identified (**Figure 3-26**): (1) areas with a more extensive stratification (similar to the control culture; hereafter referred to as Type 1); (2) areas with fewer epithelial layers and an absence of keratinization (similar to the E6/E7-transduced; Type 2); and (3) areas with unique 'pearl-like' structures embedded in the intermediate layers (Type 3). The cells in the Type 3 structures expressed basal markers CD44 and NGFR, but lacked expression of any differentiation markers tested, including CK13, CK4, involucrin and MUC1. Ki-67 expression was also absent in these non-basally located $CD44^{+}NGFR^{+}$ cells, but was observed immediately adjacent to these basal-like cells (**Figure 3-26**, arrowheads). CK19, the marker of tonsillar basal cells, was expressed by all cells regardless of the aforementioned epithelial type (**Figure 3-26**).

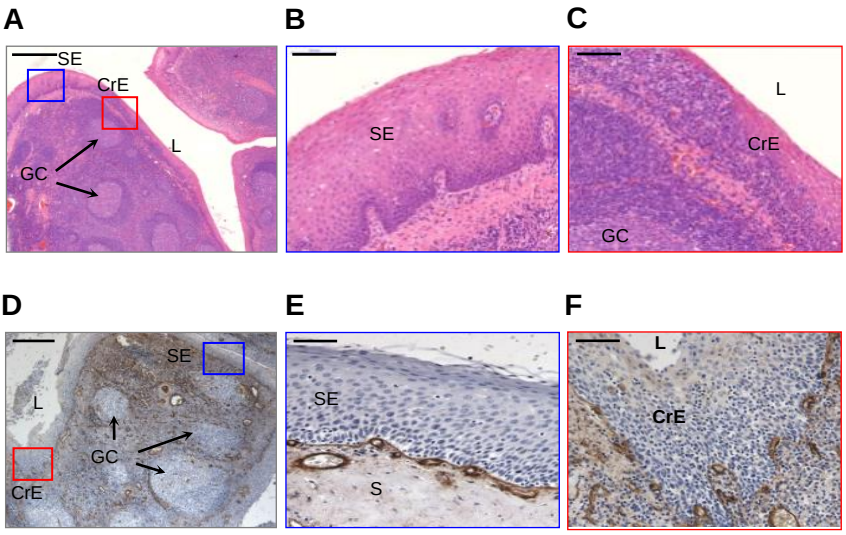


Figure 3-1 Morphology of tonsillar epithelia overlying the surface and crypts of human palatine tonsils

(A–C) H & E-stained FFPE tonsil tissue sections showing stratified surface epithelium with an organized multi-layer structure similar to oral mucosa (blue box, A and B), and the thinning epithelium overlying tonsillar crypts heavily infiltrated by lymphocytes (red box, A and C). SE, surface epithelium; CrE, crypt epithelium; S, stroma; GC, germinal center; L, lumen). Scale bars, 500 μm (A and D) and 100 μm (B and C).(D–F) IHC staining of collagen IV in FFPE tonsil tissue sections showing the presence of a continuous basement membrane in the surface epithelium (blue box, D and E) and a discontinuous and indistinct basement membrane in the crypt epithelium (red box, E and F). Scale bars, 500 μm (D) and 100 μm (E and F).

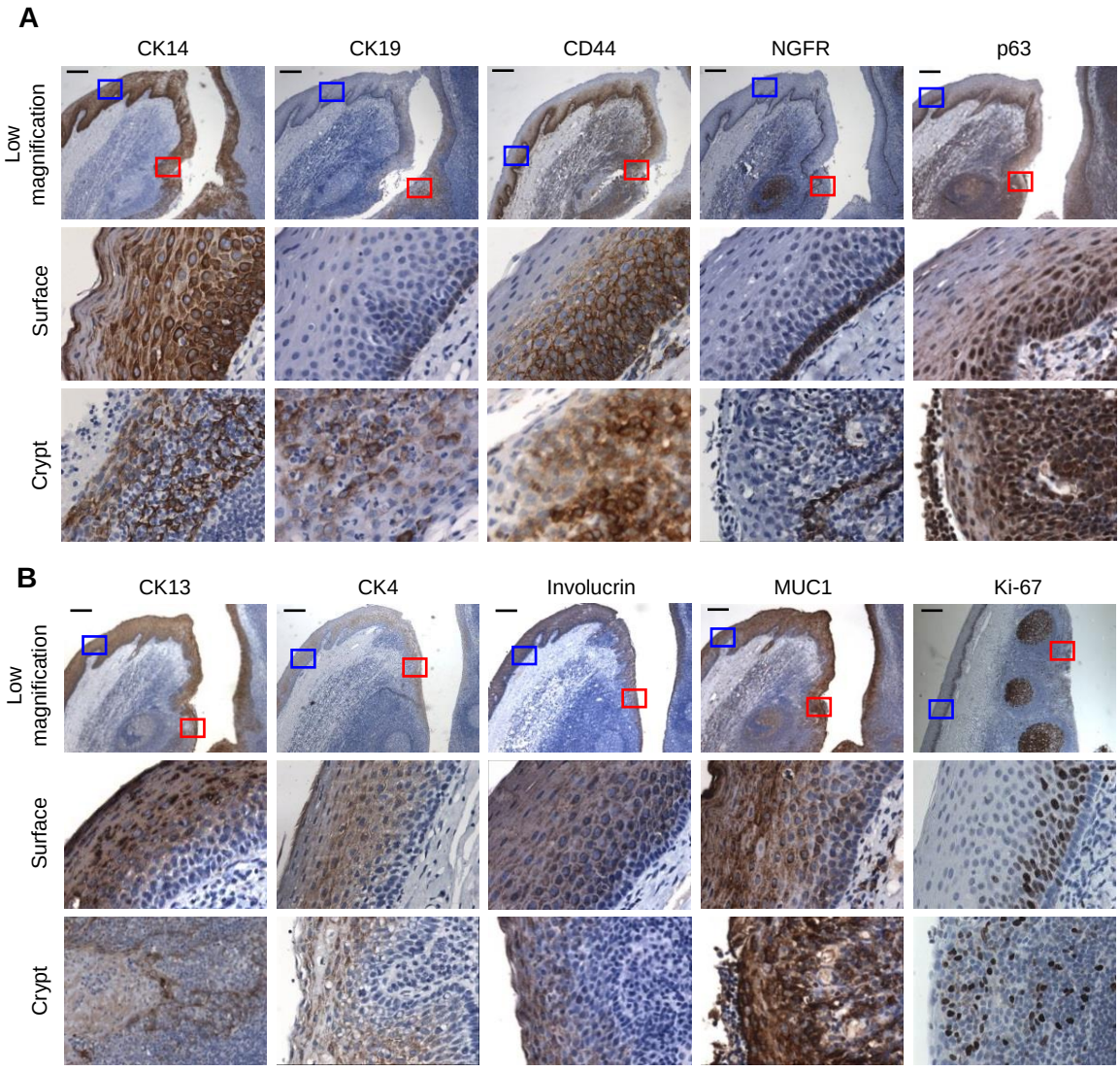


Figure 3-2. *In situ* IHC analysis of tonsillar surface and crypt epithelium
(A-B) Top rows, low-magnification images showing both surface (blue box) and crypt (red box) epithelia, portions of which are enlarged in the center and bottom rows. Scale bar, 200 µm.

Table 3-1. Immunohistological examination of human tonsillar surface epithelium

		Surface epithelium							
		Basal layer		Parabasal layer		Mid layer		Upper layer	
Marker	CL	% Pos	Intensity	% Pos	Intensity	% Pos	Intensity	% Pos	Intensity
CK5	C	4	++/+++	4	++/+++	3	+/++	0	-
CK14	C	4	+++	4	+++	4	+++	3	+++
CK19	C	4	++	1	±	1	±	0	-
CD44	M	4	+++	4	+++	3	++	0	-
NGFR	M	4	+++	0	-	0	-	0	-
P63	N	4	+++	4	+++	3	++	0	-
CK8/18	C	4	±	4	+	3	+	0	-
CK13	C	4	±	4	++	4	++	4	+++
CK4	C	0	-	1	+	4	++	3	+
Involucrin	C	0	-	3	+++	4	+++	4	+++
MUC1	M/C	1	±	3	+	4	++	4	+++
Ki-67	N	0	-	4	+++	0	-	0	-

Table 3-2. Immunohistological examination of human tonsillar crypt epithelium

		Crypt epithelium					
		Basal layer		Mid layer		Upper layer	
Marker	CL	% Pos	intensity	% Pos	intensity	% Pos	intensity
CK5	C	4	++/+++	4	++/+++	3	++/+++
CK14	C	4	+++	4	+++	4	+++
CK19	C	3	++	4	+++	2	++
CD44	M/C	4	+++	4	++	0/1	+/-
NGFR	M/C	3	++	2	+	0	-
P63	N	4	+++	3	+++	2	+++
CK8/18	C	2	++	4	+++	3	+++
CK13	C	4	±	4	++/+++	4	+++
CK4	C	0	-	4	+++	4	++/+++
Involucrin	C	4	++	4	+++	4	+++
MUC1	M/C	4	++	4	+++	4	+++
Ki-67	N	4	+++	3	+++	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.

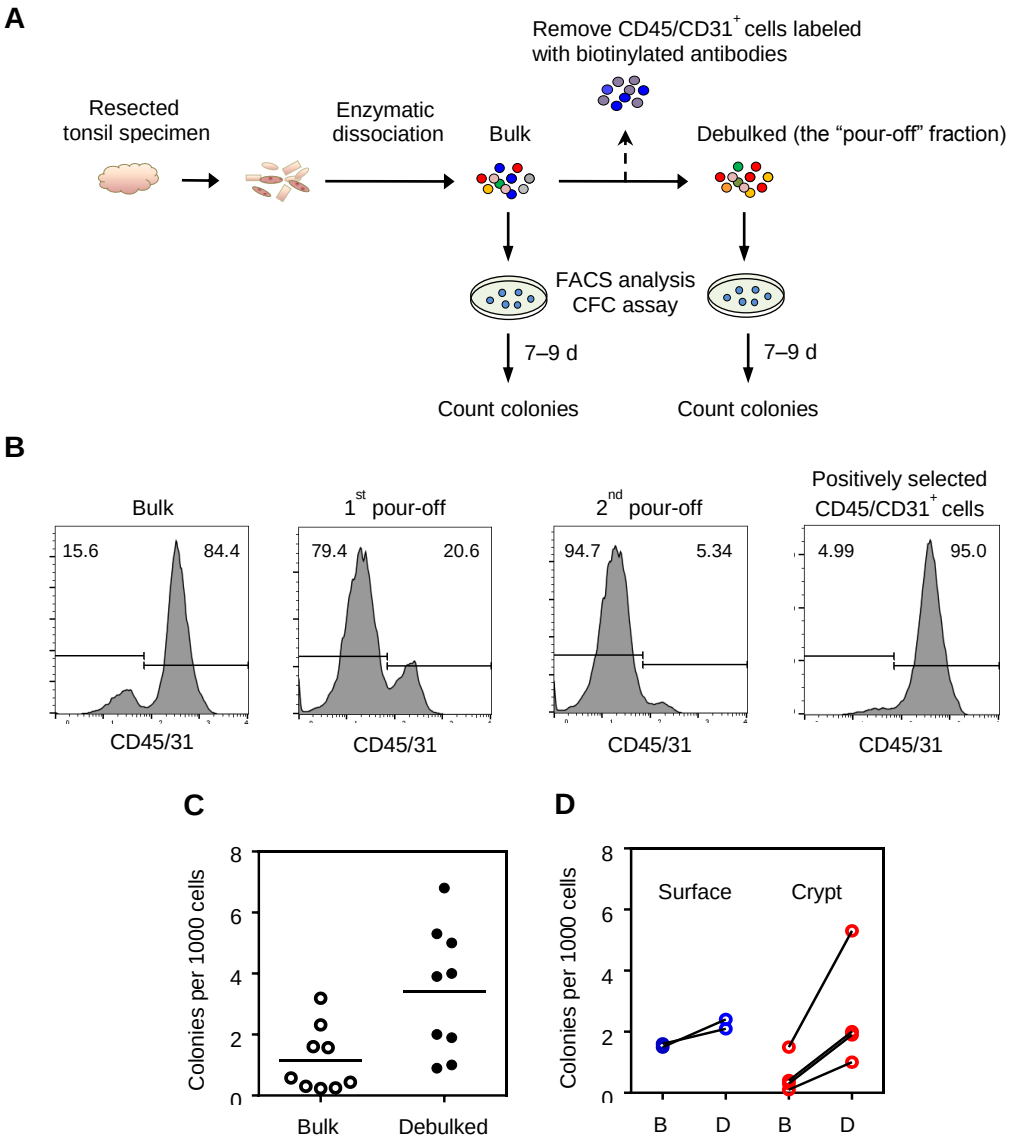


Figure 3-3. Depletion of CD45⁺ (hematopoietic) and CD31⁺ (endothelial) cells by magnetic beads-based cell separation

(A) Experimental protocol for depleting CD45⁺ and CD31⁺ cells and comparing CFC frequencies before and after the depeletion. (B) Representative histograms displaying the purity of each sorted fraction. Bulk cells (before magnetic sorting), the 1st pour-off (negatively selected CD45/31⁻ fraction), the 2nd pour-off (the repeated negative selection to further purify the 1st pour-off) and the positively selected CD45/31⁺ fraction (biotin⁺) were separately obtained and analyzed by flow cytometry. The numbers indicate the percentages of CD45/31⁻ cells (left) and CD45/31⁺ cells (right) present in each of the isolated cell fractions. (C) CFC frequencies of unseparated bulk cells compared with those of debulked cells. Data pooled from 10 experiments obtained from the whole tonsil (n=4), surface-enriched (n=2) or crypt-enriched (n=4) samples. The surface and crypt data are separately plotted in (D), showing a relative increase in CFC frequencies of surface-derived or crypt-derived cells by magnetic sorting. B, bulk; D, debulked.

Table 3-3. Expression analysis of lineage markers on freshly dissociated tonsillar single cell suspensions

Subset	Cell Type	% of total live cells ± SD[1]
CD45/CD31$^+$	Hematopoietic/endothelial cells	92.0 ± 6.1 (n=8)
CD19$^+$	B cells (subset of CD45$^+$ cells)	44.5 ± 5.0 (n=4)[2]
CD3$^+$	T cells (subset of CD45$^+$ cells)	24.5 ± 5.5 (n=4)[2]

[1]Data are expressed as percentage of positive cell of total cell population.
[2]The percentages of CD19$^+$ B cells and CD3$^+$ T cells were calculated from the flow cytometer analysis as CD19$^+$/CD3$^-$ and CD19$^+$/CD3$^-$ cell populations, respectively.

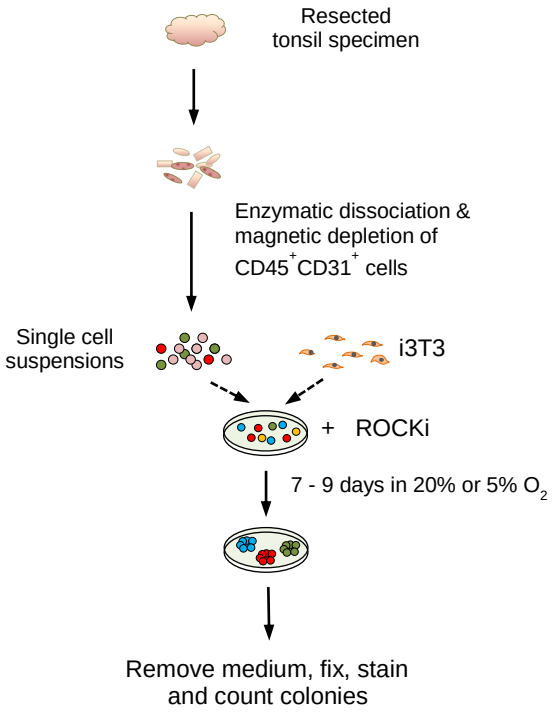


Figure 3-4. Experimental design used to optimize conditions to detect primary CFCs

Single cells obtained by enzymatic dissociation of normal tonsil specimens were cultured with or without irradiated fibroblasts (i3T3) and ROCK inhibitor (ROCKi), in either 20% or 5% O2 for 7-9 days. At the end of the assay period, the colonies were stained with Giemsa or Hematoxylin and scored using an inverted microscope.

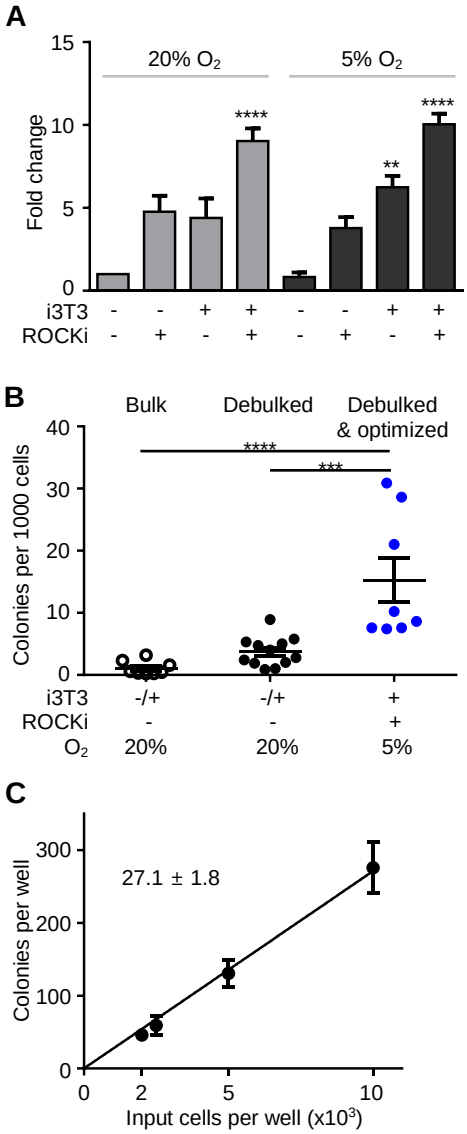


Figure 3-5. Optimization of human tonsillar CFC assay

(A) Enhancing effects of the use of feeders, a ROCK inhibitor, and a reduced O_2 level on colony growth, calculated by comparing the increased numbers of colonies obtained from a given number of cells under each test condition relative to the control condition (no added i3T3 cells or ROCKi). Values shown were derived from 10 experiments.

(B) Comparison of CFC frequencies among unseparated bulk cells (open black circles, un-optimized culture conditions), cells debulked by EasySep (closed black circles, un-optimized culture conditions) and debulked cells cultured under the optimized condition (blue circles). Values shown were derived from 18 experiments.

(C) Demonstration of a linear relationship between the number of input cells plated per well and colony yields under optimized CFC assay conditions. Values shown were derived from three experiments.

All data are expressed as means ± SEM. ** P<0.01, *** P<0.001, **** P<0.0001.

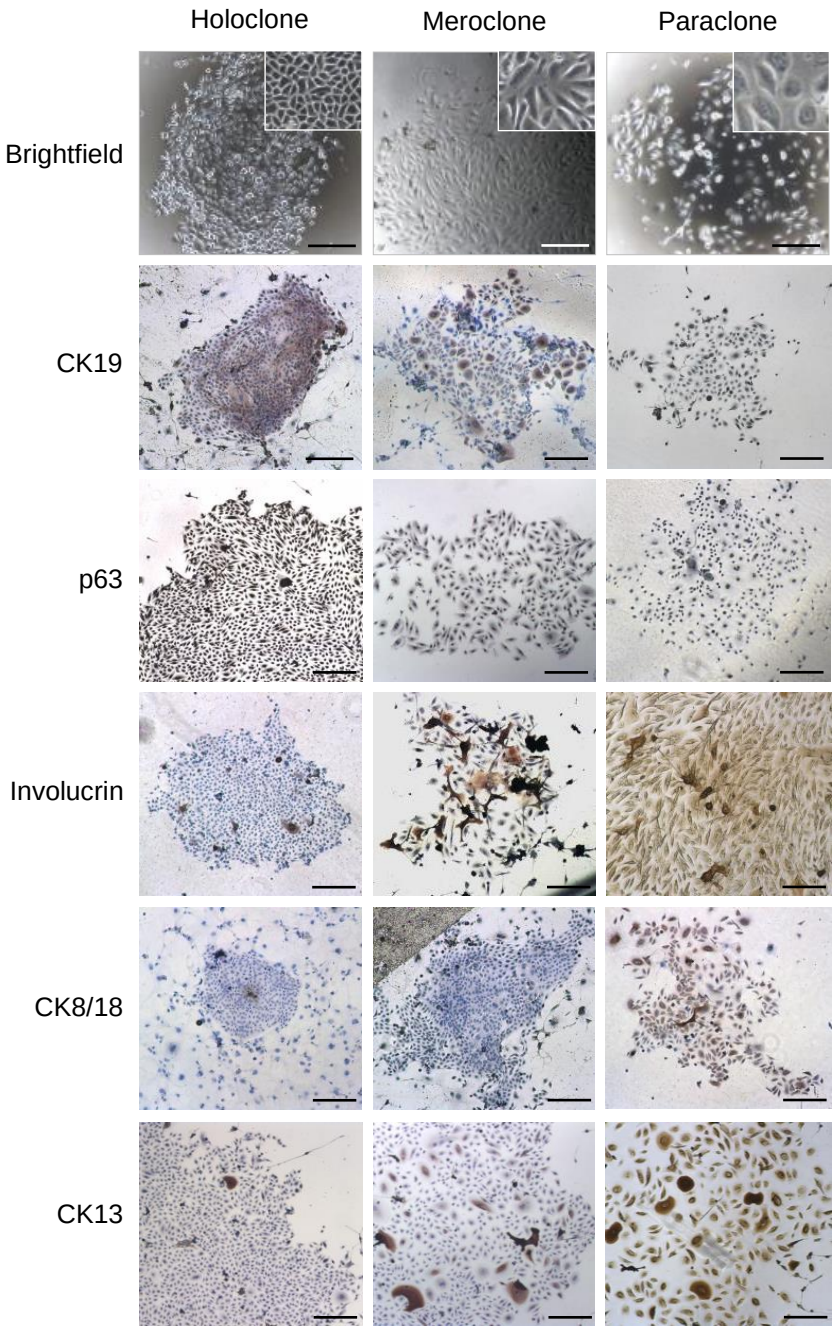
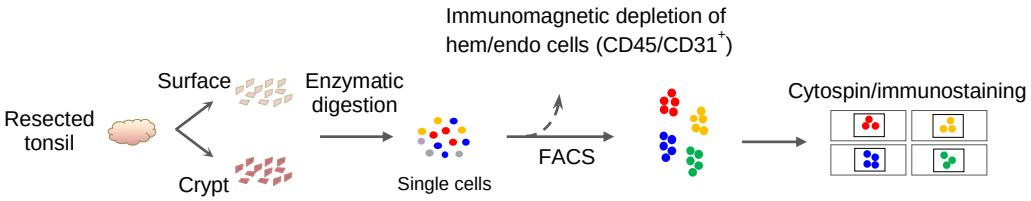


Figure 3-6. IHC characterization of the different types of colonies produced
Holoclones expressed high levels of CK19 and p63 (markers of basal cells) but infrequently or rarely expressed involucrin, CK8/18, and CK13 (markers of differentiated tonsillar epithelial cells). Many of the cells in mero/paraclones expressed differentiation markers but rarely expressed any basal markers. Scale bars, 1 mm.

A


Immunomagnetic depletion of
hem/endo cells (CD45/CD31+)
Resected
tonsil
Surface
Crypt
Enzymatic
digestion
Single cells
FACS
Cytospin/immunostaining

B

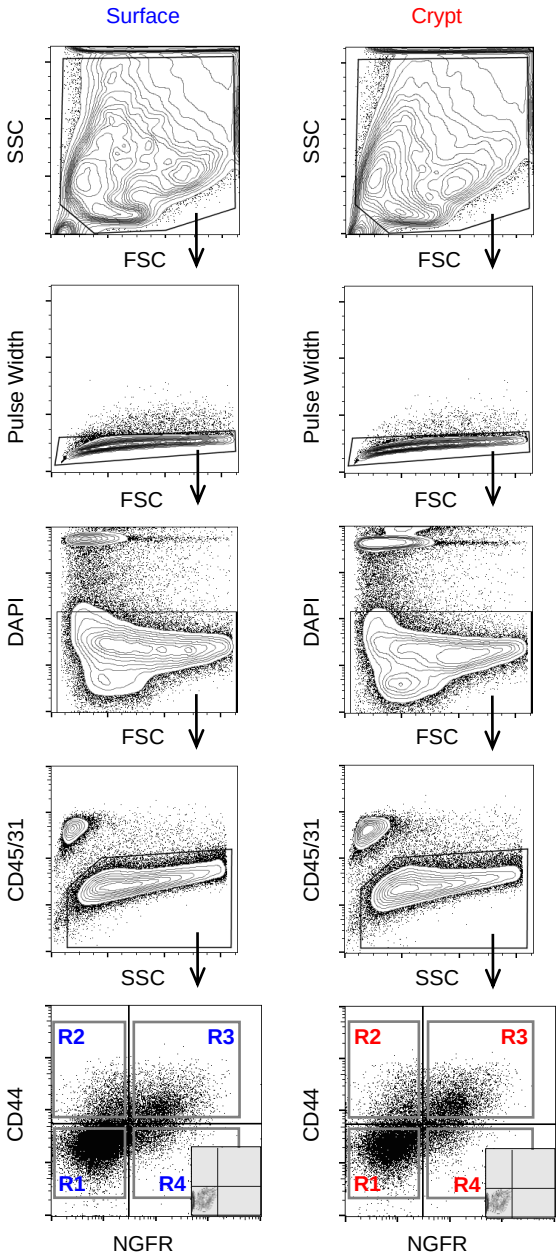
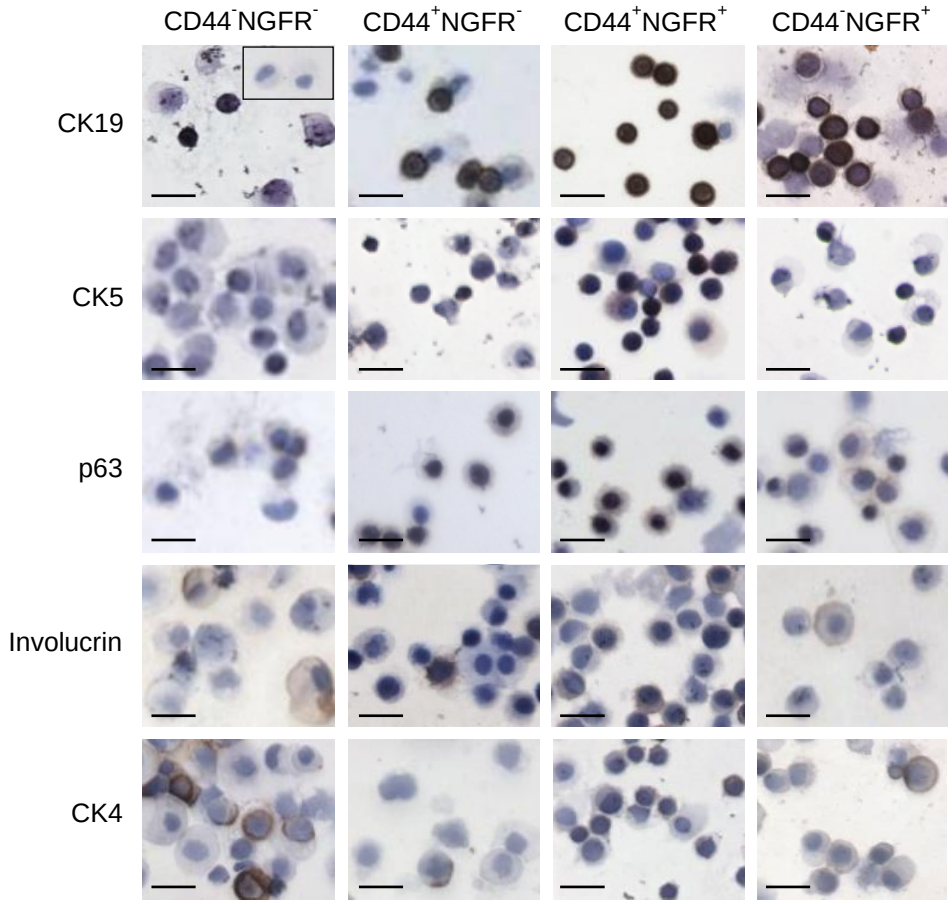

Surface
Crypt
SSC
FSC
Pulse Width
DAPI
CD45/31
CD44
NGFR
R1
R2
R3
R4

Figure 3-7

Figure 3-7. Prospective isolation of tonsillar epithelial subpopulations
(A) Experimental protocol for analyzing single CD45- (non-hematopoietic) and CD31- (non-endothelial) cells according to their expression of CD44 and NGFR with histological endpoint. (B) Representative FACS plots demonstrating the gating strategy used to isolate total cells based on forward scatter (FSC) and side scatter (SSC), and to exclude doublets and dead cells using Tigger Pulse width vs. FSC, and DAPI vs. FSC, respectively. The resultant total live singlets are analyzed by expression of CD45 and CD31 to exclude CD45+ and CD31+ cells. Further sorting includes separation of different subfractions based on expression of CD44 and NGFR. Cells were obtained from the whole (unseparated) tonsil tissues (not shown), or from the surface and crypts separately (Insets, negative controls)

A



B

Marker	CD44⁻NGFR⁻ (R1)	CD44⁺NGFR⁻ (R2)	CD44⁺NGFR⁺ (R3)	CD44⁻NGFR⁺ (R4)
CK19	25.3 ± 7.4	49.3 ± 4.8	80.3 ± 7.4	77 ± 30
CK5	6.3 ± 0.9	17.7 ± 3.9	74.7 ± 5.2	24.4 ± 2.3
P63	50 ± 23	74.7 ± 7.4	94.7 ± 2.9	40.7 ± 5.2
Involucrin	19.7 ± 1.3	15.3 ± 2.4	62 ± 22	13.0 ± 2.6
CK4	30.3 ± 6.0	17.7 ± 1.9	47 ± 20	55.3 ± 9.8

Figure 3-8. IHC analysis of CD45⁻CD31⁻ cells defined by their expression of CD44 and NGFR

(A) IHC analysis of CD45⁻CD31⁻ cells isolated by FACS at purities of >90% according and cytospun prior to fixation and staining with antibodies to CK19, CK5, p63, and CK4. Inset, negative control. Scale bar, 20 µm. (B) Percentages of cells positive for the indicated intracellular proteins. Values shown are the mean ± SD. At least three independent samples were evaluated for each marker.

Table 3-4. IHC analysis of freshly sorted and cytospun $CD45^-CD31^-$ subpopulations defined by CD44 and NGFR

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

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

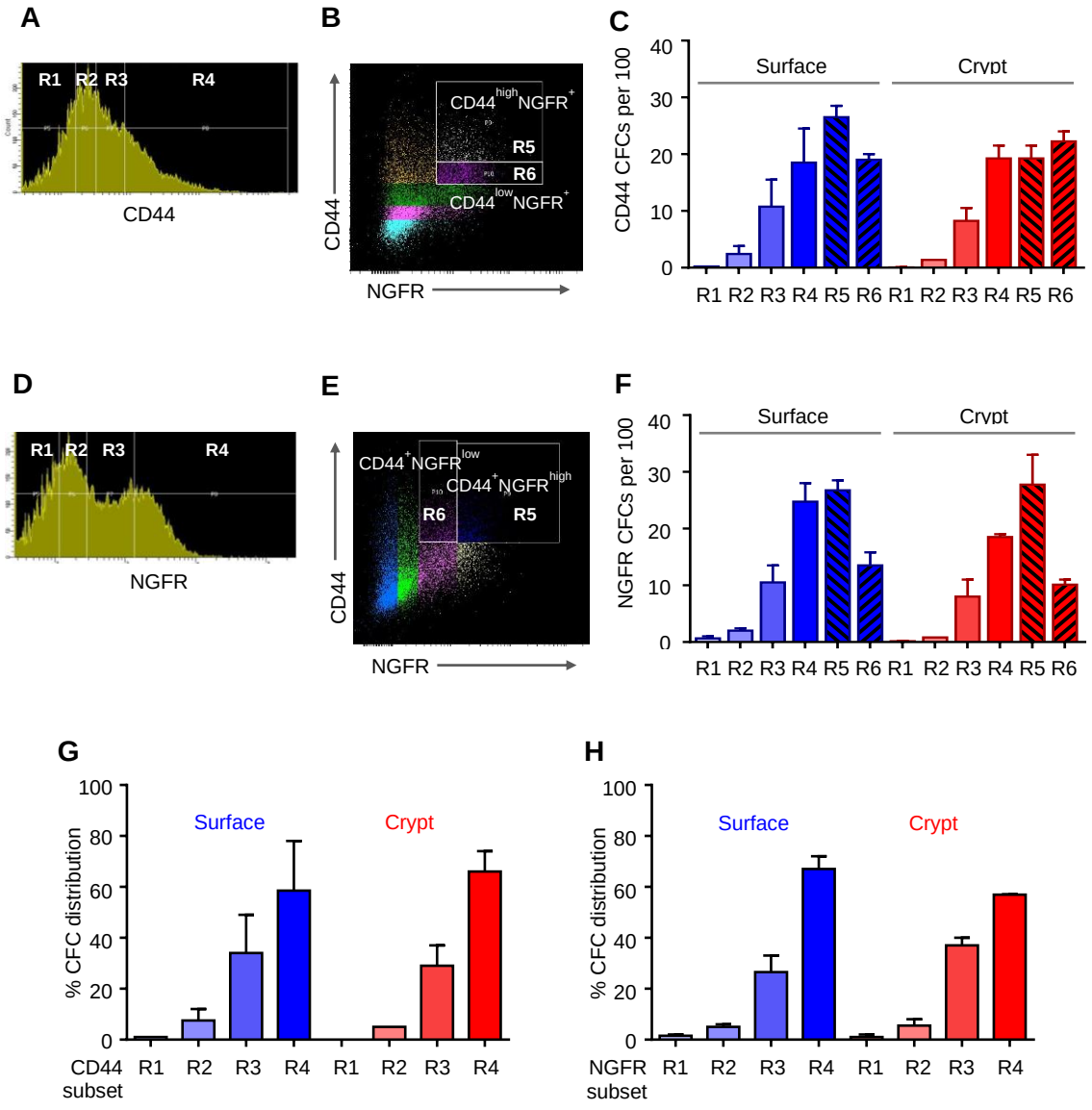


Figure 3-9. CFC frequencies and distributions in subsets defined by CD44 and/or NGFR

FACS plots showing the gating strategies used to separate 6 subsets defined by CD44 expression (A-B) and NGFR expression (D-E) within the $CD45^{-}CD31^{-}$ population. Each of the 6 subsets was subjected to CFC assays and CFC frequencies were quantified (C and F). Percentages of CFCs in each of the four subsets defined by the CD44 expression (G) or NGFR expression (H) are shown separately for both surface and crypt cells.

Table 3-5. Proportions of total cells, CFC frequencies and CFC distributions in tonsillar epithelial subsets defined by CD44 and NGFR expression

Sort region*	Subset analyzed	% of total cells	CFCs/100 cells	% CFC distribution
R1	$CD44^{--}$	24 ± 0.6	0.13 ± 0.05	0.5 ± 0.3
R2	$CD44^{-}$	26 ± 0.3	1.9 ± 0.6	6.3 ± 2.0
R3	$CD44^{low}$	26 ± 0.4	9.5 ± 2.3	31.5 ± 7.1
R4	$CD44^{high}$	25 ± 0.6	18.9 ± 2.6	62.3 ± 8.9
R5	$CD44^{high}NGFR^{+}$	16 ± 5	22.9 ± 2.4	45.3 ± 14.6
R6	$CD44^{low}NGFR^{+}$	11 ± 2	20.6 ± 1.2	22.8 ± 3.8
R1	$NGFR^{--}$	27 ± 5	0.4 ± 0.2	1.3 ± 0.5
R2	$NGFR^{-}$	25 ± 0.6	1.4 ± 0.4	5.3 ± 1.1
R3	$NGFR^{low}$	25 ± 1	9.3 ± 1.9	31.8 ± 4.2
R4	$NGFR^{high}$	21 ± 4	21.6 ± 2.2	62.0 ± 3.5
R5	$CD44^{+}NGFR^{high}$	16 ± 4	27.3 ± 2.3	54.0 ± 8.2
R6	$CD44^{+}NGFR^{low}$	7 ± 2	11.8 ± 1.4	13.0 ± 2.0

* As shown in Figure 3-8.
Data were pooled from 4 samples, 2 surface-derived and 2-crypt-derived.
All data are expressed as means ± SEM.

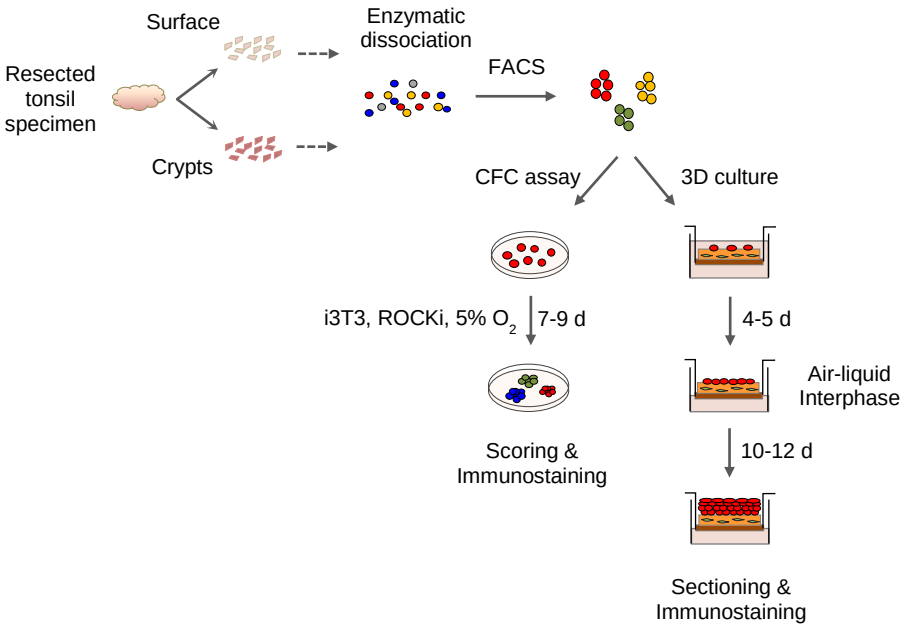


Figure 3-10. Experimental design for studying tonsillar surface or crypt epithelial progenitors

Tonsillar surface tissues were dissected out from resected tonsil specimen obtained from routine tonsillectomies and independently dissociated into single cells, fractionated into different subsets by FACS and subjected to the CFC assay and the 3D culture differentiation system to examine their ability to give rise to daughter cells and differentiate into multilayered epithelium.

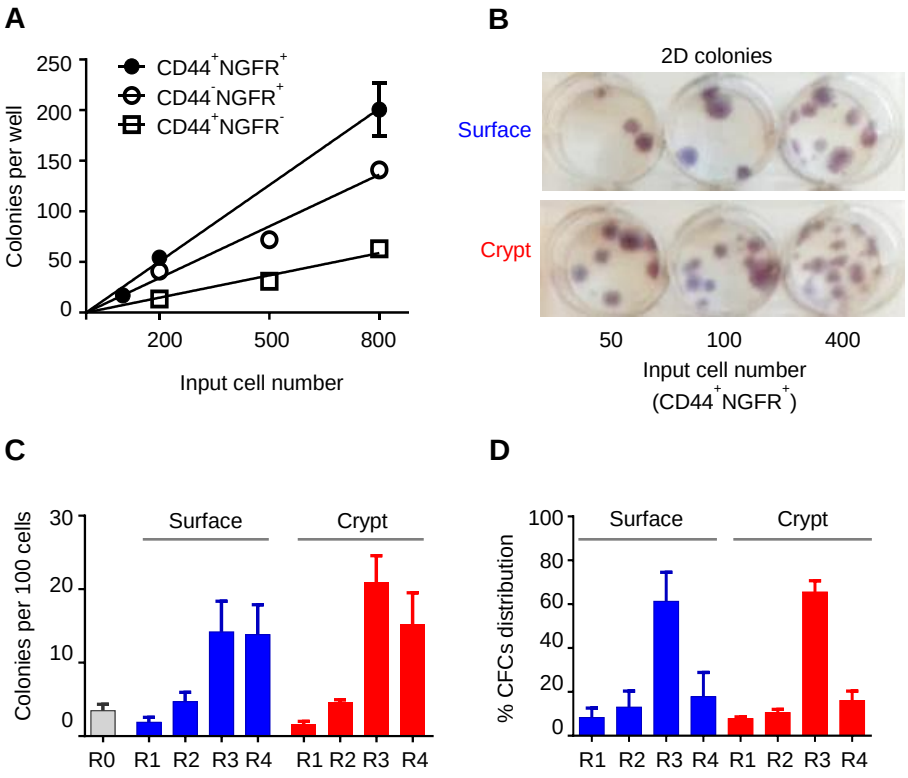


Figure 3-11. Phenotypic and functional characterization of CD44/NGFR-defined subsets of tonsillar surface and crypt epithelial cells

(A) Linear relationship between the number of cells seeded and the number of colonies formed per well in CFC assays. Conlony enumeration was performed on Days 7-9. (B) Representative images of hematoxylin staining of 14-day colonies generated in CFC assays set up in 6-well plates. The numbers represent the numbers of cells sorted by FACS that were directly seeded into each well. (C) Number of CFCs detected per 100 input cells from each of the four subpopulations analyzed (R1, $CD44^-NGFR^-$; R2, $CD44^+NGFR^-$; R3, $CD44^+NGFR^+$; and R4, $CD44^-NGFR^+$ as shown in Figure 3-7B). (D) Distribution of CFCs in each of the four surface and crypt subpopulations analyzed. Data were obtained from four independent experiments.

Table 3-6. Phenotypic and functional characterization of CD44/NGFR-defined subsets of tonsillar surface and crypt epithelial cells

Source	Sort region*	Subset analyzed[1]	% of total cells[2]	CFCs /100 cells[3]	% CFC distribution[3]
Surface	R0	Unseparated	-	3.4 ± 0.9	100
	R1	$CD44^-NGFR^-$	38 ± 5	1.9 ± 0.7	8 ± 4
	R2	$CD44^+NGFR^-$	27 ± 4	4.7 ± 1.3	13 ± 7
	R3	$CD44^+NGFR^+$	29 ± 5	14.2 ± 4.2	61 ± 13
	R4	$CD44^-NGFR^+$	7 ± 2	13.8 ± 4.1	18 ± 11
Crypt	R0	Unseparated	-	4.0 ± 1.0	100
	R1	$CD44^-NGFR^-$	43 ± 5	1.6 ± 0.4	8 ± 1
	R2	$CD44^+NGFR^-$	28 ± 5	4.7 ± 0.3	11 ± 2
	R3	$CD44^+NGFR^+$	34 ± 3	20.9 ± 3.7	66 ± 5
	R4	$CD44^-NGFR^+$	6 ± 1	15.2 ± 4.3	16 ± 4

* As shown in Figure 3-12.
[1]Subset of FACS-sorted $CD45^-CD31^-$ cells
[2]Data were pooled from 10 independent FACS profiles.
[3]Data were obtained from 4 independent experiments
All data are expressed as means ± SEM.

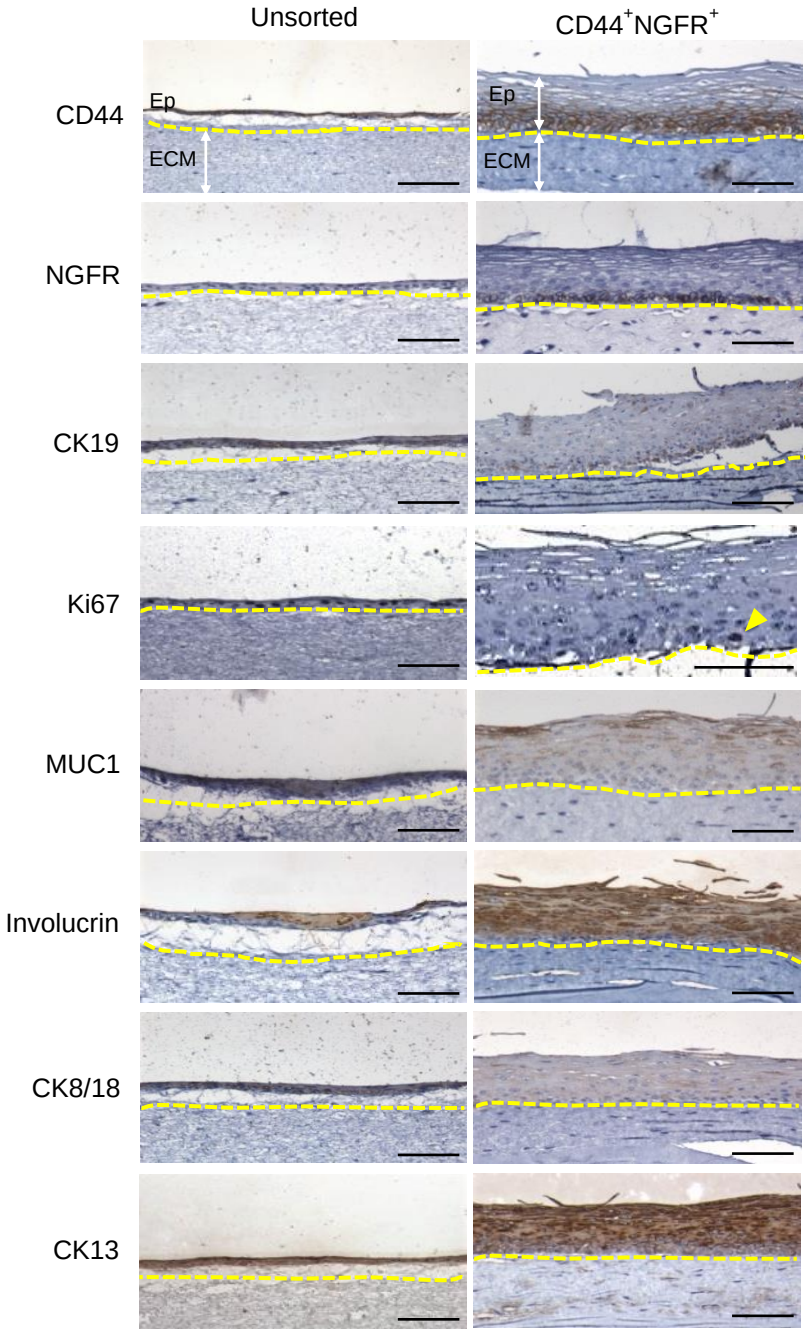


Figure 3-12. IHC staining of epithelial tissue layers generated in organotypic 3D cultures

Equal numbers (1.25 x 10^5 cells) of unsorted $CD45^-CD31^-$ cells (left panel) and sorted $CD44^+NGFR^+$ subsets (right panel) isolated from the crypt epithelium were seeded on top of extracellular matrix (ECM) in 3D organotypic cultures. Only the $CD44^+NGFR^+$ subsets successfully formed multilayered epithelium displaying the predicted marker expression patterns. Unsorted $CD45^-CD31^-$ cells did not form a complete epithelium. Dotted line indicates basement membrane. Ep, epithelial layers; ECM, extracellular matrix; Arrowhead, Ki-67^+ cell. Scale bar, 100 µm.

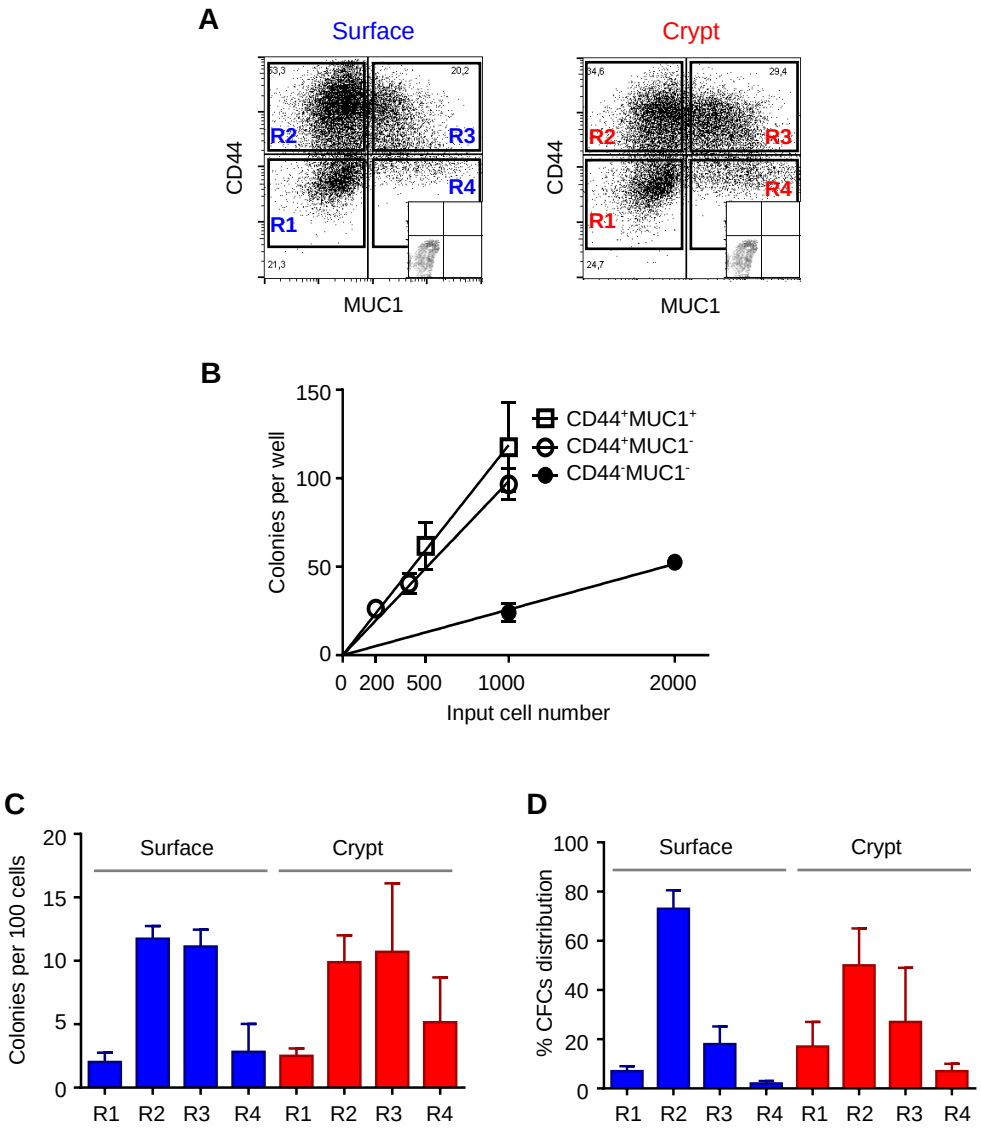


Figure 3-13. Functional characterization of tonsillar epithelial cells according to their surface expression of CD44 and MUC1

(A) Representative FACS plots demonstrating the gating strategy used to subfractionate live single CD45/CD31$^-$ cells derived from either surface or crypt into 4 subsets based on their cell surface expression of CD44 and MUC1. The gates were drawn with reference to control cells that were labeled with the viability dye (PI) alone (inset). (B) Linear relationship between the number of cells seeded and the number of colonies formed per well in CFC assays on CD44$^-$MUC1$^-$, CD44$^+$MUC1$^-$, CD44$^+$MUC1$^+$ subsets. (C) Number of CFCs detected per 100 input cells from each of the four subpopulations analyzed (R1, CD44$^-$MUC1$^-$; R2, CD44$^+$MUC1$^-$; R3, CD44$^+$MUC1$^+$; and R4, CD44$^-$MUC1$^+$) (D) Distribution of CFCs in the same 4 subpopulations. Data were pooled from 3 independent experiments.

Table 3-7. Phenotypic and functional characterization of CD44/MUC1-defined subsets of tonsillar surface and crypt epithelial cells

	Sort region*	Subset analyzed[1]	% of total cells	CFCs /100 cells	% CFC distribution
Surface	-	Unseparated	-	4.3 ± 1.4	100
(n=3)	R1	$CD44^-MUC1^-$	23 ± 5	2.0 ± 0.7	7 ± 2
	R2	$CD44^+MUC1^-$	47 ± 1	11.7 ± 1.0	73 ± 8
	R3	$CD44^+MUC1^+$	22 ± 9	11.1 ± 1.3	18 ± 7
	R4	$CD44^-MUC1^+$	7 ± 2	2.8 ± 2.2	2 ± 1
Crypt	-	Unseparated	-	3.0 ± 1.6	100
(n=2)	R1	$CD44^-MUC1^-$	38 ± 14	2.5 ± 0.6	17 ± 10
	R2	$CD44^+MUC1^-$	31 ± 1	9.9 ± 2.1	50 ± 15
	R3	$CD44^+MUC1^+$	18 ± 15	10.7 ± 5.4	27 ± 22
	R4	$CD44^-MUC1^+$	10 ± 2	5.2 ± 3.6	7 ± 3

[1]Subset of FACS-sorted $CD45^-CD31^-$ cells

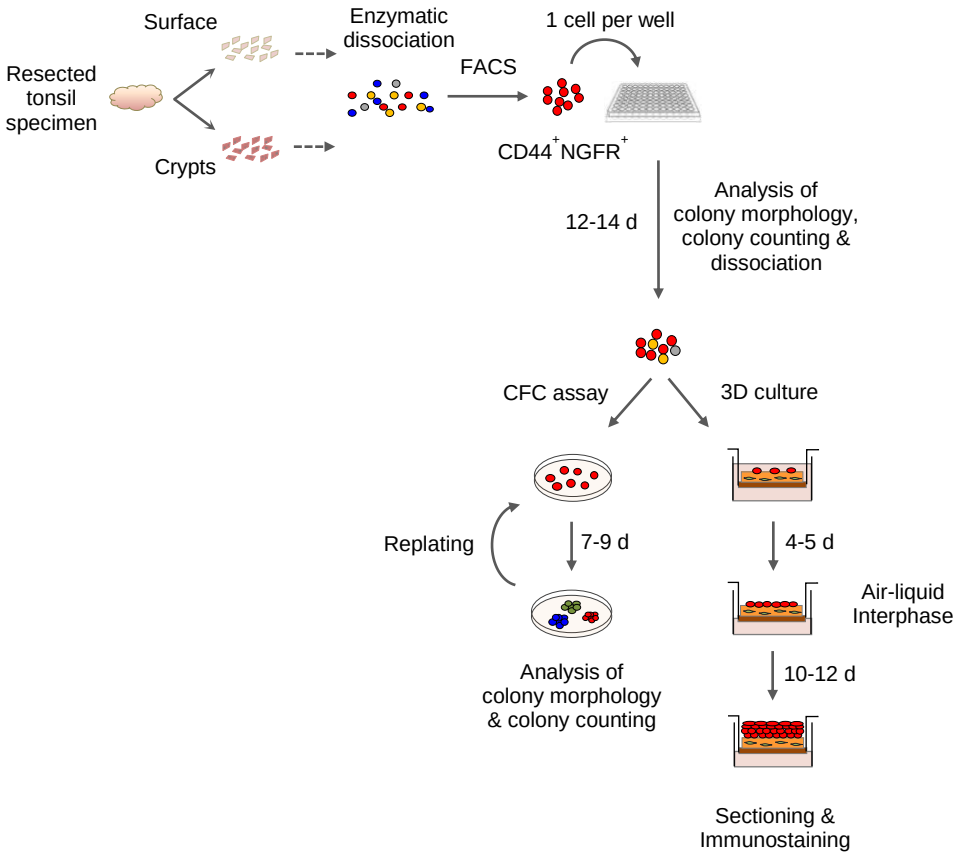


Figure 3-14. Experimental design for comparing the long-term proliferative and differentiation activity of surface- and crypt-derived CD44$^+$NGFR$^+$

CD44$^+$NGFR$^+$ cells were isolated from freshly dissociated crypt and surface epithelial cells and plated at single cell level in 96-well plates. After 12-14 days, the numbers of colonies as well as their morphology were recorded, followed by dissociation of the colonies and replating in 35 mm dishes to generate secondary colonies. The process was repeated every 7–9 days over a period of 6 weeks until all cells were terminally differentiated. A portion of the cells dissociated from the primary colonies were subjected to the 3D organotypic culture and allowed to differentiate.

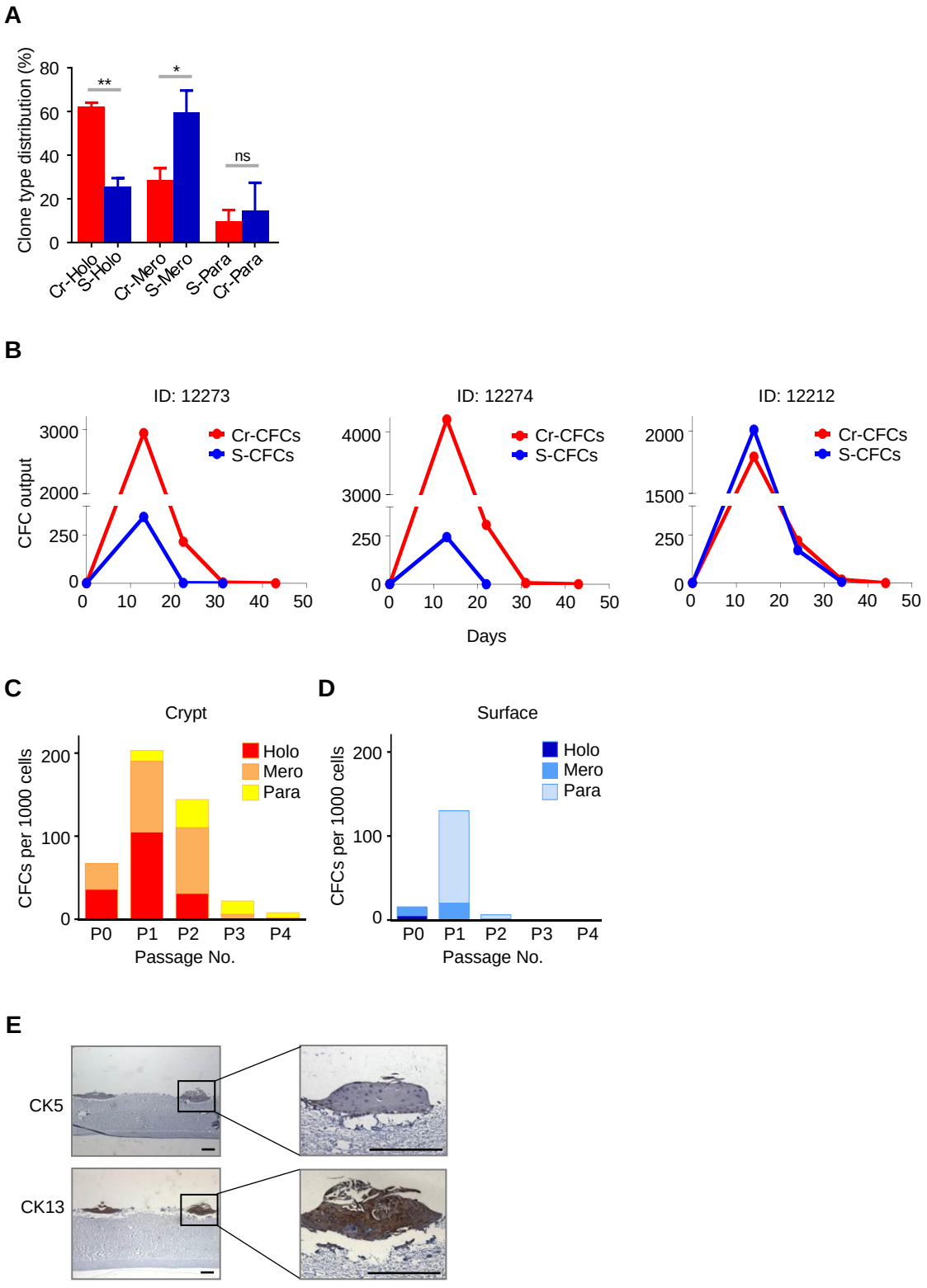

A
Clone type distribution (%)
80
60
40
20
0
**
*
ns
Cr-Holo
S-Holo
Cr-Mero
S-Mero
S-Para
Cr-Para
B
ID: 12273
ID: 12274
ID: 12212
Cr-CFCs
S-CFCs
CFC output
3000
2000
250
0
4000
3000
1500
0
10
20
30
40
50
Days
C
Crypt
D
Surface
Holo
Mero
Para
CFCs per 1000 cells
200
100
0
P0
P1
P2
P3
P4
Passage No.
E
CK5
CK13

Figure 3-15

Figure 3-15. Crypt-derived $CD44^+NGFR^+$ cells contain more primitive CFCs than their surface counterparts

(A) Proportion of single surface(S)- and crypt (Cr)-derived $CD44^+NGFR^+$ cells that produced holoclones (Holo), meroclones (Mero), and paraclones (Para) after 10–14 days of 2D cultures. Data were pooled from three independent experiments and expressed as means ± SEM. **P*<0.05, ***P*<0.01, ns: not significant, Student's t-test.

(B) CFC outputs in serially passaged CFC cultures initiated with single $CD44^+NGFR^+$ cells, showing the results from all 3 experiments.

(C-D) Changes in the proportion of holoclones, meroclones, and paraclones produced by the CFCs in harvests of serially cultured cells originating from initial $CD44^+NGFR^+$ crypt CFCs (C) and surface CFCs (D) obtained from donor 12273.

(E) IHC staining of $CK5^+$ and $CK13^+$ cells present in an organotypic 3D culture initiated from 1,400 cells produced from five crypt-derived CFCs. Scale bars, 200 µm.

Table 3-8. Summary of serial replating experiments initiated with a single tonsillar surface- or crypt-derived $CD44^+NGFR^+$ cell

Donor	Site[1]	% CFCs (Colonies/ Cells seeded)[2]	% Holo, mero, para[3]	Holoclone-forming cell per 100 CFCs[4]	Last passage (days)[5]	CFC output per 1 primary CFC[6]
12273	Cr	5.7 (11/192)	64, 36, 0	4	4th (43)	3,162
	S	1.6 (3/192)	33, 67, 0	0.5	3rd (31)	347
12274	Cr	11 (11/96)	64, 18, 18	7	4th (43)	4,516
	S	5.2 (5/96)	20, 40, 40	1	2nd (22)	246
12212	Cr	14 (27/192)	58, 32, 10	8	4th (44)	2,035
	S	13 (25/192)	24, 72, 4	2	3rd (34)	2,191

[1] Cr, Crypt; S, Surface

[2] The number of single cells that proliferated and formed a colony 11-15 days after being seeded into a culture plate at 1 cell per well of 96-well plates, with an exception to 12212 which was plated at 4 cells per well of 24-well plates

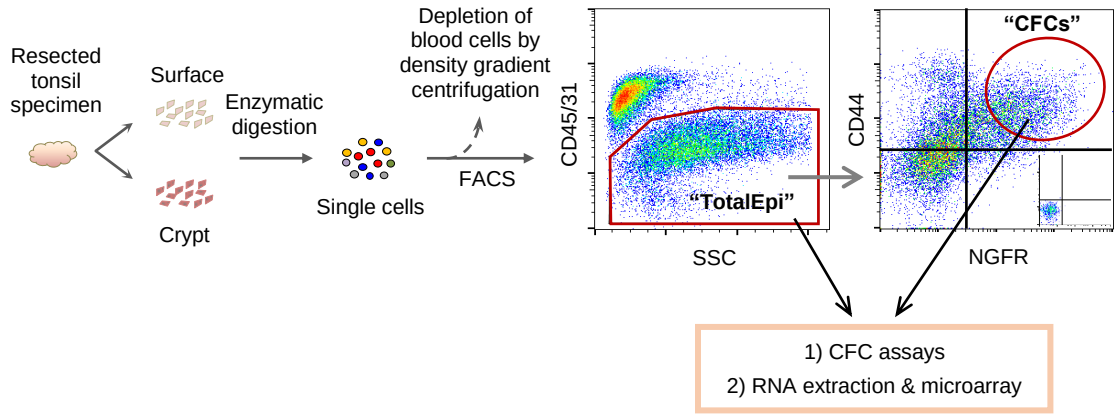
[3] The number of colonies of different types - holoclones, meroclones or paraclones.

[4] % CFCs X % holoclones / 100

[5] The number of passages carried out and the day when the last replating occurred.

[6] The sum total of the secondary, tertiary and quaternary CFCs generated by a single primary $CD44^+NGFR^+$ cell.

A



B

		CFC-enriched (CFCs)		Epi-enriched (TotalEpi)	
Donor (ID)	Age (Sex)	Crypt	Surface	Crypt	Surface
1 (12294)	7 (Male)	23.3%	20.7%	5%	8.2%
2 (12290)	20 (Male)	6.9%	3.8%	N/A	N/A
3 (12289)	24 (Male)	27.4%	17.3%	9.5%	3.2%

C

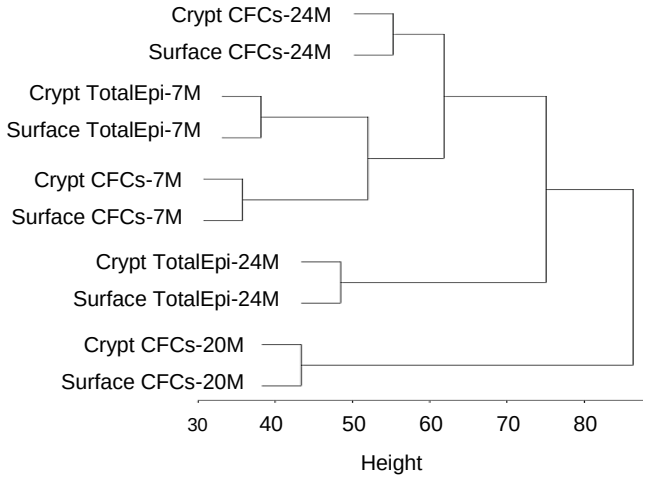
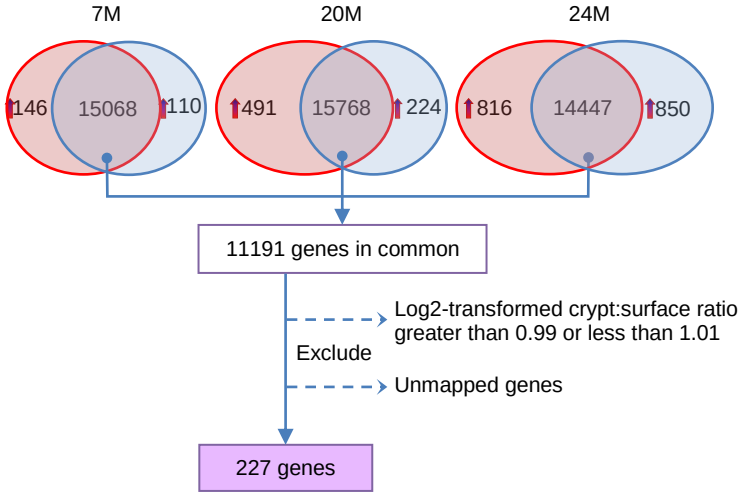


Figure 3-16. Transcriptome analysis of tonsillar epithelial cell subsets
(A) Representative FACS plots showing the subpopulations sorted for CFC assays and microarray analysis. $CD45^{-}CD31^{-}$ total epithelial cells (TotalEpi) and their $CD44^{+}NGFR^{+}$ subsets (CFCs) were isolated from each sample. (B) The age and sex of the donors, and colony-forming efficiencies of the freshly isolated subsets used for the transcriptome analyses. (C) Illumina transcriptome analyses identifying the relatedness of all four subsets of tonsillar cells derived from an unsupervised hierarchical clustering analysis. 24M, 24 year-old male; 7M, 7 year-old male; 20M, 20 year-old male.

A



B

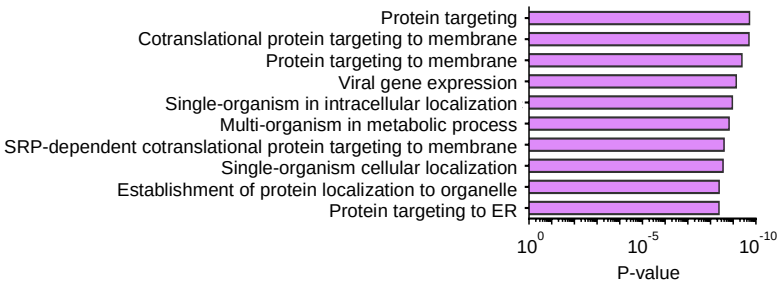


Figure 3-17. Identification of the site-independent gene expression signature of tonsillar progenitors

(A) Number of genes expressed in crypt CD44$^+$NGFR$^+$ cells and donor-matched surface TotalEpi subsets obtained from three donors. The number of genes more highly expressed in crypt or in surface by at least a 1.5-fold differential is shown in the red and blue areas, respectively. The number in the overlapping region represents the number of transcripts that displayed a <1.5-fold differential expression between the two sites. Those with crypt-to-surface ratios of log2-transformed values ≥0.99 and ≤1.01 were considered to be equally expressed and used for analysis. (B) The top ten biological processes associated with genes equally expressed in surface and crypt CD44$^+$NGFR$^+$ subsets based on gene ontology assignment. SRP, signal recognition particle. The full list of the 227 genes is shown in Table 3-9).

Table 3-9. The site-independent tonsillar progenitor signature

Gene symbol	Descriptions
ACTR3	ARP3 actin-related protein 3 homolog (yeast)
ACVR1	activin A receptor, type I
AK2	adenylate kinase 2 (AK2), transcript variant AK2B
AKT1S1	AKT1 substrate 1 (proline-rich)
ALKBH5	alkB, alkylation repair homolog 5 (E. coli)
AMPD3	adenosine monophosphate deaminase (isoform E)
ANKZF1	ankyrin repeat and zinc finger domain containing 1
AQR	aquarius homolog (mouse)
ARHGAP21	Rho GTPase activating protein 21
ARHGDIA	Rho GDP dissociation inhibitor (GDI) alpha
ARL6IP1	ADP-ribosylation factor-like 6 interacting protein 1
ATF6	activating transcription factor 6
ATG3	ATG3 autophagy related 3 homolog (S. cerevisiae
ATP2A2	ATPase, Ca++ transporting, cardiac muscle, slow twitch 2
ATXN2L	ataxin 2-like
AVPI1	arginine vasopressin-induced 1
B2M	beta-2-microglobulin
B4GALT4	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 4
BAG4	BCL2-associated athanogene 4
BCKDK	branched chain ketoacid dehydrogenase kinase
BET1L	blocked early in transport 1 homolog (S. cerevisiae)-like
BEX4	BEX family member 4
BFAR	bifunctional apoptosis regulator
BSDC1	BSD domain containing 1
BUB1B	BUB1 budding uninhibited by benzimidazoles 1 homolog beta (yeast)
BUB3	BUB3 budding uninhibited by benzimidazoles 3 homolog (yeast)
C10ORF137	chromosome 10 open reading frame 137
C10ORF72	chromosome 10 open reading frame 72
C11ORF58	chromosome 11 open reading frame 58
C14ORF79	chromosome 14 open reading frame 79
C15ORF44	mo sapiens chromosome 15 open reading frame 44
C17ORF62	chromosome 17 open reading frame 62
C17ORF71	chromosome 17 open reading frame 71
C18ORF21	chromosome 18 open reading frame 21
C1ORF183	chromosome 1 open reading frame 183
C20ORF11	chromosome 20 open reading frame 11
C22ORF9	chromosome 22 open reading frame 9
C4ORF29	chromosome 4 open reading frame 29
C6ORF170	chromosome 6 open reading frame 170
C6ORF85	chromosome 6 open reading frame 85

Table continued on subsequent page.

C7ORF41	chromosome 7 open reading frame 41
CAP1	CAP, adenylate cyclase-associated protein 1 (yeast)
CARD6	caspase recruitment domain family, member 6
CCDC49	coiled-coil domain containing 49
CCL15	chemokine (C-C motif) ligand 15
CCNA2	cyclin A2 (CCNA2), mRNA.
CCRK	cell cycle related kinase
CDV3	mo sapiens CDV3 homolog (mouse), transcript variant 3
CHUK	conserved helix-loop-helix ubiquitous kinase
CKAP4	cytoskeleton-associated protein 4
CLEC4D	C-type lectin domain family 4, member D
CLIC1	chloride intracellular channel 1
CLUAP1	clusterin associated protein 1
COMMD9	COMM domain containing 9
CREBBP	CREB binding protein
CRY2	cryptochrome 2 (photolyase-like)
CSK	c-src tyrosine kinase
CSNK2B	casein kinase 2, beta polypeptide
CSRP1	cysteine and glycine-rich protein 1
DCTN6	dynactin 6
DDX47	DEAD (Asp-Glu-Ala-Asp) box polypeptide 47
DENND3	DENN/MADD domain containing 3
DGAT2	diacylglycerol O-acyltransferase homolog 2 (mouse)
DGCR2	DiGeorge syndrome critical region gene 2
DIAPH1	diaphanous homolog 1 (Drosophila)
DKFZP779L1853	mo sapiens misc_RNA
DLG5	discs, large homolog 5 (Drosophila)
DNAL4	dynein, axonemal, light chain 4
DRG1	developmentally regulated GTP binding protein 1
DUSP3	dual specificity phosphatase 3 (vaccinia virus phosphatase VH1-related)
EEF1A1	eukaryotic translation elongation factor 1 alpha 1
EEF1AL7	eukaryotic translation elongation factor 1 alpha-like 7
EEF2	eukaryotic translation elongation factor 2
EFNB1	ephrin-B1
EIF3H	eukaryotic translation initiation factor 3, subunit H
EPN1	epsin 1
ERF	Ets2 repressor factor
ESCO1	establishment of cohesion 1 homolog 1 (S. cerevisiae)
ETS1	v-ets erythroblastosis virus E26 oncogene homolog 1 (avian)
EXOSC4	exosome component 4
FBXW7	F-box and WD repeat domain containing 7
FLJ14107	mo sapiens hypothetical protein FLJ14107
FLJ32955	hypothetical protein FLJ32955
FLJ45032	similar to F40B5.2b
FOXO4	forkhead box O4

Table continued on subsequent page.

FRG1	FSHD region gene 1
FZD3	frizzled homolog 3 (Drosophila)
GAPVD1	GTPase activating protein and VPS9 domains 1
GARS	glycyl-tRNA synthetase
GDI2	GDP dissociation inhibitor 2
GFPT1	glutamine-fructose-6-phosphate transaminase 1
GNB2L1	guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1
GNB5	guanine nucleotide binding protein (G protein), beta 5
GORASP2	golgi reassembly stacking protein 2, 55kDa
GOSR1	golgi SNAP receptor complex member 1
GPN1	GPN-loop GTPase 1
GPN2	GPN-loop GTPase 2
GRWD1	glutamate-rich WD repeat containing 1
GSTP1	glutathione S-transferase pi
GTF3C1	general transcription factor IIIC, polypeptide 1, alpha 220kDa
HOOK3	hook homolog 3 (Drosophila)
HSD11B1L	hydroxysteroid (11-beta) dehydrogenase 1-like
IKZF2	IKAROS family zinc finger 2 (Helios)
INPP5B	inositol polyphosphate-5-phosphatase, 75kDa
ISCA1	iron-sulfur cluster assembly 1 homolog (S. cerevisiae)
KBTBD4	kelch repeat and BTB (POZ) domain containing 4
KHNYN	KH and NYN domain containing
KIAA0261	KIAA0261
KIAA1688	KIAA1688 protein
KL	klotho
KLHDC3	kelch domain containing 3
KLHDC5	kelch domain containing 5
KPNA4	karyopherin alpha 4 (importin alpha 3)
KPNB1	karyopherin (importin) beta 1
LARP1B	La ribonucleoprotein domain family, member 1B
LRP10	low density lipoprotein receptor-related protein 10
MAP3K7IP1	mitogen-activated protein kinase kinase kinase 7 interacting protein 1
MAP3K7IP3	mitogen-activated protein kinase kinase kinase 7 interacting protein 3
MAPK1IP1L	mitogen-activated protein kinase 1 interacting protein 1-like
MAPK3	mitogen-activated protein kinase 3
MAPK8	mitogen-activated protein kinase 8
MAPRE1	microtubule-associated protein, RP/EB family, member 1
MARVELD3	MARVEL domain containing 3
MED12	mediator of RNA polymerase II transcription, subunit 12 homolog (yeast)
MGC71993	similar to DNA segment, Chr 11, Brigham & Womens Genetics 0434 expressed
MLEC	malectin
MPEG1	mo sapiens macrophage expressed gene 1, transcript variant 1
MTDH	metadherin
MTR	5-methyltetrahydrofolate-homocysteine methyltransferase

Table continued on subsequent page.

NCDN	neurochondrin
NDUFB10	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 10, 22kDa
NENF	neuron derived neurotrophic factor
NRBP1	nuclear receptor binding protein 1
NSUN5C	NOP2/Sun domain family, member 5C
NUBP1	nucleotide binding protein 1
NUMA1	nuclear mitotic apparatus protein 1
NUMB	numb homolog (Drosophila)
OPTN	optineurin
OR4M2	olfactory receptor, family 4, subfamily M, member 2
PA2G4	proliferation-associated 2G4, 38kDa
PARS2	prolyl-tRNA synthetase 2, mitochondrial (putative)
PDCD7	programmed cell death 7
PDE8A	phosphodiesterase 8A
PEX11B	peroxisomal biogenesis factor 11 beta
PRPF38A	PRP38 pre-mRNA processing factor 38 (yeast) domain containing A
PSEN1	presenilin 1 (Alzheimer disease 3)
PTAR1	protein prenyltransferase alpha subunit repeat containing 1
PTBP1	polypyrimidine tract binding protein 1
PTPN9	mo sapiens protein tyrosine phosphatase, non-receptor type 9
RABGEF1	RAB guanine nucleotide exchange factor (GEF) 1
RAD23A	RAD23 homolog A (S. cerevisiae)
RAF1	v-raf-1 murine leukemia viral oncogene homolog 1
RBM15B	RNA binding motif protein 15B
RBM42	RNA binding motif protein 42
RFX5	regulatory factor X, 5 (influences HLA class II expression)
RGL2	ral guanine nucleotide dissociation stimulator-like 2
RNASEK	ribonuclease, RNase K
RNASEN	ribonuclease III, nuclear
RNF38	ring finger protein 38
RPL18	ribosomal protein L18
RPL18A	ribosomal protein L18a
RPL39	ribosomal protein L39
RPL5	ribosomal protein L5
RPL6	ribosomal protein L6
RPLP2	ribosomal protein, large, P2
RPS10	ribosomal protein S10
RPS16	ribosomal protein S16
RPS19	ribosomal protein S19
RPS26	ribosomal protein S26
RPS27	ribosomal protein S27 (metallopanstimulin 1)
RPS27A	ribosomal protein S27a
RPS3A	ribosomal protein S3A
SAR1B	SAR1 homolog B (S. cerevisiae)

Table continued on subsequent page.

SARS	seryl-tRNA synthetase
SC65	synaptonemal complex protein SC65
SCAMP1	secretory carrier membrane protein 1
SCARNA10	small Cajal body-specific RNA 10
SEC31A	SEC31 homolog A (S. cerevisiae)
SEC61A1	Sec61 alpha 1 subunit (S. cerevisiae)
SF3B5	splicing factor 3b, subunit 5, 10kDa
SHQ1	SHQ1 homolog (S. cerevisiae)
SLC25A6	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 6
SNHG11	small nucleolar RNA host gene 11 (non-protein coding)
SNRPB2	small nuclear ribonucleoprotein polypeptide B''
SNX32	sorting nexin 32
SOX18	SRY (sex determining region Y)-box 18
SRP54	signal recognition particle 54kDa
SRRM2	serine/arginine repetitive matrix 2
ST7L	suppression of tumorigenicity 7 like
SURF6	surfeit 6
SYAP1	synapse associated protein 1, SAP47 homolog (Drosophila)
TADA2B	transcriptional adaptor 2B
TGFA	transforming growth factor, alpha
THAP7	THAP domain containing 7
TIMM23	translocase of inner mitochondrial membrane 23 homolog (yeast)
TMBIM1	transmembrane BAX inhibitor motif containing 1
TMEM11	transmembrane protein 11
TMEM38A	transmembrane protein 38A
TNXA	tenascin XA pseudogene
TRAPPC6A	trafficking protein particle complex 6A
TRMT12	tRNA methyltransferase 12 homolog (S. cerevisiae)
TRNT1	tRNA nucleotidyl transferase, CCA-adding, 1
TROVE2	TROVE domain family, member 2
TSC22D4	TSC22 domain family, member 4
TSGA14	testis specific, 14
TTC23	tetratricopeptide repeat domain 23
TUBGCP4	tubulin, gamma complex associated protein 4
TXNDC11	thioredoxin domain containing 11
UBB	ubiquitin B
UBC	ubiquitin C
UBE2E1	ubiquitin-conjugating enzyme E2E 1 (UBC4/5 homolog, yeast)
UBQLN1	ubiquilin 1
UNKL	unkempt homolog (Drosophila)-like
UTS2	urotensin 2
VPS33A	vacuolar protein sorting 33 homolog A (S. cerevisiae)
VPS37B	vacuolar protein sorting 37 homolog B (S. cerevisiae)

Table continued on subsequent page.

WDR64	WD repeat domain 64
XAF1	XIAP associated factor 1
XRCC6	X-ray repair complementing defective repair in Chinese hamster cells 6
YY1AP1	YY1 associated protein 1
ZFP36	zinc finger protein 36, C3H type, homolog (mouse)
ZFYVE9	zinc finger, FYVE domain containing 9
ZMYM3	zinc finger, MYM-type 3
ZNF259	zinc finger protein 259
ZNF438	zinc finger protein 438
ZNF469	mo sapiens zinc finger protein 469
ZNF649	zinc finger protein 649

Table 3-10. Gene Ontology (GO) biological processes associated with genes equally expressed in surface and crypt CD44^{+}NGFR^{+} subsets

GO biological process	No. of genes			P value
	Per GO category	Expected	Found	
protein targeting	400	3.82	26	1.90E-10
cotranslational protein targeting to membrane	111	1.06	16	2.01E-10
protein targeting to membrane	163	1.56	18	4.10E-10
viral gene expression	121	1.16	16	7.37E-10
single-organism intracellular transport	1126	10.77	41	1.08E-09
multi-organism metabolic process	127	1.21	16	1.52E-09
SRP-dependent cotranslational protein targeting to membrane	109	1.04	15	2.54E-09
single-organism cellular localization	801	7.66	34	2.81E-09
establishment of protein localization to organelle	346	3.31	23	4.16E-09
protein targeting to ER	113	1.08	15	4.22E-09
establishment of protein localization to endoplasmic reticulum	117	1.12	15	6.87E-09
protein localization to organelle	512	4.9	27	7.52E-09
cytoplasmic transport	695	6.64	31	9.97E-09
establishment of protein localization to membrane	244	2.33	19	3.36E-08
protein localization to endoplasmic reticulum	136	1.3	15	5.61E-08
intracellular transport	1319	12.61	41	1.56E-07
viral life cycle	241	2.3	18	2.45E-07
mRNA metabolic process	564	5.39	26	3.79E-07
viral transcription	111	1.06	13	7.03E-07
protein localization	1671	15.98	45	1.47E-06
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	118	1.13	13	1.47E-06
protein transport	1204	11.51	37	2.45E-06
protein localization to membrane	355	3.39	20	2.57E-06
nuclear-transcribed mRNA catabolic process	180	1.72	15	2.61E-06
viral process	667	6.38	27	2.69E-06
multi-organism cellular process	672	6.42	27	3.16E-06
cellular localization	2131	20.37	51	4.21E-06
RNA catabolic process	219	2.09	16	4.33E-06
mRNA catabolic process	192	1.84	15	6.22E-06
intracellular protein transport	704	6.73	27	8.55E-06
nucleobase-containing compound catabolic process	341	3.26	19	8.70E-06
establishment of localization in cell	1775	16.97	45	9.87E-06
translational elongation	199	1.9	15	1.00E-05
cellular component organization or biogenesis	4958	47.4	86	1.07E-05
establishment of protein localization	1279	12.23	37	1.26E-05
translation	390	3.73	20	1.26E-05
cellular protein localization	1115	10.66	34	1.75E-05

Table continued on subsequent page.

cellular macromolecule localization	1123	10.74	34	2.09E-05
interspecies interaction between organisms	743	7.1	27	2.66E-05
symbiosis, encompassing mutualism through parasitism	743	7.1	27	2.66E-05
cellular macromolecule catabolic process	748	7.15	27	3.07E-05
macromolecule localization	1992	19.05	47	3.83E-05
cellular nitrogen compound catabolic process	382	3.65	19	5.36E-05
heterocycle catabolic process	382	3.65	19	5.36E-05
aromatic compound catabolic process	392	3.75	19	8.06E-05
cellular macromolecule metabolic process	6540	62.53	101	8.71E-05
translational termination	171	1.63	13	1.18E-04
cellular process	14066	134.48	169	1.51E-04
membrane organization	826	7.9	27	2.36E-04
cellular component organization	4836	46.24	81	2.46E-04
cellular protein metabolic process	3334	31.88	63	2.61E-04
organic cyclic compound catabolic process	423	4.04	19	2.65E-04
translational initiation	228	2.18	14	4.49E-04
cellular protein complex disassembly	201	1.92	13	7.50E-04
macromolecule catabolic process	879	8.4	27	8.19E-04
gene expression	3778	36.12	67	9.31E-04
RNA metabolic process	3344	31.97	61	1.70E-03
protein complex disassembly	221	2.11	13	2.18E-03
macromolecule metabolic process	7232	69.14	104	2.34E-03
macromolecular complex disassembly	229	2.19	13	3.24E-03
single-organism membrane organization	663	6.34	22	3.85E-03
epidermal growth factor receptor signaling pathway	196	1.87	12	4.17E-03
ERBB signaling pathway	199	1.9	12	4.88E-03
localization	4644	44.4	75	5.19E-03
biological_process	16504	157.79	183	7.22E-03
cellular catabolic process	1377	13.17	33	8.08E-03
nucleic acid metabolic process	3862	36.92	65	1.07E-02
nucleobase-containing compound metabolic process	4349	41.58	70	1.77E-02
protein metabolic process	4005	38.29	66	1.80E-02
heterocycle metabolic process	4550	43.5	72	2.20E-02
neurotrophin TRK receptor signaling pathway	275	2.63	13	2.39E-02
neurotrophin signaling pathway	279	2.67	13	2.79E-02
toll-like receptor 10 signaling pathway	65	0.62	7	2.87E-02
toll-like receptor 5 signaling pathway	65	0.62	7	2.87E-02
cellular aromatic compound metabolic process	4553	43.53	71	4.61E-02
viral protein processing	12	0.11	4	4.85E-02

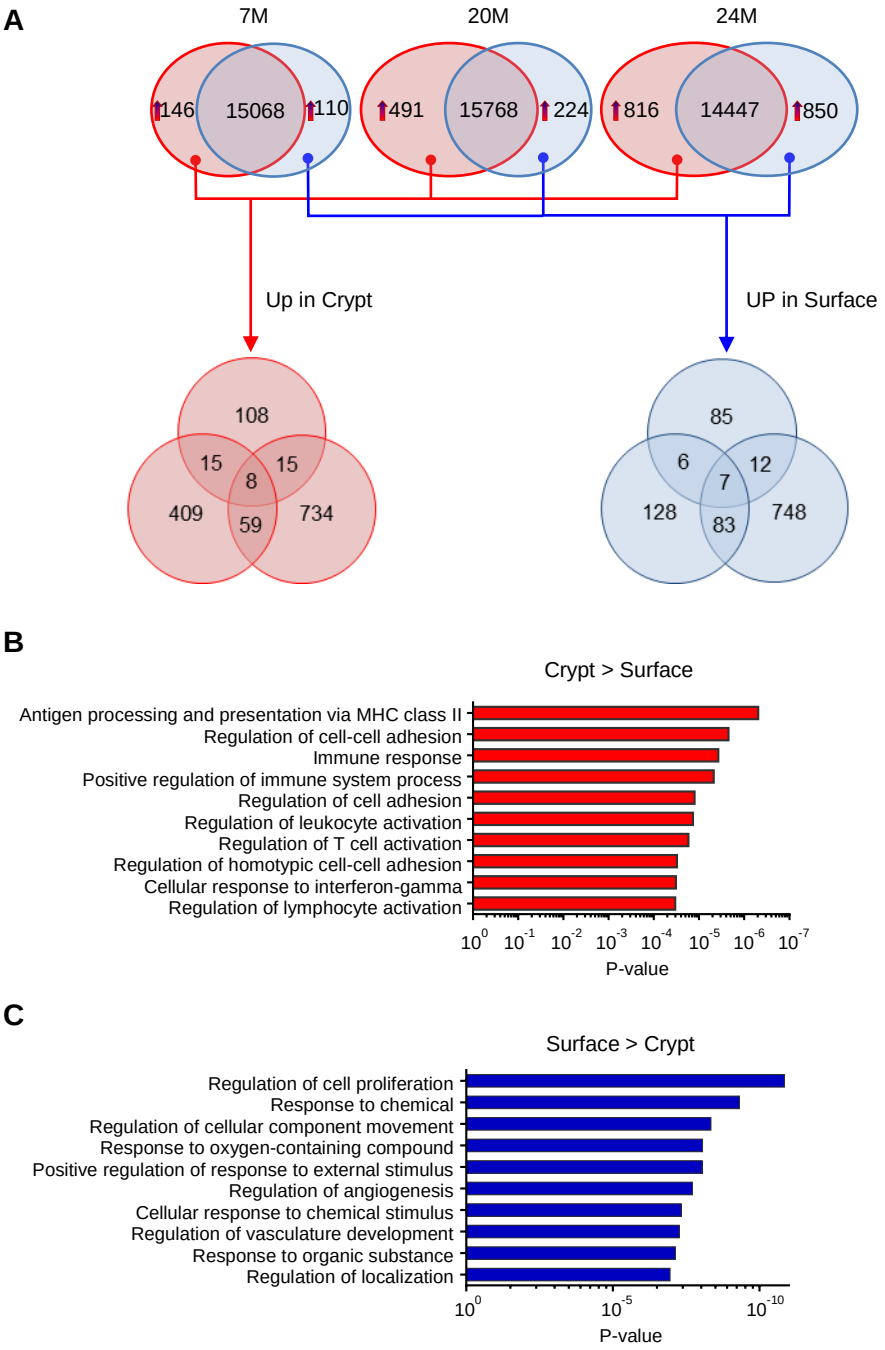


Figure 3-18. Identification of genes commonly differentially expressed between crypt and surface CD44$^+$NGFR$^+$ cells

(A) The Van diagrams shown in Figure 3-17A were further analyzed to identify the transcripts commonly overexpressed in crypt or surface progenitors obtained from two or more donors. The genes that overlap between two or more donors were pooled and categorized by Gene Ontology terms. Shown as bar graphs are top ten biological processes enriched in crypt CD44$^+$NGFR$^+$ with respect to their surface counterpart (78 of 97 genes mapped; Table 3-11) (B) and those enriched in their surface counterpart (93 of 108 genes mapped; Table 3-12) (C).

Table 3-11. Genes upregulated in crypt CFCs in comparison with their surface counterparts by ≥2-fold in 2 or more donors

Gene symbol	Description	Fold change (Cr-CFC/S-CFC)		
		7M	20M	24M
GLS	Glutaminase	2.1	2.5	
RPTN	Repetin	2.6		6.1
CD70	CD70 molecule		2.3	3.0
HLA-DMA	major histocompatibility complex, class II, DM alpha		3.5	2.3
HLA-DMB	major histocompatibility complex, class II, DM beta		2.5	2.0
HLA-DPA1	major histocompatibility complex, class II, DP alpha 1		2.8	2.1
HLA-DQA1	major histocompatibility complex, class II, DQ alpha 1		4.6	6.5
HLA-DRA	major histocompatibility complex, class II, DR alpha		2.8	3.2
HLA-DRB3	major histocompatibility complex, class II, DR beta 3		2.8	3.0
IL27RA	interleukin 27 receptor, alpha		2.0	2.3
LAYN	Layilin		2.1	2.8
LTF	Lactotransferrin		3.0	2.1
VSNL1	Visinin-like 1		2.3	2.3

Table 3-12. Genes upregulated in surface CFCs in comparison with their crypt counterparts by ≥2-fold in 2 or more donors

Gene symbol	Description	Fold change (S-CFC/Cr-CFC)		
		7M	20M	24M
CLDN5	claudin 5	2.5	2.3	
LPXN	Leupaxin	2.0		2.8
PTGS2	prostaglandin-endoperoxide synthase 2	2.1		3.0
RGS2	regulator of G-protein signaling 2	2.6		2.5
AGR2	anterior gradient homolog 2		2.6	2.3
CCL8	chemokine (C-C motif) ligand 8		2.1	2.8
CCNA1	cyclin A1		2.3	2.3
CXCL5	chemokine (C-X-C motif ligand 5		3.0	4.3
DEFB103B	defensin, beta 103B		2.5	2.1
GNG11	guanine nucleotide binding protein (G protein, gamma 11)		2.1	2.0
HSPA6	heat shock 70kDa protein 6 (HSP70B')		2.6	4.6
HSPA7	heat shock 70kDa protein 7 (HSP70B)		3.0	3.7
IL24	interleukin 24		2.0	3.2
LIF	leukemia inhibitory factor (cholinergic differentiation factor)		2.3	6.5
MGST1	microsomal glutathione S-transferase 1		2.6	3.5
NPTX2	neuronal pentraxin II		3.5	3.7
NQO1	NAD(P)H dehydrogenase, quinone 1		2.6	2.3
PRR4	proline rich 4 (lacrimal)		8.0	2.3
RGS5	regulator of G-protein signaling 5		2.5	3.0
SCGB2A2	secretoglobin, family 2A, member 1		9.2	2.6
SCGB3A1	secretoglobin, family 3A, member 1		5.3	2.0
STC1	stanniocalcin 1		2.6	4.0
TFF1	trefoil factor 1		4.0	2.3
TFF3	trefoil factor 3		9.8	2.8
THBS4	thrombospondin 4		2.8	2.3
HMOX1	heme oxygenase (decycling) 1	2.1	7.0	7.0

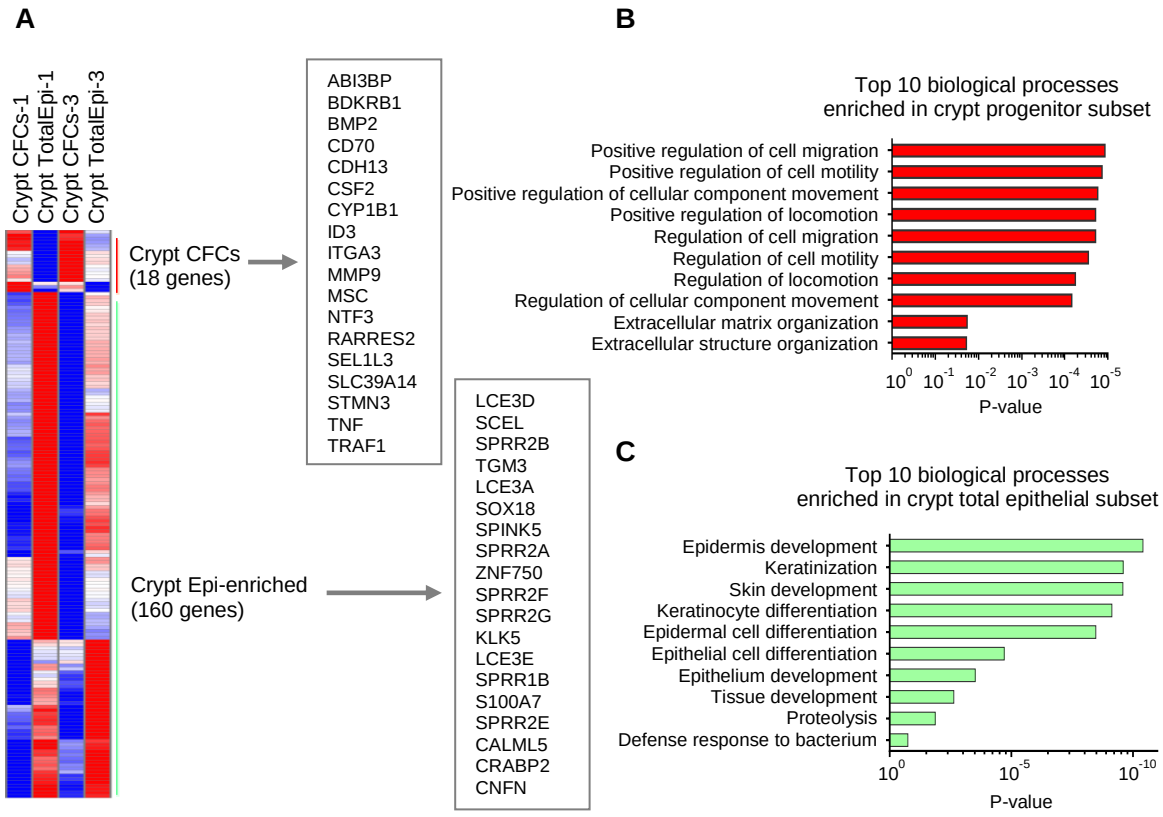


Figure 3-19. Differentially expressed genes between *crypt*-derived CD44$^+$NGFR$^+$ subsets (CFCs) and their donor-matched unseparated epithelial subsets (TotalEpi)

(A) Hierarchical cluster analysis of CD44$^+$NGFR$^+$ cells and their respective TotalEpi subsets generated by identifying the genes that displayed a 2-fold or higher differential expression in the Illumina data comparisons. Shown in red are genes that were highly overexpressed in CD44$^+$NGFR$^+$ cells in comparison with the TotalEpi populations. Shown in green are those more highly expressed in the two TotalEpi populations included in this comparison (n=2, P<0.05). (B) The top ten biological GO categories found to be upregulated in crypt CD44$^+$NGFR$^+$ cells compared with the TotalEpi populations. (C) The top ten biological GO categories upregulated in the TotalEpi populations compared with CD44$^+$NGFR$^+$ crypt cells.

Table 3-13. Genes that commonly displayed a 2-fold or higher differential expression in tonsillar crypt-derived CFCs in comparison with their TotalEpi counterpart

Gene symbol	Description	Fold change (Cr-CFC/TotalEpi)	
		Donor 1 (7M)	Donor 3 (24M)
ABI3BP	ABI Family, Member 3 (NESH) Binding Protein	2.5	3.5
BDKRB1	Bradykinin Receptor B1	2.0	3.2
BMP2	Bone Morphogenetic Protein 2	2.1	2.6
CD70	CD70 cytokine molecule (a ligand for TNFRSF27/CD27)	2.1	2.8
CDH13	Cadherin 13	2.0	2.8
CSF2	Colony Stimulating Factor 2 (Granulocyte-Macrophage)	2.1	2.3
CYP1B1	Cytochrome P450, Family 1, Subfamily B, Polypeptide 1	2.5	2.0
ID3	Inhibitor Of DNA Binding 3, Dominant Negative Helix-Loop-Helix Protein	2.1	3.0
ITGA3	Integrin, Alpha 3	2.0	3.0
MMP9	Matrix Metallopeptidase 9	2.3	2.3
MSC	Musculin	2.1	3.0
NTF3	Neurotrophin 3	2.0	2.6
RARRES2	Retinoic Acid Receptor Responder (Tazarotene Induced) 2	2.0	2.0
SEL1L3	Sel-1 Suppressor Of Lin-12-Like 3	2.0	2.8
SLC39A14	Solute Carrier Family 39 (Zinc Transporter), Member 14	2.1	2.1
STMN3	Stathmin-Like 3	2.0	3.5
TNF	Tumor Necrosis Factor	2.3	3.7
TRAF1	TNF Receptor-Associated Factor 1	2.0	2.1

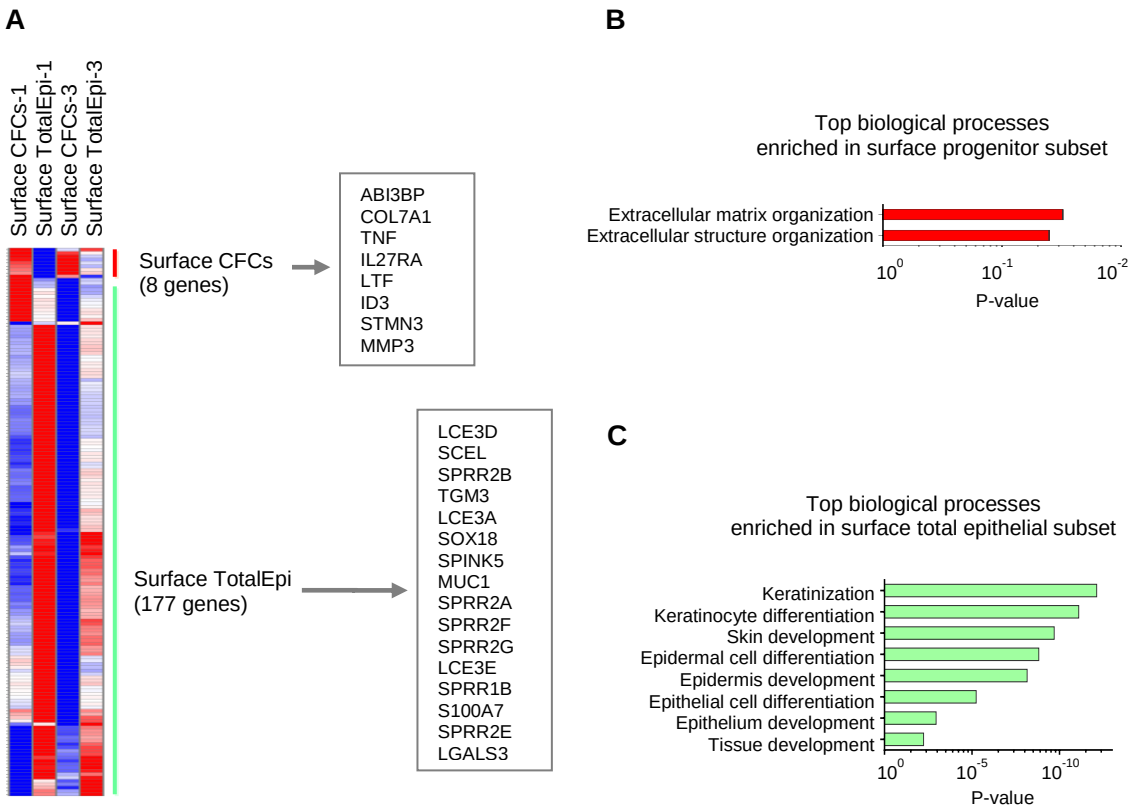


Figure 3-20. Differentially expressed genes between *surface*-derived CD44⁺NGFR⁺ subsets (CFCs) and their donor-matched unseparated epithelial subsets (TotalEpi)

(A) Hierarchical cluster analysis of CD44⁺NGFR⁺ cells and their respective TotalEpi subsets generated by identifying the genes that dis- played a 2-fold or higher differential expression in the Illumina data comparisons. Shown in red are genes that were highly overexpressed in CD44⁺NGFR⁺ cells in comparison with the TotalEpi populations. Shown in green are those more highly expressed in the two TotalEpi populations included in this comparison (n=2, P<0.05). (B) The biological GO categories found to be significantly upregulated in surface CD44⁺NGFR⁺ cells compared with the TotalEpi populations (P<0.05). (C) The biological GO categories significantly upregulated in the TotalEpi populations compared with CD44⁺NGFR⁺ crypt cells (P<0.05).

Table 3-14. Genes that commonly displayed a 2-fold or higher differential expression in tonsillar surface-derived CFCs in comparison with their Total-Epi counterpart

Gene symbol	Description	Fold change (S-CFC/TotalEpi)	
		Donor 1 (7M)	Donor 3 (24M)
ABI3BP	ABI Family, Member 3 (NESH) Binding Protein	2.5	4.3
COL7A1	Collagen, Type VII, Alpha 1	2.1	4.3
ID3	Inhibitor Of DNA Binding 3, Dominant Negative Helix-Loop-Helix Protein	2.0	2.8
IL27RA	Interleukin 27 Receptor, Alpha	2.3	2.8
LTF	Lactotransferrin	2.3	3.5
MMP3	Matrix Metallopeptidase 3	2.3	9.8
STMN3	Stathmin-Like 3	2.3	4.3
TNF	Tumor Necrosis Factor	2.3	4.3

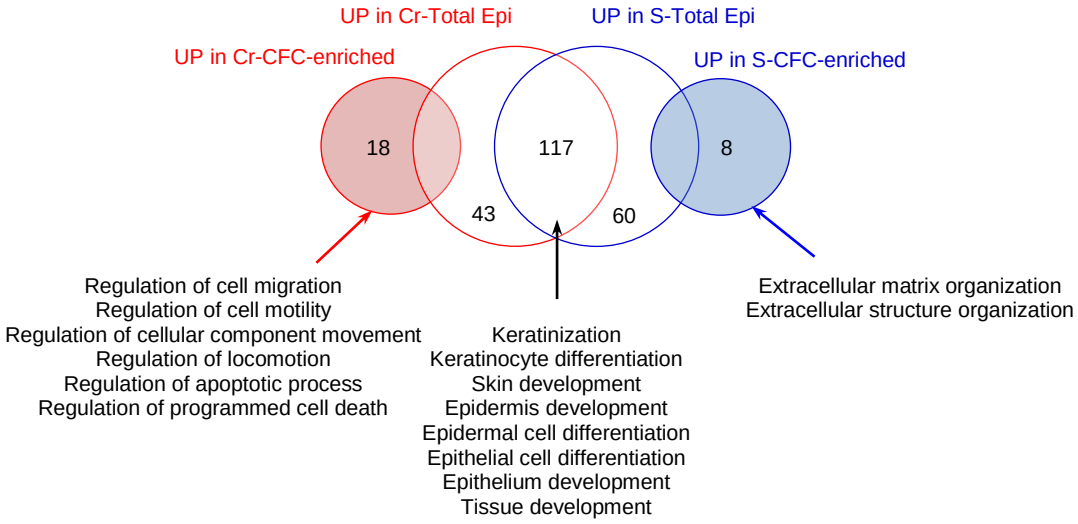


Figure 3-21. A summary of Figures 3-19A and 3-20A
The areas indicated by the red or blue arrows represent the transcripts that are more highly expressed in the $CD44^{+}NGFR^{+}$ crypt (Cr) or surface (S) cells, respectively, by >2-fold (compared to the TotalEpi cells). The region indicated by a black arrow shows the number of transcripts that are highly expressed in the TotalEpi subsets (compared to the $CD44^{+}NGFR^{+}$ cells) that are common to both the TotalEpi crypt and surface cells.

Table 3-15. List of 95 genes commonly overexpressed in TotalEpi subsets compared with their $CD44^{+}NGFR^{+}$ subset

Gene symbol	Descriptions
A2ML1	alpha-2-macroglobulin-like 1
ABLIM3	actin binding LIM protein family, member 3
AIF1L	allograft inflammatory factor 1-like
ALDH1A1	aldehyde dehydrogenase 1 family, member A1
ALOX12B	arachidonate 12-lipoxygenase, 12R type
ASPRV1	aspartic peptidase, retroviral-like 1
BCL6B	B-cell CLL/lymphoma 6, member B (zinc finger protein)
C6ORF132	mo sapiens chromosome 6 open reading frame 132
C9ORF169	chromosome 9 open reading frame 169
CAPN14	mo sapiens misc_RNA
CDH26	cadherin-like 26
CEACAM6	carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross reacting antigen)
CEBPA	CCAAT/enhancer binding protein (C/EBP), alpha
CLCA4	chloride channel, calcium activated, family member 4
CLIC3	chloride intracellular channel 3
CNFN	cornifelin
CPA3	carboxypeptidase A3 (mast cell)
CRCT1	cysteine-rich C-terminal 1
CRIP1	cysteine-rich protein 1 (intestinal)
CRNN	cornulin
CX3CL1	chemokine (C-X3-C motif) ligand 1
CYP3A5	cytochrome P450, family 3, subfamily A, polypeptide 5
DEFB1	defensin, beta 1
DEFB103A	defensin, beta 103A
DHRS9	dehydrogenase/reductase (SDR family) member 9
DKK4	dickkopf homolog 4 (Xenopus laevis)
DSC2	desmocollin 2
E2F2	E2F transcription factor 2
ECM1	extracellular matrix protein 1
EPN3	epsin 3
F2RL3	coagulation factor II (thrombin) receptor-like 3
FAM25A	mo sapiens family with sequence similarity 25, member A
FBXO32	F-box protein 32
FUT3	fucosyltransferase 3 (galactoside 3(4)-L-fucosyltransferase, Lewis blood group)
FUT6	fucosyltransferase 6 (alpha (1,3) fucosyltransferase)
GABRP	gamma-aminobutyric acid (GABA) A receptor, pi
GCNT3	glucosaminyl (N-acetyl) transferase 3, mucin type
GDF15	growth differentiation factor 15
GNG11	guanine nucleotide binding protein (G protein), gamma 11

Table continued on subsequent page.

IFI27	interferon, alpha-inducible protein 27
KLK12	kallikrein-related peptidase 12
KLK13	kallikrein-related peptidase 13
KLK5	kallikrein-related peptidase 5
KLK6	kallikrein-related peptidase 6
KLK9	kallikrein-related peptidase 9
KPRP	keratinocyte proline-rich protein
KRT4	keratin 4
KRT6C	keratin 6C
KRT78	keratin 78
KRT80	keratin 80
KRTDAP	keratinocyte differentiation-associated protein
LCE3A	late cornified envelope 3A
LCE3D	late cornified envelope 3D
LCE3E	late cornified envelope 3E
LCN2	lipocalin 2
LY6D	lymphocyte antigen 6 complex, locus D
LYZ	lysozyme (renal amyloidosis)
MAL	mal, T-cell differentiation protein
MMP7	matrix metallopeptidase 7 (matrilysin, uterine)
MUC20	mucin 20, cell surface associated
NCCRP1	non-specific cytotoxic cell receptor protein 1 homolog (zebrafish)
NDRG2	NDRG family member 2
PLA2G4B	phospholipase A2, group IVB (cytosolic)
PLAC8	placenta-specific 8
PLCXD1	phosphatidylinositol-specific phospholipase C, X domain containing 1
PROM2	prominin 2
PRSS22	protease, serine, 22
PRSS27	protease, serine 27
RHOV	ras homolog gene family, member V
RNASE7	ribonuclease, RNase A family, 7
RPTN	mo sapiens repetin
S100P	S100 calcium binding protein P
SBSN	suprabasin
SCEL	sciellin
SCNN1A	sodium channel, nonvoltage-gated 1 alpha
SERPINB13	serpin peptidase inhibitor, clade B
SIAE	sialic acid acetylesterase
SOX18	SRY (sex determining region Y)-box 18
SPINK7	serine peptidase inhibitor, Kazal type 7 (putative)
SPNS2	spinster homolog 2 (Drosophila)
SPRR1B	small proline-rich protein 1B (cornifin)
SPRR2A	small proline-rich protein 2A
SPRR2B	small proline-rich protein 2B

Table continued on subsequent page.

SPRR2E	small proline-rich protein 2E
SPRR2F	small proline-rich protein 2F
SPRR2G	small proline-rich protein 2G
TCP11L2	t-complex 11 (mouse)-like 2
TGM3	transglutaminase 3 (E polypeptide, protein-glutamine-gamma-glutamyltransferase)
TM4SF18	transmembrane 4 L six family member 18
TMEM40	transmembrane protein 40
TMPRSS11B	transmembrane protease, serine 11B
VWF	von Willebrand factor
XKRX	XK, Kell blood group complex subunit-related, X-linked
ZNF185	zinc finger protein 185 (LIM domain)
ZNF750	zinc finger protein 750

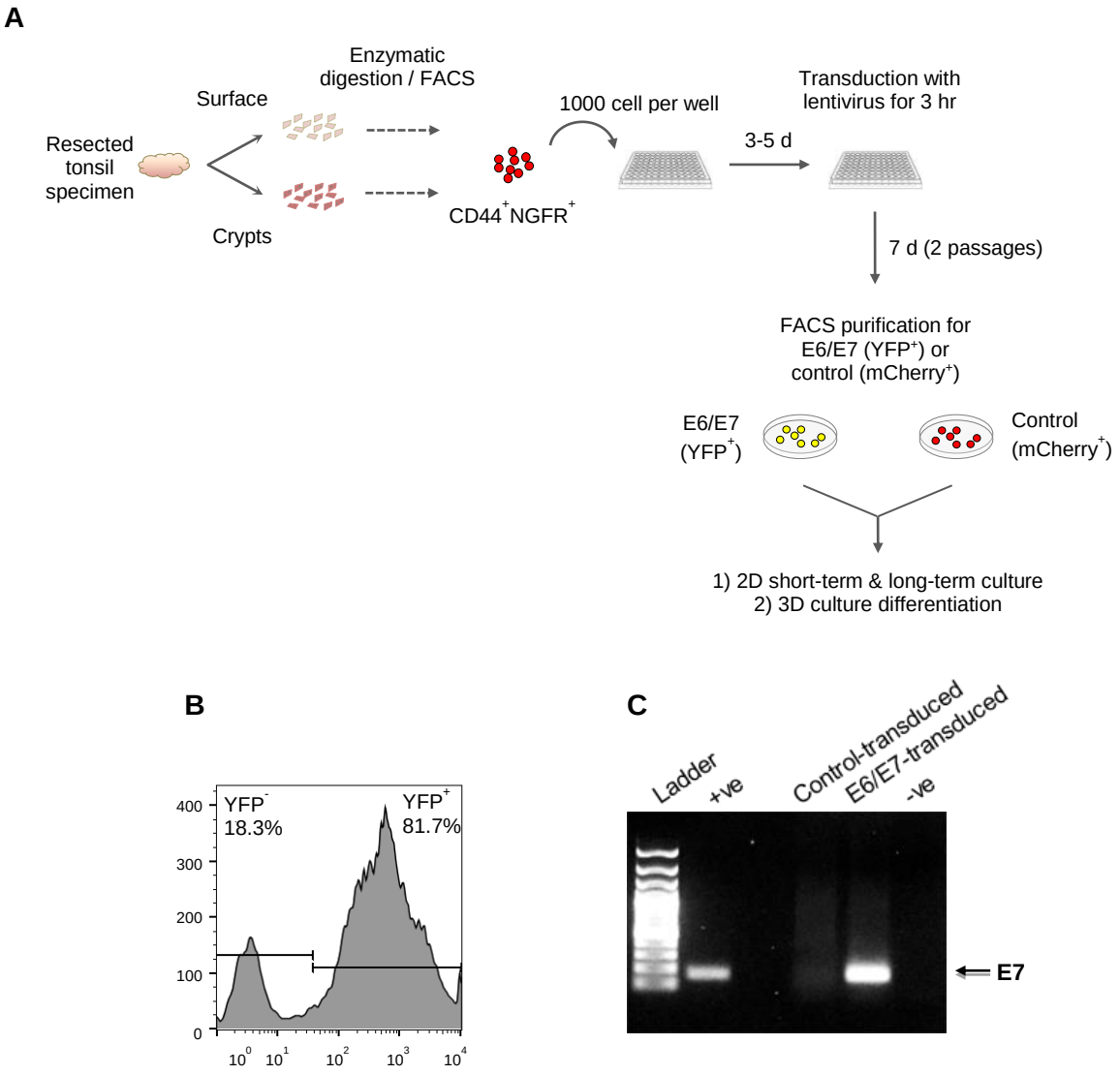


Figure 3-22. Experimental design for investigating effects of HPV16 E6/E7 oncoproteins on tonsillar progenitors

(A) Tonsillar surface and crypt tissues were independently digested, dissociated into single cells, and subsequently FACS-purified to isolate $CD44^+NGFR^+$ fractions. Cells were then directly seeded into 96-well plates at a density of 1000 cells per well. After 3-5 days in culture, cells were transduced with lentiviral vectors containing mCherry or E6/E7-YFP for 3 hours. Infected cells were expanded and then subjected to 2D *in vitro* culture (a short-term competition assay and long-term culture) or to organotypic 3D culture system (for differentiation). Both 2D and 3D cultures were performed using three different types of input cells: $mCherry^+$ alone, YFP^+ alone and 1:1 mixture of both.
(B) Confirmation of lentiviral transduction by FACS analysis of cells infected with lentiviral HPV16-E6/E7-YFP vectors on day 4 post-infection.
(C) PCR showing stable expression of Lenti-HPV16-E6/E7 in transduced crypt cells after 13 days in culture (3^{rd} passage). The ethidium bromide-stained bands represent the E7 gene. +ve, positive; -ve, negative.

Table 3-16 Effects of E6/E7 expression on proliferative activities of tonsillar progenitors

		Cells seeded ($x10^3$)	Output in 9 days ($x10^3$)	PD in 9 days[1]	Culture lifetime[2]
Crypt	Control	100	874	3.1	25 days
	E6/E7	100	9,568	6.6	>100 days
	Mix (1:1)	100 (50, 50)	3,504 (438, 2985)*	5.1 (3.1, 5.9)	ND
Surface	Control	ND			22 days
	E6/E7				>70 days
	Mix (1:1)				ND

[1]Population doubling was calculated as PD = 3.32 (log [number of cells harvested/ number of cells seeded])
[2]E6/E7-expressing cells cultured to over 100 days without any sign of senescence or reduction in growth rate were considered immortal and cryopreserved.
*The calculations were based on the FACS profile of the harvested mixed cell suspensions of which 85.2% were YFP^+ and 12.5% were $mCherry^+$.
ND, Not done

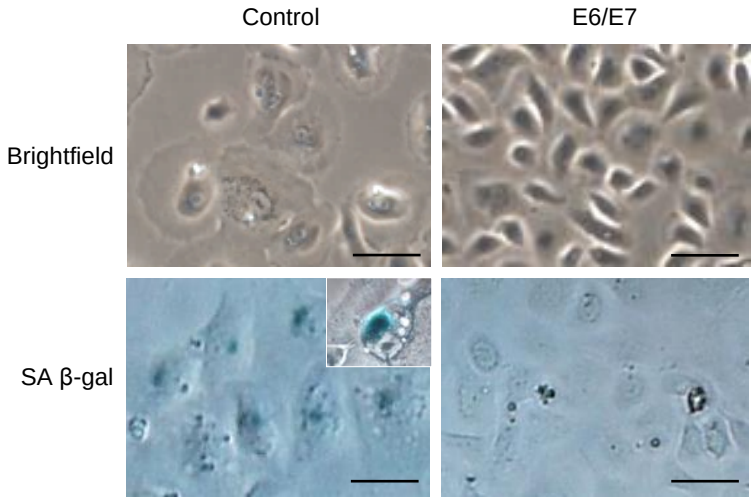


Figure 3-23. HPV16 E6/E7-induced delayed senescence of CD44^{+}NGFR^{+} cells *in vitro*

SA β-gal staining of the cells produced after 22 days in vitro under 2D CFC assay conditions by control and E6/E7-expressing cells. Scale bars, 50 µm.

Table 3-17. Summary of morphological and immunohistochemical features altered by HPV16 E6/E7 oncogprotein expression in organotypic 3D culture

		Epithelial thickness (layers)	CD44	CK19	CK4	Ki-67
Surface	Control	20-60	basal 2-3 +++	basal 1-3 +++	suprabasal +++	basal 1-3 +/++/+++
	E6/E7	10-20	basal 2-3 +++	**all layers** **+++**	suprabasal +++	**dispersed** +/++/+++
Crypt	Control	20-60	basal 2-3 +++	basal 1-3 +++	suprabasal +++	basal 2-3 +/++/+++
	E6/E7	10-16	**basal 3-6** +++	**all layers** **+++**	suprabasal **+**	**dispersed** +/++/+++

+, weakly expressed; ++, moderately expressed; +++, strongly expressed.

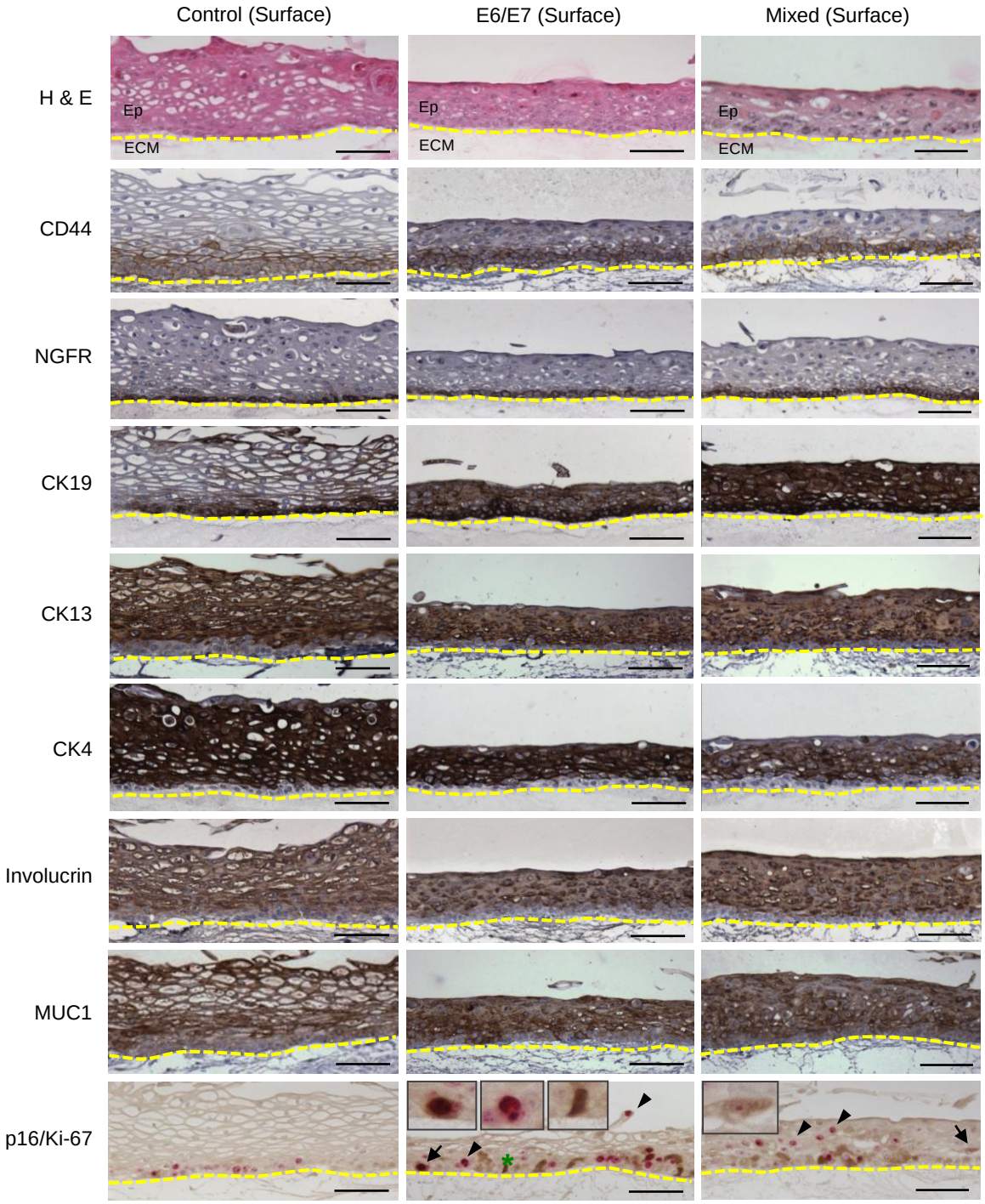


Figure 3-24. H & E and IHC staining of tissues harvested from tonsillar surface-derived 3D cultures of control cells, E6/E7-transduced cells or their 1:1 mixtures

Brown cells are indicative of the expression of CD44, NGFR, CK19, CK13, CK4 involucrin, MUC1 and p16. Ki-67-positive cells are stained red. Representative images of cells expressing both p16 and Ki-67 (arrow), solely Ki-67 (arrowhead), and solely p16 (asterisk) are shown in the insets. Scale bars, 100 µm.

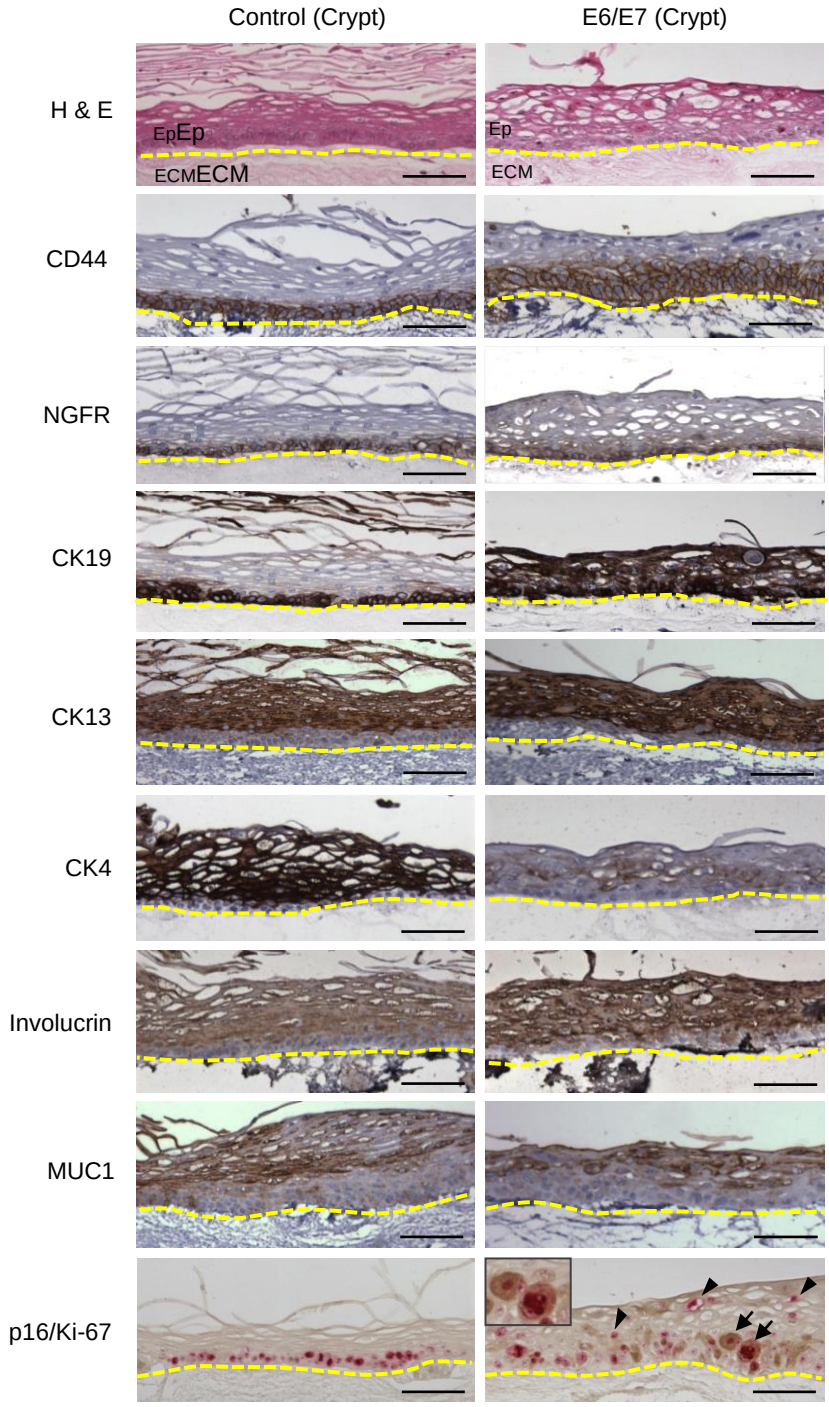


Figure 3-25. H & E and IHC staining of tissues harvested from tonsillar crypt-derived 3D cultures of control or E6/E7-transduced cells

Brown cells are indicative of the expression of CD44, NGFR, CK19, CK13, CK4 involucrin, MUC1 and p16. Ki-67-positive cells are stained red. Arrows indicate cells expressing both p16 and Ki-67 and are enlarged in the inset. Scale bars, 100 µm.

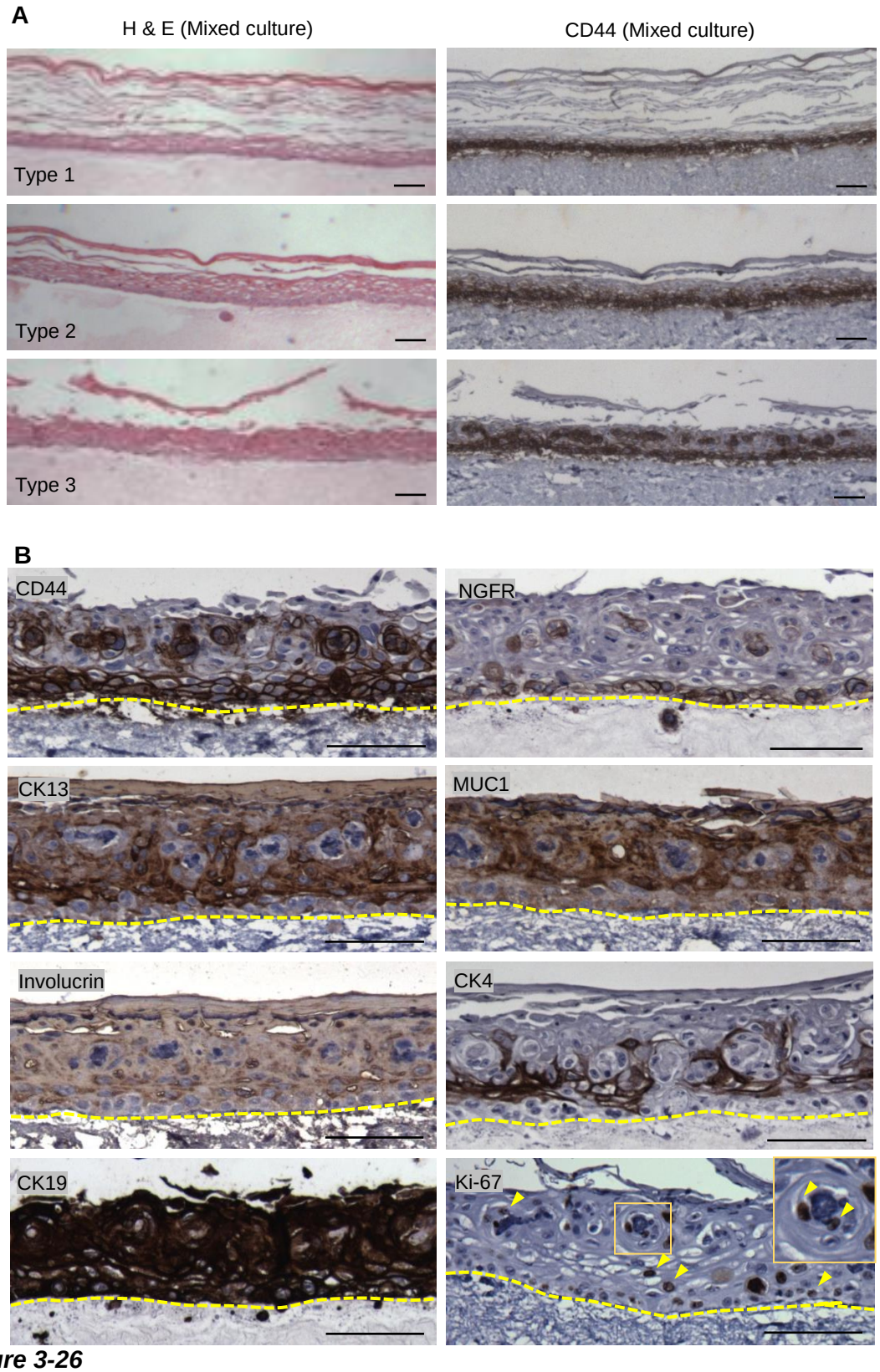

A
H & E (Mixed culture)
CD44 (Mixed culture)
Type 1
Type 2
Type 3
B
CD44
NGFR
CK13
MUC1
Involucrin
CK4
CK19
Ki-67

Figure 3-26

Figure 3-26. Tissue morphology and expression of IHC markers by 3D organotypic culture produced by crypt-derived mixed cells

Control and E6/E7-transduced crypt cells were mixed at 1:1 ratio and seeded on top of the extracellular matrix and allowed to differentiate for 14 days. (A) Areas of heterogeneous histological patterns were observed and classified into 3 types: Similar to control (Type 1), similar to E6/E7-transduced (Type 2) and unique only to the mixed culture (Type 3). (B) High magnification images showing expression patterns of CD44, NGFR, CK13, MUC1, involucrin, CK4, CK19 and Ki-67 in the areas displaying Type 3 morphology. Scale bar, 100 μm.

Table 3-18 Summary of Ki-67 expression in control, E6/E7-expressing and mixed 3D epithelial differentiation cultures

		Ki-67⁺ cells per field[1]	% Ki-67⁺ distribution			p16⁺Ki-67⁺[2] (%)
			Basal	Parabasal	Mid/Superficial	
Surface	Control	26 ± 7	82.4 ± 7.6	17.6 ± 7.6	0.0 ± 0.0	1/12 (8)
	E6/E7	30 ± 4	31.0 ± 5.1	35.8 ± 2.9	33.3 ± 4.3	18/25 (72)
	Mix	24 ± 4	35.0 ± 8.1	22.2 ± 5.5	42.6 ± 6.0	23/25 (92)
Crypt	Control	48 ± 3	71.3 ± 9.3	28.8 ± 9.3	0.0 ± 0.0	12/37 (32)
	E6/E7	34 ± 5	41.8 ± 5.7	31.4 ± 4.6	26.8 ± 3.1	15/18 (93)
	Mix	26 ± 3	52.1 ± 4.0	27.7 ± 1.6	20.4 ± 3.8	30/37 (81)

[1]The field contained a section of epithelium showing ~40 basal cells and a full thickness of the epithelium. At least five representative fields per slide were examined at 300x magnification and cells positive for Ki-67 were counted.

[2]The entire length of the embedded and immunostained epithelium was examined to identify cells showing dual staining for Ki-67 (nuclear) and p16 (cytoplasmic) throughout the epithelium. The data represent the number of microscopic fields (300x magnification) showing cells positive for both p16 and Ki-67 out of the total number of fields examined.

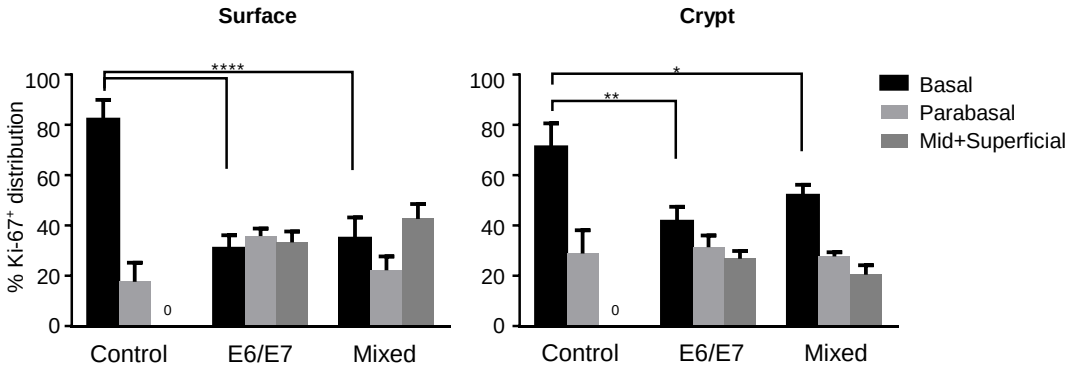


Figure 3-27. Distribution of Ki-67⁺ cells in different epithelial layer compartment of 3D culture tissues generated from controls, E6/E7-expressing, or mixed tonsillar progenitors

The number of Ki-67⁺ cells in each of the three epithelial compartments (a single basal layer, just above the ECM; Parabasal layers, two layers above the basal layer; Mid+Superficial layers, all layers above parabasal layers) was counted by examining images taken using 300x magnification and their relative proportions with respect to the total number of Ki-67⁺ cells per scored area were calculated. At least 5 independent fields per culture were scored. All data are expressed as means ± SEM. *P<0.05, **P<0.01, ****P<0.0001. Two-way ANOVA multiple comparisons test.

Chapter 4. Discussion and future directions

The epithelial cells in the crypts of human tonsils form a “reticulated” epithelium - a term that denotes their unique mesh-like organization in the crypt regions where they are interspersed with lymphocytes[256] although the degree of reticulation varies markedly between individuals[257]. The non-epithelial cells in the crypt are thought to induce multiple changes in the neighbouring epithelial cells, including a rearranged cytoskeleton of the tonofilaments as seen in squamous epithelial cell populations, increased numbers of ribosomes and mitochondria, redistributed desmosomal contacts, and changes in expression of keratin[25,258]. These effects have been described historically as “lymphoepithelial symbiosis”[259] and were thought to induce a loss of many of the typical properties of epithelial cells[25].

Mounting evidence suggests that HPV-associated head and neck carcinomas are distinct entity among the head and neck cancers, demonstrating atypical epidemiology, etiology and prognosis compared to tobacco-associated counterparts[194,260–262]. The majority of HPV-positive head and neck cancer are of tonsillar origin, with HPV16 DNA and p16 protein (a marker of transcriptionally active HPV E6/E7 oncogenes) frequently detected in tonsillar crypt regions[194,220]. The natural history of HPV-associated tonsillar cancer - how the disease arises and progresses to a life-threatening tumor - or even whether or not any distinct “premalignant” stages may exist still remains elusive due to the difficulty of detecting tonsillar malignancy in its early stages.

Progenitor cells are the life-sustaining unit that is responsible for maintaining integrity of tissues in a body. Their long-term regenerative ability, however, also places them at risk of accumulating mutations that may impact the fate of their progeny if left uncorrected or reversed. The stem cell hypothesis is also highly relevant from the standpoint of HPV-induced oncogenic transformation as it is generally thought that HPV preferentially targets stem cells in the epithelium[228]. To better understand abnormal

phenotypes associated with oncogenesis, it is essential to define the features of the "normal" progenitors that can be compared against the potential alterations of pathological significance. Nevertheless, very little is known about tonsillar progenitor cells, and whether those originating from different sites (surface or crypt) may differ in their growth and differentiation potential remains an interesting yet unanswered question.

The major objective of this thesis was therefore to develop a method to enrich for epithelial progenitor cells from both crypt and surface regions of normal human tonsils in order to characterize their functional, phenotypic and transcriptional profiles. The method was then used to investigate growth deregulation induced by lentiviral transduction of HPV16 E6/E7 oncogenes to provide insights into the early oncogenic changes.

4.1. IHC comparison of crypt and surface epithelium of human tonsils

Overall, the result from the immunohistochemical *in situ* localization of various proteins in non-malignant human tonsillar epithelium showed that cells basally located in the tonsillar surface epithelium are positive for CD44, NGFR, CK5, CK8/18, CK19 and p63, and negative or equivocal for MUC1, CK4, CK13 and involucrin (**Figure 4-1**). I have also found that the tonsil surface and crypt epithelia are differentiated by many of these markers, extending the previously reported analyses of immunophenotype of human tonsillar epithelium[243,257,263–266]. Expression patterns of CD44, NGFR, CK4 and CK13 remained relatively consistent between the surface and crypt epithelia (**Figure 4-1**)

Cytokeratins of simple epithelia, namely, CK8, CK18 and CK19, are often the first keratins to appear in embryogenesis and thus considered characteristic of precursor cells[267,268]. With the onset of epithelial stratification, other cytokeratins such as CK5/13/14 (keratins of stratified epithelia) are expressed in additional layers. Our analysis revealed that cytokeratins of simple epithlelia were either confined to the basal layers or weakly expressed in the surface epithelium, whereas they were more randomly and strongly expressed throughout the crypt epithelium, consistent with the previously reported observations[243,257,263]. CK19 is expressed in the basal layer of most epithelium including oral epithelium and is considered a marker of epithelial stem cells[269]. The

extent and intensity of CK19 has been positively correlated with an increasing degree of oral dysplasia and squamous cell carcinoma[270–274].

CK13 and its basic partner CK4 are major components of stratified non-cornifying epithelia including oral mucosa and are normally expressed by the suprabasal layers and absent in the basal layers[275,276]. Consistent with the generalization and the previous report on palatine tonsils[263], we show that CK13 and CK4 expressions are also confined to the suprabasal layers of the tonsil surface epithelium. Both antigens often displayed more intense staining pattern in the crypt epithelium (this study, ref. 263). CK14 was expressed in all tonsillar epithelial layers regardless of the location, consistent with previous reports of its ubiquitous expression in keratinocytes[275,277] as well as in tonsillar surface and crypt epithelia[278]. Clark et al., however, observed CK14 expression to be restricted to basal layers of the tonsil surface epithelium[263]. This difference might be attributed to the degree of lymphocyte infiltration or possibly other inter-individual variabilities.

The atypical expression of several basal and differentiation markers (including CK5, CK19, p63; CK8/18, involucrin, MUC1) in the crypt epithelium observed in our analysis suggests that the typical intrinsic differentiation program characteristic of stratified mucosal epithelium may not apply to the epithelium overlying the tonsillar crypts, consistent with the previously proposed “alternate differentiation pathway” model[25,243,257,263]. During the pre-natal development, ‘reticulation’ (i.e. architectural change to mesh- or spongy-like epithelial morphology) of tonsillar epithelium that was originally similar to oral epithelium begins upon infiltration of immune cells[19,244]. Since the palatine tonsils are immunological organ and a site for invasion and colonization for various pathogens, this constant disruption and interaction with non-epithelial components continue throughout a lifetime[279]. The change in the microenvironment accompanied by secretion of lymphokines and other factors lead to cellular morphological transformation as well as alterations to the expression of cytokeratins throughout the entire crypt epithelium[3].

Disruption of the epithelium covering tonsillar crypts is also evidenced by the discontinuous staining pattern of the basement membrane component collagen IV (this

study, refs. 15 and 260) as well as collagen type VII and XVII[280]. Using transmission electron microscopy and immunostaining for collagen IV, Choi et al. showed a gradual disruption of the basement membrane during the pre- and post-natal development by examining the palatine tonsils obtained from fetuses and neonates (intact basement membrane with pores), a 4-month-old infant (focal disruption of basement membrane) and children and adults (massive disruption)[16]. More detailed characterization of tonsillar epithelium based on the expression of tight junction proteins[281] revealed the expression pattern unique to the crypt epithelium, suggesting that the crypt epithelium may exhibit a mechanism for protection against the external insults distinct from the surface epithelium.

Little was known about surface antigen expression by the human tonsillar surface and crypt epithelia. Since one of the key objectives of the IHC analysis was to determine candidate markers that can be used to viably stain cells for the purpose of live cell sorting, I examined expression of transmembrane proteins that have been reported to be expressed in a differentiation-specific manner by other tissue models.

CD44 was expressed in the lower 1/3 to 1/2 of tonsillar epithelium with the strongest intensity detected in the basal layers in agreement with a number of previous studies on other stratified epithelia including skin[282], cervix[283], oral mucosa[284], esophagus[282] and urethra[246]. NGFR expression was strictly basal with 100% of cells in the basal layer shown to be $NGFR^+$ in the tonsil surface epithelium in agreement with the observations reported for oral mucosa[137,236,285], gingiva[286], esophagus[250]. In the crypt epithelium, in contrast, NGFR immunoreactivity was more widely distributed and often discontinuous, with a frequent cytoplasmic expression that was rarely observed in the surface epithelium. MUC1 was strongly expressed in most layers of tonsillar surface epithelium except for the basal/parabasal layers (either weak or negative), suggesting that basal cells are $CD44^+MUC1^{-/low}$.

Ki-67 has been widely used as a marker of proliferating cells evidenced by the observation that Ki-67 was continuously present during the cell cycle (G1, S, G2 and mitosis) but absent from quiescent or resting cells (G0)[287]. Given the slow cycling property of primitive cells[67,68], it was interesting to find that $NGFR^+$ cells and $Ki\text{-}67^+$ cells are almost completely mutually exclusive in tonsillar surface epithelium (**Figure 3-2**),

suggesting that most NGFR$^+$ cells do not cycle actively but rather divide infrequently *in vivo*. The relationship between NGFR and Ki-67 expression was less clear in the crypt epithelium, as the infiltrating lymphocytes intermingled with crypt epithelial cells were also Ki-67$^+$, and it was generally more difficult to identify basal layers in the crypts. Although NGFR/Ki-67 dual staining may provide more information, the presence of non-epithelial Ki-67$^+$ cells may still complicate the data interpretation. An alternative approach would be to purify different epithelial subsets by FACS based on their NGFR expression and stain each sorted fraction for Ki-67 to quantify the proportions of cycling cells.

4.2. Optimization of cell isolation and culture conditions

To develop optimal culture conditions to detect clonogenic tonsillar epithelial cells, I used an immunomagnetic or density centrifugation method to first deplete the large numbers of contaminating CD45$^+$ (primarily lymphoid) cells and CD31$^+$ cells. This resulted in some CFC losses but greatly increased their frequency. Further optimization was achieved using a combination of irradiated 3T3 cells, Y-27632 and 5% O_2 increasing the CFC frequency by 40-fold (~2.6%) compared with bulk cells cultured under suboptimal conditions.

The irradiated murine 3T3 fibroblasts are widely used in 2D *in vitro* epithelial culture to detect progenitors judged from their ability to form colonies on culture dishes after being plated at low cell densities[234,288]. In agreement with other tissue models, I found that feeders are indispensable for human tonsillar CFC assays as they allowed for colonies to grow as distinct clusters that can be distinguished from neighbouring colonies. Since the goal of the assay is to detect and quantify primary cells with proliferative potential based on their ability to attach to the culture surface and produce discreet colony composed of their progeny, careful attention should be paid to minimize colony merging that may lead to an inaccurate estimation of the actual number of CFCs. In the absence of feeders, cells appeared to migrate more readily and grow in a dispersed fashion. Although it was still possible to distinguish individual colonies as cells were seeded at low densities and colonies were enumerated before any merging

occurred, the use of feeders vastly improved the confidence and reproducibility in colony enumeration.

Initially introduced as a promoting agent for cloning efficiency of human embryonic stem cells, ROCK inhibitor (Y-27632) has been increasingly used in *in vitro* epithelial culture, including more recent utilization in progenitor assays studying prostate and mammary tissues[151,152]. In line with this, the present work shows that tonsillar CFCs are also sensitive to ROCK inhibition, evidenced by the >4-fold increase in plating efficiencies in the presence of ROCK inhibitor. Specific mechanisms by which this occurs have not been elucidated, although some studies propose that MYC may be responsible for the enhanced proliferation induced by Y-27632. When different types of keratinocytes from human foreskin and vaginal and cervical epithelium were cultured in the presence of 10 μM Y-27632, cells were immortalized (survived up to 150 passages for a period of 500 days) and a parallel increase in the levels of telomerase mRNA, shortened yet stable telomeres and MYC proteins were observed with continued passage[149], supported by a related study that showed a rapid upregulation of MYC mRNA expression in human keratinocytes by Y-27632[289]. Since MYC plays a positive role in keratinocyte proliferation as well as differentiation of epidermal stem cells[290], it is likely that ROCK inhibition may directly affect MYC activity, thereby promoting keratinocyte immortalization[149,289]. Interestingly, the keratinocytes immortalized by Y-27632 maintained normal p53-mediated growth arrest response as well as normal karyotypes, and differentiated normally in organotypic 3D culture[149]. Considering a number of downstream targets of ROCK, there likely are multiple pathways responsible for this phenomenon. Further work will be necessary to uncover specific mechanistic links between Rho-GTPase pathway and the enhanced progenitor activity.

Studies manipulating O_2 concentrations showed that lower tensions greatly affect both embryonic and adult stem cell biology[291–293]. More recently, low O_2 was shown to be critical for the detection of CFCs derived from the adult mouse mammary gland[152]. These observations fueled the hypothesis that a low O_2 tension might also enhance tonsillar epithelial colony formation. The significance of reduced O_2 concentration has been primarily suggested in the context of the hematopoietic, mesenchymal, and neural stem cell niche where O_2 tensions *in vivo* are considerably lower than ambient O_2

level[294], with 2%-9% O_2 concentrations considered as the physiologically normal range[295]. In 2D cultures, enhanced clonogenicity and cell proliferation as well as suppressed differentiation under hypoxia have also been documented in cultures of skin[296], corneal limbus[297] and oral epithelial cells[298]. Our data suggest the same may apply to tonsillar epithelial cells, although further investigation will be necessary to be more conclusive.

There are several limitations inherent to 2D *in vitro* CFC assays. First, since cells are cultured in an artificial environment, the extent to which the results can be extrapolated to the *in vivo* situation is limited. Second, the final readout of the CFC assay can be impacted by paracrine effects via various signals released from neighbouring cells. Also, high seeding densities may cause colonies to merge, possibly leading to misinterpretation of the actual CFC frequencies.

To minimize the technical variabilities, multiple seeding densities were initially tested to identify the range of input cell dose that are linearly related with colony output, and the upper limit of seeding densities was determined accordingly. With respect to the limited biological relevance, even though *in vitro* CFC frequency is not a direct measure of tonsillar progenitor cell frequency *in vivo*, it can serve as a relative measure of progenitor frequencies among different experimental samples, for example, those derived from different subpopulations (based on marker expression) or subsites (surface vs. crypt), as long as other experimental variables such as cell isolation protocol, culture conditions and reagents are consistent and taken into consideration. Furthermore, the fact that these cells are still clonogenic *in vitro* even after they have been removed from their natural microenvironment[299,300], producing different clonal types reflective of the varied self-renewal capacities of individual cells[126,127] indicate that a certain level of developmental hierarchy is preserved and retained in dissociated primary cells. Consistent with previous reports in skin epidermal[126,128–131,133], limbal[127] and oral keratinocytes[137], the 3 clonal types were also identified to comprise the proliferative compartment of the human tonsillar epithelium. Holoclones (i.e. tightly clustered colonies) were positive for basal markers CK19 and p63 and rarely expressed differentiation markers, involucrin, CK8/18 and CK13, whereas paraclones (i.e. colonies composed of dispersed cells) displayed the opposite expression pattern of these

markers. Meroclones were heterogeneous mixture of cells expressing basal markers and those expressing differentiation markers at varying levels.

A possible modification that could be adapted to increase the physiological relevance may be culturing cells in Matrigel, a membrane matrix derived from a mouse sarcoma line that is rich in basement membrane components such as laminin-1, collagen IV, perlecan, entactin and growth factors[301,302]. Matrigel has been used to improve conditions for growth of epithelial cells[88,102], and more recently, it was shown to substantially enhance the mammary repopulating frequency in transplantation assays when co-injected with test cells[152,303]. However, this approach should be taken with caution as multiple undefined factors may play a role in enhancing the progenitor activity. In fact, in the mammary system, it has been reported that even luminal cells that normally do not generate mammary outgrowth (the feature known to be unique to the mammary stem cells) give rise to mammary structures when transplanted in the presence of Matrigel, raising the question of whether the inclusion of Matrigel is more or less physiologically relevant[235]. Nevertheless, given that no mouse model is as yet available to study palatine tonsils *in vivo*[8,10], Matrigel is considered an attractive alternative that could be incorporated into quantitative assays to provide the microenvironment that may more closely mimic the *in vivo* niche.

4.3. Identification and isolation of discreet tonsillar cell subpopulations

Following the *in situ* IHC analysis of the candidate markers, CD44, NGFR and MUC1, and assay optimization, tonsillar epithelial subsets were then prospectively isolated by FACS and functionally characterized by subjecting each marker-defined subset to CFC assays.

We found that CFCs were most highly enriched in $CD44^{high}NGFR^{high}$ subsets (~27%) regardless of their anatomic subsites. Further analysis of CFC distribution indicated that, even though the CFC 'purity' was the highest in the $CD44^{high}NGFR^{high}$ subsets, a substantial proportion of CFCs were also $CD44^{low}NGFR^{low}$, hence, we designated all $CD44^{+}NGFR^{+}$ subpopulations as the progenitor-enriched (or CFC-

enriched) fraction. The surface and crypt-derived epithelial subsets shared similar FACS profiles and CFC distributions with a difference in their association with NGFR intensity; unlike the surface-derived CFCs shown to be more concentrated in the $NGFR^{high}$ fraction than in the $NGFR^{low}$ subset, the crypt CFCs were similarly distributed between $NGFR^{low}$ and $NGFR^{high}$ fractions. This is consistent with the *in situ* immunophenotype analysis, where NGFR was strongly and exclusively expressed in the basal layer of the surface epithelium, but only moderately expressed in the crypt epithelium and not always confined to the basal layers. Intracellular characterization of the progenitor subset revealed that $CD44^{+}NGFR^{+}$ cells are positive for basal cell markers CK19, CK5 and p63 irrespective of the site of origin, whereas differentiation markers CK4 and involucrin were highly expressed only in the crypt-derived $CD44^{+}NGFR^{+}$ cells.

The *in vitro* tissue regeneration potential has been demonstrated by subjecting different tonsillar epithelial subsets to 3D organotypic culture system. Both tonsillar surface and crypt progenitors were able to give rise to multilayered epithelium, with well-defined basal/parabasal cells ($CD44^{+}NGFR^{+}CK19^{+}$) and differentiated suprabasal cells ($MUC1^{+}Involucrin^{+}CK8/18^{+}CK13^{+}$) similar to tonsil surface epithelium *in situ*. The same number of unseparated $CD45^{-}CD31^{-}$ cells (1.25×10^{5} cells), however, failed to form a complete epithelial sheet, suggesting that the proportion of clongenic cells in the unseparated cells is very low, consistent with the finding that only ~3-5% of unseparated $CD45^{-}CD31^{-}$ cells formed colonies in 2D CFC assays.

Recognizing that MUC1 is expressed in the suprabasal layers of tonsillar epithelium *in situ*, I quantified CFC frequencies and distributions in four subpopulations defined by CD44 and MUC1 expressions, to determine whether removal of $MUC1^{+}$ cells (that are presumably differentiated/mature cells) from the total epithelial subsets may allow for further enrichment for tonsillar progenitors. The results showed that there was no difference in CFC frequencies between $CD44^{+}MUC1^{-}$ and $CD44^{+}MUC1^{+}$ subsets, suggesting that MUC1 may not be a useful marker if the purpose of prospective cell fractionation was to enrich for tonsillar progenitors. In terms of CFC distributions, ~90% of CFCs were in the $CD44^{+}$ fraction, ~80% were $MUC1^{-}$, and ~67% were $CD44^{+}MUC1^{-}$ (n=5, surface and crypt data pooled). The $CD44^{-}$ fractions, regardless of their MUC1

expression status, showed much lower CFC frequencies, again confirming the utility of CD44 as a marker of tonsillar progenitors.

On a technical note, since multiple rounds of staining and washing steps were required in order to avoid non-specific binding of the secondary fluorochrome targeting MUC1 antibody to other primary antibodies (e.g. streptavidin-conjugated CD45 and CD31), the longer staining procedure may have resulted in increased cell death, as suggested by FACS analysis of MUC1-stained cell suspensions showing a substantial proportion of PI^+ cells (dead cells) were $MUC1^+$. A possible modification may be using density gradient centrifugation as the initial debulking strategy in order to reduce the number of staining, washing and centrifugation steps that may negatively affect cell viability and quality.

A deeper interrogation of individual tonsillar CFCs was carried out by plating $CD44^+NGFR^+$ cells at a density of a single cell per well in 96-well plates. Under this condition, crypt-derived $CD44^+NGFR^+$ subsets were found to be more highly enriched for progenitors in comparison with their surface-derived $CD44^+NGFR^+$ counterpart (12.2% vs. 6.6%, n=3, $P<0.05$) contain higher proportion of holoclone-forming cells (62% vs. 26%, n=3, $P<0.05$) and can be passaged for a longer period of time (4 passages vs. 2-3 passage). Furthermore, when primary colonies were dissociated and passaged, even a 100-fold fewer cells than the number used for primary 3D culture were sufficient to produce multilayered epithelium, suggesting that passaged tonsillar CFCs are able to retain the ability to proliferate (colony formation) and differentiate (tissue regeneration), giving rise to both undifferentiated and differentiated cells even 2-3 weeks after the cells have been removed from their natural environment.

This combined approach (using both 2D CFC assays and 3D organotypic culture) has been used to characterize oral epithelial progenitor cells. One study showed that $CD49f^+CD29^+CD71^-$ keratinocytes isolated from human oral mucosa and sorted by magnetic sorting method expressed intracellular basal cell markers p63 and CK19 and were able to form epithelial multilayers when subjected to 3D organotypic culture[158]. The buccal and gingival epithelial cells expressing high level of the neurotrophin receptor p75 (NGFR) located in the tips of the dermal papillae and the rete ridges, when magnetically

isolated and subjected to 2D culture, displayed higher clonogenic efficiency than $p75^-$ subset (10.1% vs. 0.19%) and produced stratified epithelium *in vitro*[137]. The tongue epithelial cells sorted for CK5 from transgenic mouse expressing enhanced green fluorescent protein (eGFP) under the control of a CK5 promoter region ($Krt5\text{-}eGFP^{high}$; isolated by FACS) exhibited a high clonogenic potential evidenced by a generation of more holoclones (80% of the total number of colonies generated, with the report CFC frequency of 0.98%) compared with the subset lacking Krt5 expression (<20%, with CFC frequency of 0.4%), and also gave rise to differentiated epithelial tissues *in vitro*[159]. The findings from the present study that tonsillar progenitors are enriched in $CD44^+NGFR^+$ subset and that not only the tonsillar progenitor cells derived from the surface epithelium but also those from the crypt epithelium that is normally disrupted and *in vivo* is able to form a well-stratified squamous cell epithelium typical of oral mucosa are a novel addition to the existing literature.

4.4. Transcriptome profiling of tonsillar epithelial cells

There are 3 key findings from the transcriptome analyses of tonsil epithelial subsets presented in this dissertation. First, we found that tonsillar $CD44^+NGFR^+$ epithelial progenitors isolated from the crypt epithelium share common transcriptome profiles with surface-derived counterparts (**Table 4-1**). Second, although surface- and crypt-derived CFCs are highly similar with respect to their gene expression profiles, a few specific biological processes are enriched in a site-dependent manner; for example, transcripts associated with immunological functions are more highly expressed in the crypt CFCs. Third, the comparison between $CD44^+NGFR^+$ subsets and total epithelial cells reveal that the genes that were more highly expressed in the crypt $CD44^+NGFR^+$ cells were linked to cell migration and motility, whereas the same enrichment was not observed in the surface-derived subsets. Most of the genes that were upregulated in the total epithelial cells (associated with epithelial differentiation and development) were common between the two sites.

There has only been a single reported study on transcriptome profiles of tonsillar epithelium. Moutsopoulos et al.[304] conducted a comparative analysis of the tonsil *surface* epithelium and the gingival epithelium by isolating the respective epithelia by laser

capture microdissection (LCM) and generated microarray data by using an Affymatrix platform. Their study showed that many of the genes most highly expressed in both tonsil and gingival epithelium were common and associated with tissues with epithelial origins. Among them were keratin genes 6A, 13 and 14, the water channel protein aqua-porin-3, the keratinocyte factor stratifin, cystatins A and B, and S100 binding Ca^{+2}). In addition to the shared characteristics, a distinct set of genes upregulated in the tonsillar surface epithelium were identified and found to be associated with immunological functions according to GO categories, including defense response, response to biotic stimuli, immune response, and response to pathogens. This finding, combined with our data showing that the crypt CFCs are enriched with genes associated with immunological processes when compared with the surface CFCs, indicates that there is a gradual increase in the level of expression of immunologically related gene within the oral cavity and oropharynx (i.e. gingiva < tonsillar surface < tonsillar crypts).

Despite such a wide inter-individual variability of the CFC-enriched subsets, it was interesting to find that genes more highly expressed in the TotalEpi cells (compared with their respective CFC subset) were similar between the two donors of different age, with more than half of them shared between the two individuals (**Figure 3-21**). The fact that most of the genes upregulated in TotalEpi cells are not donor-specific, and yet those upregulated in CFC subsets are, suggests that perhaps the progenitors (and not the differentiated cells) evolve with time as they adapt to the dynamic tonsillar niche unique to each individual. Their progeny, however, still inherit a defined epithelial differentiation program so that the overall tissue integrity can be maintained in the manner consistent across different individuals. The idea that crypt and surface CFCs share similar developmental potentials was further supported by our observation that the CFCs isolated from either sources (surface or crypt) were able to generate stratified multilayered in organotypic 3D culture and appeared identical to typical oral mucosa and tonsillar surface epithelium (**Figures 3-12, 3-24 and 3-25**).

There are few points worth addressing moving forward. The two subsets that we have analyzed in this study, $CD44^{+}NGFR^{+}$ and $CD45^{-}CD31^{-}$, were not exclusive of each other as $CD44^{+}NGFR^{+}$ cells were a subset of $CD45^{-}CD31^{-}$ population, which may be why only a very few genes were found to be more highly expressed in the progenitor

subset in comparison with the total epithelial cells (18 for the crypt and 8 for the surface). To identify a broader and more significant list of differentially expressed genes that distinguishes progenitors from non-progenitor subsets, it will be necessary to analyze distinct epithelial subpopulations that do not overlap with one another. Such attempts have been successfully made in other epithelial models such as breast[305], prostate[306], skin[307] and airway epithelium[308], where specific set of genes and signaling pathways highly selectively enriched or downregulated in the stem/progenitor subpopulations have been identified.

4.5. Effects of E6/E7 oncoproteins on tonsillar progenitors

This last section of my thesis focused on the application of the progenitor enrichment methods and assays that I have developed in order to address several outstanding questions, including (1) how a forced expression of lentivirus-induced HPV16 E6/E7 oncoproteins in $CD44^{+}NGFR^{+}$ cells may affect the growth of tonsillar epithelial cells, (2) whether surface and crypt-derived progenitors may be affected differently by the E6 and E7 oncoproteins, and (3) whether coculturing E6/E7-expressing cells with uninfected cells may lead to the similar or unique alterations.

Using short-term and long-term 2D cultures, I showed that expression of HPV16 E6/E7 oncoproteins in tonsillar progenitors is indeed sufficient to delay the senescence and prolong their lifespan over 100 days (as opposed to 22-25 days for control cells) with over 10-fold greater total cell output compared with control cells in a 9-day 2D *in vitro* culture. Interestingly, the analysis of population doubling (PD) obtained from the 9-day output of the *mixed* culture revealed that E6/E7-transduced cells proliferated at a slower rate when cocultured with normal cells (6.6 PD when cultured alone vs. 5.9 PD when cocultured with control $mCherry^{+}$ cells), suggesting that normal cells may have an inhibitory effect on growth of E6/E7-expressing cells. Competitive interaction among cells are the basis of many homeostatic processes in biology[309,310], and has also been studied in the context of cancer, where Bourhis et al. showed in their pioneering study that non-cancerous cells (normal breast epithelia cells) cocultured with cancerous cells (MCF-7 cell line) inhibit growth of cancer cells *in vitro*[311]. Further investigation is warranted to elucidate the key factors and mechanisms that play an important role in the

intercellular communication that occur in a contact-dependent manner and -independent manner.

When subjected to 3D organotypic culture, the cultures initiated with E6/E7-transduced cells displayed abnormal, dysplastic growth with a lack of proper differentiation, evidenced by the strong, pan-epithelial expression of basal cell marker CK19, dispersed expression of the proliferation marker Ki-67 across the full thickness of the epithelium rather than exclusively to the basal/parabasal layers, the expansion of $CD44^+$ cells, and the weak expression of suprabasal cell marker CK4. The transcriptional activity of E6/E7 was confirmed by the presence of cells co-expressing p16 (cytoplasmic) and Ki-67 (nuclear) throughout the entire length of the tissues obtained from the 3D culture seeded with E6/E7-transduced (YFP^+) cells. Of note, the aberrant CD44 and CK4 expression patterns were only observed in the crypt-derived culture, and not the surface-derived. Although it is tempting to speculate that such site-dependent expression patterns may be attributed to intrinsic biological differences between the surface and crypt progenitors, further investigation and repeated experiments are warranted to confirm this possibility.

Multiple studies have shown the abnormal growth induced by expression of the E6 and E7 oncogproteins in human keratinocytes[199,200,312–314]; however, few have examined effects of co-culturing transformed cells with normal epithelial cells. In this study, when normal cells and the E6/E7-transformed cells were co-cultured in the 3D culture system, heterogeneous morphologies were observed, which could be categorized into 3 types: similar to normal differentiation patterns, dysplastic growth similar to E6/E7 culture, and a unique structure that resembles a cross section of rete ridge of the oral epithelium (**Figure 4-2**). In this last type (referred to as 'Type 3'), both the tissue morphology and IHC results suggest that these cells are in fact basal cells even though their intraepithelial location suggests otherwise (not basally located); these cells were positive for the markers of basal cells, CD44 and NGFR, and negative for differentiation markers, MUC1, CK4, CK13 and involucrin. Also, these basal-like cells did not express the proliferation-associated antigen Ki-67 but a few cells immediately adjacent to them were $Ki\text{-}67^+$, reminiscent of the tonsil surface epithelium *in situ* where $Ki\text{-}67^+$ cells are rare in the basal layer but frequently detected in the parabasal layers.

From the fact that this unique structure was seen in the mixed culture but not in the homogeneous E6/E7 culture, it may be speculated that the presence of this unique morphology may be attributed to the interaction between normal and E6/E7-transformed cells. Curiously, however, the same morphology was not observed in the mixed culture derived from tonsillar *surface* epithelium. Whether such structure was crypt-specific or whether it simply occurred by chance is unknown. Since this initial study, the mixed culture experiment has been repeated and the aberrant differentiation pattern (expansion of $CD44^+$, $CK19^+$ and $Ki\text{-}67^+$ cells) was again observed; however, the Type 3 morphology was not seen in this subsequent experiment. Possible explanations may be that the appearance of such structures is a rare event, or that it is a relatively frequent event but transient, representing a brief, yet critical time window that precedes pan-epithelial malignant transformation. Further investigation using a revised approach such as temporal analysis (fixing and visualizing them at different time points as they differentiate) may allow for capturing this particular alteration.

Attempts have been made to further characterize this Type 3 region in the context of E6/E7 expression by staining the tissue with antibodies against E6 and E7; however, IHC staining of these oncoproteins proved challenging in our hands, which has also been a well-recognized problem in the field (Dr. Miriam Reuschenbach, University of Heidelberg, personal communication). HPV DNA *in situ* hybridization (ISH) has been suggested as an alternative to determine whether these cells are E6/E7-transduced cells. This approach is currently under investigation.

With the recent and continued rise in the incidence of HPV-associated tonsil cancer globally and our inability to capture tonsillar premalignancy, a clear challenge in assessing the consequences of HPV infection in the tonsils is the lack of well-characterized markers that would enable successful detection as well as functional assays of tonsillar progenitor cells that are believed to be the target cells of HPV. The vast majority of the existing studies use culture-adapted epithelial cells such as cell lines or *in vitro*-expanded primary cells. Even though these cells do display the progenitor properties – the ability to proliferate and differentiate *in vitro* – it is difficult to assess whether they are developmentally so or whether they have acquired those characteristics as they adapted to the artificial culture environment. The ability to purify

primary cells at specific development stages and target them with HPV oncogenes before they are altered *in vitro* can serve as a powerful tool for better understanding the natural history of HPV-associated carcinogenesis.

4.6. Summary

The questions that I sought to address in this thesis were (1) Can we identify and isolate epithelial progenitor cells from human palatine tonsils? (2) Are the progenitors that maintain the surface and crypt epithelium intrinsically different? (3) What are the biological consequences of targeted manipulation of tonsillar progenitors with HPV oncogenes?

The questions were addressed majorly in 5 stages (1) Immunohistochemical *in situ* localization was first used to determine spatial distribution of various markers that allow for identification of specific subtypes of cells in tonsillar surface and crypt epithelia. (2) Different exogenous factors were tested to determine the optimal culture condition for quantitative 2D *in vitro* CFC assays. (3) Multiparameter FACS was used to prospectively isolate specific subsets defined by expression of the candidate makers (determined in stage I), followed by quantitation of their progenitor properties by using 2D CFC assays (optimized in stage II) and 3D culture differentiation system. (4) The purified subsets were molecularly characterized by transcriptome profiling, and finally, (5) the purified subsets were manipulated with HPV16 E6/E7 oncogenes to study their effects on tonsillar progenitor cells.

The work presented in this thesis describes, for the first time, a method to identify, isolate and characterize primitive subpopulations of tonsillar epithelial cells that reside in both surface and crypt sites. I show that $CD44^{+}NGFR^{+}$ cells isolated from both sites contain most of the clonogenic progenitor activity in 2D assays and are also highly enriched in their content of these cells. Individual $CD44^{+}NGFR^{+}$ cells exhibit heterogeneous clonogenic potential evidenced by the emergence of different types of colonies. $CD44^{+}NGFR^{+}$ cells from both sites are able to form multilayered epithelium in the 3D organotypic culture system, displaying a complex differentiation spectrum that recapitulates the morphology and immunohistophenotype exclusively seen in the surface

of the tonsil *in vivo*. The finding that even the crypt-derived $CD44^+NGFR^+$ cells are able to generate well-structured epithelium that is dissimilar to their *in vivo* reticulated morphology indicates that tonsillar $CD44^+NGFR^+$ cells are biologically similar independent of their sites of origin. This biological similarity is further supported by their molecular similarity revealed by the transcriptome profiling of surface-derived and crypt-derived $CD44^+NGFR^+$ subsets.

In addition to the shared, site-independent, hierarchical organization, differentiation potential and transcriptional profile between surface- and crypt-derived $CD44^+NGFR^+$ cells, a few differences have also been identified. First, crypt-derived $CD44^+NGFR^+$ cells appear to be more highly enriched in progenitors with greater proliferative potential (higher CFC frequencies and proportion of holoclone-forming cells), accompanied by the reduced propensity and speed with which their progeny became terminal differentiated and incapable of further growth. Second, the transcriptome analysis reveals that crypt-derived $CD44^+NGFR^+$ cells more highly express the genes associated with immunological functions compared with their surface counterparts, and that the molecular variations between surface and crypt progenitors appear to be dependent on the inflammation status of individual tonsils.

Taken together, these findings indicate that epithelial disruption and the atypical marker expression observed in tonsillar crypts *in vivo* are largely caused by the infiltrating lymphocytes which may suppress their developmental potentials that can be unveiled once they are exposed to an environment that allows them to be so. This suggests that functional and phenotypic expressions of tonsillar progenitor cells must be highly context-dependent.

As the first step toward investigating HPV-induced perturbations in tonsillar progenitors, purified $CD45^-CD31^-CD44^+NGFR^+$ cells from both surface and crypt regions were transduced with lentivirus coding for HPV16 E6 and E7 (**Figure 4-2**). The results suggest that expression of E6 and E7 oncoproteins confers enhanced proliferative potential to $CD44^+NGFR^+$ cells, which is passed on to their progeny, leading to an extended lifespan and perturbed epithelial regeneration *in vitro*. Interestingly, when E6/E7-transduced cells were mixed with control cells at 1:1 ratio and cocultured in the

same culture vessel, E6/E7-transduced cells proliferated at a slower rate. In 3D cultures, the cocultured mixed cells gave rise to heterogeneous epithelial structures displaying three different forms including the structure seen neither in normal nor E6/E7 3D cultures (**Figure 4-2**). The results from this study suggest that the normal cells may have inhibitory effects on neighboring E6/E7-expressing cells and their ability to induce malignant transformation. Further investigation is warranted to confirm these findings and elucidate the exact mechanisms and key factors responsible for these alterations.

In conclusion, the system developed in this thesis opens doors to further investigating the HPV-induced oncogenesis in human palatine tonsils, especially in the crypt regions where HPV-associated oropharyngeal cancers arise but pre-cancerous changes remain a mystery, as the anatomic location of crypt epithelium precludes early detection. The ability to purify specific subsets of tonsillar epithelial cells and to selectively manipulate them either with lentivirus or by coculturing them with various exogenous factors or other cell types will allow for elucidating complex biological interactions. This work sets the stage for asking many unanswered questions pertaining to the natural history of tonsillar cancer, which can ultimately lead to identification of effective intervention strategies.

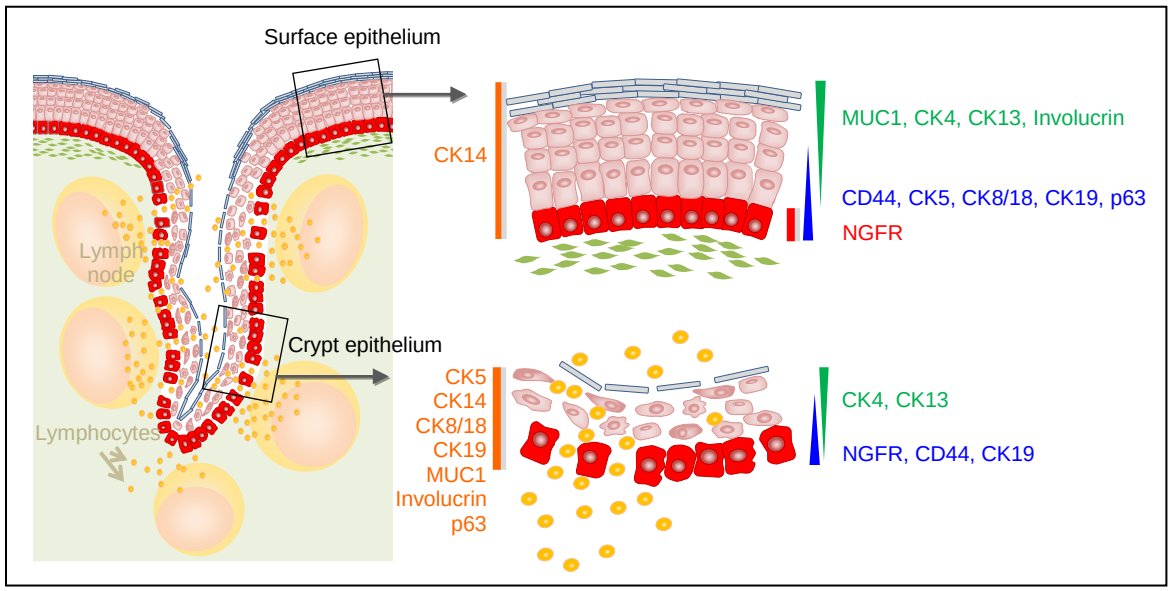


Figure 4-1. Graphical illustration of protein expression patterns in human tonsillar surface and crypt epithelia

Table 4-1. Summary of the features of $CD45^{-}CD31^{-}CD44^{+}NGFR^{+}$ subpopulations isolated from surface or crypt epithelia of human palatine tonsils

	Surface	Crypt
Basal markers	$CK19^{+}CK5^{+}p63^{+}$	
Differentiation markers	$CK4^{-}$Involucrin^{-}	$CK4^{+}$Involucrin^{+}
% CFCs[1]	17.1% ± 2.6%	21.7% ± 2.1%
% Holoclones / Total colonies[2]	25.7% ± 3.8%	62.0% ± 2.0%
Organotypic 3D culture	Stratified; Recapitulated typical tonsillar surface epithelium	
Site-independent progenitor signature[3]	• Protein targeting to membrane and organelle • Viral gene expression • Single-organism in intracellular localization • Multi-organism in metabolic process	
Site-dependent[4]	• Regulation of cell proliferation • Response to chemical • Regulation of cellular component movement	• Antigen processing and presentation via MHC class II • Regulation of cell-cell adhesion • Immune response
Upregulated in CFCs[5]	• Extracellular matrix / structure organization	• Extracellular matrix / structure organization • Positive regulation of cell migration/ cell motility / cellular component movement

[1] Frequency of colonies per 100 $CD44^{+}NGFR^{+}$ cells. Data were pooled from 11 surface/crypt-matched samples. The values shown are mean ± SEM, and represent a ~1000-fold enrichment and a ~2000-fold enrichment for CFCs (surface and crypt, respective) compared with bulk cells cultured under suboptimal conditions.
[2] Proportions of holonclones with respect to the total number of colonies scored.
[3] Based on GO Biological Process enrichment analysis of genes similarly expressed between surface and crypt $CD44^{+}NGFR^{+}$ subsets (Genes with log2-transformed crypt:surface ratio greater than 0.99 or less than 1.01 were excluded)
[4] Transcripts commonly differentially expressed (≥1.5-fold) between surface and crypt $CD44^{+}NGFR^{+}$ subsets
[5] Transcripts commonly upregulated (≥2-fold) in the CFC subsets compared with their matched unseparated $CD45/CD31^{-}$ cells

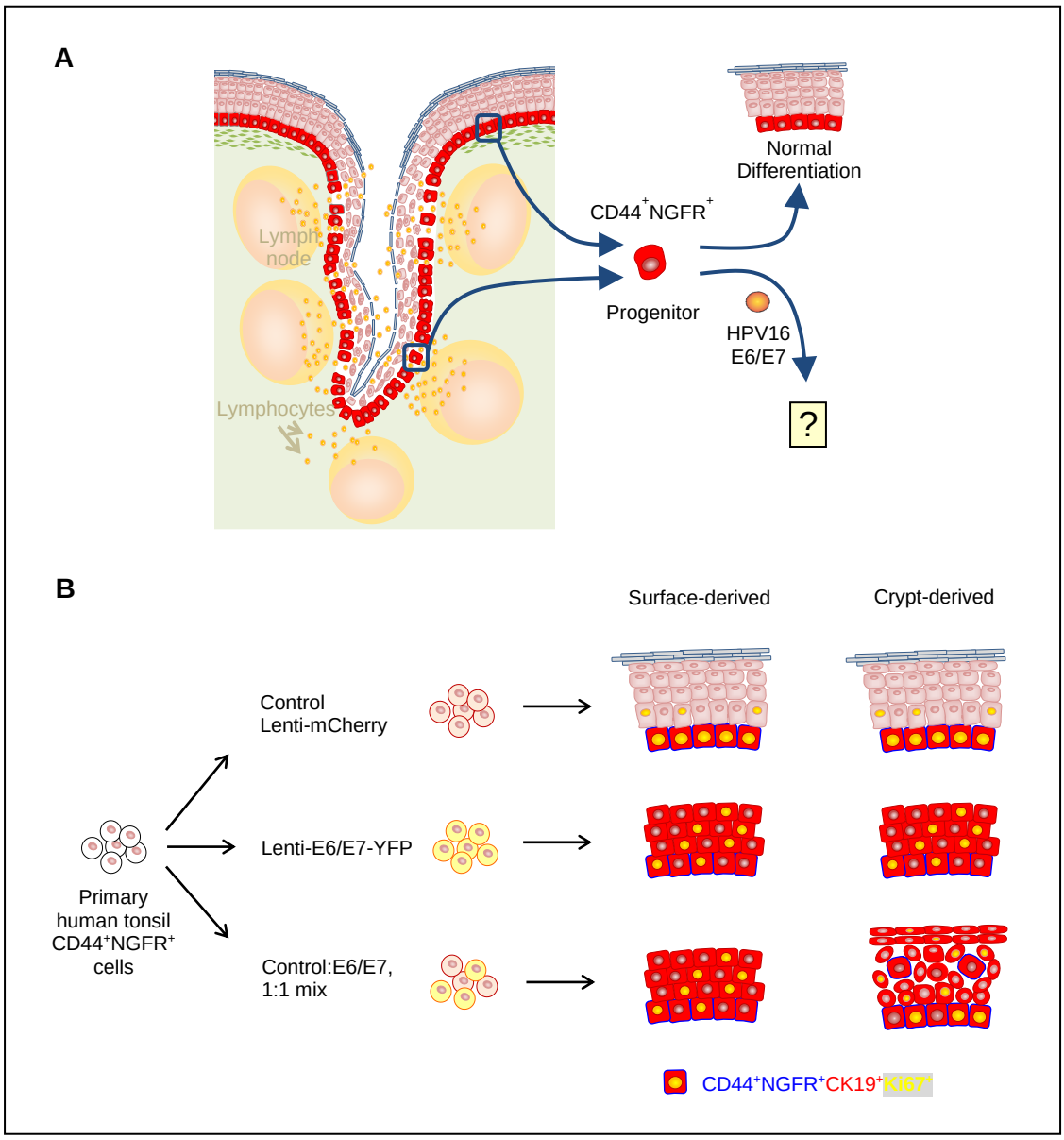


Figure 4-2. HPV16 E6/E7-induced perturbation of tonsillar epithelial differentiation

(A) Single $CD44^{+}NGFR^{+}$ cells isolated from both surface and crypt epithelia have the ability to regenerate stratified epithelium *in vitro* 3D culture system. (B) When these cells are infected with HPV16 E6/E7 oncogenes and subjected to 3D cultures, they do not fully differentiate into normal epithelium, evidenced by pan-epithelial expression of CK19 (shown as red cytoplasm) and dispersed expression of Ki-67 (yellow nuclei). When control cells and E6/E7-infected cells derived from crypt epithelium are cocultured, a distinct subtype arises where cells of basal phenotype ($CD44^{+}NGFR^{+}Ki67^{-}CK4^{-}CK13^{-}MUC1^{-}Involucrin^{-}$) are found in the intermediate layer of the regenerated epithelium.

References

1. Cesta, M. F. Normal structure, function, and histology of mucosa-associated lymphoid tissue. *Toxicol. Pathol.* **34,** 599–608 (2006).

2. von Waldeyer-Hartz. Ueber den lymphatischen Apparat des Pharynx. *Dtsch. Medizinische Wochenschrift* **10,** (1884).

3. Perry, M. & Whyte, A. Immunology of the tonsils. *Immunol Today* **19,** 414–421 (1998).

4. Gebert, a, Rothkötter, H. J. & Pabst, R. M cells in Peyer's patches of the intestine. *Int. Rev. Cytol.* **167,** 91–159 (1996).

5. Brandtzaeg, P. & Prydz, H. Direct evidence for an integrated function of J chain and secretory component in epithelial transport of immunoglobulins. *Nature* **311,** 71–73 (1984).

6. Bernstein, J. M. Waldeyer's ring and otitis media: the nasopharyngeal tonsil and otitis media. *Int J Pediatr Otorhinolaryngol* **49 Suppl 1,** S127–32 (1999).

7. Brook, I. The clinical microbiology of Waldeyer's ring. *Otolaryngol. Clin. North Am.* **20,** 259–272 (1987).

8. Effat, K. G. & Milad, M. Comparative study of palatine tonsil histology in mammals, with special reference to tonsillar salivary glands. *J. Laryngol. Otol.* **121,** 468–71 (2007).

9. Belz, G. T. & Heath, T. J. The epithelium of canine palatine tonsils. *Anat Embryol* **192,** 189–194 (1995).

10. Casteleyn, C., Breugelmans, S., Simoens, P. & Van den Broeck, W. The tonsils revisited: review of the anatomical localization and histological characteristics of the tonsils of domestic and laboratory animals. *Clin. Dev. Immunol.* **2011,** 472460 (2011).

11. Slipka, J. & Slipka, J. Palatine tonsils-are they branchiogenic organs? *Proc. 5th Int. Symp. Tonsils Mucosal Barriers Up. Airways* **1257,** 71–74 (2003).

12. Winning, T. & Townsend, G. Oral mucosal embryology and histology. *Clin Dermatol* **18,** 499–511 (2000).

13. Rothova, M., Thompson, H., Lickert, H. & Tucker, A. S. Lineage tracing of the endoderm during oral development. *Dev. Dyn.* **241,** 1183–91 (2012).

14. Kaltschmidt, B., Kaltschmidt, C. & Widera, D. Adult craniofacial stem cells: sources and relation to the neural crest. *Stem Cell Rev.* **8,** 658–71 (2012).

15. Le Douarin, N. M. The avian embryo as a model to study the development of the neural crest: A long and still ongoing story. *Mech. Dev.* **121,** 1089–1102 (2004).

16. Choi, G., Suh, Y., Lee, H., KY, J. & Hwang, S. Prenatal and postnatal changes of the human tonsillar crypt epithelium. *Acta oto-laryngologica. Suppl.* **523,** 28–33 (1996).

17. Noussios, G., Xanthopoulos, J., Zaraboukas, T., Vital, V. & Konstantinidis, I. Morphological study of development and functional activity of palatine tonsils in embryonic age. *Acta Otorhinolaryngol Ital* **23,** 98–101 (2003).

18. von Gandecker, B. Development and functional anatomy of the human tonsilla palatina. *Acta Otolaryngol* **454(Suppl),** (1988).

19. Isaacson, G. & Parikh, T. Developmental anatomy of the tonsil and its implications for intracapsular tonsillectomy. *Int J Pediatr Otorhinolaryngol* **72,** 89–96 (2008).

20. Larsen, W. J. *Anatomy: Development, Function, Clinical Correlations*. (Philadelphia: WB Saunders, 2002).

21. Wu, T. *et al.* Morphogenesis of Rete Ridges in Human Oral Mucosa: A Pioneering Morphological and Immunohistochemical Study. *Cells Tissues Organs* **197,** 239–248 (2013).

22. Moore, K. L. *Clinically oriented anatomy*. (Philadelphia: Lippincott Williams & Wilkins, 1999).

23. Squier, C. a & Kremer, M. J. Biology of oral mucosa and esophagus. *J. Natl. Cancer Inst. Monogr.* **52242,** 7–15 (2001).

24. Papagerakis, S. *et al.* Oral epithelial stem cells — Implications in normal development and cancer metastasis. *Exp. Cell Res.* **325,** 111–129 (2014).

25. Perry, M. E. The specialised structure of crypt epithelium in the human palatine tonsil and its functional significance. *J Anat* **185 (Pt 1,** 111–127 (1994).

26. Stohr, P. Die Entwicklung des adenoiden Gewebes, der Zungenbälge und der Mandeln des Menschen. *Anat. Anz.* **6,** 545–546 (1891).

27. Reibel, J., Kenrad, B. & Schwartz, O. Architectural organization of human oral epithelium as visualized by keratin staining pattern in tobacco-associated leukoplakias. *J Oral Pathol Med* **20,** 265–270 (1991).

28. Morrison, S. J., Shah, N. M. & Anderson, D. J. Regulatory Mechanism in stem cell biology. *Cell* **88,** 287–298 (1997).

29. Weissman, I. L. Stem cells: units of development, units of regeneration, and units in evolution. *Cell* **100,** 157–168 (2000).

30. Li, L. & Clevers, H. Coexistence of quiescent and active adult stem cells in mammals. *Science* **327,** 542–5 (2010).

31. Simons, B. D. & Clevers, H. Strategies for homeostatic stem cell self-renewal in adult tissues. *Cell* **145,** 851–862 (2011).

32. Deneault, E. *et al.* A Functional Screen to Identify Novel Effectors of Hematopoietic Stem Cell Activity. *Cell* **137,** 369–379 (2009).

33. Morrison, S. J. & Weissman, I. L. The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity* **1,** 661–673 (1994).

34. Rock, J. R. & Hogan, B. L. M. Epithelial progenitor cells in lung development, maintenance, repair, and disease. *Annu. Rev. Cell Dev. Biol.* **27,** 493–512 (2011).

35. Abou-Khalil, R. & Brack, A. S. Muscle stem cells and reversible quiescence: The role of sprouty. *Cell Cycle* **9,** 2575–2580 (2010).

36. Dhawan, J. & Rando, T. A. Stem cells in postnatal myogenesis: Molecular mechanisms of satellite cell quiescence, activation and replenishment. *Trends Cell Biol.* **15,** 666–673 (2005).

37. Blanpain, C. & Fuchs, E. Epidermal homeostasis: a balancing act of stem cells in the skin. *Nat. Rev. Mol. Cell Biol.* **10,** 207–217 (2009).

38. Saleh, M. & Elson, C. O. Experimental inflammatory bowel disease: Insights into the host-microbiota dialog. *Immunity* **34,** 293–302 (2011).

39. He, S., Nakada, D. & Morrison, S. J. Mechanisms of stem cell self-renewal. *Annu. Rev. Cell Dev. Biol.* **25,** 377–406 (2009).

40. Shea, K. L. *et al.* Sprouty1 Regulates Reversible Quiescence of a Self-Renewing Adult Muscle Stem Cell Pool during Regeneration. *Cell Stem Cell* **6,** 117–129 (2010).

41. Biteau, B., Hochmuth, C. E. & Jasper, H. Maintaining tissue homeostasis: Dynamic control of somatic stem cell activity. *Cell Stem Cell* **9,** 402–411 (2011).

42. Gage, F. H. Mammalian Neural Stem Cells. *Science (80-.).* **287,** 1433–1438 (2000).

43. Tajbakhsh, S. Stem cell: what's in a name? *Nat. Reports Stem Cells* (2009). doi:10.1038/stemcells.2009.90

44. Schofield, R. The relationship between the spleen colony-forming cell and the haemopoietic stem cell. *Blood Cells* **4,** 7–25 (1978).

45. Kimble, J. E. & White, J. G. On the control of germ cell development in Caenorhabditis elegans. *Dev. Biol.* **81,** 208–219 (1981).

46. Xie, T. & Spradling, A. C. A niche maintaining germ line stem cells in the \emph{Drosophila} ovary. *Science (80-.).* **290,** 328–330 (2000).

47. Yamashita, Y. M. Orientation of Asymmetric Stem Cell Division by the APC Tumor Suppressor and Centrosome. *Science (80-.).* **301,** 1547–1550 (2003).

48. Song, X. & Xie, T. DE-cadherin-mediated cell adhesion is essential for maintaining somatic stem cells in the Drosophila ovary. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 14813–8 (2002).

49. Puklin-Faucher, E. & Sheetz, M. P. The mechanical integrin cycle. *J Cell Sci* **122,** 179–186 (2009).

50. Hall, P. A. & Watt, F. M. Stem cells: the generation and maintenance of cellular diversity. *Development* **106,** 619–633 (1989).

51. Jones, D. L. & Wagers, A. J. No place like home: anatomy and function of the stem cell niche. *Nat. Rev. Mol. Cell Biol.* **9,** 11–21 (2008).

52. Blanpain, C., Horsley, V. & Fuchs, E. Epithelial Stem Cells: Turning over New Leaves. *Cell* **128,** 445–458 (2007).

53. Corthésy, B. Roundtrip ticket for secretory IgA: role in mucosal homeostasis? *J. Immunol.* **178,** 27–32 (2007).

54. Van Keymeulen, A. & Blanpain, C. Tracing epithelial stem cells during development, homeostasis, and repair. *J. Cell Biol.* **197,** 575–584 (2012).

55. Clayton, E. *et al.* A single type of progenitor cell maintains normal epidermis. *Nature* **446,** 185–9 (2007).

56. Blanpain, C., Lowry, W. E., Geoghegan, A., Polak, L. & Fuchs, E. Self-Renewal, Multipotency, and the Existence of Two Cell Populations within an Epithelial Stem Cell Niche. *Cell* **118,** 635–648 (2004).

57. Horsley, V. *et al.* Blimp1 Defines a Progenitor Population that Governs Cellular Input to the Sebaceous Gland. *Science (80-.).* **126,** 597–609 (2006).

58. Ito, M. *et al.* Stem cells in the hair follicle bulge contribute to wound repair but not to homeostasis of the epidermis. *Nat. Med.* **11,** 1351–1354 (2005).

59. Levy, V., Lindon, C., Harfe, B. D. & Morgan, B. A. Distinct stem cell populations regenerate the follicle and interfollicular epidermis. *Dev. Cell* **9,** 855–861 (2005).

60. Blanpain, C. & Fuchs, E. Epidermal stem cells of the skin. *Annu. Rev. Cell Dev. Biol.* **22,** 339–73 (2006).

61. Potten, C. S. The epidermal proliferative unit: the possible role of the central basal cell. *Cell Tissue Kinet* **7,** 77–88 (1974).

62. Ghazizadeh, S. & Taichman, L. B. Multiple classes of stem cells in cutaneous epithelium: A lineage analysis of adult mouse skin. *EMBO J.* **20,** 1215–1222 (2001).

63. Kolodka, T. M., Garlick, J. a & Taichman, L. B. Evidence for keratinocyte stem cells in vitro: long term engraftment and persistence of transgene expression from retrovirus-transduced keratinocytes. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 4356–4361 (1998).

64. Mackenzie, I. C. Retroviral transduction of murine epidermal stem cells demonstrates clonal units of epidermal structure. *J. Invest. Dermatol.* **109,** 377–383 (1997).

65. Ro, S. & Rannala, B. H. A stop-EGFP transgenic mouse to detect clonal cell lineages generated by mutation. *EMBO Rep.* **5,** 914–920 (2004).

66. Morris, R. J. & Potten, C. S. Highly persistent label-retaining cells in the hair follicles of mice and their fate following induction of anagen. *J. Invest. Dermatol.* **112,** 470–475 (1999).

67. Cotsarelis, G., Cheng, S.-Z., Dong, G., Sun, T.-T. & Lavker, R. M. Existence of slow-cycling limbal epithelial basal cells that can be preferentially stimulated to proliferate: Implications on epithelial stem cells. *Cell* **57,** 201–209 (1989).

68. Bickenbach, J. R. Identification and behavior of label-retaining cells in oral mucosa and skin. *J. Dent. Res.* **60 Spec No,** 1611–1620 (1981).

69. Cotsarelis, G., Sun, T. T. & Lavker, R. M. Label-retaining cells reside in the bulge area of pilosebaceous unit: Implications for follicular stem cells, hair cycle, and skin carcinogenesis. *Cell* **61,** 1329–1337 (1990).

70. Asaka, T., Akiyama, M., Kitagawa, Y. & Shimizu, H. Higher density of label-retaining cells in gingival epithelium. *J. Dermatol. Sci.* **55,** 130–132 (2009).

71. Huang, Y. L., Tao, X., Xia, J., Li, C. Y. & Cheng, B. Distribution and quantity of label-retaining cells in rat oral epithelia. *J. Oral Pathol. Med.* **38,** 663–667 (2009).

72. Asaka, T., Akiyama, M., Kitagawa, Y. & Shimizu, H. Higher density of label-retaining cells in gingival epithelium. *J. Dermatol. Sci.* **55,** 130–132 (2009).

73. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449,** 1003–1007 (2007).

74. Kretzschmar, K. & Watt, F. M. Lineage tracing. *Cell* **148,** 33–45 (2012).

75. Jaks, V. *et al.* Lgr5 marks cycling, yet long-lived, hair follicle stem cells. *Nat. Genet.* **40,** 1291–1299 (2008).

76. Okubo, T., Clark, C. & Hogan, B. L. M. Cell lineage mapping of taste bud cells and keratinocytes in the mouse tongue and soft palate. *Stem Cells* **27,** 442–450 (2009).

77. Arnold, K. *et al.* Sox2 + adult stem and progenitor cells are important for tissue regeneration and survival of mice. *Cell Stem Cell* **9,** 317–329 (2011).

78. Tsai, Y. C. *et al.* Contiguous patches of normal human mammary epithelium derived from a single stem cell: implications for breast carcinogenesis. *Cancer Res* **56,** 402–404 (1996).

79. Bedi, G. C., Westra, W. H., Gabrielson, E., Koch, W. & Sidransky, D. Multiple head and neck tumors: Evidence for a common clonal origin. *Cancer Res.* **56,** 2484–2487 (1996).

80. Califano, J., Koch, W., Sidransky, D. & Westra, W. H. Inverted sinonasal papilloma : a molecular genetic appraisal of its putative status as a Precursor to squamous cell carcinoma. *Am. J. Pathol.* **156,** 333–7 (2000).

81. Greaves, L. C. *et al.* Mitochondrial DNA mutations are established in human colonic stem cells, and mutated clones expand by crypt fission. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 714–719 (2006).

82. Taylor, R. W., Taylor, G. a, Durham, S. E. & Turnbull, D. M. The determination of complete human mitochondrial DNA sequences in single cells: implications for the study of somatic mitochondrial DNA point mutations. *Nucleic Acids Res.* **29,** E74–E74 (2001).

83. Taylor, R. W. *et al.* Mitochondrial DNA mutations in human colonic crypt stem cells. *J Clin Invest* **112,** 1351–1360 (2003).

84. McDonald, S. a C. *et al.* Mechanisms of Field Cancerization in the Human Stomach: The Expansion and Spread of Mutated Gastric Stem Cells. *Gastroenterology* **134,** 500–510 (2008).

85. Fellous, T. G. *et al.* A methodological approach to tracing cell lineage in human epithelial tissues. *Stem Cells* **27,** 1410–1420 (2009).

86. Tárnok, A., Ulrich, H. & Bocsi, J. Phenotypes of stem cells from diverse origin. *Cytom. Part A* **77,** 6–10 (2010).

87. Bryder, D., Rossi, D. J. & Weissman, I. L. Hematopoietic stem cells: the paradigmatic tissue-specific stem cell. *Am. J. Pathol.* **169,** 338–46 (2006).

88. Stingl, J. *et al.* Purification and unique properties of mammary epithelial stem cells. *Nature* **439,** 993–997 (2006).

89. Lawson, D. A., Xin, L., Lukacs, R. U., Cheng, D. & Witte, O. N. Isolation and functional characterization of murine prostate stem cells. *Proc Natl Acad Sci U S A* **104,** 181–186 (2007).

90. Li, A., Simmons, P. J. & Kaur, P. Identification and isolation of candidate human keratinocyte stem cells based on cell surface phenotype. *Proc Natl Acad Sci U S A* **95,** 3902–3907 (1998).

91. Kim, C. F. B. *et al.* Identification of bronchioalveolar stem cells in normal lung and lung cancer. *Cell* **121,** 823–35 (2005).

92. Whitsett, J. A. & Kalinichenko, V. V. Notch and basal cells take center stage during airway epithelial regeneration. *Cell Stem Cell* **8,** 597–598 (2011).

93. Watt, F. M. & Jones, P. H. Expression and function of the keratinocyte integrins. *Dev Suppl* 185–192 (1993). at <http://www.ncbi.nlm.nih.gov/pubmed/8049472>

94. Kaur, P. & Li, A. Adhesive properties of human basal epidermal cells: An analysis of keratinocyte stem cells, transit amplifying cells, and postmitotic differentiating cells. *J. Invest. Dermatol.* **114,** 413–420 (2000).

95. Legg, J., Jensen, U. B., Broad, S., Leigh, I. & Watt, F. M. Role of melanoma chondroitin sulphate proteoglycan in patterning stem cells in human interfollicular epidermis. *Development* **130,** 6049–6063 (2003).

96. Michel, M. *et al.* Keratin 19 as a biochemical marker of skin stem cells in vivo and in vitro: keratin 19 expressing cells are differentially localized in function of anatomic sites, and their number varies with donor age and culture stage. *J. Cell Sci.* **109 (Pt 5,** 1017–1028 (1996).

97. Lyle, S. *et al.* The C8/144B monoclonal antibody recognizes cytokeratin 15 and defines the location of human hair follicle stem cells. *J. Cell Sci.* **111,** 3179–3188 (1998).

98. Pellegrini, G. *et al.* p63 identifies keratinocyte stem cells. *Proc Natl Acad Sci U S A* **98,** 3156–3161 (2001).

99. Senoo, M., Pinto, F., Crum, C. P. & McKeon, F. p63 Is essential for the proliferative potential of stem cells in stratified epithelia. *Cell* **129,** 523–536 (2007).

100. Nguyen, L. V, Vanner, R., Dirks, P. & Eaves, C. J. Cancer stem cells: an evolving concept. *Nat Rev Cancer* **12,** 133–143 (2012).

101. Shackleton, M. *et al.* Generation of a functional mammary gland from a single stem cell. *Nature* **439,** 84–88 (2006).

102. Spike, B. T. *et al.* A mammary stem cell population identified and characterized in late embryogenesis reveals similarities to human breast cancer. *Cell Stem Cell* **10,** 183–197 (2012).

103. Makarem, M. *et al.* Stem cells and the developing mammary gland. *J. Mammary Gland Biol. Neoplasia* **18,** 209–219 (2013).

104. Pouliot, N., Redvers, R. P., Ellis, S., Saunders, N. A. & Kaur, P. Optimization of a transplant model to assess skin reconstitution from stem cell-enriched primary human keratinocyte populations. *Exp. Dermatol.* **14,** 60–69 (2005).

105. Reynolds, B. a. & Weiss, S. Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. *Science (80-.).* **255,** 1707–10 (1992).

106. Dontu, G. *et al.* In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. *Genes Dev.* **17,** 1253–1270 (2003).

107. Yokoo, S. *et al.* Human corneal endothelial cell precursors isolated by sphere-forming assay. *Invest. Ophthalmol. Vis. Sci.* **46,** 1626–1631 (2005).

108. Toma, J. G., McKenzie, I. a, Bagli, D. & Miller, F. D. Isolation and characterization of multipotent skin-derived precursors from human skin. *Stem Cells* **23,** 727–37 (2005).

109. Seaberg, R. M. *et al.* Clonal identification of multipotent precursors from adult mouse pancreas that generate neural and pancreatic lineages. *Nat Biotechnol* **22,** 1115–1124 (2004).

110. 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 (2009).

111. Ponti, D. *et al.* Isolation and in vitro propagation of tumorigenic breast cancer cells with stem/progenitor cell properties. *Cancer Res.* **65,** 5506–5511 (2005).

112. Ricci-Vitiani, L. *et al.* Identification and expansion of human colon-cancer-initiating cells. *Nature* **445,** 111–115 (2007).

113. Fang, D. *et al.* A tumorigenic subpopulation with stem cell properties in melanomas. *Cancer Res.* **65,** 9328–9337 (2005).

114. Collins, A. T., Habib, F. K., Maitland, N. J. & Neal, D. E. Identification and isolation of human prostate epithelial stem cells based on α2β1-integrin expression. *J. Cell Sci.* **114,** 3865–3872 (2001).

115. Krishnamurthy, S. & Nör, J. E. Orosphere assay: a method for propagation of head and neck cancer stem cells. *Head Neck* **35,** 1015–21 (2013).

116. Pastrana, E., Silva-Vargas, V. & Doetsch, F. Eyes Wide Open: A Critical Review of Sphere-Formation as an Assay for Stem Cells. *Cell Stem Cell* **8,** 486–498 (2011).

117. Coles-Takabe, B. L. *et al.* Don't look: growing clonal versus nonclonal neural stem cell colonies. *Stem Cells* **26,** 2938–2944 (2008).

118. Ferron, S. R. *et al.* A combined ex/in vivo assay to detect effects of exogenously added factors in neural stem cells. *Nat. Protoc.* **2,** 849–859 (2007).

119. Chojnacki, A. & Weiss, S. Production of neurons, astrocytes and oligodendrocytes from mammalian CNS stem cells. *Nat. Protoc.* **3,** 935–40 (2008).

120. Singec, I. *et al.* Defining the actual sensitivity and specificity of the neurosphere assay in stem cell biology. *Nat. Methods* **3,** 801–806 (2006).

121. Mori, H. *et al.* Effect of neurosphere size on the growth rate of human neural stem/progenitor cells. *J. Neurosci. Res.* **84,** 1682–1691 (2006).

122. Louis, S. A. *et al.* Enumeration of Neural Stem and Progenitor Cells in the Neural Colony-Forming Cell Assay. *Stem Cells* **26,** 988–996 (2008).

123. Manuel Iglesias, J. *et al.* Mammosphere Formation in Breast Carcinoma Cell Lines Depends upon Expression of E-cadherin. *PLoS One* **8,** (2013).

124. Memmi, E. M. *et al.* p63 Sustains self-renewal of mammary cancer stem cells through regulation of Sonic Hedgehog signaling. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 3499–504 (2015).

125. Spike, B. T. *et al.* CRIPTO/GRP78 Signaling Maintains Fetal and Adult Mammary Stem Cells Ex Vivo. *Stem cell reports* **2,** 427–39 (2014).

126. Barrandon, Y. & Green, H. Three clonal types of keratinocyte with different capacities for multiplication. *Proc. Natl. Acad. Sci.* **84,** 2302–2306 (1987).

127. Pellegrini, G. *et al.* Location and Clonal Analysis of Stem Cells and Their Differentiated Progeny in the Human Ocular Surface. *J. Cell Biol.* **145,** 769–782 (1999).

128. Mathor, M. B. *et al.* Clonal analysis of stably transduced human epidermal stem cells in culture. *Proc. Natl. Acad. Sci.* **93,** 10371–10376 (1996).

129. Barrandon, Y. & Green, H. Cell size as a determinant of the clone-forming ability of human keratinocytes. *Proc Natl Acad Sci U S A* **82,** 5390–5394 (1985).

130. Tudor, D., Locke, M., Owen-Jones, E. & Mackenzie, I. C. Intrinsic patterns of behavior of epithelial stem cells. *J. Investig. Dermatol. Symp. Proc.* **9,** 208–14 (2004).

131. Tudor, D., Chaudry, F., Harper, L. & Mackenzie, I. C. The in vitro behaviour and patterns of colony formation of murine epithelial stem cells. *Cell Prolif.* **40,** 706–720 (2007).

132. Murayama, K. *et al.* Akt activation induces epidermal hyperplasia and proliferation of epidermal progenitors. *Oncogene* **26,** 4882–8 (2007).

133. Rochat, a, Kobayashi, K. & Barrandon, Y. Location of stem cells of human hair follicles by clonal analysis. *Cell* **76,** 1063–73 (1994).

134. Shortt, A. J. *et al.* Characterization of the limbal epithelial stem cell niche: novel imaging techniques permit in vivo observation and targeted biopsy of limbal epithelial stem cells. *Stem Cells* **25,** 1402–9 (2007).

135. O'Hare, M. J., Ormerod, M. G., Monaghan, P., Lane, E. B. & Gusterson, B. A. Characterization in vitro of luminal and myoepithelial cells isolated from the human mammary gland by cell sorting. *Differentiation* **46,** 209–221 (1991).

136. Emerman, J. T., Stingl, J., Petersen, A., Shpall, E. J. & Eaves, C. J. Selective growth of freshly isolated human breast epithelial cells cultured at low concentrations in the presence or absence of bone marrow cells. *Breast Cancer Res. Treat.* **41,** 147–59 (1996).

137. Nakamura, T., Endo, K. & Kinoshita, S. Identification of human oral keratinocyte stem/progenitor cells by neurotrophin receptor p75 and the role of neurotrophin/p75 signaling. *Stem Cells* **25,** 628–38 (2007).

138. Rheinwald, J. G. & Green, H. Formation of a keratinizing epithelium in culture by a cloned cell line derived from a teratoma. *Cell* **6,** (1975).

139. Stanley, M. a & Parkinson, E. K. Growth requirements of human cervical epithelial cells in culture. *Int. J. Cancer* **24,** 407–14 (1979).

140. Taylor-Papadimitriou, J., Shearer, M. & Stoker, M. G. Growth requirements of human mammary epithelial cells in culture. *Int. J. Cancer* **20,** 903–8 (1977).

141. Stingl, J., Eaves, C. J., Zandieh, I. & Emerman, J. T. Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue. *Breast Cancer Res Treat* **67,** 93–109 (2001).

142. Hall, A. Rho GTPases and the Actin Cytoskeleton. *Science (80-.).* **279,** 509–514 (1998).

143. Kaibuchi, K., Kuroda, S., Fukata, M. & Nakagawa, M. Regulation of cadherin-mediated cell-cell adhesion by the Rho family GTPases. *Curr. Opin. Cell Biol.* **11,** 591–6 (1999).

144. Watanabe, K. *et al.* A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nat. Biotechnol.* **25,** 681–686 (2007).

145. Koyanagi, M. *et al.* Inhibition of the Rho/ROCK pathway reduces apoptosis during transplantation of embryonic stem cell-derived neural precursors. *J. Neurosci. Res.* **86,** 270–80 (2008).

146. Gauthaman, K., Fong, C. Y. & Bongso, A. Effect of ROCK inhibitor Y-27632 on normal and variant human embryonic stem cells (hESCs) in vitro: Its benefits in hESC expansion. *Stem Cell Rev. Reports* **6,** 86–95 (2010).

147. Li, X., Meng, G., Krawetz, R., Liu, S. & Rancourt, D. E. The ROCK inhibitor Y-27632 enhances the survival rate of human embryonic stem cells following cryopreservation. *Stem Cells Dev.* **17,** 1079–1085 (2008).

148. van den Bogaard, E. H. *et al.* Rho kinase inhibitor Y-27632 prolongs the life span of adult human keratinocytes, enhances skin equivalent development, and facilitates lentiviral transduction. *Tissue Eng. Part A* **18,** 1827–36 (2012).

149. Chapman, S., Liu, X., Meyers, C., Schlegel, R. & McBride, A. A. Human keratinocytes are efficiently immortalized by a Rho kinase inhibitor. *J. Clin. Invest.* **120,** 2619–26 (2010).

150. Liu, X. *et al.* ROCK inhibitor and feeder cells induce the conditional reprogramming of epithelial cells. *Am. J. Pathol.* **180,** 599–607 (2012).

151. Zhang, L. *et al.* ROCK inhibitor Y-27632 suppresses dissociation-induced apoptosis of murine prostate stem/progenitor cells and increases their cloning efficiency. *PLoS One* **6,** e18271 (2011).

152. Makarem, M. *et al.* Developmental changes in the in vitro activated regenerative activity of primitive mammary epithelial cells. *PLoS Biol.* **11,** e1001630 (2013).

153. Asselineau, D. & Prunieras, M. Reconstruction of 'simplified' skin: control of fabrication. *Br. J. Dermatol.* **111 Suppl ,** 219–22 (1984).

154. Green, H., Kehinde, O. & Thomas, J. Growth of cultured human epidermal cells into multiple epithelia suitable for grafting. *Proc. Natl. Acad. Sci. U. S. A.* **76,** 5665–5668 (1979).

155. Fuchs, E. Epidermal differentiation: The bare essentials. *J. Cell Biol.* **111,** 2807–2814 (1990).

156. Kalabis, J. *et al.* A subpopulation of mouse esophageal basal cells has properties of stem cells with the capacity for self-renewal and lineage specification. *J Clin Invest* **118,** 3860–3869 (2008).

157. Ridky, T. W., Chow, J. M., Wong, D. J. & Khavari, P. a. Invasive three-dimensional organotypic neoplasia from multiple normal human epithelia. *Nat. Med.* **16,** 1450–1455 (2010).

158. Calenic, B. *et al.* Characterization of oral keratinocyte stem cells and prospects of its differentiation to oral epithelial equivalents. *Rom. J. Morphol. Embryol.* **51,** 641–645 (2010).

159. Luo, X., Okubo, T., Randell, S. & Hogan, B. L. M. Culture of endodermal stem/progenitor cells of the mouse tongue. *In Vitro Cell. Dev. Biol. Anim.* **45,** 44–54 (2009).

160. Andrei, G. Three-dimensional culture models for human viral diseases and antiviral drug development. *Antiviral Res.* **71,** 96–107 (2006).

161. Klausner, M. *et al.* Organotypic human oral tissue models for toxicological studies. *Toxicol. Vitr.* **21,** 938–949 (2007).

162. Franceschi, S., Muñoz, N., Bosch, X. F., Snijders, P. J. & Walboomers, J. M. Human papillomavirus and cancers of the upper aerodigestive tract: a review of epidemiological and experimental evidence. *Cancer Epidemiol. Biomarkers Prev.* **5,** 567–75 (1996).

163. Ferlay, J. *et al.* GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase. No. 11 [Internet]. *Lyon, Fr. Int. Agency Res. Cancer.* **11,** http://globocan.iarc.f (2013).

164. Guay, M. E. & Lavertu, P. Tonsillar carcinoma. *Eur Arch Otorhinolaryngol* **252,** 259–64 (1995).

165. Mizono, G. S., Diaz, R. F., Fu, K. K. & Boles, R. Carcinoma of the tonsillar region. *Laryngoscope* **96,** 240–244 (1986).

166. Golas, S. M. Trends in palatine tonsillar cancer incidence and mortality rates in the United States. *Community Dent. Oral Epidemiol.* **35,** 98–108 (2007).

167. López-Guillermo, A. *et al.* Diffuse large B-cell lymphoma: Clinical and biological characterization and outcome according to the nodal or extranodal primary origin. *J. Clin. Oncol.* **23,** 2797–2804 (2005).

168. Vita, V., Lawrence, T. & Rosenberg, S. *Cancer: Principles & Practice of Oncology*. (2008).

169. Chaturvedi, A. K. *et al.* Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *J. Clin. Oncol.* **29,** 4294–4301 (2011).

170. Pytynia, K. B., Dahlstrom, K. R. & Sturgis, E. M. Epidemiology of HPV-associated oropharyngeal cancer. *Oral Oncol.* **50,** 380–386 (2014).

171. Marur, S., D'Souza, G., Westra, W. H. & Forastiere, A. A. HPV-associated head and neck cancer: a virus-related cancer epidemic. *Lancet Oncol* **11,** 781–789 (2010).

172. Nasman, A. *et al.* Incidence of human papillomavirus (HPV) positive tonsillar carcinoma in Stockholm, Sweden: an epidemic of viral-induced carcinoma? *Int J Cancer* **125,** 362–366 (2009).

173. Chaturvedi, A. K. *et al.* Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *J Clin Oncol* **29,** 4294–4301 (2011).

174. Chaturvedi, A. K. *et al.* Worldwide trends in incidence rates for oral cavity and oropharyngeal cancers. *J. Clin. Oncol.* **31,** 4550–4559 (2013).

175. Moody, C. A. & Laimins, L. A. Human papillomavirus oncoproteins: pathways to transformation. *Nat. Rev. Cancer* **10,** 550–560 (2010).

176. Bravo, I. G. & Felez-Sanchez, M. Papillomaviruses: viral evolution, cancer and evolutionary medicine. *Evol. Med. Public Heal.* 32–51 (2015). doi:10.1093/emph/eov003

177. Parkin, D. M. The global health burden of infection-associated cancers in the year 2002. *Int J Cancer* **118,** 3030–3044 (2006).

178. Walboomers, J. M. M. *et al.* Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J. Pathol.* **189,** 12–19 (1999).

179. Muñoz, N., Castellsagué, X., de González, A. B. & Gissmann, L. Chapter 1: HPV in the etiology of human cancer. *Vaccine* **24,** (2006).

180. Parkin, D. M. & Bray, F. Chapter 2: The burden of HPV-related cancers. *Vaccine* **24,** 11–25 (2006).

181. Ndiaye, C. *et al.* HPV DNA, E6/E7 mRNA, and p16INK4a detection in head and neck cancers: A systematic review and meta-analysis. *Lancet Oncol.* **15,** 1319–1331 (2014).

182. Kreimer, A. R., Clifford, G. M., Boyle, P. & Franceschi, S. Human papillomavirus types in head and neck squamous cell carcinomas worldwide: a systematic review. *Cancer Epidemiol Biomarkers Prev* **14,** 467–475 (2005).

183. D'Souza, G. *et al.* Case-control study of human papillomavirus and oropharyngeal cancer. *N Engl J Med* **356,** 1944–1956 (2007).

184. Gillison, M. L. *et al.* Distinct Risk Factor Profiles for Human Papillomavirus Type 16-Positive and Human Papillomavirus Type 16-Negative Head and Neck Cancers. *JNCI J. Natl. Cancer Inst.* **100,** 407–420 (2008).

185. Pintos, J. *et al.* Human papillomavirus infection and oral cancer: A case-control study in Montreal, Canada. *Oral Oncol.* **44,** 242–250 (2008).

186. Kreimer, A. R. *et al.* Evaluation of human papillomavirus antibodies and risk of subsequent head and neck cancer. *J. Clin. Oncol.* **31,** 2708–2715 (2013).

187. Sapp, M. & Day, P. M. Structure, attachment and entry of polyoma- and papillomaviruses. *Virology* **384,** 400–409 (2009).

188. Cheng, S., Schmidt-Grimminger, D. C., Murant, T., Broker, T. R. & Chow, L. T. Differentiation-dependent up-regulation of the human papillomavirus E7 gene reactivates cellular DNA replication in suprabasal dlfferentiated keratinocytes. *Genes Dev.* **9,** 2335–2349 (1995).

189. Doorbar, J. The papillomavirus life cycle. *J. Clin. Virol.* **32,** 7–15 (2005).

190. zur Hausen, H. Papillomaviruses causing cancer: evasion from host-cell control in early events in carcinogenesis. *J. Natl. Cancer Inst.* **92,** 690–698 (2000).

191. Ghittoni, R. *et al.* The biological properties of E6 and E7 oncoproteins from human papillomaviruses. *Virus Genes* **40,** 1–13 (2009).

192. Münger, K. *et al.* Mechanisms of human papillomavirus-induced oncogenesis. *J. Virol.* **78,** 11451–11460 (2004).

193. Deshpande, A., Sicinski, P. & Hinds, P. W. Cyclins and cdks in development and cancer: a perspective. *Oncogene* **24,** 2909–2915 (2005).

194. Begum, S. Tissue Distribution of Human Papillomavirus 16 DNA Integration in Patients with Tonsillar Carcinoma. *Clin. Cancer Res.* **11,** 5694–5699 (2005).

195. Klingelhutz, A., Foster, S. & McDougall, J. Telomerase activation by the E6 gene product of human papillomavirus type 16. *Nature* **380,** 79–82 (1996).

196. Klingelhutz, A. J. & Roman, A. Cellular transformation by human papillomaviruses: Lessons learned by comparing high- and low-risk viruses. *Virology* **424,** 77–98 (2012).

197. McLaughlin-Drubin, M. E. & Münger, K. Oncogenic activities of human papillomaviruses. *Virus Res.* **143,** 195–208 (2009).

198. Halbert, C. L., Demers, G. W. & Galloway, D. A. The E7 Gene of Human Papillomavirus Type 16 Is Sufficient for Immortalization of Human Epithelial Cells. **65,** 473–478 (1991).

199. Hawley-Nelson, P., Vousden, K. H., Hubbert, N. L., Lowy, D. R. & Schiller, J. T. HPV16 E6 and E7 proteins cooperate to immortalize human foreskin keratinocytes. *EMBO J.* **8,** 3905–10 (1989).

200. Munger, K., Phelps, W. C., Bubb, V., Howley, P. M. & Schlegel, R. The E6 and E7 genes of the human papillomavirus type 16 together are necessary and sufficient for transformation of primary human keratinocytes. *J Virol* **63,** 4417–4421 (1989).

201. McCance, D. J., Kopan, R., Fuchs, E. & Laimins, L. a. Human papillomavirus type 16 alters human epithelial cell differentiation in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **85,** 7169–7173 (1988).

202. Li, S. L., Kim, M. S., Cherrick, H. M., Doniger, J. & Park, N. H. Sequential combined tumorigenic effect of HPV-16 and chemical carcinogens. *Carcinogenesis* **13,** 1981–1987 (1992).

203. Liu, X., Han, S., Baluda, M. a & Park, N. H. HPV-16 oncogenes E6 and E7 are mutagenic in normal human oral keratinocytes. *Oncogene* **14,** 2347–53 (1997).

204. Park, N. H. *et al.* Combined oral carcinogenicity of HPV-16 and benzo(a)pyrene: an in vitro multistep carcinogenesis model. *Oncogene* **10,** 2145–2153 (1995).

205. Al Moustafa, A.-E. *et al.* E6/E7 proteins of HPV type 16 and ErbB-2 cooperate to induce neoplastic transformation of primary normal oral epithelial cells. *Oncogene* **23,** 350–8 (2004).

206. Herber, R., Liem, a, Pitot, H. & Lambert, P. F. Squamous epithelial hyperplasia and carcinoma in mice transgenic for the human papillomavirus type 16 E7 oncogene. *J. Virol.* **70,** 1873–1881 (1996).

207. Summersgill, K. F. *et al.* Human papillomavirus in the oral cavities of children and adolescents. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.* **91,** 62–69 (2001).

208. Smith, E. M. *et al.* Prevalence of human papillomavirus in the oral cavity/oropharynx in a large population of children and adolescents. *Pediatr. Infect. Dis. J.* **26,** 836–840 (2007).

209. D'Souza, G., Agrawal, Y., Halpern, J., Bodison, S. & Gillison, M. L. Oral sexual behaviors associated with prevalent oral human papillomavirus infection. *J. Infect. Dis.* **199,** 1263–1269 (2009).

210. Hansson, B. G. *et al.* Strong association between infection with human papillomavirus and oral and oropharyngeal squamous cell carcinoma: a population-based case-control study in southern Sweden. *Acta Otolaryngol.* **125,** 1337–1344 (2005).

211. Schwartz, S. M. *et al.* Oral cancer risk in relation to sexual history and evidence of human papillomavirus infection. *J Natl Cancer Inst* **90,** 1626–1636 (1998).

212. Edelstein, Z. R. *et al.* Rates and Determinants of Oral Human Papillomavirus Infection in Young Men. *Sex. Transm. Dis.* **39,** 860–867 (2012).

213. Gillison, M. L. *et al.* Prevalence of oral HPV infection in the United States, 2009-2010. *JAMA* **307,** 693–703 (2012).

214. Kreimer, A. R. *et al.* Oral human papillomavirus in healthy individuals: a systematic review of the literature. *Sex Transm Dis* **37,** 386–391 (2010).

215. Kreimer, A. R. *et al.* Oral human papillomavirus infection in adults is associated with sexual behavior and HIV serostatus. *J. Infect. Dis.* **189,** 686–698 (2004).

216. PR, do S. *et al.* The prevalence of human papillomavirus in the oropharynx in healthy individuals in a Brazilian population. *J Med Virol* **78,** 614–618 (2006).

217. Franceschi, S. *et al.* Deep brush-based cytology in tonsils resected for benign diseases. *Int. J. Cancer* **137,** 2994–2999 (2015).

218. Sanjosé, S., Díaz, M., Castellsagué, X., Clifford, G. & Bruni, L. Worldwide prevalence and genotype distribution of cervical HPV in women with normal cytology. *Lancet infect Dis* **7,** 453–459 (2007).

219. zur Hausen, H. Papillomaviruses and Cancer: From Basic Studies to Clinical Application. *Nature* **2,** 342–350 (2002).

220. Kim, S. H. *et al.* HPV integration begins in the tonsillar crypt and leads to the alteration of p16, EGFR and c-myc during tumor formation. *Int J Cancer* **120,** 1418–1425 (2007).

221. Pai, S. I. & Westra, W. H. Molecular pathology of head and neck cancer: implications for diagnosis, prognosis, and treatment. *Annu. Rev. Pathol.* **4,** 49–70 (2009).

222. El-Mofty, S. K. & Patil, S. Human papillomavirus (HPV)-related oropharyngeal nonkeratinizing squamous cell carcinoma: characterization of a distinct phenotype. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.* **101,** 339–45 (2006).

223. Begum, S. & Westra, W. H. Basaloid squamous cell carcinoma of the head and neck is a mixed variant that can be further resolved by HPV status. *Am. J. Surg. Pathol.* **32,** 1044–1050 (2008).

224. Maglennon, G. A., McIntosh, P. & Doorbar, J. Persistence of viral DNA in the epithelial basal layer suggests a model for papillomavirus latency following immune regression. *Virology* **414,** 153–163 (2011).

225. Nicholls, P. K. *et al.* Detection of viral DNA and E4 protein in basal keratinocytes of experimental canine oral papillomavirus lesions. *Virology* **284,** 82–98 (2001).

226. Michael, S., Lambert, P. F. & Strati, K. The HPV16 oncogenes cause aberrant stem cell mobilization. *Virology* **443,** 218–225 (2013).

227. da Silva-Diz, V. *et al.* Progeny of Lgr5-expressing hair follicle stem cell contributes to papillomavirus-induced tumor development in epidermis. *Oncogene* 3732–3743 (2012). doi:10.1038/onc.2012.375

228. López, J., Ruíz, G., Organista-Nava, J., Gariglio, P. & García-Carrancá, A. Human papillomavirus infections and cancer stem cells of tumors from the uterine cervix. *Open Virol. J.* **6,** 232–40 (2012).

229. Herfs, M. *et al.* A discrete population of squamocolumnar junction cells implicated in the pathogenesis of cervical cancer. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 10516–21 (2012).

230. Webb, A., Li, A. & Kaur, P. Location and phenotype of human adult keratinocyte stem cells of the skin. *Differentiation.* **72,** 387–95 (2004).

231. Goldstein, A. S. *et al.* Trop2 identifies a subpopulation of murine and human prostate basal cells with stem cell characteristics. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 20882–20887 (2008).

232. Garraway, I. P. *et al.* Human prostate sphere-forming cells represent a subset of basal epithelial cells capable of glandular regeneration in vivo. *Prostate* **70,** 491–501 (2010).

233. Wang, F. *et al.* Isolation and characterization of intestinal stem cells based on surface marker combinations and colony-formation assay. *Gastroenterology* **145,** 321–383 (2013).

234. Eirew, P. *et al.* A method for quantifying normal human mammary epithelial stem cells with in vivo regenerative ability. *Nat. Med.* **14,** 1384–9 (2008).

235. Lim, E. *et al.* Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat. Med.* **15,** 907–13 (2009).

236. Gemenetzidis, E., Costea, D. E., Parkinson, E. K. & Waseem, A. Induction of Human Epithelial Stem / Progenitor Expansion by FOXM1. *Cancer Res.* **70,** 9515–9526 (2011).

237. Baak, J. P. A. *et al.* Dynamic behavioural interpretation of cervical intraepithelial neoplasia with molecular biomarkers. *J. Clin. Pathol.* **59,** 1017–1028 (2006).

238. Bergeron, C. *et al.* The clinical impact of using p16INK4a immunochemistry in cervical histopathology and cytology: An update of recent developments. *Int. J. Cancer* **136,** 2741–2751 (2015).

239. Rheinwald, J. G. & Green, H. Formation of a keratinizing epithelium in culture by a cloned cell line derived from a teratoma. *Cell* **6,** 317–330 (1975).

240. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2014). at <http://www.r-project.org/>

241. Imren, S. *et al.* High-level beta-globin expression and preferred intragenic integration after lentiviral transduction of human cord blood stem cells. *J. Clin. Invest.* **114,** 953–962 (2004).

242. Owen, R. L., Nemanic, P. C. & Stevens, D. P. Ultrastructural observations on giardiasis in a murine model. I. Intestinal distribution, attachment, and relationship to the immune system of Giardia muris. *Gastroenterology* **76,** 757–769 (1979).

243. Sato, Y., Wake, K. & Watanabe, I. Differentiation of crypt epithelium in human palatine tonsils: the microenvironment of crypt epithelium as a lymphoepithelial organ. *Arch Histol Cytol* **53,** 41–54 (1990).

244. Výborná, E. THE SEQUENCE OF RETICULARISATION OF EPITHELIUM OF HUMAN PALATINE TONSIL : SCANNING ELECTRON MICROSCOPIC STUDY Eva Výborná. *Acta Univ. Palacki. Olomuc., Fac. Med.* **142,** 12–15 (1999).

245. Howie, A. J. Scanning and transmission electron microscopy on the epithelium of human palatine tonsils. *J Pathol* **130,** 91–8 (1980).

246. Desai, S. *et al.* Relationship of cytokeratin 20 and CD44 protein expression with WHO/ISUP grade in pTa and pT1 papillary urothelial neoplasia. *Mod. Pathol.* **13,** 1315–1323 (2000).

247. Guo, C. *et al.* Epcam, cd44, and cd49f distinguish sphere-forming human prostate basal cells from a subpopulation with predominant tubule initiation capability. *PLoS One* **7,** (2012).

248. Louderbough, J. M. V, Brown, J. A., Nagle, R. B. & Schroeder, J. A. CD44 Promotes Epithelial Mammary Gland Development and Exhibits Altered Localization during Cancer Progression. *Genes Cancer* **2,** 771–81 (2011).

249. Paik, D. Y. *et al.* Stem-like epithelial cells are concentrated in the distal end of the fallopian tube: a site for injury and serous cancer initiation. *Stem Cells* **30,** 2487–97 (2012).

250. Okumura, T., Shimada, Y., Imamura, M. & Yasumoto, S. Neurotrophin receptor p75(NTR) characterizes human esophageal keratinocyte stem cells in vitro. *Oncogene* **22,** 4017–4026 (2003).

251. Linden, S. K., Sutton, P., Karlsson, N. G., Korolik, V. & McGuckin, M. A. Mucins in the mucosal barrier to infection. *Mucosal Immunol.* **1,** 183–197 (2008).

252. Patton, S., Gendler, S. J. & Spicer, A. P. The epithelial mucin, MUC1, of milk, mammary gland and other tissues. *Biochim. Biophys. Acta* **1241,** 407–23 (1995).

253. Bradley, T. R., Hodgson, G. S. & Rosendaal, M. The effect of oxygen tension on haemopoietic and fibroblast cell proliferation in vitro. *J. Cell. Physiol.* **97,** 517–522 (1978).

254. Smalley, M. J., Titley, J. & O'Hare, M. J. Clonal characterization of mouse mammary luminal epithelial and myoepithelial cells separated by fluorescence-activated cell sorting. *In Vitro Cell. Dev. Biol. Anim.* **34,** 711–21 (1998).

255. Sahasrabuddhe, V. V, Luhn, P. & Wentzensen, N. Human papillomavirus and cervical cancer: biomarkers for improved prevention efforts. *Future Microbiol.* **6,** 1083–1098 (2011).

256. Yamamoto, Y., Okato, S., Takahashi, H., Takeda, K. & Magari, S. Distribution and morphology of macrophages in palatine tonsils. *Acta Otolaryngol Suppl* **454,** 83–95 (1988).

257. Reibel, J. & Sorensen, C. H. Association between keratin staining patterns and the structural and functional aspects of palatine tonsil epithelium. *Apmis* **99,** 905–915 (1991).

258. Sato, Y., Wake, K. & Watanabe, I. Changes in cell shapes and cytokeratins of epithelial cells during the infiltration of lymphocytes in the human palatine tonsils. *Acta Otolaryngol Suppl* **105,** 48–52 (1988).

259. Fioretti, A. *La Tonsilla Palatina*. (1957).

260. Gillison, M. L. *et al.* Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. *J Natl Cancer Inst* **92,** 709–720 (2000).

261. Klussmann, J. P. *et al.* Expression of p16 protein identifies a distinct entity of tonsillar carcinomas associated with human papillomavirus. *Am. J. Pathol.* **162,** 747–53 (2003).

262. Jordan, R. C. *et al.* Validation of Methods for Oropharyngeal Cancer HPV Status Determination in US Cooperative Group Trials. *Am. J. Surg. Pathol.* **36,** 945–954 (2012).

263. Clark, M. A., Wilson, C., Sama, A., Wilson, J. A. & Hirst, B. H. Differential cytokeratin and glycoconjugate expression by the surface and crypt epithelia of human palatine tonsils. *Histochem Cell Biol* **114,** 311–321 (2000).

264. Nagle, R. B., Moll, R., Weidauer, H., Nemetschek, H. & Franke, W. W. Different patterns of cytokeratin expression in the normal epithelia of the upper respiratory tract. *Differentiation* **30,** 130–140 (1985).

265. Mischke, D., Genka, T., Lobeck, H., Wille, G. & Wild, A. G. Keratins as molecular markers of epithelial differentiation: Differential expression in crypt epithelium of human palatine tonsils. *Ann. Otol. Rhinol. Laryngol.* **100,** 372–377 (1991).

266. Favre, A., Baracchini, P., Giampalmo, A. & Rossi, F. Immunohistochemical investigations of the human palatine tonsil from surgical specimens using polyclonal and monoclonal antikeratin antibodies. *Z Mikrosk Anat Forsch* **103,** 909–24 (1989).

267. Dale, B. A., Holbrook, K. A., Kimball, J. R., Hoff, M. & Sun, T. T. Expression of epidermal keratins and filaggrin during human fetal skin development. *J. Cell Biol.* **101,** 1257–1269 (1985).

268. Moll, R., Divo, M. & Langbein, L. The human keratins: Biology and pathology. *Histochem. Cell Biol.* **129,** 705–733 (2008).

269. Larouche, D., Tong, X., Fradette, J., Coulombe, P. a & Germain, L. Vibrissa hair bulge houses two populations of skin epithelial stem cells distinct by their keratin profile. *FASEB J.* **22,** 1404–1415 (2008).

270. Lindberg, K. & Rheinwald, J. G. Suprabasal 40 kd keratin (K19) expression as an immunohistologic marker of premalignancy in oral epithelium. *Am. J. Pathol.* **134,** 89–98 (1989).

271. Takeda, T. *et al.* Immunohistological evaluation of Ki-67, p63, CK19 and p53 expression in oral epithelial dysplasias. *J. Oral Pathol. Med.* **35,** 369–375 (2006).

272. Coltrera, M. D., Zarbo, R. J., Sakr, W. a & Gown, a M. Markers for dysplasia of the upper aerodigestive tract. Suprabasal expression of PCNA, p53, and CK19 in alcohol-fixed, embedded tissue. *Am. J. Pathol.* **141,** 817–825 (1992).

273. Zhong, L. ping, Chen, W. tao, Zhang, C. ping & Zhang, Z. yuan. Increased CK19 expression correlated with pathologic differentiation grade and prognosis in oral squamous cell carcinoma patients. *Oral Surgery, Oral Med. Oral Pathol. Oral Radiol. Endodontology* **104,** 377–384 (2007).

274. Babiker, A. Y. *et al.* Expressional analysis of p16 and cytokeratin19 protein in the genesis of oral squamous cell carcinoma patients. *Int. J. Clin. Exp. Med.* **7,** 1524–1530 (2014).

275. Moll, R., Franke, W. W. & Schiller, D. L. The catalog of human cytokeratins: Patterns of expression in normal epithelia, tumors and cultured cells. *Cell* **31,** 11–24 (1982).

276. Quinlan, R. A. *et al.* Patterns of expression and organization of cytokeratin intermediate filaments. *Ann N Y Acad Sci* **455,** 282–306 (1985).

277. Cooper, D., Schermer, A. & Sun, T.-T. in *Laborary Investig.* 243–255 (1985).

278. Ihrler, S., Zietz, C., Sendelhofert, A., Riederer, A. & L??hrs, U. Lymphoepithelial duct lesions in Sjogren-type sialadenitis. *Virchows Arch.* **434,** 315–323 (1999).

279. Ruco, L. P. *et al.* The lymphoepithelial organization of the tonsil: an immunohistochemical study in chronic recurrent tonsillitis. *J Pathol* **176,** 391–398 (1995).

280. Määttä, M. *et al.* Differential expression of basement membrane components in lymphatic tissues. *J. Histochem. Cytochem.* **52,** 1073–81 (2004).

281. Go, M. *et al.* Expression and function of tight junctions in the crypt epithelium of human palatine tonsils. *J. Histochem. Cytochem.* **52,** 1627–38 (2004).

282. Mackay, C. R. *et al.* Expression and modulation of CD44 variant isoforms in humans. *J. Cell Biol.* **124,** 71–82 (1994).

283. Woerner, S. M. *et al.* Expression of CD44 splice variants in normal, dysplastic, and neoplastic cervical epithelium. *Clin. Cancer Res.* **1,** 1125–1132 (1995).

284. Oliveira, D. T. & Odell, E. W. Expression of CD44 variant exons by normal oral epithelia. *Eur. J. Cancer Part B Oral Oncol.* **33,** 260–262 (1997).

285. Kiyosue, T. *et al.* Immunohistochemical location of the p75 neurotrophin receptor (p75NTR) in oral leukoplakia and oral squamous cell carcinoma. *Int. J. Clin. Oncol.* **18,** 154–63 (2013).

286. Hatakeyama, S., Yaegashi, T., Takeda, Y. & Kunimatsu, K. Localization of bromodeoxyuridine-incorporating, p63- and p75(NGFR)- expressing cells in the human gingival epithelium. *J Oral Sci* **49,** 287–291 (2007).

287. du Manoir, S., Guillaud, P., Camus, E., Seigneurin, D. & Brugal, G. Ki-67 labeling in postmitotic cells defines different Ki-67 pathways within the 2c compartment. *Cytometry* **12,** 455–463 (1991).

288. Rheinwald, J. G. & Green, H. Serial cultivation of strains of human epidermal keratinocytes: the formation of keratinizing colonies from single cells. *Cell* **6,** 331–343 (1975).

289. Honma, M., Benitah, S. A. & Watt, F. M. Role of LIM kinases in normal and psoriatic human epidermis. *Mol. Biol. Cell* **17,** 1888–96 (2006).

290. Watt, F. M., Frye, M. & Benitah, S. A. MYC in mammalian epidermis: how can an oncogene stimulate differentiation? *Nat. Rev. Cancer* **8,** 234–42 (2008).

291. Eliasson, P. & Jönsson, J.-I. The hematopoietic stem cell niche: low in oxygen but a nice place to be. *J Cell Physiol* **222,** 17–22 (2010).

292. Panchision, D. M. The role of oxygen in regulating neural stem cells in development and disease. *J Cell Physiol* **220,** 562–568 (2009).

293. Silván, U. *et al.* Hypoxia and pluripotency in embryonic and embryonal carcinoma stem cell biology. *Differentiation* **78,** 159–168 (2009).

294. Mohyeldin, A., Garz??n-Muvdi, T. & Qui??ones-Hinojosa, A. Oxygen in stem cell biology: A critical component of the stem cell niche. *Cell Stem Cell* **7,** 150–161 (2010).

295. Simon, M. C. & Keith, B. The role of oxygen availability in embryonic development and stem cell function. *Nat. Rev. Mol. Cell Biol.* **9,** 285–96 (2008).

296. Ngo, M. A., Sinitsyna, N. N., Qin, Q. & Rice, R. H. Oxygen-Dependent Differentiation of Human Keratinocytes. *J. Invest. Dermatol.* **1275700522,** 354–361 (2007).

297. Bath, C. *et al.* Hypoxia is a key regulator of limbal epithelial stem cell growth and differentiation. *Stem Cell Res.* **10,** 349–360 (2013).

298. Kato, H. *et al.* Hypoxia induces an undifferentiated phenotype of oral keratinocytes in vitro. *Cells Tissues Organs* **199,** 393–404 (2014).

299. Jones, P. H. & Watt, F. M. Separation of human epidermal stem cells from transit amplifying cells on the basis of differences in integrin function and expression. *Cell* **73,** 713–724 (1993).

300. Jones, P. H., Harper, S. & Watt, F. M. Stem cell patterning and fate in human epidermis. *Cell* **80,** 83–93 (1995).

301. Kleinman, H. K., Philp, D. & Hoffman, M. P. Role of the extracellular matrix in morphogenesis. *Curr. Opin. Biotechnol.* **14,** 526–532 (2003).

302. Kleinman, H. K. & Martin, G. R. Matrigel: Basement membrane matrix with biological activity. *Semin. Cancer Biol.* **15,** 378–386 (2005).

303. Vaillant, F., Lindeman, G. J. & Visvader, J. E. Jekyll or Hyde: does Matrigel provide a more or less physiological environment in mammary repopulating assays? *Breast Cancer Res.* **13,** 108 (2011).

304. Moutsopoulos, N. M. *et al.* Tonsil epithelial factors may influence oropharyngeal human immunodeficiency virus transmission. *Am. J. Pathol.* **171,** 571–579 (2007).

305. Raouf, A. Transcriptome Analysis of the Normal Human Mammary Cell Commitment and Differentiation Process. *Cell Stem Cell* (2008). doi:10.1016/j.stem.2008.05.018

306. Oudes, A. J. *et al.* Transcriptomes of human prostate cells. *BMC Genomics* **7,** 92 (2006).

307. Ohyama, M. *et al.* Characterization and isolation of stem cell-enriched human hair follicle bulge cells. *J. Clin. Invest.* **116,** 249–260 (2006).

308. Hackett, N. R. *et al.* The human airway epithelial basal cell transcriptome. *PLoS One* **6,** e18378 (2011).

309. Johnston, L. a. Competitive interactions between cells: death, growth, and geography. *Science* **324,** 1679–1682 (2009).

310. Díaz, B. & Moreno, E. The competitive nature of cells. *Exp. Cell Res.* **306,** 317–322 (2005).

311. Bourhis, X. D. Le *et al.* Effect of stromal and epithelial cells derived from normal and tumorous breast tissue on the proliferation of human breast cancer cell lines in co-culture. *Int. J. Cancer* **71,** 42–48 (1997).

312. Münger, K. & Howley, P. M. Human papillomavirus immortalization and transformation functions. in *Virus Res.* **89,** 213–228 (2002).

313. Miller, J. *et al.* HPV16 E7 protein and hTERT proteins defective for telomere maintenance cooperate to immortalize human keratinocytes. *PLoS Pathog.* **9,** (2013).

314. Sdek, P. *et al.* Alteration of cell-cycle regulatory proteins in human oral epithelial cells immortalized by HPV16 E6 and E7. *Int. J. Oral Maxillofac. Surg.* **35,** 653–657 (2006).