CELL-INTRINSIC MECHANISMS GOVERNING STEM CELL QUIESCENCE AND REGULATING CANCER PROGRESSION IN MAMMALIAN SKIN

by

LI WANG

B.A., Fudan University, 2009

M.A., University of Colorado, 2015

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Molecular, Cellular and Developmental Biology

2015

This thesis entitled:

Cell-intrinsic mechanisms governing stem cell quiescence and regulating cancer

progression in mammalian skin

written by Li Wang

has been approved for the

Department of Molecular, Cellular and Developmental Biology

_____

Dr. Joaquin Espinosa
(Chair)

_____

Dr. Rui Yi

Dr. Robin Dowell

Dr. Tom Blumenthal

Dr. Julie Siegenthaler

Dr. Bradley Olwin

Dr. Xuedong Liu

Date_____

The final copy of this thesis has been examined by the signatories, and we
find that both the content and the form meet acceptable presentation standards
of scholarly work in the above mentioned discipline.

IACUC protocol # _____YI 1408.01_____

Cell-intrinsic mechanisms governing stem cell quiescence and regulating cancer progression in mammalian skin

Li Wang (Ph.D Molecular, Cellular and Developmental Biology)

Thesis directed by Associate Professor Rui Yi and Assistant Professor Robin Dowell

Self-renewal of quiescent stem cells requires reentry of the cell cycle and activates cell division machinery. Molecular mechanisms that govern quiescence or promote self-renewal of tissue stem cells have been extensively studied. However, it is unclear how quiescent stem cells maintain identity and reinforce quiescence when they transition to self-renewal. Using mouse genetics and bioinformatic analysis, I showed that when hair follicle stem cells are activated, the mouse hair follicle bulge compartment induces transcription factor Foxc1. Importantly, deletion of Foxc1 in the activated but not quiescent stem cells fails to re-establish quiescence in these stem cells and subsequently drives premature stem cell activation. In contrast, deletion of Foxc1 in the stem cell niche leads to loss of the club hair without impairing quiescence. Genomic analysis reveals that Foxc1 activates critical gene networks required for quiescence including BMP signaling pathways and the transcription factor Nfatc1. These findings illustrate that quiescent and self-renewing stem cells occupy distinct cellular states and uncover an adaptive, cell-intrinsic mechanism employed by hair follicle stem cells to reinforce stemness in response to activation.

Enlightened by the success of systematic analysis of transcription networks, I further investigated cell-intrinsic mechanisms regulating cancer progression in the skin. Cancer derived from different cell-of-origins within a single tissue could develop distinct tumor

types, leading to tumor heterogeneity. Using genetically engineered mouse models, my collaborators and I assessed whether the cancer cell-of-origin controls tumor heterogeneity in mouse skin squamous cell carcinoma (SCC). We observed a considerably higher portion of invasive, mesenchymal tumors formed in hair follicle-derived SCCs while tumors developed in interfollicular epidermis consisted of well-differentiated tumor epithelial cells. Systematic transcriptome and chromatin landscape analysis revealed a set of hair follicle-specific genes that were epigenetically primed for developing more invasive tumor types. This study unraveled the importance of the cancer cell of origin in controlling tumor heterogeneity and uncovers key intrinsic regulators promoting multi-lineage differentiation during cancer progression.

Finally, I examined mRNA polyadenylation patterns in embryonic and neonatal skin lineages. Alternative polyadenylation (APA) of mRNAs plays an important role in post-transcriptional regulation by controlling binding sites for miRNAs and RNA binding proteins at the 3' untranslated regions (3'UTR) of mRNAs. To decipher this layer of gene regulation, I developed a bioinformatic pipeline that quantitatively measures mRNA expression in a 3'UTR-specific manner in epidermal stem cell lineages. My analysis suggests an evolving diversity instead of switch-like dynamics in 3'end formation in closely related epidermal cell types during skin development.

In conclusion, combining genetic mouse models with genomic analysis, I have examined dynamic regulation by transcriptional networks in mammalian skin. My study has provided new insights into cell-intrinsic mechanisms that govern stem cell quiescence, control differentiation and regulate tumorigenesis.

DEDICATION

This thesis is dedicated to my family for their continuous support and encouragement.

ACKNOWLEDGEMENT

experiments for me and is incredibly helpful for all the issues. I also would like to thank Jamie Prior and Katrina Diener for their work on the sequencing facility. Last but not the least, the staff in the MCDB department, especially Karen and Kathy, helped me with all the administrative processes and made my graduate school life much easier.

Friends and family have provided a much-needed network of support during my thesis. I cannot thank them enough for dragging me out of the lab and reminding me of the world outside of my work. I especially want to thank my parents for their unwavering support of my work, even when I couldn't explain to them what I was doing or why it was important. Lastly, I want to thank my wife Xian for always being there for me. I look forward to our life together.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1   INTRODUCTION

The extended lifespan of mammals necessitates constant tissue regeneration during homeostasis and upon injury. Tissue stem cells usually reside in specialized niches within each tissue. These stem cells both replenish their differentiated progenies during homeostasis and repair damaged tissues upon injury. To fulfill these important functions, tissue stem cells have two fundamental properties: 1) self-renewal: the ability to maintain undifferentiated stem cell populations throughout life and 2) multipotency: the potential to generate differentiated cell lineages within a tissue (Leushacke and Barker, 2011; Li and Clevers, 2010; Weissman, 2000). A major interest for stem cell research focuses on understanding the balance of self-renewal and differentiation as well as harnessing the potential of stem cells to benefit human health.

In this chapter, I will first discuss properties of adult stem cells and models of stem cell self-renewal. The research progress on different models of stem cell self-renewal establishes a context for my work concerning regulation of stem cell quiescence. I will next introduce the anatomy and development of mouse hair follicles, which are the model system for my research. Finally, I will briefly review current understanding of stem cell activation and quiescence in mouse hair follicles.

1.1   The slow-cycling adult stem cells

The enormous diversity of cellular turnover rates of different tissues suggests that distinct tissue stem cells must have specific control of the proliferation timing. For example, inner tissues such as brain and skeletal muscle exhibit relatively low proliferation rate during adulthood (Cheung and Rando, 2013; Fuchs, 2009). In contrast, tissues that

constantly encounter physical and chemical stresses, such as intestinal epithelium and skin, sustain regular tissue damage throughout normal life. To respond to these stresses, these tissues must undergo self-renewal and differentiation on a regular basis (Barker et al., 2008).

However, many tissue stem cells are slow cycling and often arrested in the G0 phase for a prolonged period of time. Consequently, these stem cell populations are referred as quiescent stem cells (Arai et al., 2004; Cotsarelis et al., 1990; Potten et al., 1997). In fact, many tissue stem cell populations were discovered based on their ability to retain DNA labels for an extended period of time, a property of retaining labels that will be otherwise diluted through the cell cycle during continuous DNA replication and cell division (Blanpain et al., 2004; Cotsarelis et al., 1990; Potten et al., 1997; 1978; Tumbar et al., 2004). However, to achieve self-renewal and embark on differentiation, stem cells must enter the cell cycle and divide. After cell division, at least one of the daughter cells must maintain an undifferentiated state and return to quiescence. In this manner, self-renewal and differentiation of tissue stem cells are accomplished through an activation-quiescence cycle. Defects in controlling this activation cycle can lead to stem cell depletion and eventually tissue malfunction.

1.2   Models of stem cell self-renewal

A simple model of self-renewal can be illustrated by single stem cell proliferation, where self-renewal is accomplished by an asymmetric cell division. During asymmetric cell division, cell fate-defining components are polarized and unequally segregated into two daughter cells during mitosis (Fig 1.1A). At the end of anaphase, one daughter cell receives

factors associated with stem cell maintenance and remains undifferentiated, whereas the other daughter cell receives factors associated with lineage commitment and embarks on a differentiation program. However, this simplified model is insufficient to fully explain the dramatic disparity in self-renewal frequency of stem cells in different tissue types. Even within the same tissue, this simple model may not be applicable to all heterogeneous stem cell populations. To account for stem cell heterogeneity, an alternative approach proposes to divide tissue stem cells into several subpopulations that co-exist within the same tissue and study stem cell activities in each subpopulation. One of the standards to classify stem cells is based on whether stem cells are quiescent or actively dividing (Fig 1.1B) (Li and Clevers, 2010). In this model, active stem cells enter the cell cycle and undergo asymmetric or symmetric cell division, and quiescent stem cells function as a reserved subpopulation to maintain the pool of stem cells without cell division. This reserved population can be activated by multiple mechanisms such as a stochastic mechanism (McKenzie et al., 2006), feedback upon loss of active stem cells or tissue injury. Different assays have been developed to analyze quiescent and active stem cells. For example, stem cell transplantation assays test the potential of isolated stem cells to generate an entire tissue in a new environment. In vivo lineage-tracing assays, in contrast, detect the potential of genetically labeled stem cells in giving rise to differentiated cell types in intact organs.

Interestingly, recent discoveries suggest that the quiescent state is not an inert state. Instead, it may also be heterogeneous and actively regulated by various intrinsic mechanisms. From the cell cycle point of view, quiescent stem cells are in G0 phase while active stem cells enter the cell cycle. Previously, Rando and colleagues observed that when a mouse limb was injured, quiescent stem cells in the injured skeleton muscle were
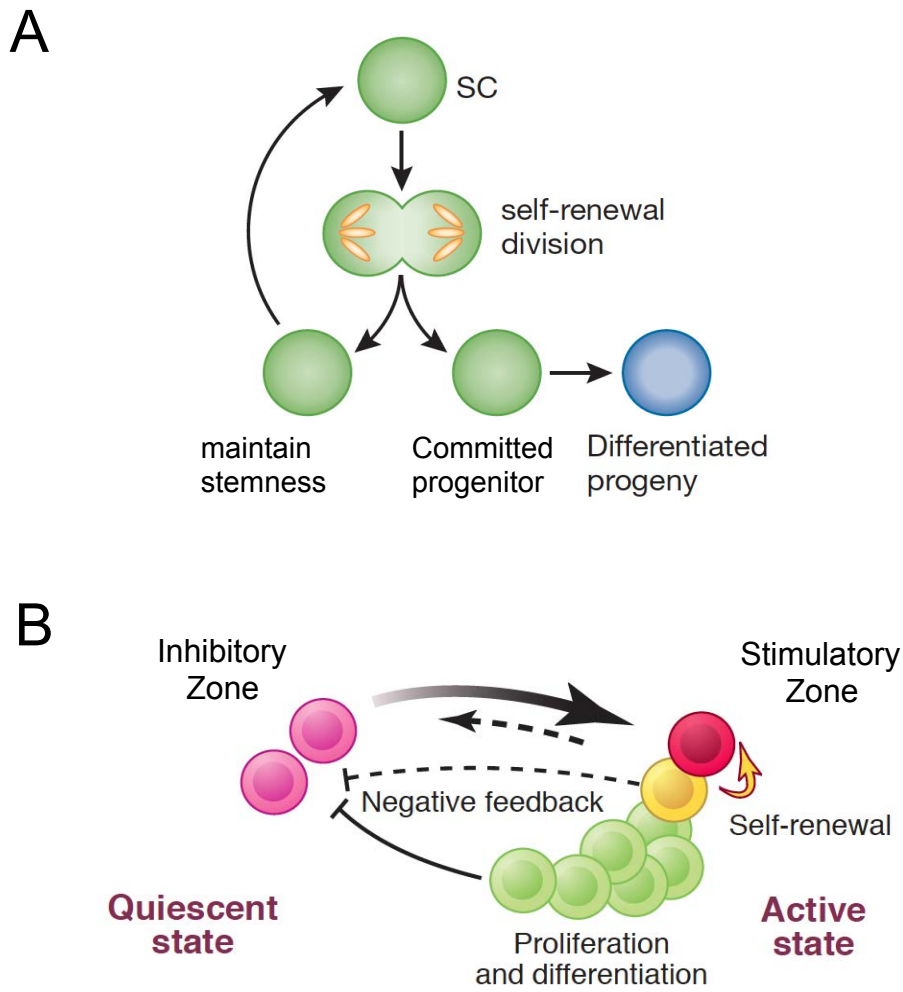
Figure 1.1 Models of stem cell self-renewal (A) Single stem cell model..
Asymmetric division is the key mechanism to maintain balance between self-
renewal and differentiation. Modified from Fuchs and Chen, *EMBO Rep*, 2013 (B)
Alternative, bi-complement model: co-existing quiescent and active stem cell
populations located in adjacent zones determined by inhibitory and stimulatory
environment. Quiescent stem cells give rise to active stem cells (heavy arrow).
Conversely, active stem cells may replenish lost quiescent stem cells (light
arrow). A negative feedback from either active stem cells (dashed line) or their
progeny (solid line) may contribute to prevention of quiescent stem cells from
activation. Modified from Li and Clevers, *Science*, 2010

activated to initiate tissue regeneration. Meanwhile, the stem cells in the opposite limb (the uninjured muscle) also changed their cell cycle profile and entered an intermediate state between activation and quiescence (Rodgers et al., 2014). The stem cells in the intermediate state show a minor but significant enlargement of cell volume, and more importantly, have a different transcriptional program than their quiescent counterparts. Since these intermediate state cells were not proliferative, the authors referred them as in a G-alert state, in which cells are poised to proliferation. Interestingly, muscle injury also leads to the emergence of G-alert quiescent stem cells in non-muscle tissues, such as fibro-adipogenic progenitors and hematopoietic stem cells (Baumann, 2014; Rodgers et al., 2014). Therefore, the G-alert state seems to be an adaptive response of tissue stem cells to injury and stress, similar to a strategy used by immune cells, in which the response to injury is influenced by previous experiences. In addition, the existence of G-alert stem cells amongst quiescent stem cells further underscores stem cell heterogeneity.

What could be the benefits by maintaining a heterogeneous pool of stem cells? One possible explanation is that animals with long lifespan need to increase the longevity of adult stem cell pools without increasing the risk of accumulating tumorigenic mutations, particularly in the rapidly regenerating tissues such as intestine and skin. Maintaining quiescent stem cell pools in adult tissues may provide a mechanism to minimize stem cells turnover when it is not necessary. When activated, however, quiescent stem cells are expected to replenish active stem cells and prevent the active pool from becoming exhausted. In turn, the active stem cells give rise to differentiated lineages but are not kept in the activated state long term. This strategy avoids accumulating DNA mutations in tissue stem cells by regularly replenishing active stem cells and thereby limiting the number of

cell divisions incurred by the population. When the quiescent pool is damaged, however, active stem cells can also replenish them (Fig 1.1B) (Rompolas et al., 2012; 2014). The combination of these reciprocal backup systems provides a robust mechanism that allows not only highly efficient self-renewal and rapid damage repair upon injury but also minimizes stem cell turnover.

1.3   Hair follicle structure and formation

The mammalian skin is the first barrier to protect an animal against the harsh environment. Skin is a complex tissue containing specialized structures and cell types derived from ectoderm and mesoderm of embryogenesis (Blanpain and Fuchs, 2009). The epidermis, an ectodermal derivative, lies on the surface and provides the barrier function of the skin. Beneath the epidermis, mesenchymal cells form the dermis, which are rich in collagen and other extracellular matrix components. These dermal cells provide a physical support for the epidermis and produce mechanical strength for the skin. At the interface of epidermis and the dermis, a layer of epidermis cells form the basement membrane by secreting and depositing extracellular matrix proteins between the epidermis and the dermis. These cells are called the basal cells of the epidermis. Basal cells consist of the stem cells of the epidermis and produce differentiated daughter cells, which migrate upward and continually renew the outer layers of the epidermis (Fuchs, 2008).

1.3.1 Architecture of hair follicle

The hair follicle is an epidermal appendage that originates from embryonic epidermis and extends downwards into the dermis. A hair follicle consists of an upper,

permanent portion and a lower portion that produces the hair shaft during the active phase of hair cycle (Fig 1.2A). The upper permanent portion includes the bulge, sebaceous gland and isthmus, a short connecting channel between the hair follicle and the interfollicular epidermis. The lower portion changes its morphology through a growth-degeneration hair cycle and is therefore called the cycling portion of the hair follicle. I will discuss the hair cycle later. For a mature hair follicle, the cycling portion consists of an extended tube structure, with the outmost layer called outer root sheath (ORS), which is contiguous with and biochemically similar to the basal layer of the epidermis. The inner layers of the hair follicle include three concentric layers of inner root sheath (IRS) and three concentric layers of hair shaft-producing cells. At the base of the hair follicle lies the matrix, which contains rapidly dividing, transit-amplifying cells that produce all of the layers of the IRS and the hair shaft itself. Dermal papilla (DP), a cluster of dermal cells and encircled by the matrix, are the signaling source of mesenchymal cell to the hair follicle. Because their profound influence to the hair follicle, dynamic signals emanating from the DP are considered one of the major driving forces for the cycling behavior of the hair follicle.

1.3.2 Hair follicle formation

Hair follicle morphogenesis has been extensively investigated in genetically engineered mouse models. In the murine skin, morphogenesis of the hair follicle starts during late embryonic development at ~E14.5, when a subset of embryonic epidermal progenitors are specified to form hair placode, the progenitor of hair follicle (Fig 1.2B) (Fuchs, 2007; Hardy, 1992; Paus and Cotsarelis, 1999). Placode specification causes dermal cells to form a condensate cell cluster directly underneath the placode, which is the
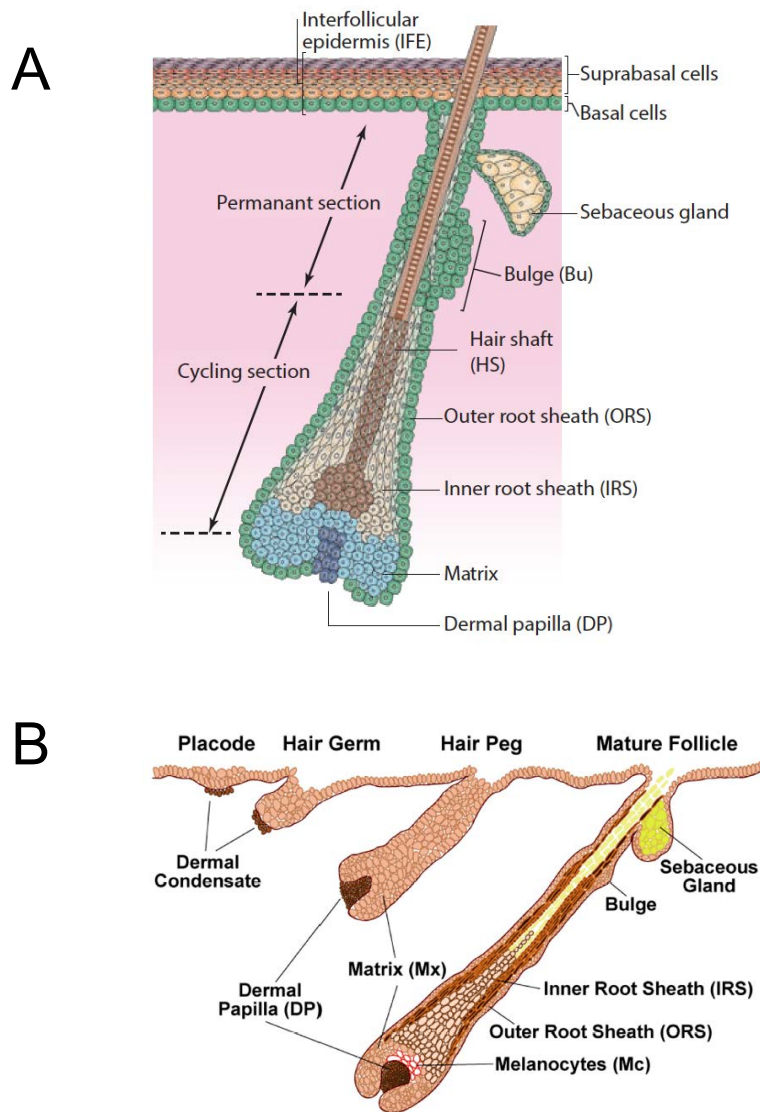
Figure 1.2 Hair follicle structure and formation. (A) Architecture of hair follicle Modified from Blanpain and Fuchs, *Nat Rev Mol Cell Biol*., 2009 (B) Embryonic stages of hair follicle morphogenesis. Discrete waves of placodes are seen beginning at approximately E14.5 and ending near birth. Once initiated, newly formed follicles continue to mature postnatally until approximately day 9. from Fuchs *J of Cell Biol*., 2008

progenitor of dermal papilla (Hardy, 1992; Paus and Cotsarelis, 1999). Upon the specification, the highly proliferative cells within the placode continue to grow and rapidly transform into the hair germ, which is a structure that grows more deeply into the dermis and subsequently becomes the hair peg (Fig 1.2B) (Hardy, 1992; Paus and Cotsarelis, 1999). Beginning at the peg stage, regional variation in proliferative status appears in the follicle. Cells at the leading edge of the hair peg rapidly proliferate and give rise to the matrix population as they surround the dermal condensate, which then becomes the dermal papilla (DP) (Hardy, 1992; Paus and Cotsarelis, 1999). Reciprocal signaling exchanges between the DP and transit-amplifying matrix cells allow the matrix progeny to engage in a distinct differentiation programs that generates the full spectrum of terminal differentiated lineages in the hair follicle, including the the inner root sheath (IRS) and the hair shaft (Fig 1.2B) (Fuchs, 2007; Hardy, 1992; Paus and Cotsarelis, 1999). Both of IRS and hair shaft are internal to the outer root sheath (ORS), whose cells are in direct contact with the basement membrane and are topologically contiguous with the basal layer of the interfollicular epidermis (Fuchs, 2007; Hardy, 1992; Paus and Cotsarelis, 1999). At birth, precursor cells to the sebaceous gland, another skin appendage, appear in the upper portion of the ORS, and give rise to the sebaceous gland shortly thereafter (Horsley et al., 2006). For most back skin follicles in mouse, hair follicle formation is completed by the end of the first postnatal week (Müller-Röver et al., 2001).

1.4 **Hair cycle**

The hair follicle is one of the few organs within the body that undergoes cyclic bouts of degeneration and regeneration throughout life (Fig 1.3A). During hair cycle, anagen is

the growth phase; catagen is the destructive or degenerating phase; and telogen is the resting or quiescent phase (Fuchs, 2007; Müller-Röver et al., 2001). Each of these phases can be further dissected into several morphologically distinguishable sub-phases. There is also a unique shedding phase called exogen, particularly prominent in human, in which the hair shaft from previous cycles shed off from the body (Milner et al., 2002). Thus, exogen hair follicles represent the hair follicles that exit the hair cycle. Hair cycle initiates with catagen after completion of hair follicle formation. Structures below the bulge, including the matrix, ORS, IRS and the hair shaft, undergo apoptosis and gradually diminish, forming a secondary hair germ, which is different from the hair germ formed during embryonic hair morphogenesis and in between the bulge and DP. The formation of the secondary hair germ morphologically marks the entry of telogen phase.

One of the distinct features for murine hair cycle is that the progression in the back skin is largely synchronized for the first 2 cycles although the exact time-scale varies depending on gender and genetic background. Typically at postnatal day 20 (P20), most hair follicles in the back skin are at a telogen phase, which last 2-6 days. The next anagen phase starts at P22-P24 for males, and P25-P27 for females. Sub-phases of anagen are of particular interest in this study because the progression of sub-phases correlates with activity of stem cells in the hair follicle (Fig 1.3B) (Greco et al., 2009; Hsu et al., 2014; Müller-Röver et al., 2001). At anagen I, hair follicle morphology is almost identical to telogen, but cell proliferation can be detected in the hair germ. At anagen II, both hair germ and DP are significantly enlarged, reflecting the continuous growth and morphological change of the hair follicle. Starting at Anagen III, signs of lineage differentiation at the bottom of the follicle become visible, which includes specification of the matrix and IRS. A

10
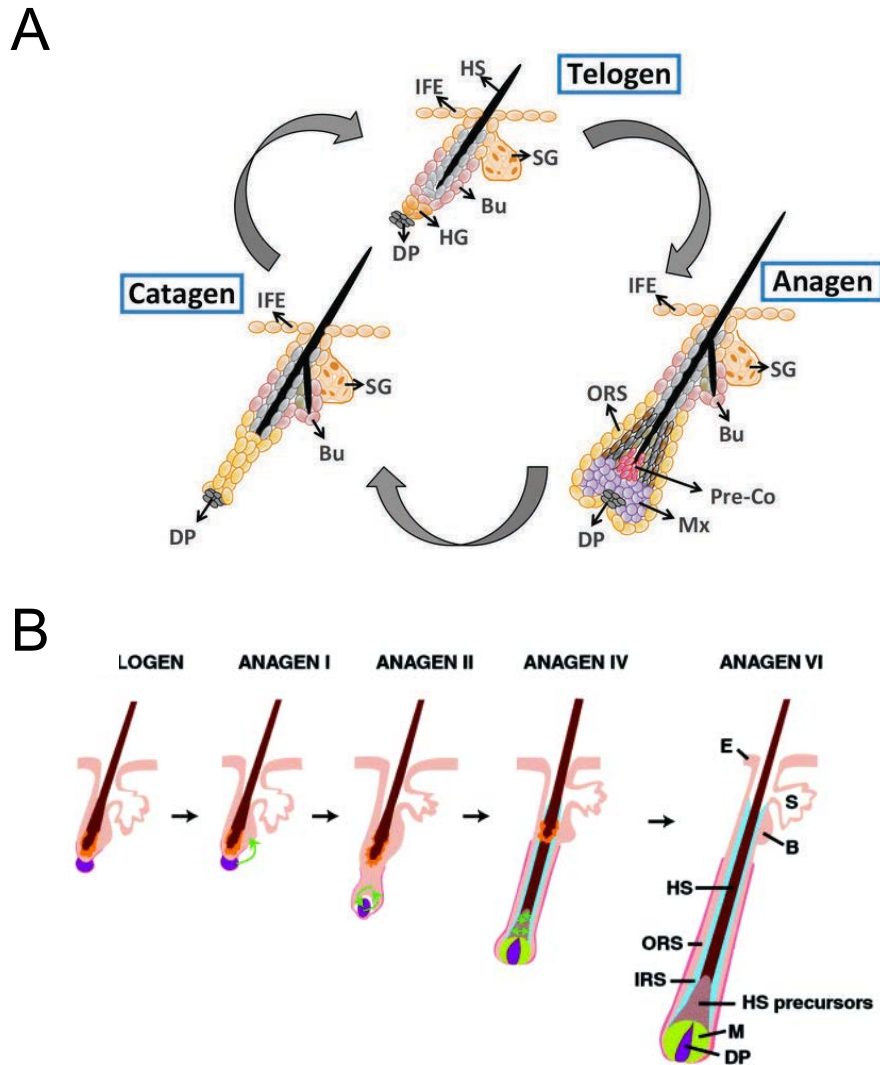
Figure 1.3 The hair cycle. (A) During the resting phase (telogen), hair follicle remain minimum size. At the onset of the growing phase (anagen), activated HFSCs located in hair germ (HG) proliferate and initiate HF regeneration. At the end of anagen, HFs enter a destructive phase (catagen) and much of the lower part of the HF undergo apoptosis. DP is drawn upward towards the bulge and the HF re-enters telogen. From Lien et al., *Cell Stem Cell.*, 2014 (B) Morphology anagen sub-phases. Selective sub-phases are shown with main markers of matrix and IRS development. IRS, inner root sheath; E, epidermis; HS, hair shaft; ORS, outer root sheath; DP, dermal papilla; M, matrix; B, bulge; S, sebaceous gland, IFE, interfollicular epidermis. from Miller, *J Invest Dermatol.*, 2008

very distinct histology marker for this sub-anagen phase is the deposition of melanin from the melanogenic activity of the follicular melanocytes (Slominski and Paus, 1993). The sub-phase of anagen III can be further separated based on parameters such as the extent of downward growth of the new hair bulb (structure below bulge) and the upward growth of the new IRS and hair shaft. The length of the anagen hair follicle reaches its maximum at the end of anagen III, after which the new IRS and hair shaft will keep extending towards epidermis. Anagen IV is classified as the tip of the new IRS and hair shaft reach the hair canal, which is the junction zone connecting the epidermis and hair follicle. The main hair follicle formation can be considered complete at the end of anagen VI, but the hair shaft will continue to grow in anagen V until it emerges through the epidermis and reaches its maximum length in the subsequent anagen VI (Müller-Röver et al., 2001) stage. The time-scale of anagen completion is approximately at P35 in murine back skin. After the completion of anagen, the hair follicle enters another catagen and degenerates the lower cycling segment. The remaining cells at the permanent portion of the newly formed hair follicle form a new bulge structure, which harbors the newly formed hair shaft and lies directly next to the old bulge. As a result, the old bulge harboring the hair shaft formed during the previous hair cycle is called the club hair. The next, extended telogen is the last synchronized hair cycle in murine back skin. The duration of this telogen can be as short as 4 weeks starting at P40 but often extends over 6 weeks in some areas of the back skin. The termination of this extended telogen also varies at different locations of the back skin. The posterior area (towards the tail) usually has a shorter telogen than the anterior area (towards the head). Different entry of the next anagen also leads to patches of hair growth

at the end of this telogen phase. However, in all WT animals that I have examined during my study, telogen termination and initiation of the next anagen is always later than P60.

1.5 Hair follicle stem cells and model of activation

1.5.1 Identification of Hair follicle stem cells

The bulge region of the follicle is located at the bottom portion of the permanent portion of the hair follicle, making it anatomically ideal for a stem cell pool that drives the cycling behavior of the hair follicle. Cotsarelis and colleagues were the first to identify label-retaining cells (LRCs) in the bulge region (Cotsarelis et al., 1990). However, because of the inability to isolate live [^{3}H] tritiated thymidine+ LRCs, they could not determine whether bulge LRCs were stem cells. Fuchs and colleagues developed an elegant method to isolate live LRCs by generating a doxycycline-controlled green fluorescent protein fused with histone H2B (H2BGFP) mouse line (Tumbar et al., 2004). Expression of H2BGFP was seen in almost all of the epidermal cells without adding doxycycline. When doxycycline was administered to young mice for 4 weeks to 4 month, most cells diluted out the H2BGFP label by cell division and only the bulge cells retained fluorescence units, reflecting their slow cycling behavior. Therefore bulge cells were isolated by flow cytometry based on the slow cycling property that leads to accumulation of the H2BGFP signal. Interestingly, the majority of these H2BGFP LRCs express high levels of a cell surface marker, CD34. (Tumbar et al., 2004). This enables purification of the bulge stem cells based on CD34 and integrin alpha6 (Blanpain et al., 2004) in any mouse skin, independent of the H2BGFP mouse line.

The multipotency of stem cells within the bulge was first suggested from transplantation experiments in which a dissected bulge region was grafted onto

immunodeficient mice (Oshima et al., 2001). Lineage tracing experiments using bulge-specific Krt15 (keratin 15)-promoter to drive LacZ reporter expression demonstrated that cells in the bulge can contribute to all epithelial lineages in the skin, including hair follicles, sebaceous glands and interfollicular epidermis (Morris et al., 2004). Transplantation experiments using a single FACS-isolated CD34(+) bulge cell that was subsequently expanded in culture demonstrated that individual bulge cells are multipotent to give rise to all skin epithelial lineages, including the bulge stem cell niche (Blanpain et al., 2004). Similar results were reported in which micro-dissected single cells isolated from rat whiskers gave rise to all hair follicle lineages (Claudinot et al., 2005). Together, these results demonstrated that bulge cells are multipotent rather than a heterogeneous mix of unipotent progenitors of specific lineages.

1.5.2 Molecular control of HFSC activation and quiescence

Since the bulge cells were determined to be the authentic hair follicle stem cells, RNA profiling assays have identified several factors as stem cell markers to these cells, such as Sox9, Lhx2, NFATc1, Tcf3, and Lgr5 (Horsley et al., 2008; Jaks et al., 2008; Nguyen et al., 2006; Nowak et al., 2008; Rhee et al., 2006; Vidal et al., 2005). It was well-accepted that HFSC activation is bulge-centric, in which interacting signals from the bulge and DP is sufficient to drive the telogen-anagen transition (Blanpain et al., 2004; Morris et al., 2004; Tumbar et al., 2004). However, careful examination of proliferation dynamics at the onset of anagen indicated that the bulge is not the first compartment to respond to the activation cue. Instead, the hair germ, which lies below the bulge and is in direct contact with the DP, begins proliferation prior to the bulge and then gives rise to the transit-amplifying (TA)

14

matrix. The bulge starts to proliferate after the hair germ proliferation (Greco et al., 2009; Zhang et al., 2009). These observations suggest a two-step model for HFSC activation (Fig. 1.4A).

Another intriguing finding from the proliferation analysis of the bulge cells is that some label-retaining bulge cells retain labels over many months, suggesting that they stay in quiescence and are not involved in hair cycle progression (Waghmare et al., 2008). This observation suggests that the bulge cells are heterogeneous. At the onset of anagen, hair germ first expands and generates the TA cells, and the bulge cells are activated later and undergo self-renewal cell division to main the stem cell pool as well as replenishing the hair germ through migration (Greco et al., 2009; Zhang et al., 2009). This bi-compartmental organization suggests that the hair follicle contains both quiescent and active stem cell populations in separate yet adjacent locations. Direct intravital microscopy observation further elaborated that the specific localization of the bulge cells can predict whether a bulge cell is likely to remain uncommitted, to generate precursors or to commit to a differentiated fate (Rompolas et al., 2014). Taken together, these studies have revealed the heterogeneity of the bulge stem cells and established a series of sequential events during stem cell activation.

Two signaling pathways are known to participate in driving HFs from telogen to anagen: BMP signaling and Wnt/β-catenin signaling (Fig 1.4B-C) (Blanpain and Fuchs, 2009; Kandyba et al., 2013; Kobielak et al., 2007; Plikus and Chuong, 2014; Plikus et al., 2011). β-catenin is a cell adhesion molecule and is required to transmit Wnt signals from cytoplasmic membrane to the nucleus when stabilized. In HF morphogenesis, Wnt/β-catenin signaling specifies the HF fate for epidermal progenitors. Ectopic stabilization of β-
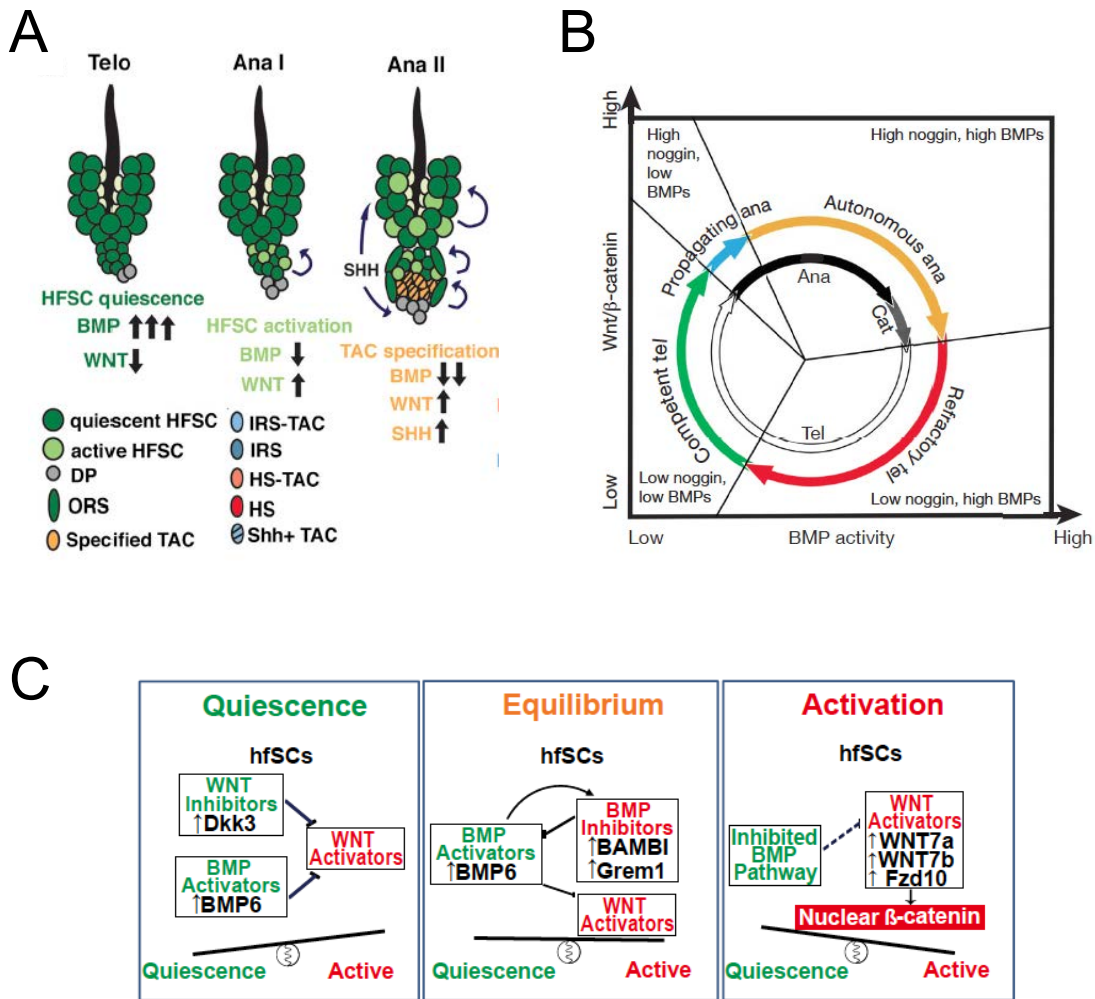
Figure 1.4 Molecular control of hair follicle stem cell (HFSC) activation. (A) Two-step model of activation of HFSCs from Genander et al., *Cell Stem Cell*., 2014 (B) Oscillation of BMP activity and Wnt/β-catenin activities establish the cycling activation-inhibitory microenvironment for HFSCs. Cumulative BMP activity characterizes refractory (inhibitor high) and competent telogen (inhibitor low). from Plikus et al., *Nature* 2008, (C) Intrinsic constant competitive cycling between activator (Wnt) and inhibitor (BMP) from Kandyba et al., *PNAS*, 2013

catenin triggers de novo HF formation in the skin epithelium (DasGupta and Fuchs, 1999; Gat et al., 1998; Silva-Vargas et al., 2005). In the adult hair follicle, nuclear β-catenin accumulates in the hair germ at the telogen to anagen transition (Greco et al., 2009). Constitutive stabilization of β-catenin triggers precocious entry of anagen in resting HFs (Lowry et al., 2005; Van Mater et al., 2003). The opposite force that maintains the bulge quiescence is BMP signaling (Fig 1.4B-C) (Blanpain et al., 2004; Horsley et al., 2008). Inhibition of BMP signaling induces hair growth and proliferation in HFSCs (Botchkarev et al., 2001; Kobielak et al., 2007). Dermal BMP signals progressively diminish during the telogen, leading to hair follicle growth (Plikus et al., 2008). Transcription factor Nfatc1 has also been shown to be activated by BMP signaling and, in turn, Nfact1 targets Cdk4 to maintain SC quiescence (Horsley et al., 2008). In addition to BMP, quiescence in the telogen is also maintained by Fgf18 signaling. Importantly, disruption of either BMP or Fgf18 signaling pathway is sufficient to shorten telogen duration and activate the anagen entry (Kimura-Ueki et al., 2012; Plikus et al., 2008). It has also been reported that the inner bulge layer, a putative stem cell niche, secretes Fgf18 and Bmp6 to provide a constant inhibitory microenvironment to maintain bulge cell quiescence in telogen (Hsu et al., 2011; Kimura-Ueki et al., 2012).

Additional regulators controlling the activation-quiescence balance in HFSCs include Fgf-7/10, Tgf-β and Shh. Under influence of all of these regulatory pathways, HFSCs constantly compare the sum of activating and inhibitory inputs from the DP and the stem cell niche (Chen and Chuong, 2012), and make the decision to become activated or to remain quiescent. Collectively, these early studies have established an important framework for me to study additional mechanisms that govern the quiescence and

activation of the bulge stem cells.

In Chapter 2, I investigate the largely synchronized HFSC populations that exist during early adult life of mouse and reveal an adaptive, cell-autonomous mechanism of regulating quiescence in the HFSCs. Integrating genetic and molecular analyses of transcription factor Foxc1, I demonstrated a lineage-specific function of Foxc1 in HFSCs as well as in the SC niche. I also discussed the future direction of phenotypic analysis to extend the understanding of Foxc1's function in other hair follicle lineages.

In Chapter 3, through systematic analysis of the transcriptome and chromatin accessibility profiling, I revealed that activated HFSCs reside at a distinct cellular state different than either their quiescent counterparts or the committed hair germ (HG) progenitors. Furthermore, I showed that adaptive expression of Foxc1 in the activated HFSCs is required to reinforce the quiescent cellular state and maintain the SC identity. These findings uncover an unexpectedly dynamic response of quiescent SCs to self-renewal and provide new possibilities for examining SC maintenance. In the discussion, I focused on the functional impact on the stem cell state in Foxc1 cKO context and future direction to further characterize both the upstream and other cooperative signaling pathways of Foxc1-dependent regulatory network.

Chapter 4 and Chapter 5 describe additional works that focused on epigenetic and transcriptional control of cancer progression and stem cell differentiation in skin lineages. Because of their relative independency to the main focus of stem cell quiescence, I included an introduction section in each of the chapters. Discussion is also provided at the end of each chapter.

CHAPTER 2 STEM CELL AND NICHE-SPECIFIC FUNCTIONS OF FOXC1 IN HAIR FOLLICLE BULGE

All of the work described in this chapter is included in Wang L, Siegenthaler JA, Dowell RD and Yi R. Foxc1 reinforces quiescence in self-renewing hair follicle stem cells, under review at *Science*.

In mammalian tissues, notably hair follicle (HF) of the skin, blood, and muscle, adult stem cells acquire quiescence and infrequently divide for self-renewal. Mechanisms that govern quiescence or self-renewal in these model systems have been extensively investigated. Deletion of CDK inhibitor p21 leads to increased proliferation of hematopoietic stem cells (HSCs) and it causes premature depletion of the HSC pool (Cheng et al., 2000). In the skin, Lhx2 and Nfatc1, both are transcription factors (TFs) highly enriched in the SCs, as well as the BMP signaling pathway are shown to be critical in governing quiescence and maintenance of hair follicle stem cells (HFSCs) (Botchkarev et al., 1999; Folgueras et al., 2013; Horsley et al., 2008; Keyes et al., 2013; Kobielak et al., 2007; Oshimori and Fuchs, 2012). Constitutive deletion of these critical regulators in the skin causes loss of quiescence, premature HFSC activation, and impaired HFSC functions, highlighting the requirement of quiescence in SC maintenance throughout the life of HFSCs (Folgueras et al., 2013; Horsley et al., 2008; Keyes et al., 2013; Kobielak et al., 2007). In addition to mechanisms intrinsic to stem cells for their maintenance, microenvironment created by stem cell niche have been implicated to play essential roles in controlling cellular state of stem cells. Specific deletion of Cxcl12 in endothelial cells or perivascular stromal cells that constitute the HSC niche depletes HSCs whereas deletion of Cxcl12 in the

HSCs themselves has little effects (Ding and Morrison, 2013). Expression of Foxc1 in marrow mesenchymal cells but not in endothelial cells or hematopoietic cells is recently shown to be required for development and maintenance of the HSC pool (Omatsu et al., 2014). In the skin, the K6+ inner bulge layer (Hsu et al., 2011), dermal papillae and more broadly, dermal fibroblasts (Clavel et al., 2012; Driskell et al., 2009; Plikus and Chuong, 2008), and adipocytes (Festa et al., 2011) are all found to contribute to the micro environment regulating quiescence and activation of HFSCs. These studies have identified important mechanisms that are constitutively required for maintaining quiescent stem cells in adult tissues. However, a critical issue remains largely unexplored - how do quiescent SCs cope with an ever-changing demand for balancing quiescence and self-renewal through cell division? Since quiescent stem cells receive activation cues from their microenvironment, it is widely assumed that activated stem cells return to quiescence and properly resume their slow-cycling identity when activating signals wanes. It is unknown for quiescent SCs whether the process of self-renewal and cell cycle reentry induces a significant change to the cellular state and how dividing SCs return to quiescence cell-autonomously.

2.1 Dynamic epigenetic control Foxc1 locus in hair cycle progression

To search for important molecules involved in cell-autonomous control of stem cell quiescence, I performed a bioinformatic screen for interesting patterns of various histone modification marks in quiescent and activated HFSCs to search for potential transcription factors involved in SC activation, using published ChIP-seq data (Lien et al., 2011). H3K4me3 marks promoter, both H3K36me4 and H3K79me3 marks active transcriptional

elongation and H3K27me3 marks repressive chromatin (Pokholok et al., 2005; Vakoc et al., 2006). Combination of several key histone modification marks demarcate gene expression in the chromosome (Mikkelsen et al., 2007). For example, accumulation of both H3K4me3 and H3K27me3 at gene promoter is an indicator of poised gene that under bivalent control (Bernstein et al., 2006). ChIP-Seq data of H3K4me3 and H3K36me4 from embryonic epidermis that I previous generated also provide a comparison to screen for gene specifically expressed in adult hair follicle stem cells.

The Foxc1 locus came to my notice because its dynamic pattern of histone modifications in HFSCs. This single-exon locus shows no sign of active transcription in embryonic epidermis; however, the gene body is heavily marked with H3K4me3 indicating it might be not strictly silenced (Figure 2.1A). Moreover, in adult, H3K27me3 shows very interesting pattern among quiescent and active HFSCs. H3K27me3 mark is specifically removed in quiescent stem cells while remains high in active HFSCs and TACs, which indicates Foxc1 might be specifically up-regulated in quiescent HFSCs (Figure 2.1B). However, H3K79me3 marks are unambiguously deposited and do not differentiate quiescent HFSCs from the active HFSCs, suggesting transcriptional elongation might be at comparable level between both populations, which is contradictory to the first hypothesis. Considering the size of this locus, the H3K27me3 marks spanning the gene body are of the size of gene promoter for multi-exon genes, thus qualify as a promoter pattern. Under such circumstances, this locus might be under an "unusual" regulation mechanism in HFSCs, but not necessarily be on/off in quiescent/active HFSCs.
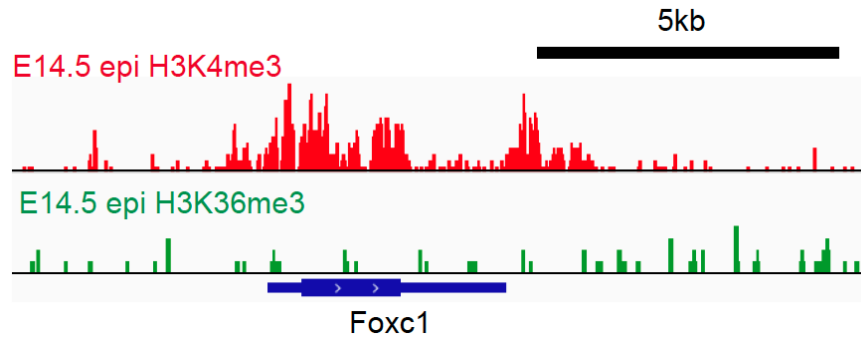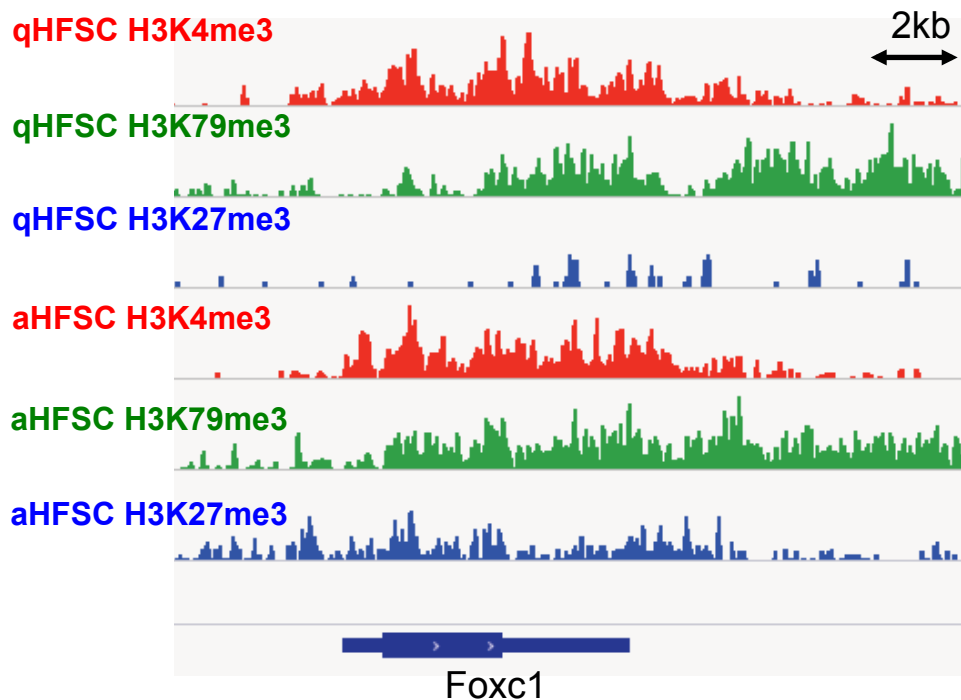
Figure 2.1 Dynamic Histone modification pattern at Foxc1 locus. H3K4me3 marks gene promoter or transcription start site; both H3K36me3 and H3K79me3 mark active Pol II elongation and H3K27me3 is associated with gene repression. (A) H3K4me3 and H3K36me3 ChIP-Seq data from E14.5 epidermis indicates absence of active transcriptional at Foxc1 locus. (B) H3K27me3 is absent in quiescent HFSCs (qHFSC) but accumulated in active HFSCs (aHFSC), while active Pol II elongation is comparably present in both populations.

2.2   Foxc1 is dynamically induced in activated HFSC compartments

Using a Foxc1-LacZ knock-in mouse model (Kume et al., 1998), I monitored the expression pattern of Foxc1 during adult hair cycle. In all stages examined, I did not observe any expression of Foxc1 in the interfollicular epidermis (Fig 2.2B). In the first telogen (~P18), Foxc1 was robustly expressed in the infundibulum and sebaceous gland progenitors but completely absent from the outer bulge layer of HFSCs, the inner bulge layer (IBL) and the hair germ (HG) progenitors (Fig. 2.2A). The first sign of Foxc1 expression in the bulge region was detected in cells located in the middle, upper portion of the HG, just below the bulge SC compartment, at the onset of hair follicle growth (Fig. 2.2A). By anagen III, when the HG further extended, Foxc1 was robustly detected in the middle layer, the precursors to the inner root sheath (IRS), but not in the outer root sheath (ORS) of the nascent HFs, which largely consists of committed progenitors (Fig 2.2A). In the bulge region, Foxc1 was strongly induced in the K6+ IBL that is sandwiched between the HFSCs at the outer bulge layer and the hair shaft, a putative niche for HFSCs (Hsu et al., 2011). The HFSCs in the outer layer of the bulge showed relatively mild expression of Foxc1 but the ORS still showed no sign of Foxc1 expression (Fig. 2.2A). When HF growth continued, Foxc1 expression became very strong in both IBL and the HFSCs, comparable to the expression levels observed in the IRS layer of the mature HFs (Fig. 2.2). In mature HFs that have fully extended into the dermis and developed the hair shaft, Foxc1 was strongly expressed in both the SC and niche in the bulge compartment and the IRS (Fig. 2.2A, 2.2C). To confirm faithful recapitulation of Foxc1 expression by the lacZ signals, I detected Foxc1 with immunofluorescent (IF) staining, which showed bulge expression of Foxc1 in anagen is at both the outer HFSC layer and inner niche layer (Fig. 2.3). When HFs progress to catagen,
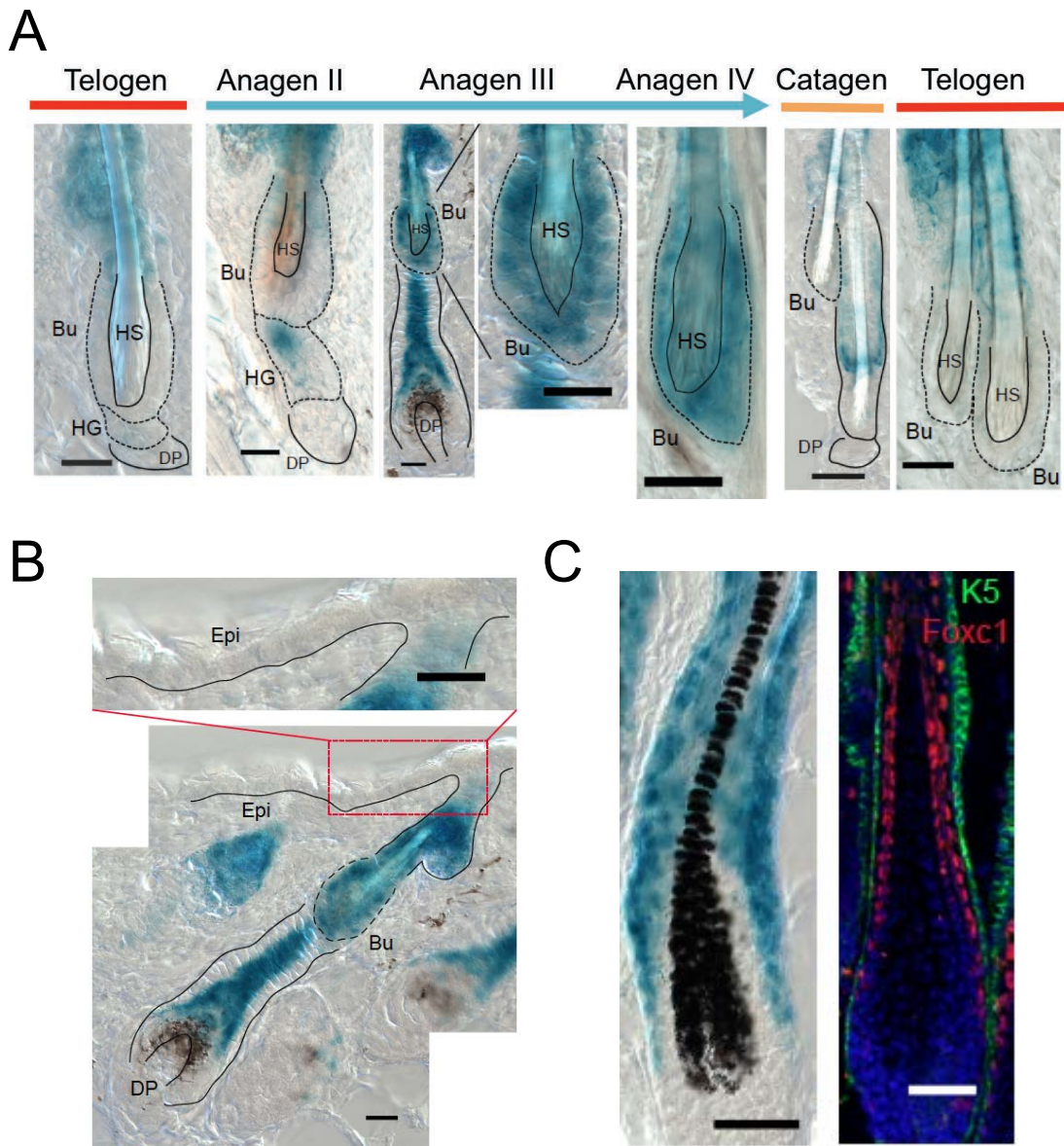
Figure 2.2 Foxc1 expression pattern through hair cycle. Foxc1 expression is detected by LacZ staining and IF in skin sections of Foxc1-LacZ knock-in mice. (A) Dynamic expression of *Foxc1* in the bulge. *Foxc1* expression is absent in telogen bulge, induced in bulge and the IRS at early anagen and remain high in bulge of mature HF. No *Foxc1* expression in the bulge during catagen and the next telogen. (B) Foxc1 expression is absent in epidermis. (C) Foxc1 expression in IRS of anagen IV. Bu, bulge; HG, hair germ; DP, dermal papillae; HS, hair shaft; Scale bar 20μm in A,B; 10μm in C.
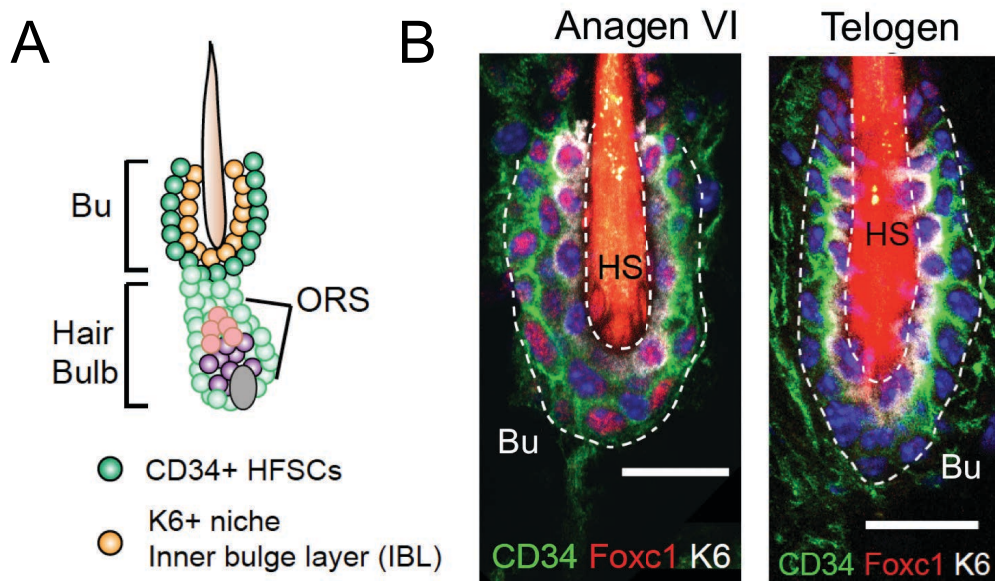
Figure 2.3 Foxc1 is expressed in both HFSCs and niche in anagen bulge. (A) Schematic structure of anagen HF. HFSCs (dark green) and niche (orange) are labeled in the bulge. (B)IF indicates that *Foxc1* is robustly induced in both CD34+ HFSCs and K6+ IBL in anagen IV, but not detectable in telogen. Bu, bulge; HS, hair shaft, with auto fluorescence; Scale bar 20μm in B.

Foxc1 was already absent from the bulge compartment of the old HF and was still expressed in the IRS layer of the new HF (Fig. 2.2A). By the early second telogen (~P47), Foxc1 completed the cycle of HF expression: it was again absent in the bulge and only expressed in the infundibulum and sebaceous gland region (Fig. 2.2 and 2.3). I also validated that the dynamic expression of Foxc1 correlates with hair cycle but not with developmental stages of mice by examining old Foxc1-LacZ animals (15-month old) in which hair cycle became asynchronous and all hair cycle stages were present in a single animal (Fig 2.4). Together, our data revealed that Foxc1 was dynamically induced in both the inner bulge layer and the HFSCs when HFs began to grow during the anagen but not expressed in either compartment during the quiescent telogen.

2.3   Foxc1 is required to maintain stemness of HFSCs

The highly dynamic expression of Foxc1 during adult hair cycle was unique among known TFs that are expressed in the HFSC compartment. I then investigated Foxc1 function in the HF with a conditional knockout (cKO) model using a well-established K14-Cre that deletes Foxc1 in all skin lineages. Both WT and cKO animals were born at the expected Mendelian ratio and developed normal skin neonatally (Fig 2.5), consistent with the lack of Foxc1 expression in embryonic skin progenitors (Fig 2.1A). I first confirmed with IF staining the deletion of Foxc1 in both the bulge HFSCs and the inner bulge layer of the SC niche (Fig. 2.6). To monitor progression of hair cycle, I referred the skin pigmentation as indicator of anagen entry (Fig 2.7A) (Plikus et al., 2008; Slominski and Paus, 1993). By monitoring the skin color that reflects hair growth in gender matched littermates, I found that the cKO HFs had a significantly shortened telogen and entered the
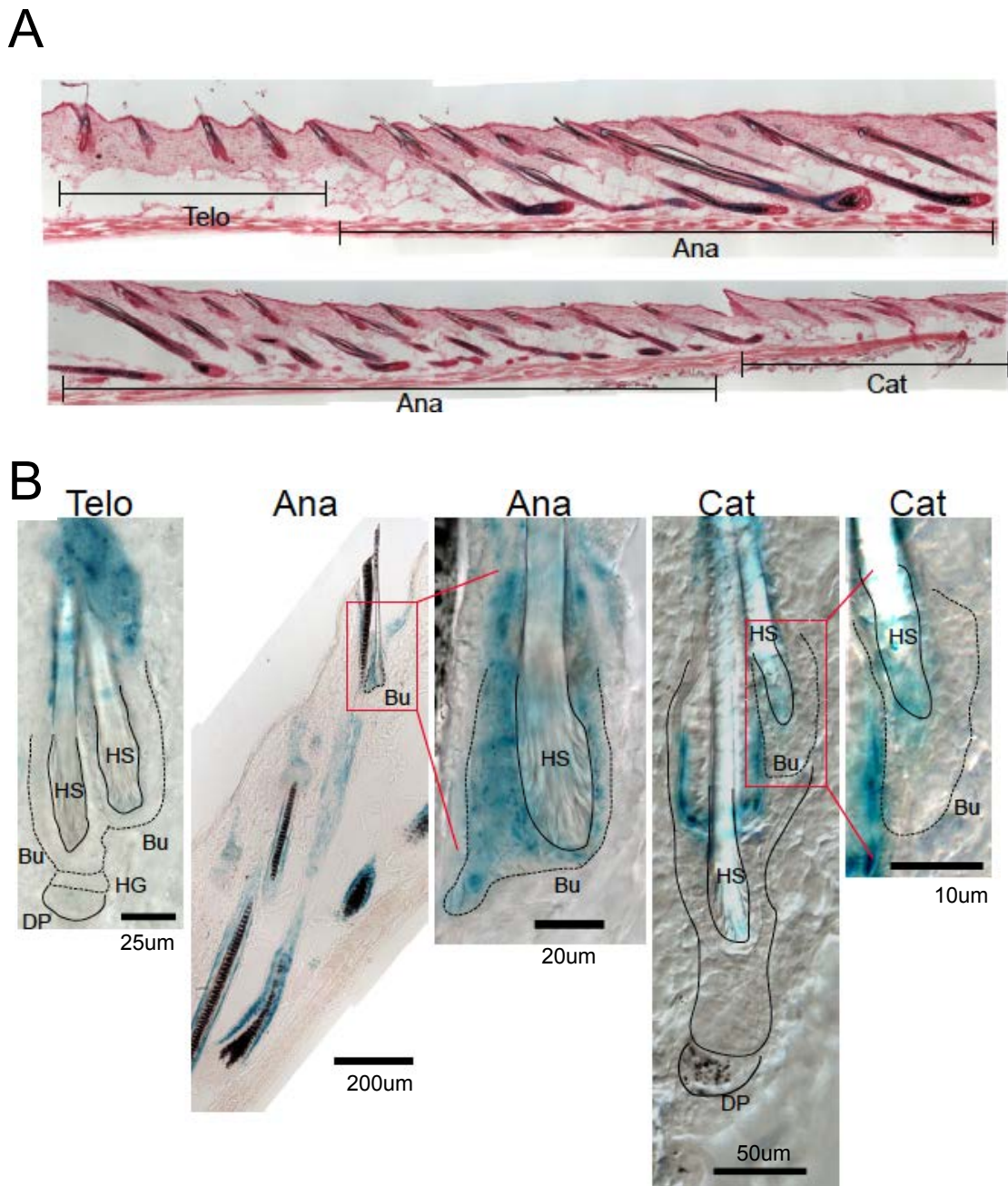
Figure 2.4 Foxc1 Expression levels correlates with hair cycle progression.
Foxc1 expression is detected by LacZ staining. (A), hair follicle morphology of a
15-month old mouse. Section was stained with LacZ and nuclear fast red. (B) In
15-month old mouse, Foxc1 expression is absent in telogen bulge, detectable in
anagen and wanes in catagen bulge. Red dashed square indicates amplified
region in the next panel. Telo, Telogen; Ana, Anagen; Cat, Catagen; Bu, bulge; HG,
hair germ; DP, dermal papillae; HS, hair shaft; Scale bar is labeled in the panel.
25μm in C, 200μm in D, 20μm in B and E, 50um in F and 10um in G.
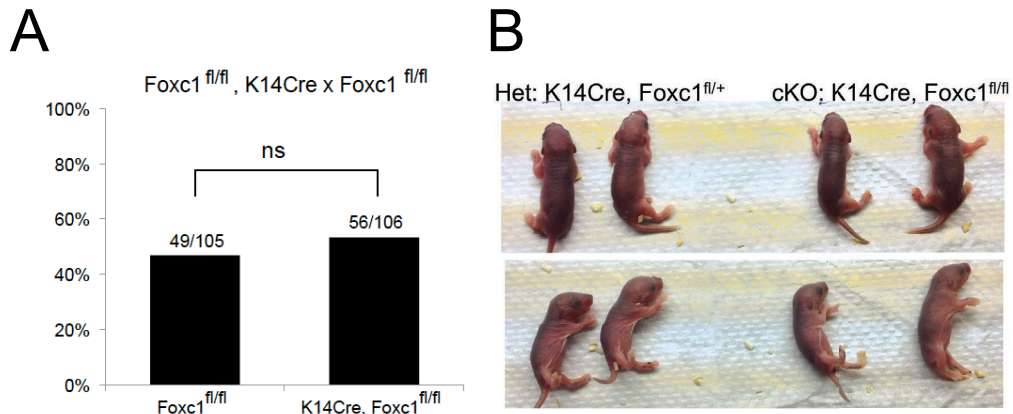
Figure 2.5 No defects in K14Cre Foxc1 cKO new born mice (A) Both WT and cKO animals were born at the expected Mendelian ratio. (B) Neonatal cKO mice is phenotypically indistinguishable from heterozygous littermates.
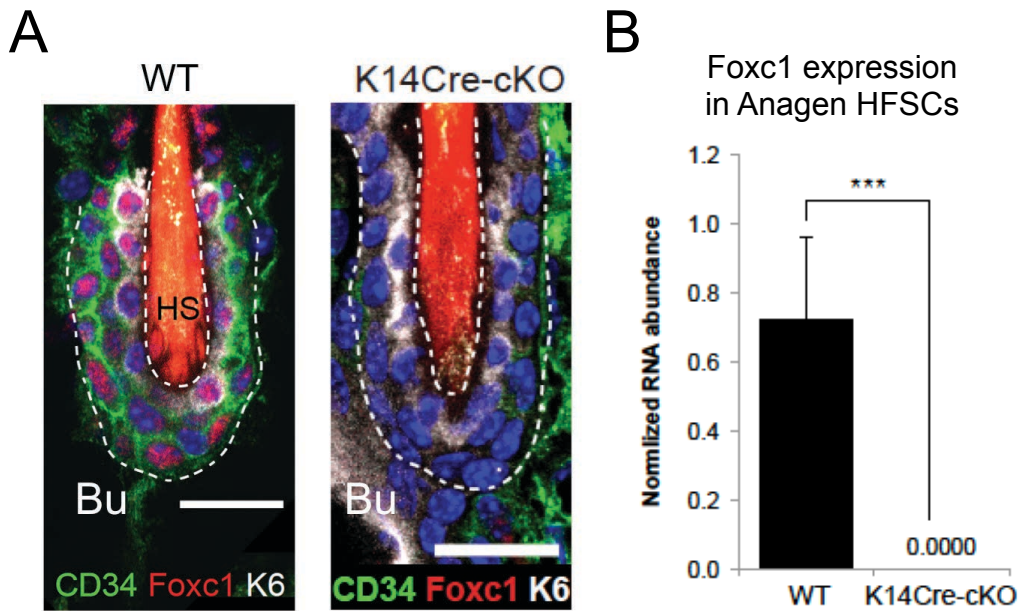
Figure 2.6 Confirmation of Foxc1 deletion in K14Cre cKO mice. (A) Foxc1 expression is not detectable in both HFSCs and IBL niche in anagen bulge of K14Cre cKO mice. Scale bar 20µm. (B) Foxc1 transcripts are not detectable in anagen HFSCs by qPCR.

anagen prematurely by the middle of the second telogen at ∼P60 (Fig. 2.7B). In comparison, the similar HF growth was typically observed not until P75 in the WT controls. These phenotypes prompted me to examine the timing of HFSC activation closely during the second adult telogen between P47 and P64. At P47, when the skin color was light in both WT and cKO animals, I did not observe any sign of HF growth and HFSC division (Fig. 2.8). However, it was evident for the club hair in the cKO is lost (Fig. 2.8). At P60, however, HF growth was initiated in the cKO as indicated by widespread BrdU incorporation in both the HG and HFSCs (Fig. 2.8). The widespread loss of the old bulge was further confirmed by unbiased, flow cytometry analysis in which the whole population of HFSCs in the back skin was detected as H2BGFP^{hi}/CD34^{hi}/α6^{hi} and H2BGFP^{hi}/CD34^{hi}/α6^{lo} populations in the WT skin (Blanpain et al., 2004). Of note, the H2BGFP^{hi}/CD34^{hi}/α6^{lo} population was indicative a cell population sandwiched between the old bulge and the new hair follicle, characteristic of the two-bulge formation in the WT skin (Blanpain et al., 2004) (Fig. 2.9A). Comparing P30 (anagen) and P47 (telogen) FACS profiles, I observed a progressive loss of the H2BGFP^{hi}/CD34^{hi}/α6^{lo} population, indicating only the new bulge was present and the old bulge was gradually lost in the absence of Foxc1 (Fig. 2.9B). To track the timing of the old bulge loss, I tracked the loss of the corresponding club hair by dyeing the hair coat at P21 prior to the anagen entry, and then observing the hair coat growth in the following hair cycle. The K14Cre cKO mice lost the whole club hair shafts between P27 and P40, which corresponds to mid to late anagen (Fig 2.10A). The timing is further confirmed by tracking the club hair loss and hair cycle progression in a single K14Cre cKO (Fig 2.10B). By shaving half the dyed hair coat and monitoring the progression of hair shaft growth, it turned out the major club hair loss happens at anagen V around P33-P35, during which the new hair
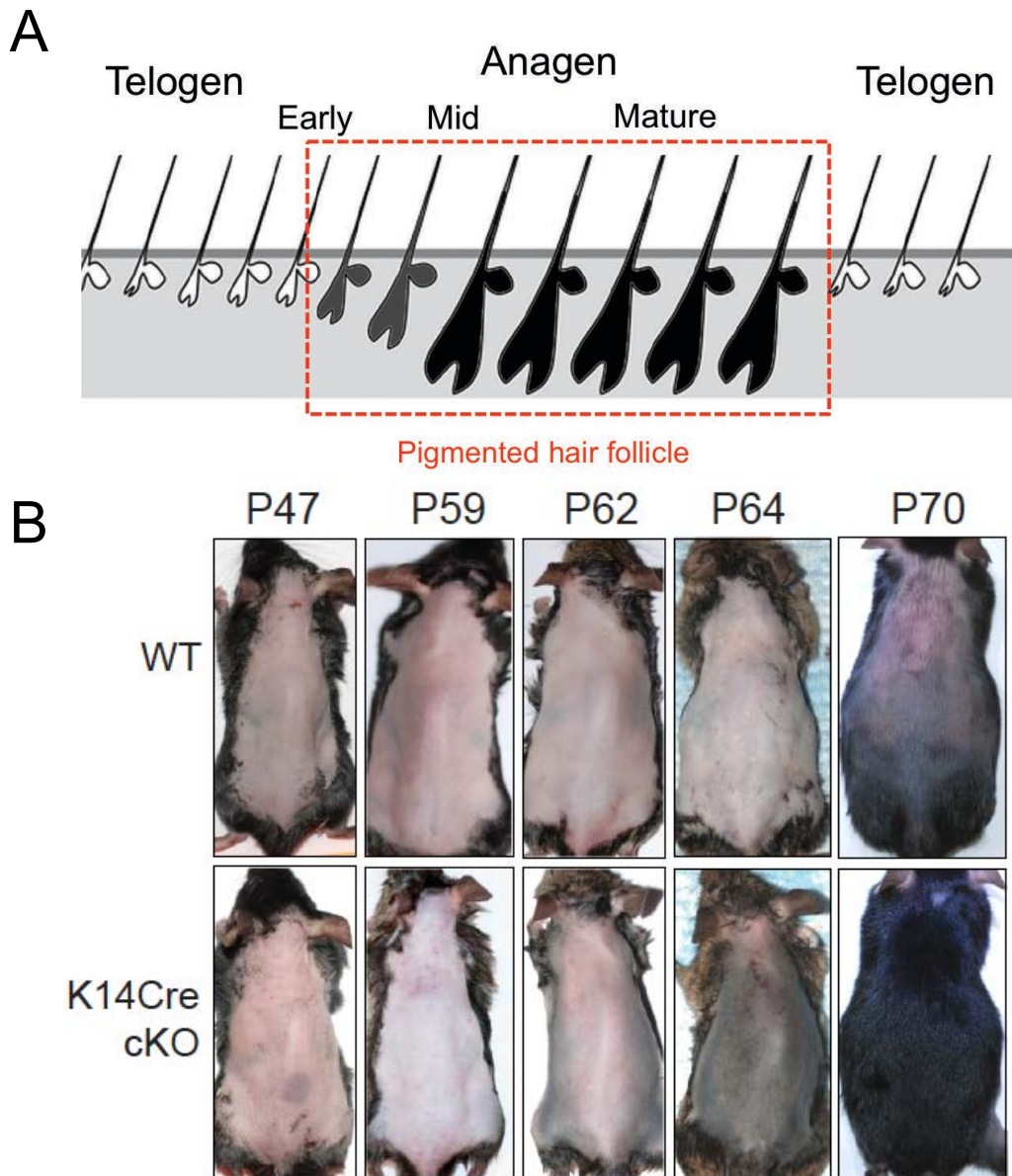
Figure 2.7 Early anagen entry for K14Cre Foxc1 cKO mice . (A) Schematic summary of hair follicle pigmentation patterns during hair cycle progression. Modified from Plikus et al., *Nature*, 2008. (B) Precocious anagen entry in K14Cre cKO mice indicated by skin pigmentation pattern and hair coat growth at P47 to P70.
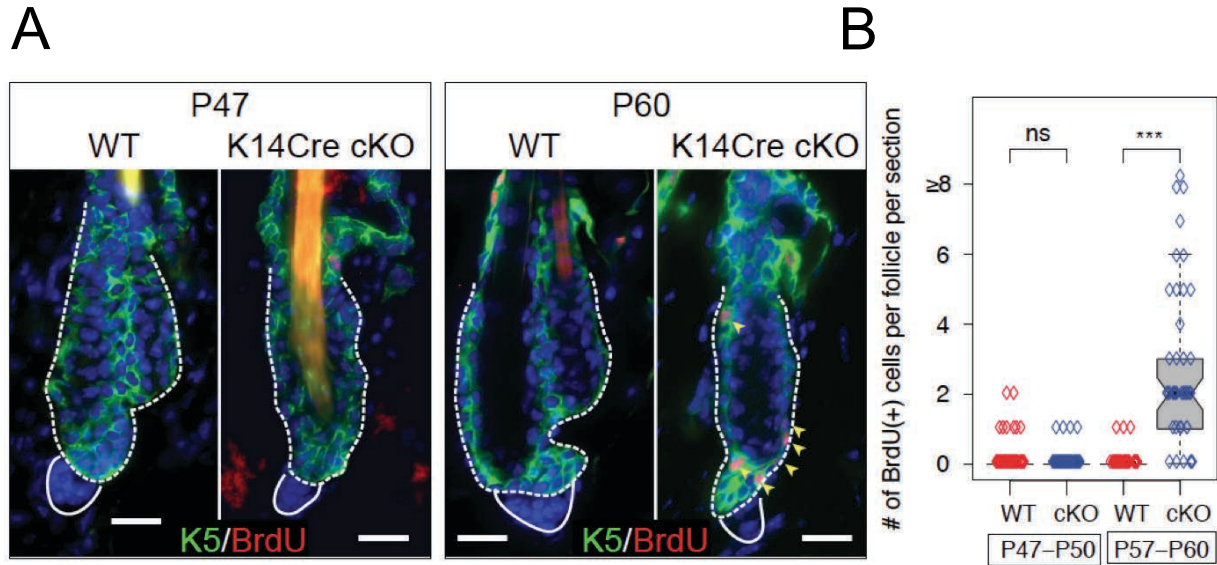
Figure 2.8 Premature HFSC activation in K14Cre Foxc1 cKO mice. (A)Detecting HFSC activation by BrdU incorporation and HF growth and the loss of club hair in the *K14-Cre* cKO. (B) Quantification of BrdU+ cells before (P47-P50, n=3) and after (P57-P60, n=3) the premature HFSC activation (***,p<0.001).
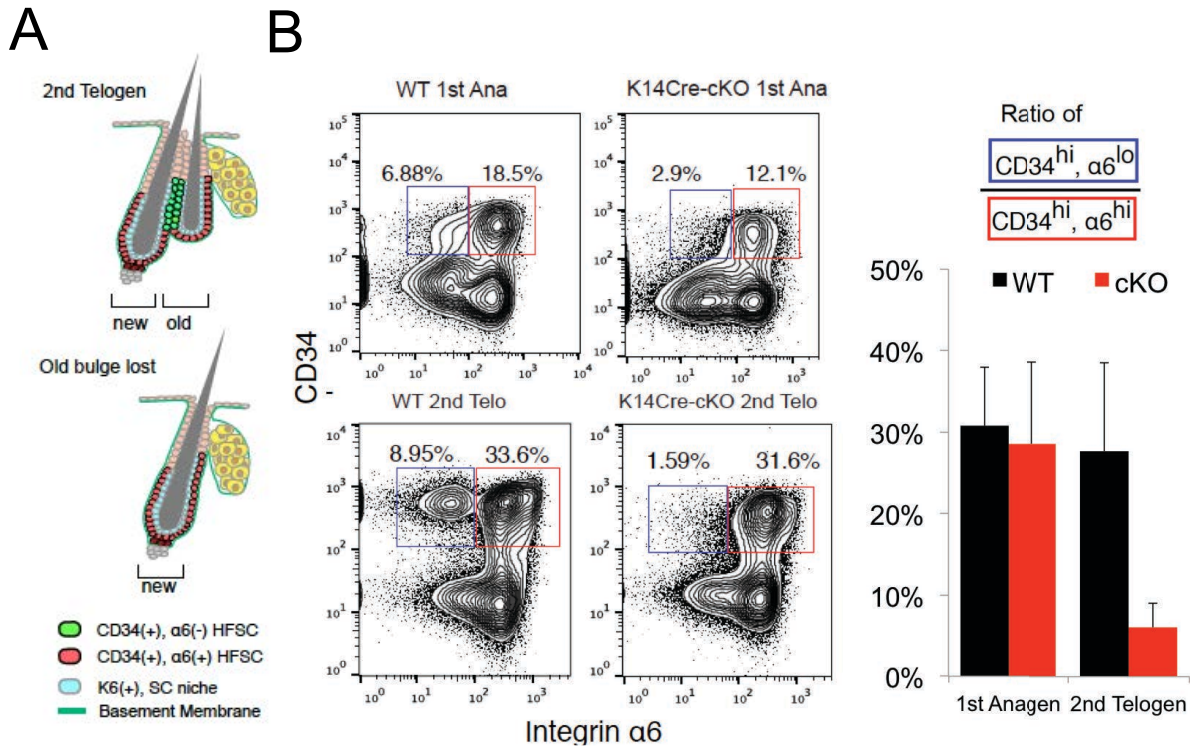
Figure 2.9 Flow cytometry confirms loss of club hair in Foxc1 cKO mice. (A) Schematic of the loss of club hair. (B) Flow cytometry analysis and quantification of the loss of club hair by the second telogen. (anagen: n=3; telogen: n=8; *** p<0.001). ns: not significant.

shaft reaches the hair canal and potentially increase the physical stress for club hair anchorage (Fig 2.10B). Morphological examination showed a flipped location between the club hair bulge and sebaceous gland, indicating the loss of club hair was correlated with upward movement of the club hair (Fig 2.10C-D). Because no abnormal apoptosis or proliferation of the bulge was observed (Fig. 2.11), this aberrant bulge movement is likely caused by compromised cell-cell adhesion. Together, these data reveal an intrinsic requirement of Foxc1 by hair follicle bulge and suggest that temporal induction of Foxc1 in the bulge during anagen may be essential to reinforce quiescence and maintain proper anchor of the club hair.

Furthermore, Immunofluorescence staining showed that the CD34 levels of the KO HFSCs were significantly reduced (Fig. 2.6A). Since multiple transcriptional factor cKO models in HFSCs downregulates CD34 expression along with compromise of the stemness properties ((Chen et al., 2012; Folgueras et al., 2013; Keyes et al., 2013), I was prompted to examine whether the HFSC identity is compromised in the Foxc1 KO HFSCs.

2.4 Dissection of SC and niche-specific function of Foxc1 in hair follicles

The model of stem cell self-renewal indicates an inhibitory micro-environment in the stem cell niche, which establishes a "default" quiescent state for adult stem cells (Fig 1.1B) (Li and Clevers, 2010). The inhibitory signals are maintained by the niche cells as well as extra inhibitory signal from the active stem cells and dermal fibroblasts (Hsu et al., 2014; Li and Clevers, 2010). Under this construction, the premature activation of Foxc1 cKO HFSCs could be explained by two mechanisms. The first explanation is a niche-dependent model, in which a less inhibitory niche in Foxc1 cKO bulge leads to precocious

32
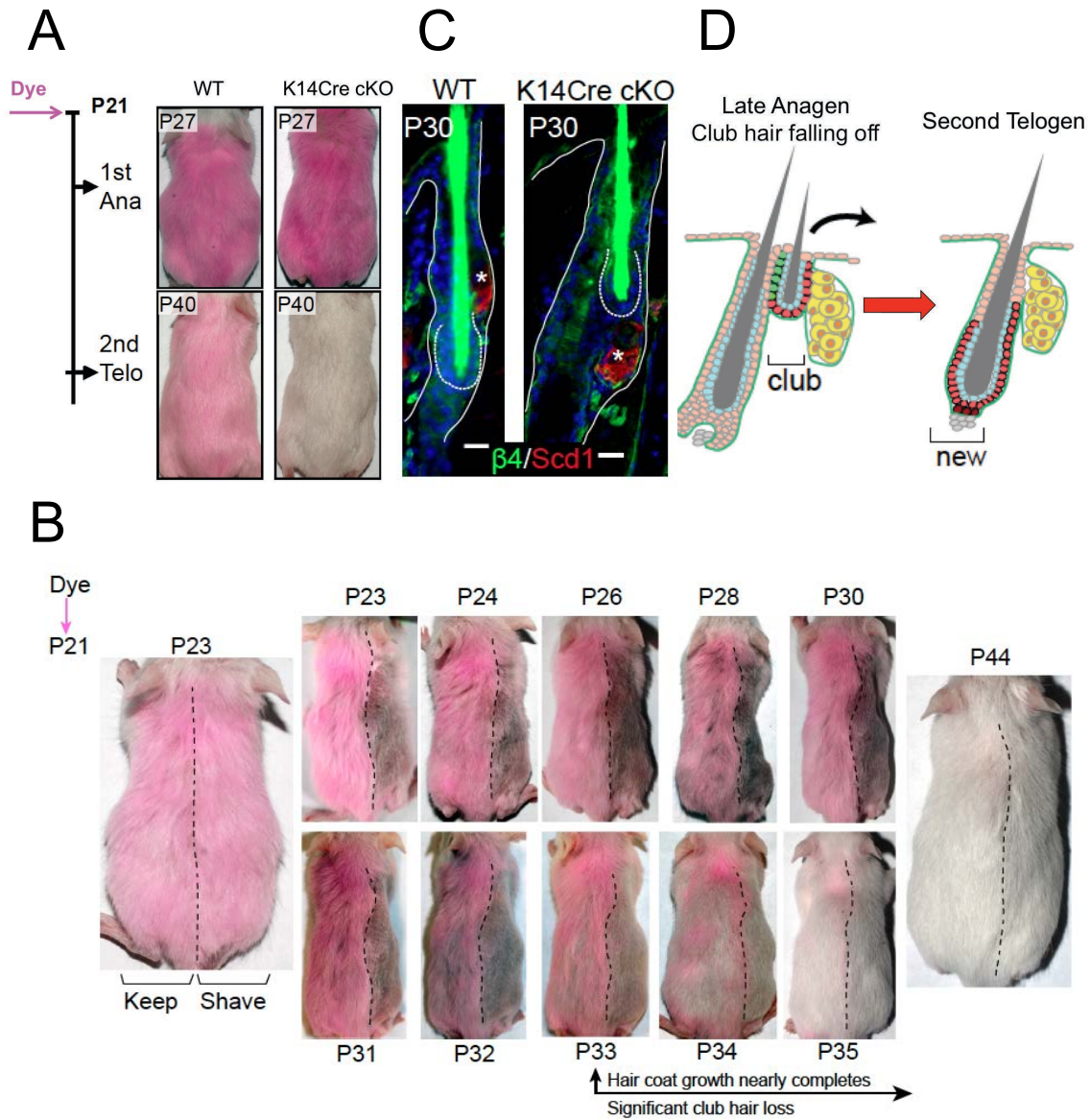
Figure 2.10 K14Cre Foxc1 cKO HFs fail to anchor the club hair. (A) Dyed hair coat shows the timing of club hair loss. (B) After dyed at P21, the right half of back hair coat was shaved and monitored to the next cycle. Significant amount of dyed hair started to shed at P33-P35, when the new hair shafts nearly complete the growth phase which corresponds to anagen V. (C) Loss of club hair correlates with upward movement of the hair. Note the relative position of the club hair and the sebaceous gland. (D) Schematic summary of the loss of club hair in cKO mice.
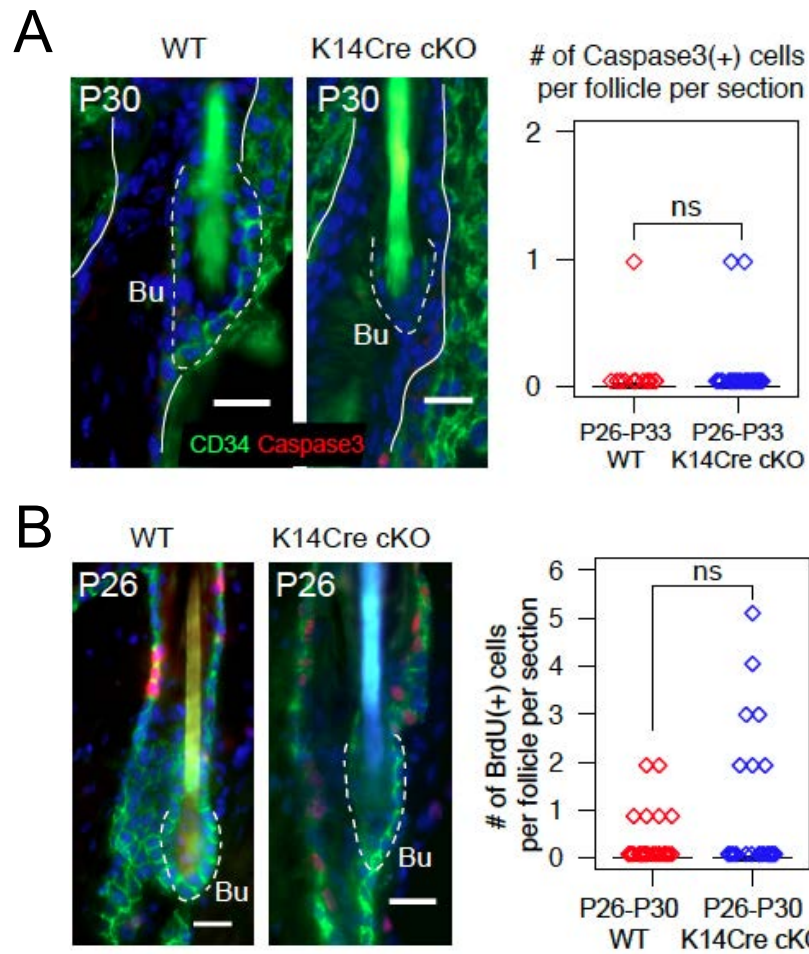
Figure 2.11 Examine apoptosis and proliferation in cKO bulge. No defects in apoptosis (B) or proliferation (C) were observed in mid to late anagen in K14Cre cKO Bulge compared to WT. n=4 in B and n=3 in C for each genotypes.

quiescence exit. The IBL is believed to partially serve as stem cell niche and send inhibitory signal to the stem cell by paracrine of Fgf18 and Bmp6 (Hsu et al., 2011). Artificial plucking of telogen hair removes the IBL layer and subsequently leads to immediate HFSC activation (Hsu et al., 2011). Therefore it is highly likely that Foxc1 expression in IBL is involved in quiescence control. An alternative niche-independent, SC-autonomous model is that HFSCs are over-sensitized, or aberrantly primed for activation in telogen when Foxc1 is deleted. Moreover, the premature activation of HFSCs in the cKO was observed in the middle of the telogen when Foxc1 was largely absent in both the niche and the HFSCs. It was unclear when precisely Foxc1 expression is required by the HFSCs. A final unsolved puzzle is whether expression of Foxc1 in IBL is sufficient to fulfill Foxc1's function to anchor the club hair. To unravel all these mysteries, SC and niche-specific deletion of Foxc1 is essential to elaborate how the loss of Foxc1 in each compartment contributes to the phenotypic defects.

Foxn1-Cre induces Cre recombination only in the inner bulge layer and the IRS. I validate the IBL-specific Cre expression by examining Tdtomato reporter in anagen bulge (Fig. 2.12A-B). With IF staining, I further validated the specific loss of Foxc1 in the inner bulge and IRS layers but not in the HFSCs at anagen (Fig. 2.13). Specific deletion of Foxc1 in the inner bulge layer resulted similar club hair/old bulge loss, as detected during mid telogen by both hair dyeing (Fig. 2.14) and flow cytometry profile (Fig. 2.15). This defect was nearly identical to that of the cKO mediated by K14-Cre, suggesting the expression of Foxc1 in the inner bulge niche is required for maintaining the proper anchorage of the old bulge. However, despite the loss of the club hair, I did not observe any precocious anagen entry at P60 in Foxn1Cre cKO mice (Fig 2.16). I cannot detect bulge BrdU incorporation
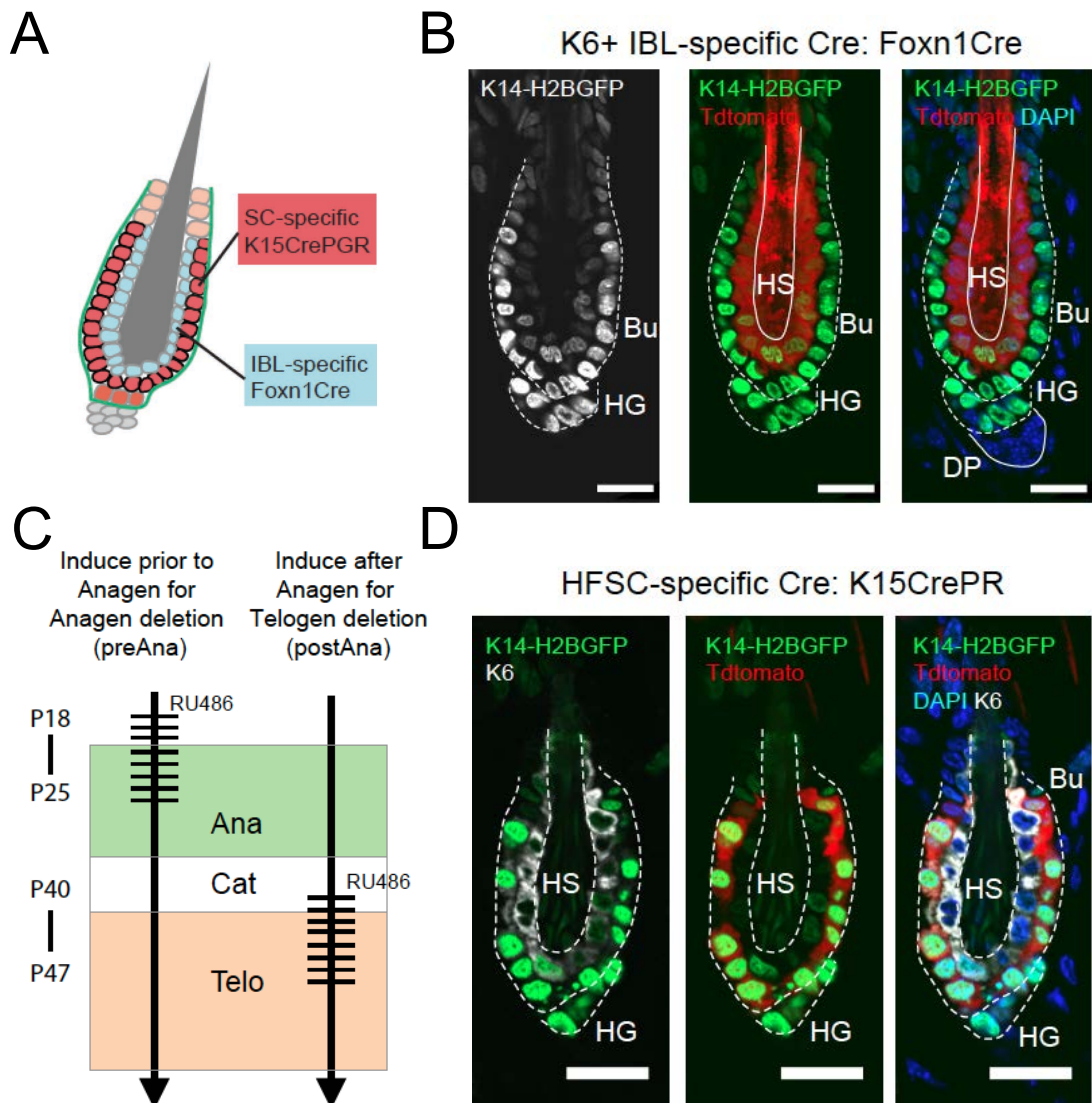
Figure 2.12 Lineage-specific cre recombinase in HF bulge. (A) Schematic summary of SC and IBL niche-specific Cre lines. (B) Foxn1Cre specifically induce Cre expression in K6+ inner bulge layer (IBL). Note that K14-H2BGFP transgene has high GFP expression in CD34+ HFSC layer and low GFP expression in IBL. (C) K15CrePR induction scheme. horizontal bars across the time arrow indicate single RU486 topical treatment. (D) K15CrePR, upon RU486 induction, specifically induce Cre expression in CD34+, HFSC layer. Bu, bulge; HG, hair germ; DP, dermal papillae; HS, hair shaft; Scale bar: 20μm.
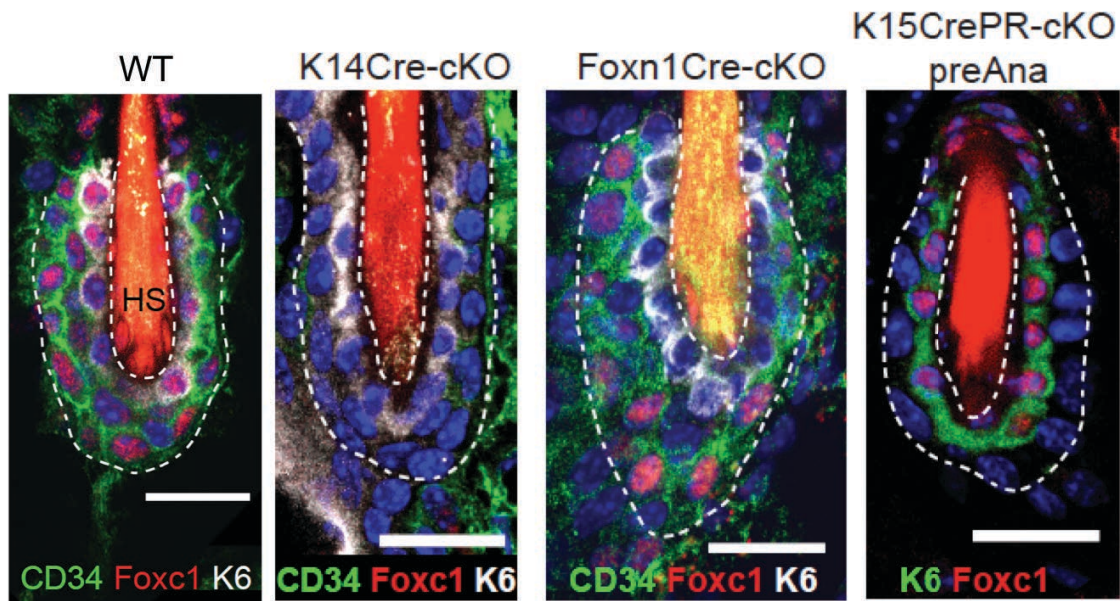
Figure 2.13 Confirmation of deletion efficiency of lineage-specific cKO. Immunofluorescence of Foxc1 expression in anagen bulge in WT and cKO of K14Cre, Foxn1Cre and K14CrePR anagen deletion animals. Bu, bulge; HG, hair germ; DP, dermal papillae; HS, hair shaft; Scale bar: 20μm.
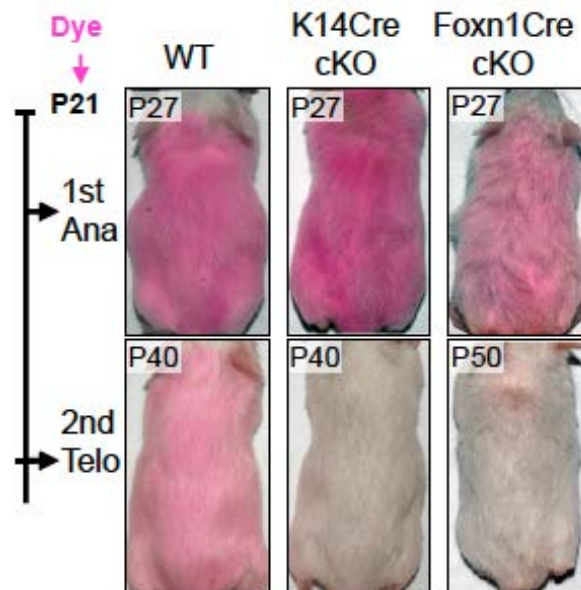
Figure 2.14 Both K14Cre and Foxn1Cre cKO mice fail to maintain old hair shaft. Hair shafts were dyed at P21 and observed through the next telogen. The Foxn1Cre-cKO mouse displayed delayed hair cycle progress and did not show complete club hair loss until P50.

even at P70 time point, indicating the timing of anagen entry for Foxn1Cre cKO is comparable to WT mice. This result suggests that the premature HFSC activation is not simply a result of compensation to the loss of the old bulge and therefore against the niche-dependent model.

To confirm the cell-autonomous role of Foxc1 in restricting HFSCs activation, I utilized a well-established K15-CrePR line (Fog 2.12A, 2.12C-D) (Morris et al., 2004), to delete Foxc1 only in the HFSCs. I first induced Foxc1 deletion in the HFSCs between P18 and P25 when the HFs enters the first adult anagen and the HFSCs become activated with strong Foxc1 expression to guarantee Foxc1 deletion throughout anagen (preAna, fig 2.12C). I did not observe significant loss of the club hair at early telogen (fig 2.15). However, by P60, I observed the premature HFSC activation as seen in the K14-Cre cKO animals (Fig. 2.17). The anagen entry timing (P60) is slightly earlier than K14Cre cKO (P62-P64), which is likely due to genetic background. Because Foxc1 is strongly induced in the anagen HFSCs and quickly wane in the telogen HFSCs, it suggests that the expression of Foxc1 in the activated HFSCs during the anagen may be sufficient to determine the SC state and modulate their behavior in the subsequent telogen stage. To confirm this, I induced Foxc1 deletion in the HFSCs between P37 and P45, when the HFSCs return to quiescence and when Foxc1 expression is already waned in the WT (postAna, Fig 2.12C). In these cKO animals, I did not observe any sign of premature HFSC activation or the loss of the club hair in the next telogen between P42 and P60 (Fig. 2.15, 2.17). Taken together, these results suggest that the temporal induction of Foxc1 in the activated HFSCs during the anagen is required to reinforce quiescence and maintain SC identity in a cell-autonomous manner.
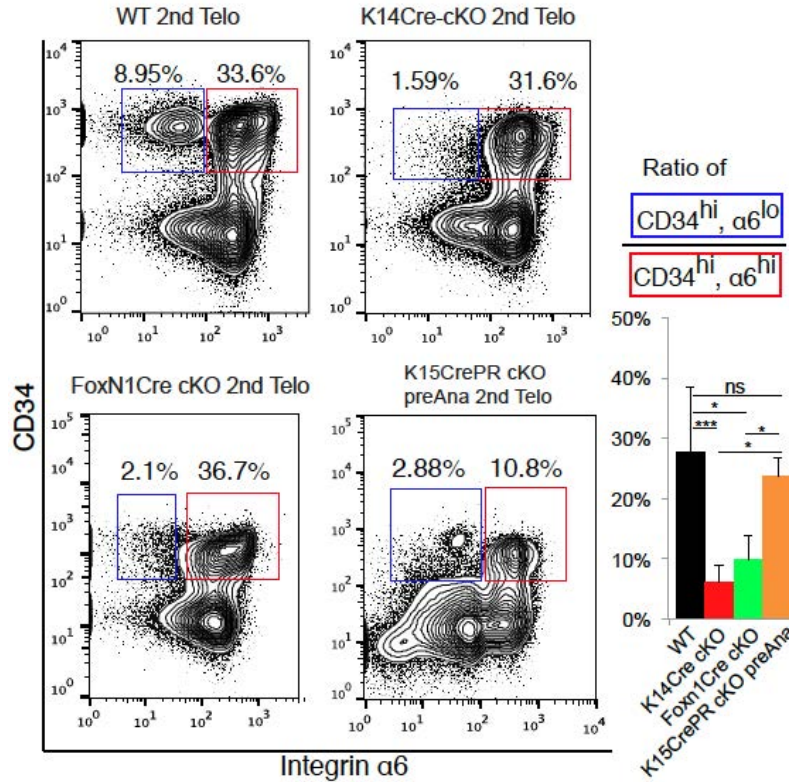
Figure 2.15 Flow cytometry confirms loss of club hair in K14Cre-cKO and Foxn1Cre-cKOmice. Flow cytometry analysis and quantification of ratio of CD34-high, α6-low over CD34-high, α6-high cells. * p<0.05; *** p<0.001, ns: not significant. WT and K14Cre-cKO are same as Fig 2.11.
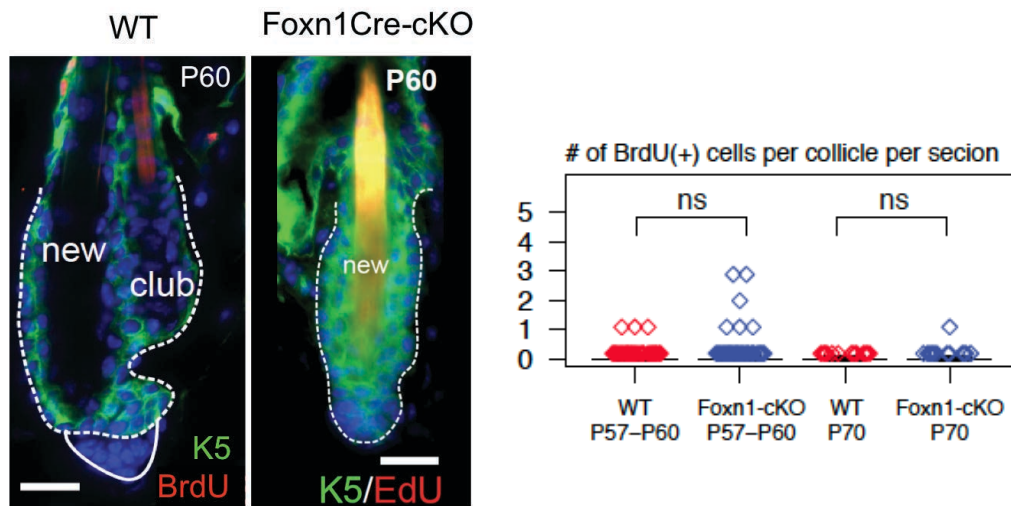
Figure 2.16 No premature activation for HFSCs in Foxn1-cKO mice. BrdU or EdU pulse labeling detects no premature activation of the HFSCs in the middle (P60, n=4) and late (P70, n=2) telogen. K5 staining as HF morphology marker. Scale bar: 20µm. ns: not significant.

2.5   Discussion

In this chapter, I examined a transcription factor, Foxc1, whose expression is dynamically regulated during hair cycle progression. The structure of hair follicle can be divided into a permanent section and a cycling section (Fig 1.2). The cycling section undergoes a growth-destruction-rest cycle corresponding to the anagen, catagen and telogen phases during hair cycle. Interestingly, zones of Foxc1 expression in hair follicle can also be categorized into permanent and cycling sections. The cycling part of Foxc1 expression has been described in details in this chapter. But Foxc1 also expresses in sebaceous gland, isthmus (the section between sebaceous gland and bulge) and part of the infundibulum throughout every phase of hair cycle without any apparent fluctuation of expression levels (Fig 2.2, 2.4). By comparing phenotypes of K14Cre, Foxn1Cre and K15CrePR cKO mice, it is evident that the phenotypes observed in K14Cre cKO can be largely interpreted as the sum of phenotypes observed in the Foxn1Cre and K15CrePR cKO mice, which reveals Foxc1 functions in the bulge and IRS, respectively. Therefore, whether Foxc1 plays any important role in isthmus, sebaceous gland or infundibulum is not clear, since no significant abnormity was observed in these regions. Two additional studies may provide more insights. Firstly, although Foxn1Cre-cKO mice lost the club hair in a similar fashion as K14Cre-cKO mice, I cannot exclude the possibility that Foxc1 is also required in isthmus and infundibulum region for anchoring the club hair. Because hair shaft is anchored in the hair cannel through the isthmus, infundibulum and the IBL, it seems possible that cell adhesion through these layers are collectively required for the anchorage of the club hair. To test this hypothesis, I need to find microscopic, molecular markers as readout for the potential phenotype. For example, if an abnormal cell-cell junction
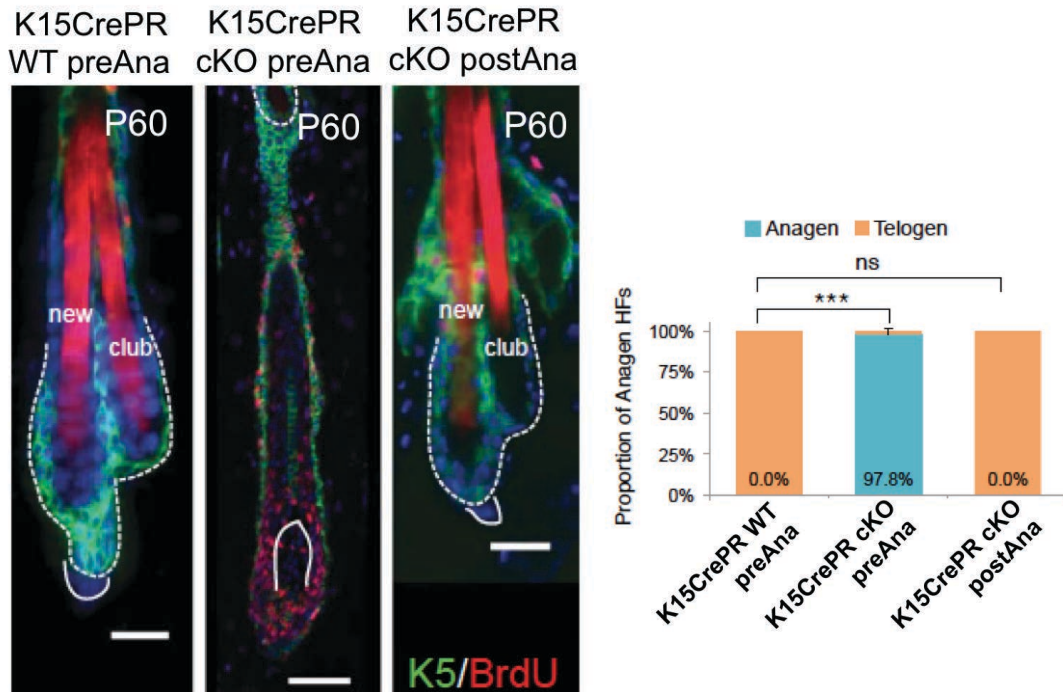
Figure 2.17 Premature activation for HFSCs in K15CrePR-cKO preAna mice but not postAna mice. Anagen but not telogen deletion of Foxc1 leads to premature HFSC activation by P60. In both cases, however, the old bulge is largely intact. Quantification of anagen HFs at P60 is shown in j (n=3, ***, p<0.001). Scale bar: 20µm. ***: p<0.001, ns: not significant.

structure can be observed as a possible explanation for the anchorage failure, I can compare K14Cre-cKO and Foxn1Cre cKO and examine whether alterations to this sub-cellular structure appear only in the IBL or in both IBL and isthmus, or infundibulum. In the latter case, I can tentatively conclude that cell adhesion through all of these regions contributes to the hair anchorage. The second potential study is to examine the sebaceous gland in the absence of Foxc1. Sebaceous gland secretes sebum and plays important roles in the skin such as barrier function, immunomodulation, antimicrobial peptide production, lipid metabolism as well as various signaling pathways involving steroid and peptide hormone synthesis and metabolism (Hinde et al., 2013). Immuno-staining of cell type markers, such as Muc1, Hmfg2, Fasn and Prdm1 can be performed to check the differentiation of sebocytes (Hinde et al., 2013). I noticed that the sebaceous gland of the K14-Cre cKO has more BrdU+ cells than that of the WT, indicative of dysregulation of proliferation in this skin appendage. However, by Scd-1 staining, I did not notice any hyperplasia in the cKO sebaceous gland. Therefore, careful quantification is required to draw a conclusion in future studies.

The induction of Foxc1 was observed in both the HFSCs and the inner bulge niche. Specific deletion of Foxc1 in each compartment leads to distinct defects. These observations indicate important but distinct roles of Foxc1 in each compartment. The mechanism of Foxc1-mediated niche maintenance remains unclear. I examined cell proliferation and apoptosis in mid to late anagen between WT and K14Cre-cKO but cannot find significant differences (Fig 2.11). However, this study can be further improved by performing more precise quantification between the cKO and WT samples in early, mid and

late telogen. The current sampling numbers are not large enough to draw conclusions separately in each phase.

Another area that can benefits from future studies is to determine whether the CD34 SC layer is lost together with the IBL. My current model speculates this is the case without robust evidence (Fig 2.10D). Since CD34 expression is downregulated in the K14Cre cKO bulge, it is no longer a suitable marker for making the HFSC layer in the cKO. However, in the Foxn1Cre cKO, CD34 expression in the HFSC is intact and the club hair loss is still evident. Thus, I can stain CD34 in Foxn1Cre cKO mice in the mid and late anagen and determine whether this layer is lost together with the club hair.

I have demonstrated that loss of Foxc1 in the HFSCs leads to premature activation of these cells in the telogen. However, I did not investigate the impact of loss of Foxc1 in activated HFSCs. Our data indicate the cKO HFSCs fail to reinforce quiescence, which occurs in mid to late anagen in WT HFSCs. It will be interesting to comprehensively examine cell proliferation levels in every sub-phases throughout the mid to late anagen in WT and cKO mice and determine whether the cKO HFSCs extend the duration of their self-renewal proliferation.

2.6   Methods and Materials

2.6.1 Mice

Mice were bred and housed according to the guidelines of IACUC at a pathogen-free facility at the University of Colorado (Boulder, CO, USA). The following mouse lines were used: K14Cre (Yi et al., 2006), K14-H2BGFP (Wang et al., 2013), Foxc1$^{fl/fl}$ (Siegenthaler et al., 2009),K15CrePGR (Jackson, 005249), Foxn1Cre (Nancy Manley, University of Georgia) and

tdtomato (Jackson, 021876). K15-CrePGR induction was perfromed by topical application of 4% RU486 (dissolved in ethanol) for 7–9 days. For anagen deletion, RU486 treatment started at P18-P20 and proceeded for 7 days or 3-4 days after the back skin turned dark color, whichever came first. For telogen deletion, RU486 treatment started after back skin turn pink/white color, which was around P37-P40 and proceeded for 7-9 days. BrdU (50 mg/g body weight) or EdU (25 mg/g body weight) was injected i.p. twice within 24h before lethal administration of CO2. Gender matched littermates were analyzed for histology and FACS. Hair dyeing experiments was performed following descriptions in Folgueras et al, 2013(Folgueras et al., 2013). For counting BrdU/EdU (+) cells in Bulge+hair germ related to Fig 2G, 2H, at least 10 hair follicles with a median of 23 hair follicles counted for each animal.

2.6.2   Histology and Immunofluorescence

Immunofluorescence was performed as described previously (Wang et al., 2013). Back skins from mice were embedded in OCT (Tissue Tek), frozen, crysosectioned (8-10μm) and fixed for 10 min in 4% paraformaldehyde (PFA) in phosphate buffered saline (PBS). Sections were permeabilized with 0.1% Triton X-100 in PBS for 10min and then blocked for 1 hr in gelatin block (2.5% normal donkey serum, 2.5% normal goat serum, 1% BSA, 1% fish gelatin, 0.1% Triton X-100). If staining with mouse antibodies, I used reagents and protocols from MOM Basic kit (Vector Laboratories). For BrdU immune-staining, slides were pretreated with DNase I for 1 hr at 37 ºC (for each panel on the slide: 0.5ul β-ME, 89ul 100mM Tris-HCL pH 8.0, 1oul 19mM $MgCl_2$, 200ul H_2O and 4ul 2mg/ml DNase I) before adding the anti-BrdU antibody. EdU incorporation was detected by Click-It EdU plus Kit

(Life Technology). Primary antibodies were diluted in blocking buffer and incubated at 4ºC overnight (O/N). After washing with PBS, secondary antibodies, conjugated with Alexa488, Alexa594, or Alexa647 (Invitrogen), were added for 1-2 hr at room temperature (RT). Slides were washed with PBS, counterstained with 4'6'-diamidino-2-phenylindole (DAPI). The following antibodies and dilutions were used: Foxc1 (rabbit, 1:200, Cell Signaling D8A6), Keratin 5 (K5, chicken, 1:2000, Covance, SIG-3475), BrdU (1:500, rat, Abcam, ab6326), Keratin 6 (K6, Ganna Bilousova and Dennis Roop, University of Colorado Anschutz Medical Campus), CD34 (1:200, rat, eBiosciences, 13-0341), β4-integrin (β4, rat, 1:200, BD Biosciences, 553745) Scd1 (1:200, Santa Cruz, sc-14719), active Caspase3 (1:200, R&D Systems, AF835), P-cadherin (1:200, R&D Systems, AF761). Nuclei were stained using 4060-diamidino-2-phenylindole (DAPI). For β-gal staining, thick sections (25μm) were fixed at 0.2% glutaraldehyde for 10min at 4°C and washed twice with detergent rinse, incubated in staining solution containing 1mg/ml X-gal at 37°C for 24h. Nuclear fast red were used as counter stain. Microscopy images were obtained using a Leica DM5500B microscope with either a Leica camera (bright field) or Hamamatsu C10600-10B camera (fluorescence) and processed with MetaMorph (MDS Analytical Technologies) and Fiji software.

2.6.3 FACS

Back skin was dissected and subcutaneous fat was removed by a scalpel. Hair cycle stage was confirmed by examining sample under dissection microscope based K14-H2BGFP-labeled hair follicle morphology, or by embedding part of the sample in OCT and examining morphology by cryosection. The remaining skin sample was minced and incubated with

0.25% collagenase (Worthington, LS004188) in 1x HBSS buffer at 37°C for 2h and Trypsin (GIBCO) at 37°C for 8min. Single cell suspensions were obtained by filtration and centrifugation, and cells were incubated with the appropriate antibodies for 30-45 min on ice. DAPI was used to exclude dead cells. Hair follicle stem cells (HFSCs) of K14Cre-based and Foxn1Cre-based experiments were isolated by enriching DAPI^{neg}, K14-H2BGFP^{hi}, Sca1^{lo}, α6^{hi} and CD34^{hi} cells. HFSCs of K15CrePGR-based experiments were isolated by enriching DAPI^{neg}, K14-H2BGFP^{hi} (if available), tdtomato^{hi}, Sca1^{lo}, α6^{hi} (if available) and CD34^{hi} cells. The following antibodies were used: integrin α6 (CD49f, 1:75; eBioscience, PE-conjugated, 12-0495; APC-conjugated, 17-0495), CD34 (1:50; eBioscience, eFluor 660-conjugated, 50-0341), Sca1 (Ly-6A/E, 1:500; eBiosciences, PerCP-Cy5.5-conjugated, 45-5981). FACS was performed on MoFlo XDP machine (Beckman Coulter). FACS data were analyzed with FlowJo.

2.6.4 Statistics

The following statistical tests were performed: two-tailed Student t-test and two sample Z-test. The 0.05 level of confidence was accepted for statistical significance. For all measurement, at least 3 biological replicates (gender-match, age-matched animals, usually littermates) were chosen as the minimum number necessary to achieve statistically significant differences between groups. No statistical method was used to pre-determine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

CHAPTER 3   MECHANISM OF FOXC1 IN REGULATION OF STEM CELL QUIESCENCE IN HAIR FOLLICLE

All of the work described in this chapter is included in Wang L, Siegenthaler JA, Dowell RD and Yi R. Foxc1 reinforces quiescence in self-renewing hair follicle stem cells, under review at *Science.*

The stem cell-specific requirement of Foxc1 to reinforce quiescence in anagen raises several important questions. I have shown in chapter 2 that ablation of Foxc1 in activated HFSCs leads to their premature activation at around P60, which is at least 20 days after Foxc1 expression wanes in the bulge. This dramatic delay in the manifestation of phenotypes suggests a lasting effect of Foxc1 on the HFSC, an effect similar to immune memory. Therefore, it is intriguing to ask what is the impact of loss of Foxc1 on the HFSCs, either epigenetically or transcriptionally. In addition, how to explain the temporal gap between Foxc1 expression and the occurrence of the corresponding phenotype? Mechanistically, what are the targeted genes of Foxc1? To answer these questions, I resorted to global, transcription and chromatin profiling for characterization of cellular state dynamics and identification of Foxc1 targets. In this chapter, I will analyze the data from transcriptome and epigenetic profiling to determine the mechanism of Foxc1 in quiescence regulation.

3.1   HFSCs fail to re-establish quiescence in the absence of Foxc1

In the early telogen, although cKO HFSCs are not active, it is possible that the apparently quiescent stem cells are transformed to an over-sensitized state. To

47

characterize the molecular impact of Foxc1 on the quiescence of HFSCs, I performed RNA-seq at P47, when the stem cells lost Foxc1 in the previous anagen and were poised for premature activation. I used the same cell sorting strategy in Fig 2.9 to isolate H2BGFP^{hi}/Sca1^{neg}/CD34^{hi}/α6^{hi} HFSCs. I purified HFSCs at P47, when both WT and KO HFSCs were still morphologically quiescent and lack any sign of cell division (see Fig 2.8). Strikingly, genes associated with quiescence of HFSCs including Bmp6 and Fgf18 (Greco et al., 2009; Kimura-Ueki et al., 2012; Oshimori and Fuchs, 2012) and HFSC markers (Blanpain et al., 2004; Morris et al., 2004; Tumbar et al., 2004) including CD34 and S100a4 were among the downregulated genes (Fig. 3.1A). Furthermore, signature genes characteristic of quiescent SCs (Cheung and Rando, 2013) were generally dysregulated (Fig. 3.1A). Gene Ontology revealed that genes involved in cell cycle were the most significantly upregulated genes and signaling molecules, adhesion molecules and cell membrane proteins were the most significantly downregulated genes in the KO HFSCs, reflecting an altered cellular state from the quiescent, WT counterparts (Fig. 3.1B). I next confirmed by qPCR the downregulation of Bmp6, CD34, and Fgf18 and the upregulation of cell cycle genes in the K14Cre cKO HFSCs, in which Foxc1 was constitutively deleted since specification of the epithelial lineage. Consistent with cell-intrinsic requirement of Foxc1 in the HFSCs during anagen, these changes were only pheno-copied in K15-CrePR preAna models, in which Cre was induced before anagen entry and Foxc1 was specifically deleted in the activated HFSCs. Neither the Foxn1-Cre cKO mice nor K15-CrePR postAna mice show this pattern (Fig 3.1C). Since Foxn1-Cre mice delete Foxc1 in the IBL but not in the HFSCs, and K14CrePR postAna mice delete Foxc1 in the HFSC after anagen, these data clearly demonstrated the cell-intrinsic requirement of Foxc1 reinforce quiescence in activated HFSCs.

A

HFSC marker	CD34 (-2.3x), S100a4 (-2.2x)
Negative regulator of HFSC activation	Fgf18 (-2.5x), Bmp6 (-3.3x), Grem1 (-2.0x)
Positive regulator of HFSC activation	Ptch2 (+8.4x), Gli1 (+2.2x)
Quiescent SC signature (Enriched)	Zyx (-2.5x), Smarca2 (-1.9x)
Quiescent SC signature (Depleted)	Top2a (+4.7x), Ccnb1 (+3.8x), Ccna2 (+3.0x), Anln (+3.9x), Sgol1 (+3.9x)

B

Upregulated
cell cycle
cell division
-log10(FDR)

signal
glycoprotein
cell adhesion
membrane
ECM
-log10(FDR)

Downregulated

C

HFSC / Quiescence Markers

Normalized RNA Abundance

Foxc1 Bmp6 CD34 Fgf18

WT
K14Cre
FoxN1Cre
K15Cre preAna
K15Cre postAna

Cell-cycle Genes

Normalized RNA Abundance

Ccna2 Cdc20 Foxm1 Kif23 Top2a

Figure 3.1 Up-regulation of cell cycle genes and down-regulation of HFSC quiescence regulators (A) Functional classification of selected differentially expressed gene in P47 KO HFSCs. (B) Gene Ontology of all differentially expressed genes in P47 KO HFSCs. (C) qPCR validation of downregulated and upregulated genes in the HFSCs from different cKO models (n≥3, *, p<0.05).

49

To globally evaluate the cellular state of the cKO HFSCs, I performed Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005) to compare Foxc1 KO HFSCs in early telogen (P47) to quiescent HFSCs and HG progenitors (Greco et al., 2009). This analysis showed that bulge signature genes from telogen HFs are significantly depleted in Foxc1 KO HFSCs while HG signature genes are potently enriched, indicating Foxc1 KO HFSCs in early telogen are more similar to the HG progenitors that are activated or poised to be activated than the quiescent HFSCs (Fig. 3.2). I also compared the differentially expressed genes identified in the Foxc1 KO HFSCs to those identified in the Lhx2 KO HFSCs (Folgueras et al., 2013) and observed a strong similarity in the molecular signatures between these KO HFSCs (Fig. 3.2). Interestingly, Lhx2 cKO mice display significant similar phenotypes as Foxc1 cKO mice, including defects in maintaining SC quiescence, compromised SC identity and failure of club hair anchor (Folgueras et al., 2013). Moreover, when I expanded the comparison to multiple datasets using previously published microarray profiling data for HFSCs (Greco et al., 2009; Lien et al., 2011), gene signature of HGs in early and late telogen and transient-amplifying cells (TACs) in anagen matrix are all significantly over-represented while bulge signatures from early and late telogen are both depleted (Fig 3.2), further confirming the altered quiescence program in cKO HFSCs. To directly visualize the relationship and potential hierarchy of cellular state during HFSC activation-quiescence cycle, I mapped the cellular states of Foxc1 KO HFSCs along with datasets from the HFSCs residing at the quiescent or activated states (Greco et al., 2009; Lien et al., 2011; Zhang et al., 2009), the HGs isolated during the telogen (Greco et al., 2009) and the TACs (Lien et al., 2011) using principle component analysis (PCA). Quiescent HFSCs from both the first (P20 and P23) and the second (P56, P69) telogen were grouped

together (Fig. 3.3). All HGs (P43, P56 and P69) were grouped closely and clearly separated from the HFSC groups and the TACs, consistent with their progenitor but non-SC status (Fig. 3.3). Interestingly, activated HFSCs (P28) and dividing HFSCs destined for differentiation (P20) (Zhang et al., 2009) were located between the quiescent HFSCs and the HGs, defining an intermediate state between the HFSCs and the HGs (Fig. 3.3). Strikingly, the P47 Foxc1 KO HFSCs were grouped with the activated HFSCs and clearly distinct from the quiescent HFSCs by K-means clustering and linear discriminant analysis (Fig 3.3). This meta-analysis of cellular state based on global mRNA profiling data delineates that the activated HFSCs reside at a transitioning state between the quiescent HFSCs and the HG progenitors. When Foxc1 is ablated in the activated HFSCs, these HFSCs fail to return to the quiescent state. Instead, they continuously occupy the intermediate, ready-for-activation state. Such sensitized cellular state leads to premature HFSC activation in the subsequent telogen. Taken together, Foxc1's function of quiescence regulation is taken place specifically in activated HFSCs while Foxc1 in the IBL function to maintain proper anchor of hair shaft.

3.2   Global impact of transcriptome upon Foxc1 deletion in Anagen

Transcriptome profiling at early telogen HFSCs show enhanced activation features, but differentially expressed genes identified in this dataset are likely mixed with lasting primary targets with significantly amount of secondary effects, since Foxc1 expression is absent since catagen. To determine the mechanism of Foxc1's action and identify Foxc1 controlled targets in the HFSCs, I analyzed differentially expressed genes in Foxc1 WT and KO HFSCs between P29 and P31, when the HFSCs were activated and when Foxc1 was

51
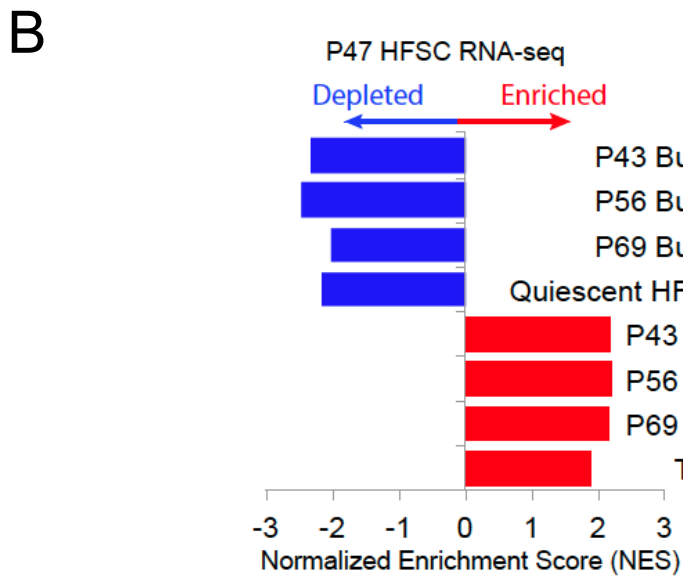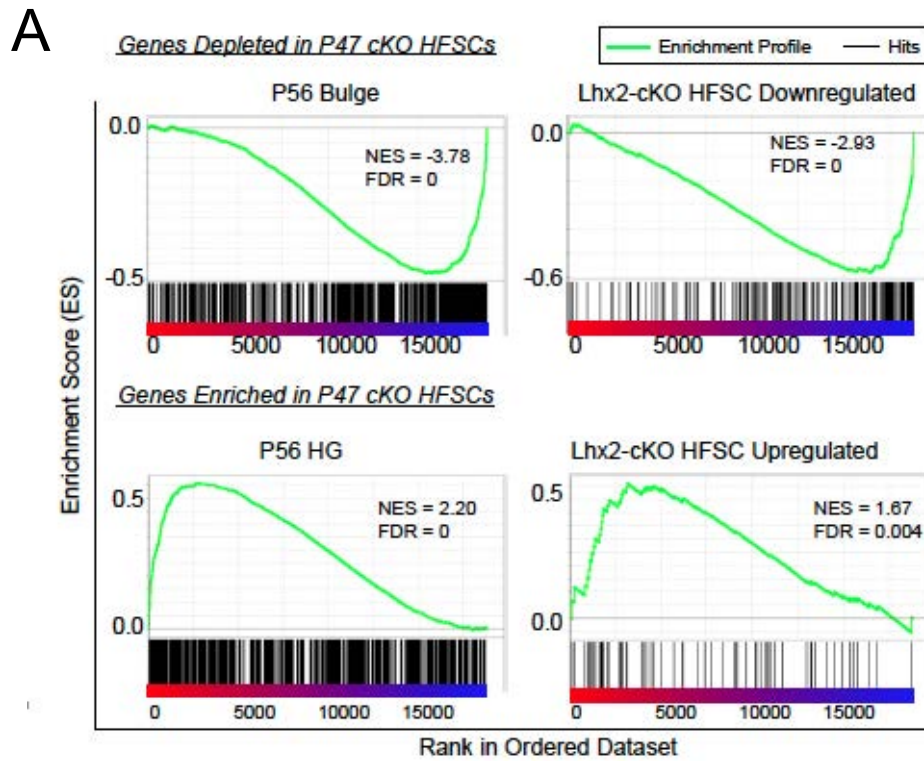
Figure 3.2 Over-representation of activation signature and depletion of quiescence signature in Foxc1 K14Cre-cKO HFSCs. (A) GSEA comparison of P47 Foxc1 KO transcriptome to telogen (P56) bulge and hair germ (HG) signature genes as well differentially expressed genes in Lhx2 HFSC cKO mice. (B) Quantification GSEA results of NES of bulge, HG and trans-amplifying cells (TAC) signature dataset. All depletion/enrichment has FDR<0.001.
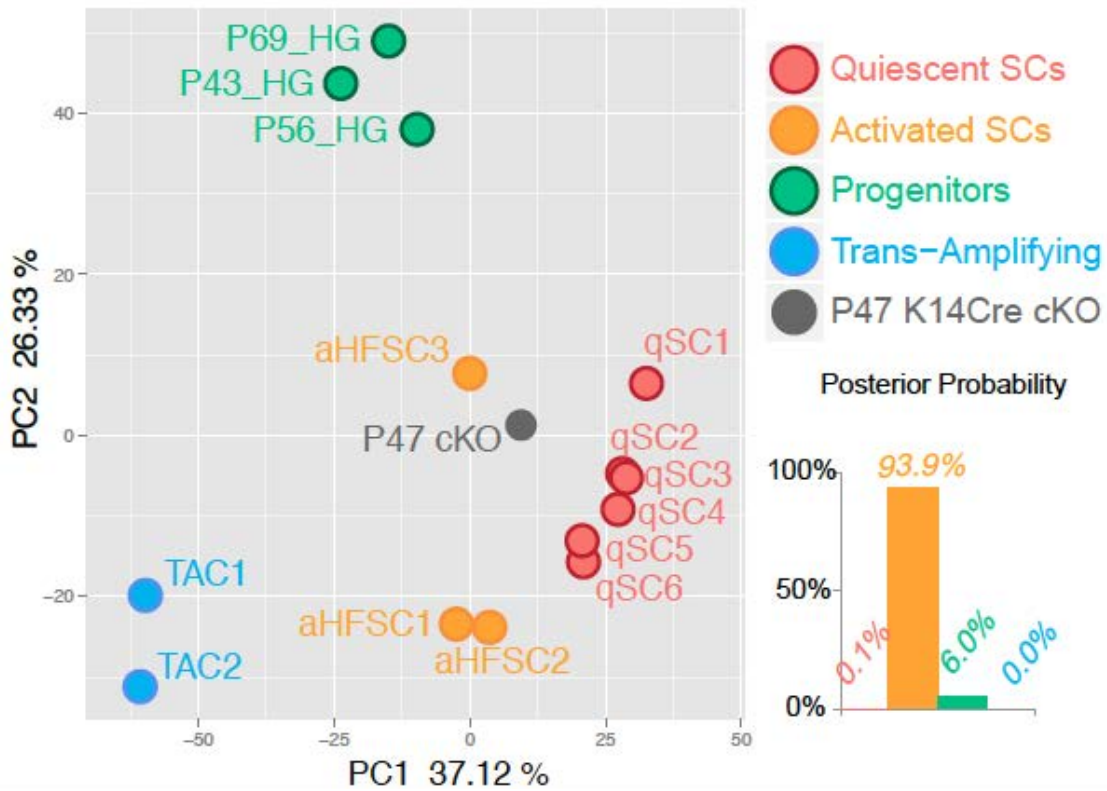
Figure 3.3 PCA of Foxc1 KO transcriptome and profiling data from WT HFSCs, HGs and TACs. Posterior probability calculated using linear discriminant analysis groups Foxc1 KO transcriptome to activated SCs. aHFSC3: self-renewing HFSCs at P18-P22; qSC1: HFSCs at P69 (late telogen); qSC2: Non-dividing HFSCs at P18-P22; qSC3: Self-renewal dividing HFSCs at P22-P25; qSC4: Non-dividing HFSCs at P22-P25; qSC5 and qSC6: HFSCs at mid telogen.

highly expressed. Genes associated with HFSC quiescence were notably downregulated including Fgf18 (Greco et al., 2009; Kimura-Ueki et al., 2012), Bmp2 (Plikus et al., 2008), Bmp6 (Oshimori and Fuchs, 2012), Nfatc1 (Horsley et al., 2008; Keyes et al., 2013) and Foxp1 (Leishman et al., 2013) in the cKO (Fig. 3.4A), among which Nfatc1, Bmp2 and Bmp6 levels are not affected in telogen RNA-seq data. HFSC markers including Krt15, CD34, S100a4 (Blanpain et al., 2004; Morris et al., 2004; Tumbar et al., 2004), and Lgr6 (Snippert et al., 2010) were also significantly downregulated (Fig. 3.4A). Among upregulated genes, the strongest groups were genes directly involved in the regulation of cell cycle, cell division and kinetochore, reflecting a more proliferative state of the cKO HFSCs (Fig. 3.4B). In addition, signaling and secreted molecules, adhesion molecules and basement membrane genes were significantly enriched among the downregulated genes (Fig. 3.4B). Intriguingly, GSEA strongly suggested a prominent similarity of Foxc1-dependent transcriptome with BMP signaling pathway-dependent program, as both up and down-regulated genes in Bmpr1a cKO HFSCs (Genander et al., 2014; Kobielak et al., 2007) are similarly dysregulated in Foxc1 cKO HFSCs (Fig. 3.5). Together, the transcriptome profiling data suggests a potential co-regulation mechanism of BMP targets through Foxc1 and BMP signaling pathway.

3.3   Strategy of transcription factor target prediction using ATAC-seq

To globally define Foxc1's impact on the HFSC genome, I characterized open chromatin regions in Foxc1 WT and KO HFSCs with ATAC-Seq (Buenrostro et al., 2013). To validate whether our ATAC-Seq data can robustly identify occupied DNA regions by TFs, I analyzed published ChIP-Seq data for Sox9 (Kadaja et al., 2014), Tcf3/4 (Lien et al., 2014)
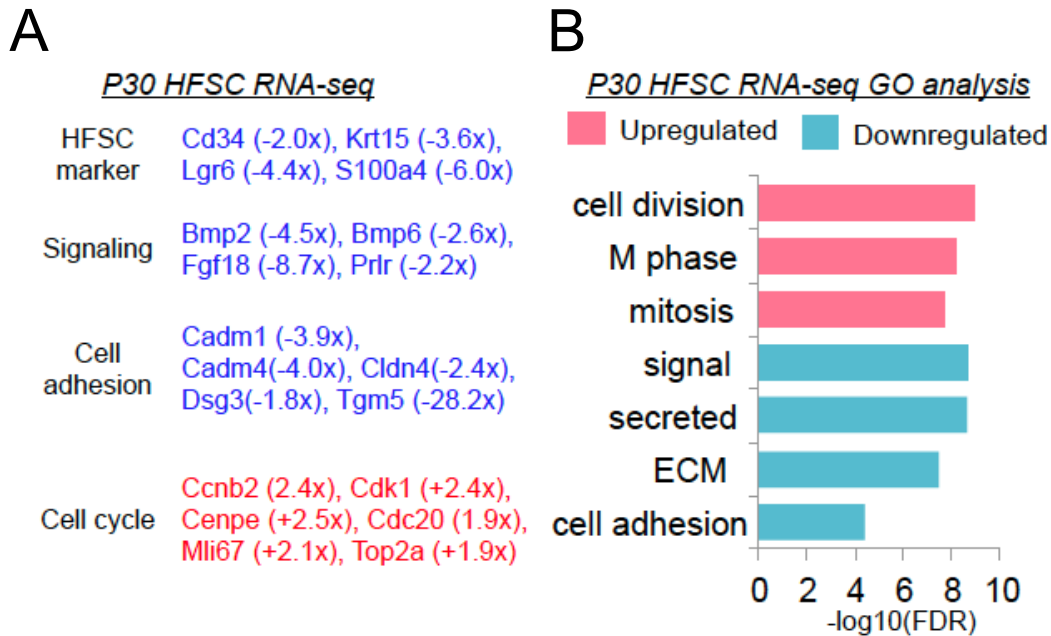
Figure 3.4 Differentially expressed gene in P30 Foxc1 cKO HFSCs. (A) Functional classification of selected differentially expressed gene in P30 KO HFSCs. (B) Gene Ontology of all differentially expressed genes in P30 KO HFSCs.
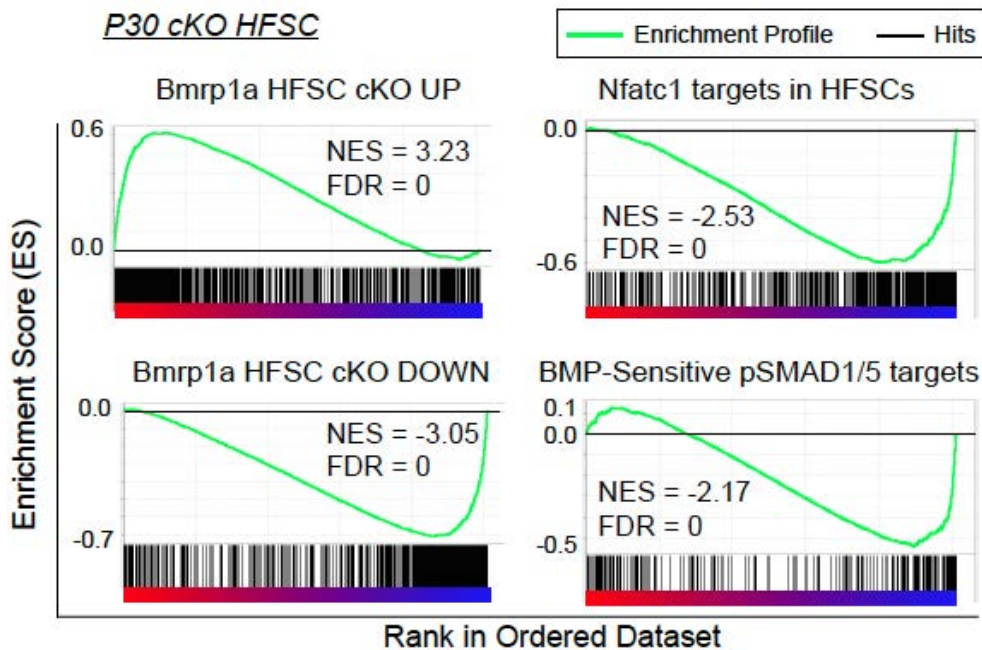
Figure 3.5 GSEA of BMP responsive genes and Nfatc1 targets in P30 KO HFSCs. Left: similar transcriptome response upon Foxc1 deletion and Bmpr1a deletion in HFSCs. Right: preferential down-regulation of Nfatc1 targets and BMP-sensitive pSmad1/5 targets in P30 Foxc1 cKO HFSCs.

and Lhx2 (Folgueras et al., 2013), all of them are critical TFs in the HFSCs. I observed 97.5%, 91.7%, and 100% overlapping between ChIP-Seq peaks for Sox9, Tcf3/4, Lhx2 and our ATAC-Seq peaks, respectively (Fig. 3.6A). Furthermore, 96.1% of the ATAC-Seq peaks, which overlapped with the TF ChIP-Seq peaks, only harbored a single event (Fig. 3.6B). Reversely, 96.1% of ATAC-seq peaks only intersect with a single ChIP-seq peak for a given TF (Fig 3.6C). It indicates that ATAC-Seq can robustly resolve individual TF occupancy genome-wide, consistent with reports in other cell types (Lara-Astiaso et al., 2014). Although the number of FACS-purified HFSCs that are required to carry out Foxc1 ChIP-Seq (>1x10^7 cells) exceeded our ability to isolate these cells, I analyzed a Foxc1 ChIP-Seq dataset that was generated from E11.5 mouse embryos (Amin et al., 2015), for Foxc1 binding sites. De novo motif discovery retrieved Foxc1 binding motifs that were best matched by known Foxa2 and Foxa1 recognition motifs (p=1e-152 and 1e-24, respectively) (Fig. 3.7A). Importantly, Foxa2 and Foxc1 were experimentally validated to bind to the same enhancer sequences of Tbx1 with a similar affinity in a gel shift assay (Tamagishi et al., 2003), consistent with our genomic findings. In addition, both Foxa1 and Foxa2 loci are epigenetically silenced in the HFSCs (Fig 3.7B-C). Therefore, I intersected ATAC-seq peaks harboring the Foxc1 motif with differentially expressed genes (hits) in anagen RNA-seq data, to achieve a list of high-confidence Foxc1 binding sites in the genome (Fig 3.8).

3.4 Foxc1 coordinates with BMP-signaling pathway and Nfatc1 in regulation of HFSC quiescence

Target prediction generated a list of 104 genes that are likely targeted by Foxc1. Among these putative Foxc1 targets, 85 (81.7%) genes were downregulated in the Foxc1
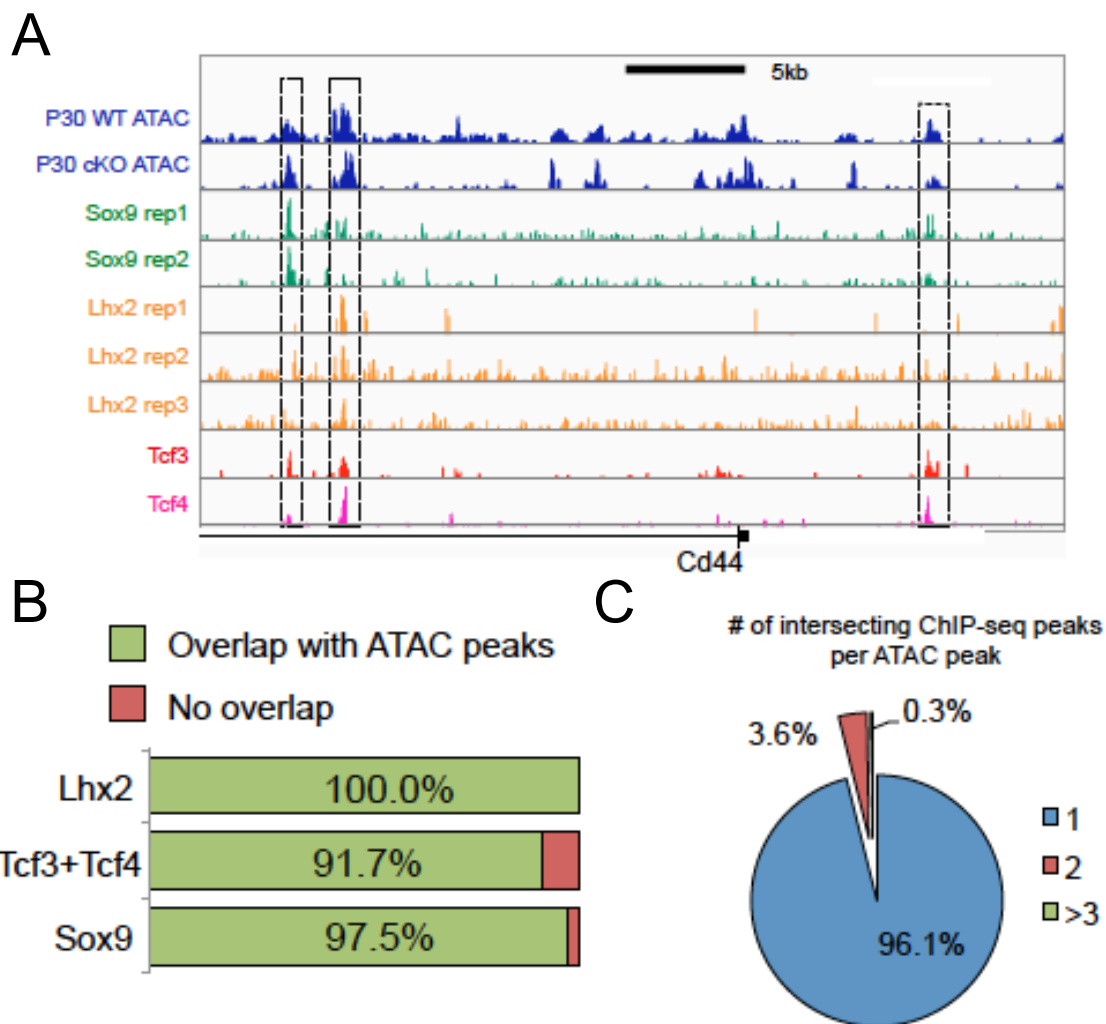
Figure 3.6 ATAC-seq as a union of all TF ChIP-Seq (A) Cd44 locus shows the overlapping peaks of ATAC-seq and ChIP-Seq signals of Sox9, Lhx2, Tcf3/4. (B) Percentage of ChIP-Seq peaks of Lhx2, Tcf3/4, and Sox9 that were detected by ATAC-seq. (C) Number of intersecting ChIP-Seq peaks per ATAC peak in All ChIP seq peaks examined in A and B.
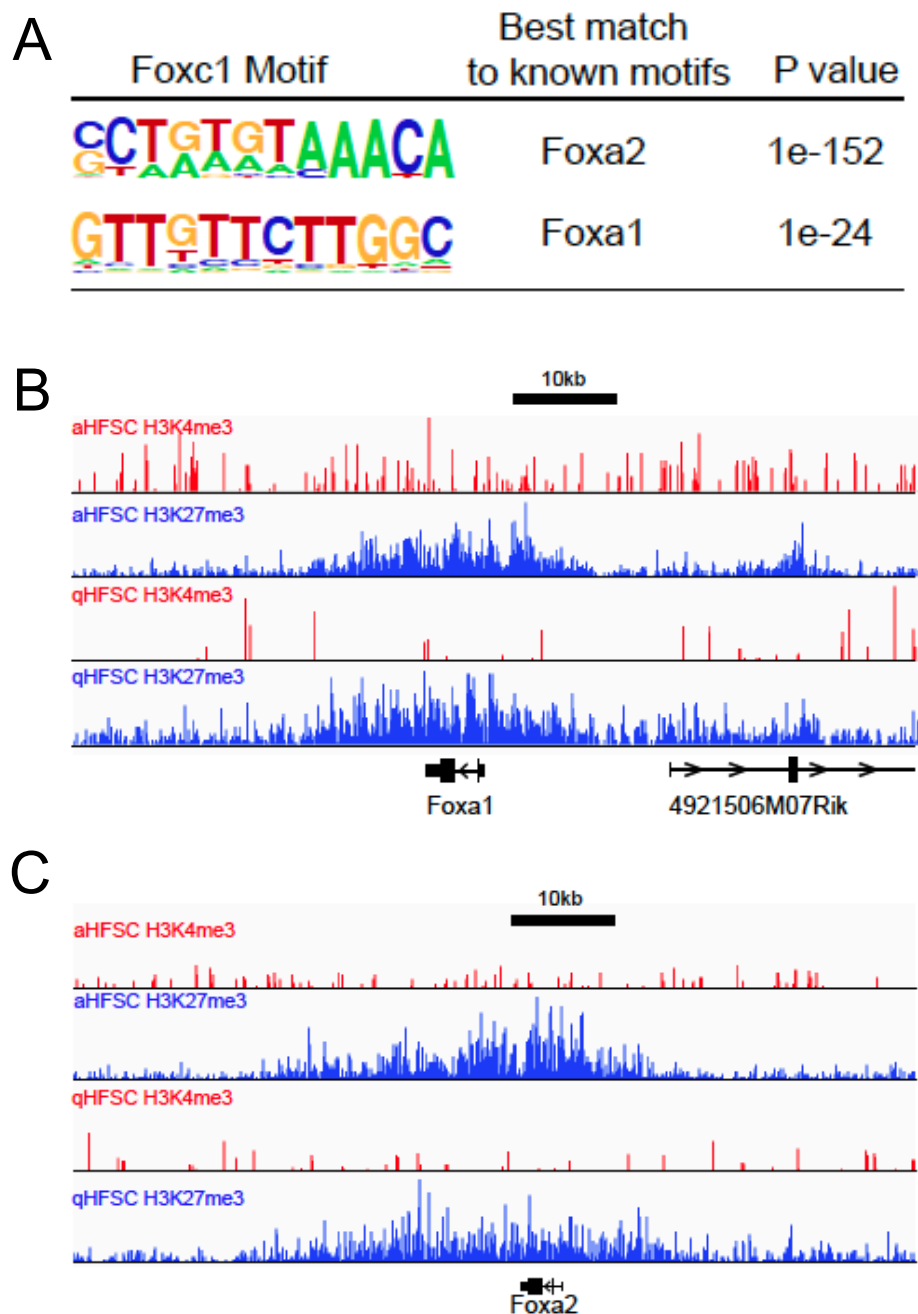
Figure 3.7 Foxc1 motif resembles motif of Foxa1 and Foxa2. (A) De novo
motif discovery of Foxc1 ChIP-Seq data identifies Foxc1 recognition motifs (B-
C)Snapshot of H3K4me3 and H3K27me3 ChIP-Seq of Foxa1 (A) and Foxa2 (B)
loci in activated HFSCs (aHFSC) and quiescence HFSCs (qHFSC)) shows
depletion of H3K4me3 marks accumulation of H3K27me3 marks at both locus,
indicating strong gene silencing.

KO HFSCs between P29 and P31, which is significantly high than the RNA-seq only (71.3%) and even more significant than the telogen RNA-seq data (44.7%). These data suggest that Foxc1 functions as a transcriptional activator, which is consistent with previously published reporter assays (Zarbalis et al., 2007). Importantly, several genes critically involved in the regulation of HFSC quiescence were found to contain Foxc1 binding sites in their promoter or enhancer regions, including Bmp2 (Plikus et al., 2008) Foxp1 (Leishman et al., 2013), Nfatc1 (Goldstein et al., 2014; Horsley et al., 2008), Prlr (Craven et al., 2001; Goldstein et al., 2014) (Fig. 3.11). I also noticed that Arid5b (Hata et al., 2013), a transcriptional co-regulator of Sox9, Trim16 (Cheung et al., 2012), a potent cell cycle inhibitor and a tumor suppressor, and Zbtb16 (also known as Plzf, (Hobbs and Pandolfi, 2010; Lu et al., 1994)), a key regulator for self-renewal of male germ cells, were also found as putative Foxc1 targets (Fig. 3.11).

Because differentially expressed genes in the Foxc1 cKO HFSCs could be regulated by Foxc1 either directly or indirectly, I was interested in further delineating critical pathways that are downstream of Foxc1. Since ATAC-Seq provides a comprehensive survey for all TF binding sites in the genome, I analyzed top TF hits in our ATAC-Seq peaks that associate with differentially expressed genes identified in anagen RNA-seq and obtained from the P30 WT and K14Cre Foxc1 cKO HFSCs. When searched against the genomic sequence, motifs representing AP-1/c-Jun, p63 and CTCF were the most enriched binding sites for transcriptional factors that were detected by ATAC-Seq, consistent with high expression levels and critical roles of these TFs in the skin epithelium (Lu et al., 1994; Mills et al., 1999) (Fig. 3.9). However, these common TFs were also the top enriched motifs when using all ATAC-seq peaks as input, indicating that they are universally enriched in HFSCs,
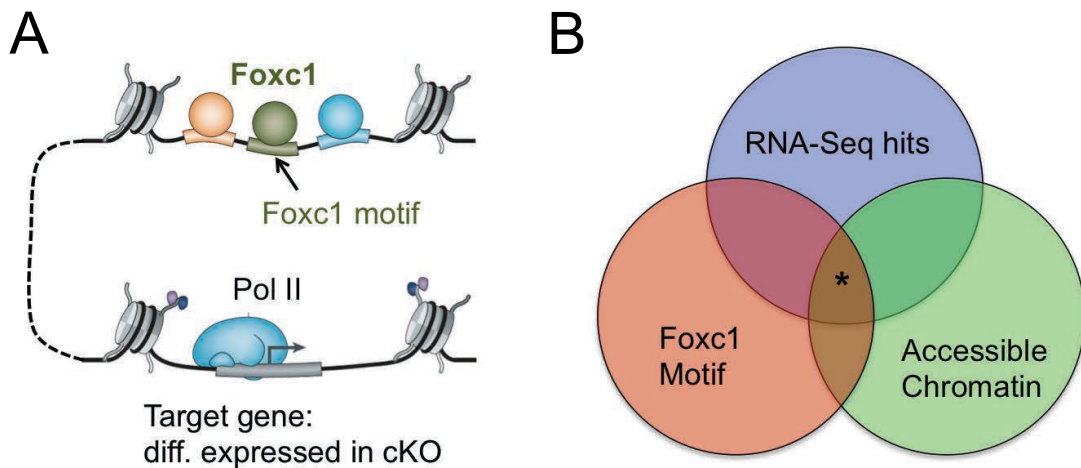
Figure 3.8 Strategy of Foxc1 target prediction. (A) Schematic summary of Foxc1-dependent gene regulation. Foxc1 protein binds to an accessible chromatin region harboring a specific motif and leads to target gene expression. (B) Intersection of RNA-seq data, ATAC-seq data and motif scan generates the list of Foxc1 targets (asterisk).
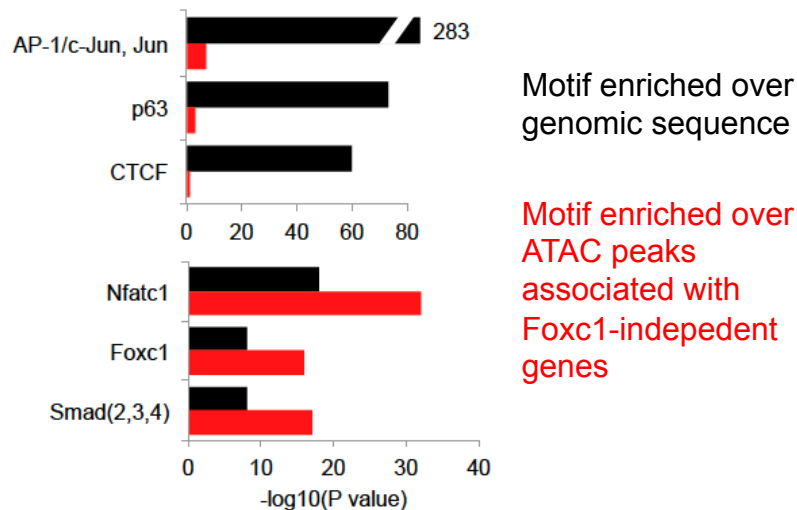
Figure 3.9 Motif analysis in differentially expressed genes. Comparison of enriched TF motifs over genomic sequence and in the peaks covering the differentially expressed genes in P30 KO HFSCs.

which is a sub-lineage of epithelium and thus unlikely involved in Foxc1-dependent regulatory network. To search for Foxc1-dependent TF motifs enriched in our dataset, I use the same set of peaks as input, which represent accessible regions of Foxc1-dependent genes, and search against background composed of accessible regions of Foxc1-independent genes. The Foxc1-dependent, preferentially enriched motifs include Foxa1/2, likely reflecting the motifs for Foxc1, Nfatc1, and pSmad2/3/4, reflecting the TGFβ and BMP signaling pathway, were significantly enriched on the promoter or enhancer regions of these differentially expressed genes in the absence of Foxc1 (Fig. 3.9). To map the coverage of these three TFs in the Foxc1 target list, I scanned the specific motif across all Foxc1-dependent genes (differentially expressed gene in cKO). Overall, 82% (343/421) of differentially expressed genes are likely controlled by Foxc1 directly (104) or by Nfatc1 (170) and Smads (328) indirectly, indicative of a collaborative gene regulatory network between Foxc1 and Nfatc1 and Bmp signaling, two critical regulatory networks governing the quiescence of HFSCs (Fig 3.10) (Botchkarev et al., 1999; Greco et al., 2009; Horsley et al., 2008; Kobielak et al., 2007; Oshimori and Fuchs, 2012; Plikus et al., 2008).

To test whether Foxc1 directly binds to Nfatc1 and component of BMP signaling pathway, I performed Foxc1 ChIP-qPCR. To overcome the limit of cell numbers, I sorted all GFP+ cells from K14-H2BGFP mice at early to mid anagen. I also specifically scrape away the hair follicle matrix during dissection to further enrich the bulge population. The direct regulation of Nfatc1, Bmp6 and BMP-sensitive Smad targets Hsbp8 (Genander et al., 2014) by Foxc1 was tested. I designed two sets of primers for Nfatc1 and Bmp6, in which one was against the ATAC peak predicted as Foxc1 binding site, i.e. containing a Foxc1 motif and one nearby ATAC peak without a motif. In both cases of Nfatc1 and Bmp6, the peak containing
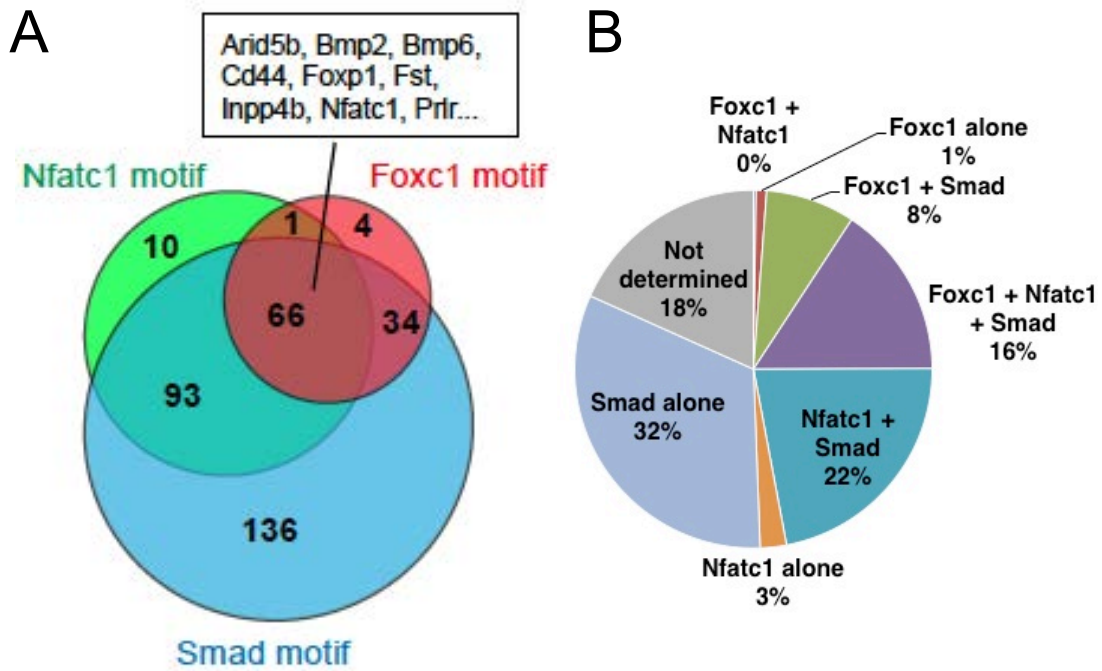
Figure 3.10 Co-operative regulation of Foxc1, Nfatc1 and Smad. (A)Venn diagram of the differentially expressed genes in P30 KO HFSCs that contain motifs of Foxc1, Smad and Nfatc1. (B) Breakdown of regulation modules of Foxc1, Nfatc1 and Smad in Foxc1 target genes

Foxc1 motif showed strong Foxc1 ChIP enrichment and the peak without motif has ChIP signal comparable with control gene Krt5 (Fig. 3.11). The promoter region of Hsbp8, which has motif-containing ATAC peak, also showed strong Foxc1 ChIP signal. Given the fact that all Nfatc1, Bmp6 and Hsbp8 were downregulated in P30 Foxc1 cKO HFSCs, The ChIP results suggest Foxc1 directly activates Nfatc1, Bmp6 and Hsbp8 (Fig 3.11).

In conclusion, I uncovered a cell-intrinsic and adaptive mechanism mediated by Foxc1 that operates on-demand to maintain SC identity in response to SC activation and division (Fig 3.12). The HFSC activation at the onset of anagen simultaneously activates Foxc1 expression, as part of the activation program. Foxc1 protein then turned on the target gene expression, including Nfatc1, Bmp6 and other genes that co-regulated by BMP signaling pathway. The cooperative effects of Foxc1-dependent program inhibit cell-cycle entry and reinforce quiescence of HFSCs. After the arrival of quiescence, Foxc1 expression is turned off but the secondary targets regulated by Nfatc1 and pSmads likely maintain the inhibitory program against cell cycle entry and thus maintain stem cell quiescence (Fig 3.12).

3.5   Discussion

In this study, I identified Foxc1 as a key regulator to reinforce stem cell quiescence when hair follicle stem cells are activated to self renew and to initiate HF growth. My findings discover an unexpectedly dynamic mechanism to maintain SC identity in response to SC activation.

My principle component analysis of the cellular state reveals that the activated, self-renewing HFSCs reside at an intermediate state between the quiescent HFSCs and the
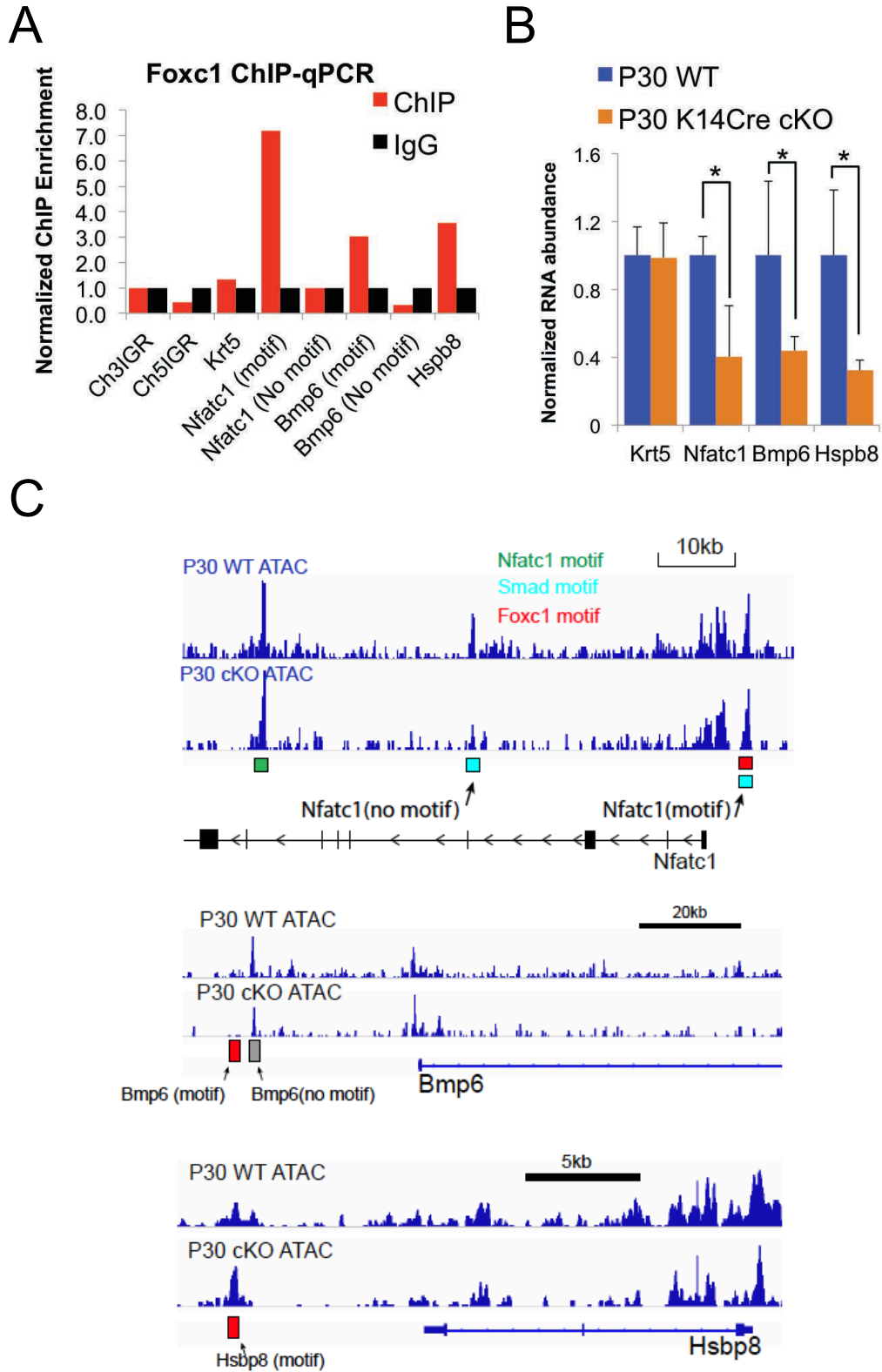
Figure 3.11 Foxc1 directly binds Nfatc1, Bmp6 and Smads target Hsbp8
(Continued in the next page)
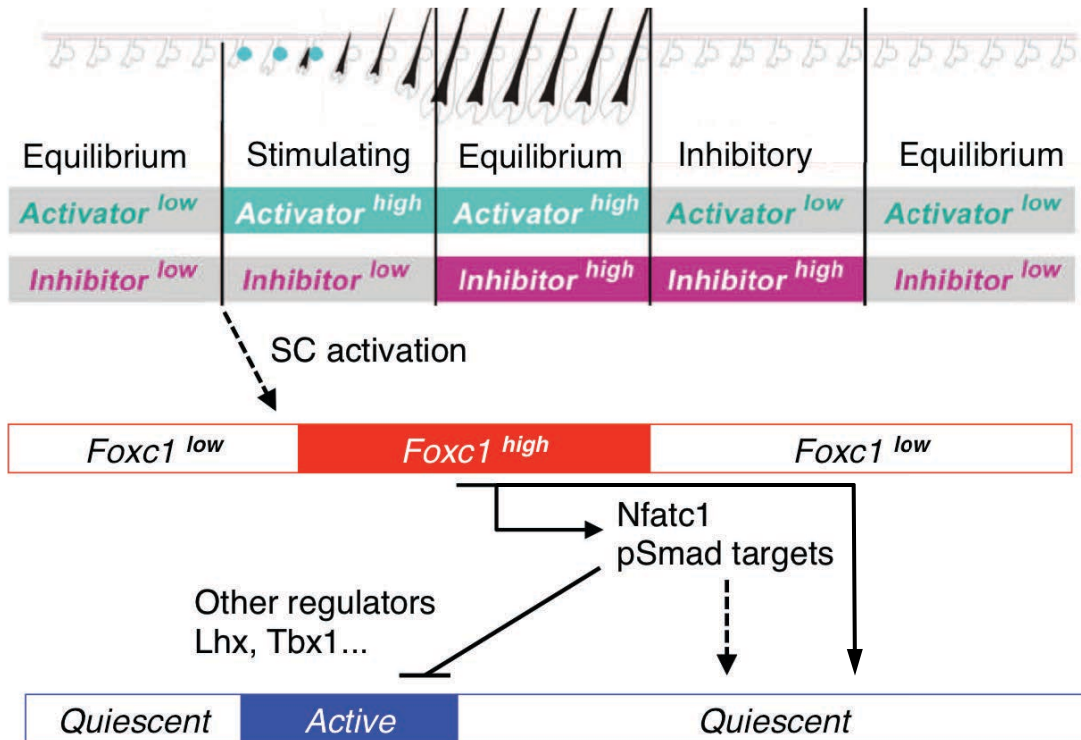
Figure 3.12 Foxc1 orchestrates an adaptive control of HFSC quiescence
Hair cycle progression is depicted as hair follicle morphology and pigmentation and is break down into phases according to extrinsic environment signals. After HFSC activation, Foxc1 expression is induced. Foxc1 directly activate Nfatc1 and Smad targets as well as other target genes directly to reinforce quiescence in HFSCs. When HFs reach the end of anagen, quiescence is restored and Foxc1 expression wanes.

Figure 3.11 Foxc1 directly binds Nfatc1, Bmp6 and Smads target Hsbp8
(A) Foxc1 ChIP-PCR of intergeric region in chr3 (Chr3IGR) and chr5 (Chr5IGR), Krt5, Nfatc1, Bmp6 and Hspb8. ChIP quantities were normalized to Chr3IGR and fold change in ChIP over IgG control is displayed. (B) Normalized RNA abundance of Krt5, Nfatc1, Bmp6 and Hspb8 from P30 RNA-seq. * $p < 0.05$, two-tail t test. (C) ATAC-Seq track of P30 WT and K14Cre cKO HFSCs of Nfatc1, Bmp6 (upstream and 5' end, C) and Hsbp8 loci. Primers targeting ATAC-Seq peaks that contains Foxc1 motif (red rectangle) or does not contain a motif (grey rectangle) are labeled.

committed HGs. When activated HFSCs fail to express Foxc1, they stay at this intermediate, activated state and fail to re-establish quiescence. This intermediate state is also analogous to the G-alert state of stem cell proposed by Rando and colleagues (Rodgers et al., 2014). Between quiescent, G0 stem cells and activated cells that enter the cell cycle, stem cells at the G-alert state are primed for activation. The priming process includes increased mitochondrial activity and increased cell volume. The ability to transition between G0 and G-alert is critical to the ability of stem cell populations to respond rapidly during tissue homeostasis and repair. This ability also allows maintenance of a pool of deeply quiescent, reserve stem cells for an extended period of time. In Foxc1 cKO HFSCs, upregulation of cell cycle-related gene may serve as an indicator for the entry of G-alert state. However, functional evidence is still needed to confirm this view. For example, if the cKO HFSCs are primed to activation, they are expected to respond to activation cues more readily than their WT counterparts. One way to test this hypothesis is to observe the colony formation dynamics for single, isolated HFSCs from the WT and cKO mice. I would expect that the cKO HFSCs initiate cell division more rapidly than the WT cells. To overcome potential artifacts introduced in *in vitro* culture, I can perform hair plucking experiment in early telogen and examine the timing of HFSC activation *in vivo*. It is likely that the cKO HFSCs will have a shorter response time to the activation cue.

Data collected using cKO mice can only demonstrate the necessity for Foxc1 to promote HFSC quiescence. To examine the sufficiency of Foxc1 to promote HFSC quiescence, a mouse model in which Foxc1 is constitutively expressed in all phases of hair cycle will be useful. If Foxc1 were sufficient to promote quiescence, sustaining high levels

of Foxc1 expression throughout the telogen should prevent, or at least delay the activation of HFSCs at the onset of next anagen.

I noticed although Foxc1 is dynamically induced in the HFSC compartment when the HFSCs are activated, Foxc1 is most constitutively expressed in the infundibulum area and in the differentiated IRS layers of HFs. In this regard, Foxc1 may not function as canonical stem-cell genes like Oct4 (Nichols et al., 1998) and Nanog (Mitsui et al., 2003) for embryonic SCs and Sox9 (Kadaja et al., 2014) and Lhx2 (Folgueras et al., 2013) for HFSCs. These stem-cell genes are highly enriched in the SC population of their corresponding cell lineages and their constitutive expression is required to maintain stemness throughout the lifespan of SCs. On the contrary, Foxc1 is the only transcription factor that has been shown to have such an dynamic expression pattern correlating with HFSC activation. Thus, Foxc1 may represent an adaptive mechanism "borrowed" by the HFSCs from their differentiated lineages to specifically regulate activated HFSCs. It is not known what factor or pathway activates Foxc1 at the onset of anagen entry. One candidate is sonic-hedgehog (Shh). Shh pathway is activated at the onset of anagen entry and is the initiating factor for HFSC activation, because ablation of Shh in DP or deletion of Shh effectors such as Gli1 and Gli2 in HFSCs completely blocked HFSC activation (Hsu et al., 2014). It would be interesting to examine whether Foxc1 expression is affected in Gli1 cKO HFSCs. The absence of ORS expression of Foxc1 is also very intriguing. Although Foxc1 is expressed in the hair germ in anagen I, it is clearly silenced in ORS, the direct progeny of hair germ differentiation. I speculate that the absence of Foxc1 in ORS cells might be essential to maintain the high proliferation and migration levels for the ORS cells. Overexpression of Foxc1 in the ORS would be very informative to test this hypothesis.

Finally, my ATAC-seq analysis reveals a cooperative network for HFSC quiescence regulation. As Foxc1 is mainly an activator for gene expression, the repression of cell cycles gene is likely an indirect effect for Foxc1-dependent program. Since Nfatc1 directly represses Cdk4 to maintain SC quiescence (Horsley et al., 2008) and Nfatc1 is significantly down-regulated in Foxc1 cKO HFSCs, it serves as a favorable candidate to mediate this secondary effect. My data shows Foxc1 also activates BMP-sensitive pSmad1/5 target genes. However, whether ablation of Foxc1 directly affects the functional pSmad1/5 level in activated HFSC is unclear and thus worth testing. For example, immuno-staining of pSmad1/5 levels in the anagen of the WT and cKO samples will provide important information to answer this question. Taken together, my findings demonstrate a dynamic and adaptive mechanism that operates on-demand in the HFSCs during self-renewal. It will be crucial to identify underlying mechanisms that directly link SC activation to the induction of Foxc1. Such knowledge may allow identification of other regulatory pathways that govern stem cell state.

3.6 Materials and Methods

3.6.1 Chromatin Immunoprecipitation

Cells were pelleted by centrifugation in a cold centrifuge at 500g for 5 minutes. Carefully resuspend the pellet with 1ml hypotonic lysis buffer (10 mM Tris pH 7.5, 10 mM NaCl, and 0.5% NP40) 1 mL, keep on ice for 30min, gently vortex every 5min. Final pelleted nuclei were resuspended in 200ul lysis Buffer B (1% SDS, 10 mM EDTA, 50 mM Tris-HCL pH 8.1) with 1x EDTA-free Complete mini® protease inhibitors (Roche), 1ul PMSF, Na-butyrate (optional) added just before use. Nuclei were incubated with lysis Buffer B for 10

minutes on ice before sonication. Sonicate samples in 1ml volume in a 15ml conical tube, with Biorupter, for 30 cycles of 30" ON +30" ON to reach total processing time of 15min. Spin at 14,000rpm for 10min at 4°C, save the sup carefully (leave 50ul in the tube). ChIP lysates keep for up to one year but around one year later the efficiency of pulldown seemed reduced and much of the chromatin appeared to have precipitated out of solution. Perhaps brief sonication would correct this problem but I have not verified this. The high level of SDS is important to keep chromatin in suspension. However, these high levels would prevent antibodies from interacting with their protein antigens. For this reason, 100 μL of lysate in diluted in 900 μL of IP Buffer (0.01% SDS, 1.1% Triton-X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8.1, 167 mM NaCl) with 1x Complete mini® protease inhibitor cocktail (Roche) added just before use. To this is added either 2 μg of anti-Foxc1 antibody (Cell Signaling) or 2 μg of mouse IgG for negative control. Lysate and IP buffer with incubated with antibody at 4 ºC overnight with rotation.

The next day 60 or 70 μL of protein G beads are added to each sample and incubated at 4 ºC with rotation for at least 2 hours. Beads are recovered using magnetic stands and proceed to wash steps. Washes are mixed by hand every 2 minutes for a total of 6 minutes and then beads are pelleted by spinning at 2000 rpm for 5 minutes, washes removed, and new wash added. Five washes are performed with the fifth wash being optional. The first wash is low salt (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.1, 150 mM NaCl). The second is high salt wash, which is low salt wash but with 500 mM NaCl. The third wash is a LiCl wash (0.25M LiCl, 1% NP40, 1% deoxycholate w/v, 1 mM EDTA, 10 mM Tris-HCl pH 8.1). The last two washes are simply Tris-EDTA pH 8.0 (Ambion).

Following washing, protein is eluted off beads at room temperature for 30 minutes with 400 µL of Elution Buffer (1% SDS, 0.1M NaHCO3, in water). NaHCO3 is stored as 1M stocks in -20 ºC and Elution Buffer is made fresh. 20 µL of lysate is added to 380 µL elution buffer for "input" samples. Following elution 16 µL of 5M NaCl is added to a final concentration of 200 mM and samples are incubated at 65 ºC for at least 2 hours for reverse crosslink. DNA can be incubated overnight but not RNA. At any step during washing, the wash can be removed and the beads stored at -80 ºC overnight. Elute before or after reversing crosslinking can be stored at -80 ºC overnight.

After reversing crosslinking, to each sample is added 32 µL of 1M Tris-HCl pH 6.5, 16 µL of 0.5M EDTA pH 8.0, 1 µL RNaseA, and 1 µL proteinase K. Samples are incubated at 40 ºC for 1 hour. Following proteinase K (Invitrogen) digestion, 400 µL of phenol:chloroform are added, samples are shaken, and centrifuged at maximum speed for 15 minutes. After proteinase digestion, DNA can be purified by Qiagen PCR purification kit and the eluted DNA is ready for PCR or sequencing library preparation.

3.6.2   Real-Time PCR

Total RNAs from FACS-purified cells were isolated using TRIZOL reagent (Invitrogen). 200ng of RNA (when available) was reverse-transcribed and qPCR was performed using the iQ SYBR Green Supermix system (BioRad) and BioRad CFX-384 machine. Fold-changes were computed using the ΔΔCt formula normalized to Hprt values. In all qPCR figures error bars denote standard errors of the mean.

3.6.3 RNA-seq assay

Total RNAs from FACS-purified cells RNA integrity numbers (RIN)>8 were used to perform RNA-seq assay. Libraries were prepared using NEBNext Ultra Directional RNA Library Prep Kit and sequenced at BioFrontiers Next Generation Sequencing Facility at University of Colorado Boulder and Genomics and Microarray Core Facility at University Of Colorado Denver on Illumina HiSeq 2000 and 2500 machines.

3.6.4 ATAC-seq assay

Assay for transposase accessible chromatin followed by sequencing (ATAC-seq) was performed as previously described (Buenrostro et al., 2013) with following modifications: An average of 100,000 FACS-sorted HFSCs were collected PBS containing 3% chelated fetal bovine serum (FBS) and pelleted by centrifugation for 5min at 500g at 4°C. Cell pellets were re-suspended in 50ul of lysis buffer per 100,000 cells processed (10mM Tris-HCl pH 7.4, 10mM NaCl, 3mM MgCl2, 0.1% Igepal CA-630) and nuclei were pelleted by centrifugation for 25min at 500g, 4°C using a swing rotor with low acceleration and brake settings. Supernatant was carefully discarded and nuclei were re-suspended in 50 ul reaction buffer containing 2ul of Tn5 transposase and 22.5ul of TD buffer (Nextera DNA Sample Preparation Kit, Illumina). The reaction was incubated at 37°C for 30min and terminated by adding 10ul of clean up buffer (900mM NaCl, 300mM EDTA) and immediately purified using MinElute PCR Purification Kit (QIAGEN). Library amplification was performed following the manufacturer's protocol (Nextera DNA Sample Preparation Kit, Illumina) except using 2.5ul of each primer and 2min of extension time in PCR reaction. PCR cycle number was determined by a pilot 15ul qPCR reaction containing 1x EvaGreen

(Biotium) on 10% of unamplified library. Libraries were sized-selected to enrich 250-800bp of size and sequenced as single-end or paired-end for at least 40 million reads per sample.

3.6.5   RNA-seq analysis

RNA-seq reads (101 bp, single-ended) were aligned to the mouse genome (NCBI37/mm10) using TopHat2 (version 2.0.13 (Kim et al., 2013)). Expression measurement of each gene was calculated from the resulting alignment bam file by HT-seq (Anders et al., 2015) Differentially expressed genes were determined using DEseq2(Love et al., 2014) with inclusion of littermates factors and p-value cutoff of 0.05 (B-H adjustment).

3.6.6   GO-term and GSEA analysis

GO-term enrichment analysis was performed by DAVID Bioinformatics Resources 6.7 using gene symbols as input. GSEA analysis was performed using ranked fold change values (cKO vs. WT) for the displayed dataset, and gene sets selected from the referenced publications.

3.6.7   Principle component analysis (PCA)

Transcriptome profiling of HFSCs and stem cell lineages were obtained from microarray dataset in the following publications:(Greco et al., 2009) (Zhang et al., 2009) (Lien et al., 2011). For each dataset, fold change values were calculated using following normalization sample: (Greco et al., 2009) (Zhang et al., 2009): P18-P21 Bulge 0 division replicate #1; Lien et al: quiescent HFSC replicate #1. Presence/absence call was performed and genes with both counterparts of the fold change comparison classified as presence were included

in downstream analysis. Fold change were calculated using the difference of log transformed intensity measurements. If multiple probes were mapped to a single gene, fold change of each individual probes were first calculated and the median of all values was selected as the representative fold change of the gene. Fold change data of P47 Foxc1 cKO were obtained from DEseq2 output. All fold change data are merged in to a table, and lines containing missing values were excluded. The remaining data were further down-sized by selecting genes with >2 fold change in at least one of the following samples: aHFSC, qHFSC, TAC1 and TAC2 from Lien et al.; P56 bulge, P69 bulge, P43 HG, P56 HG and P69 HG from Greco et al. A total of 2684 genes were selected for PCA plotting. Data matrix was standardized by subtracting the mean and divided by the standard deviation of each sample. PCA was performed using the function of prcomp in R and the first three principal components were used for plotting. Linear Discriminant Analysis (LDA) was performed using R library MASS (Venables and Ripley, 2013).

3.6.8 ATAC-seq, ChIP-seq and motif analysis

ATAC-seq reads (single-end or paired-end) and ChIP-seq reads (Amin and Bobola, 2014) were aligned to mouse genome (NCBI37/mm10) using Bowtie2 (version 2.2.3, (Langmead and Salzberg, 2012)). Duplicate reads were removed by Picard tools (Broad Institute). Mitochondrial reads were removed and peak calling was performed on each individual sample by MACS (version 2.0.9) (Zhang et al., 2008)). Peaks from different ATAC-seq samples were merged for downstream analysis. De novo motif searching was performed using the HOMER(Heinz et al., 2010). For Foxc1 motif discovery, the top 500 peaks (based on MACS2 output p value) were used with regions of 500bp around peak center. For motif

73

analysis using ATAC-peaks, all ATAC peaks are first used for motif discovery, with mm10 genome sequence as background. Then ATAC peaks associated with differentially expressed gene in P30 cKO-vs-WT RNA-seq were used as target sequence set and motif discovery was performed against sequenced of ATAC peaks that associated with non-RNAseq hits genes. In both cases, enriched motifs from the default HOMER motif were analyzed, as shown in Fig 5h. Sequencing data tracks were presented by Integrative Genomics Viewer (IGV) (Robinson et al., 2011).

3.6.9 Data Accession

Sequencing data have been deposited to GEO with the accession number GSE67404.

CHAPTER 4   TRANSCRIPTIONAL CHARACTERIZATION OF LINEAGE-SPECIFIC TUMORIGENESIS IN SKIN SQUAMOUS CELL CARCINOMA

Part of the work described in this chapter is included in Latil M and Nassar et al., The cancer cell of origin controls EMT-related tumour heterogeneity, under review at *Cell*.

Tumor heterogeneity describes the phenotypic differences between tumor cells within a given tumor (intratumor heterogeneity) and between different tumors (intertumor heterogeneity) (Marusyk et al., 2012). These differences have major implications for the diagnosis, prognosis, and therapy of cancer patients. Different mechanisms have been proposed to account for tumor heterogeneity including genetic differences (Marusyk et al., 2012)(Burrell et al., 2013), the influence of the microenvironment (Junttila and de Sauvage, 2013), the existence of a hierarchical organization of tumor growth (Beck and Blanpain, 2013; Marusyk et al., 2012), or the phenotypic plasticity of cancer cells (Meacham and Morrison, 2013). While the cancer cell of origin has been also suggested to control tumor heterogeneity (Visvader, 2011), no study has demonstrated so far that in carcinoma, the most frequent cancers in humans, the cancer cell of origin controls tumor heterogeneity. Depending on the cell lineage within a given tissue that is initially targeted by the oncogenic hits (multipotent and unipotent stem cells, progenitors and differentiated cells), different tumor phenotypes may arise, represented as their levels of differentiation, aggressiveness and prognosis.

The skin epidermis is an ideal model to study whether the cancer cell of origin controls tumor heterogeneity, as it is composed of spatially distinct cell lineages including

the interfollicular epidermis (IFE), the hair follicle (HF) and its associated sebaceous glands, and the infundibulum that connects the HF to the IFE (Fig. 4.1A). Each of these distinct epidermal lineages is self-sustained by their own pool of resident stem cells (SCs) during homeostasis. Each epidermal compartment can be genetically targeted by different inducible Cre mouse lines (Schepeler et al., 2014) (Blanpain and Fuchs, 2014), allowing the conditional expression of oncogenes or deletion of tumor suppressor genes in different epidermal lineages and assess their capacity to induce tumor formation (Blanpain, 2013). In this chapter, I will describe a collaboration project in which our collaborators in ULB, Belgium utilized genetically engineered mouse model to induce tumor in IFE and HF stem cells, and demonstrated that tumors arising different cell-of-origin consist of different type of tumors. Specifically, tumors from IFE consisted essentially of well-differentiated tumor epithelial cells whereas tumors arising from HF lineages displayed considerable heterogeneity ranging from well-differentiated skin squamous cell carcinoma to tumor mesenchymal cells exhibiting epithelial to mesenchymal transition (EMT), which are more invasive and metastasis-prone. Based on these data, I performed chromatin profiling using ATAC-seq on the two different tumor lineages and identified the key transcription factors that likely promote the EMT-related tumor heterogeneity. Clustering analysis revealed a branching model for HF-derived tumors to commit either epithelial fate or EMT fate. I also tested whether the hair follicle is epigenetically and transcriptionally primed to undergo EMT in HF while to induce well-differentiated tumors in IFE. Global analysis chromatin landscape suggests a significant similarity between normal HF and transformed mesenchymal tumor cells, which may represent the group of primed genes that facilitate EMT in HF-derived tumors.

4.1   The cell-of-origin of skin SCC controls EMT-related tumor heterogeneity

To assess the EMT-related tumor heterogeneity in SCC, our collaborators induced KRasG12D expression and p53 deletion in the IFE and the infundibulum using K14CreER mice with low dose tamoxifen (TAM) administration, which preferentially targets these epidermal compartments (K14CreER/KRas^{G12D-}/p53$^{fl/fl}$/Rosa-YFP) or in HFSCs and their progeny using Lgr5CreER mice (Jaks et al., 2008; Lapouge et al., 2011) (Lgr5CreER/KRas^{G12D}/p53$^{fl/fl}$/Rosa-YFP) (Fig. 4.1B). The Rosa-YFP reporter gene was included in these different mouse models to enable lineage tracing and isolation of the induced tumor cells. Our collaborators have previously described that the kinetic of tumor appearance (around 6-9 weeks) and initial tumor growth rate were similar in K14- and Lgr5-derived tumors. Histological analyses revealed that K14-derived SCCs were mostly well-differentiated SCC, composed of tumor epithelial cells (TECs) and containing numerous keratin pearls (KP) (Fig. 4.1C). In contrast tumors arising from Lgr5CreER/KRasG12D/p53$^{fl/fl}$/Rosa-YFP mice were composed of variously differentiated tumor phenotypes, including well-differentiated SCCs resembling to the SCCs arising in K14CreER mice, mixed tumors containing well-differentiated TECs and mesenchymal-like cells (TMCs) and tumors that were completely mesenchymal (Fig. 4.1C). Well-differentiated SCCs, regardless whether they originate from K14CreER or Lgr5CreER cells, expressed classical epidermal markers including K14, Epcam and E-cadherin. Moreover, no YFP+ TECs expressed the mesenchymal marker Vimentin (Vim), which was only expressed by YFP- cancer associated fibroblasts (CAFs) that composed the tumor stroma. In Lgr5-derived mixed SCCs, the well-differentiated part of the tumor expressed epithelial markers K14 and E-cadherin, comparable as K14Cre-derived SCCs (Fig. 4.1C). However, many cells

located in the underlying mesenchymal part of the tumor were in fact YFP+ TMCs that completely lost the expression of all classical epithelial markers such as K14, Epcam and E-cadherin and expressed mesenchymal marker Vimentin (Fig. 4.1C). Likewise, in Lgr5-derived mesenchymal tumors, YFP+ TMCs and mesenchymal stromal cells were morphologically indistinguishable and were all K14, Epcam and E-cadherin negative (Fig. 4.1C). FACS analysis of Epcam expression showed that the vast majority of YFP+ TECs in well-differentiated SCCs, whether K14- or Lgr5-derived, expressed a high level of Epcam (Fig. 4.1D). Epcam expression was always bimodal in mixed tumors with two distinct YFP+ Epcam+ TEC and Epcam- YFP+ TMC populations (Fig. 4.1D), while in mesenchymal tumors all YFP+ TMCs were Epcam- (Fig. 4.1D). Quantification of the proportion of YFP+ Epcam+ and YFP+ Epcam- cells in a large number of tumors showed that the vast majority of K14 derived tumors (n=63) were almost exclusively composed of Epcam+ cells well-differentiated SCCs. In contrast, Lgr5 derived tumors (n=192), only on small proportion of tumor presented high percentage of Epcam cells, the most frequent tumors were composed of moderate to low proportion of Epcam+ (mixed tumors) and the remaining tumors (25%) were composed of tumor cells that are exclusively Epcam- (mesenchymal tumors). Together, these data demonstrated tumor induced from HFSCs developed significantly higher level of EMT-related heterogeneity than tumor induced from IFE.

Our collaborators further demonstrated that HF-derived TECs are intrinsically primed to give rise to EMT tumors. Dilutions of tumor cells were transplanted into immunodeficient mice and the their ability to reform secondary tumors were evaluated. FACS isolated Epcam- TMCs isolated from the same mixed Lgr5-derived SCCs form secondary tumor more efficiently than Epcam+ TECs after their transplantation into

immunodeficient mice (Fig. 4.2A). However, secondary tumors arising from the transplantation of Epcam- TMCs were exclusively composed of Epcam- TMCs. In contrast, transplantation of Epcam+ TECs from the same Lgr5 mixed tumors gave rise to well-differentiated, mesenchymal and mixed tumors that recapitulated the heterogeneity found in the primary tumors containing both Epcam+ and Epcam- cells (Fig. 4.2B), suggesting that Epcam+ TECs are an heterogeneous tumor cell population consisting of multipotent, and lineage restricted unipotent TECs that can only sustain either Epcam+ or Epcam- populations. Interestingly, transplantation of FACS isolated YFP+ Epcam+ TECs from a well-differentiated K14-derived SCCs into immunodeficient mice gave rise to a majority (75%) of well-differentiated SCCs, 25% were either mixed or mesenchymal SCCs, whereas in Lgr5 derived tumors, more than 75% were either mixed or mesenchymal (Fig 4.2B). These data indicate that the cancer cell of origin influences the intrinsic ability of tumor initiating cells to undergo EMT, although the existence of secondary mixed and mesenchymal tumors suggests that the TECs from K14-derived SCCs are not irreversibly committed to the epithelial fate in transplantation assays, suggesting a certain degree of plasticity of these cells in a different microenvironment.

4.2   ATAC-seq captures known tumorigenesis and EMT signatures

Inext wanted to define the global change of the chromatin landscape that occurred during tumorigenesis and EMT and how the IFE. To this end, I performed ATAC-seq, which I have shown as a powerful technique to depict regulatory network of transcription factors (see Chapter 3), on FACS isolated normal HF and IFE cells and their respective tumors including well-differentiated K14-Epcam+ and Lgr5 Epcam+ and Epcam-
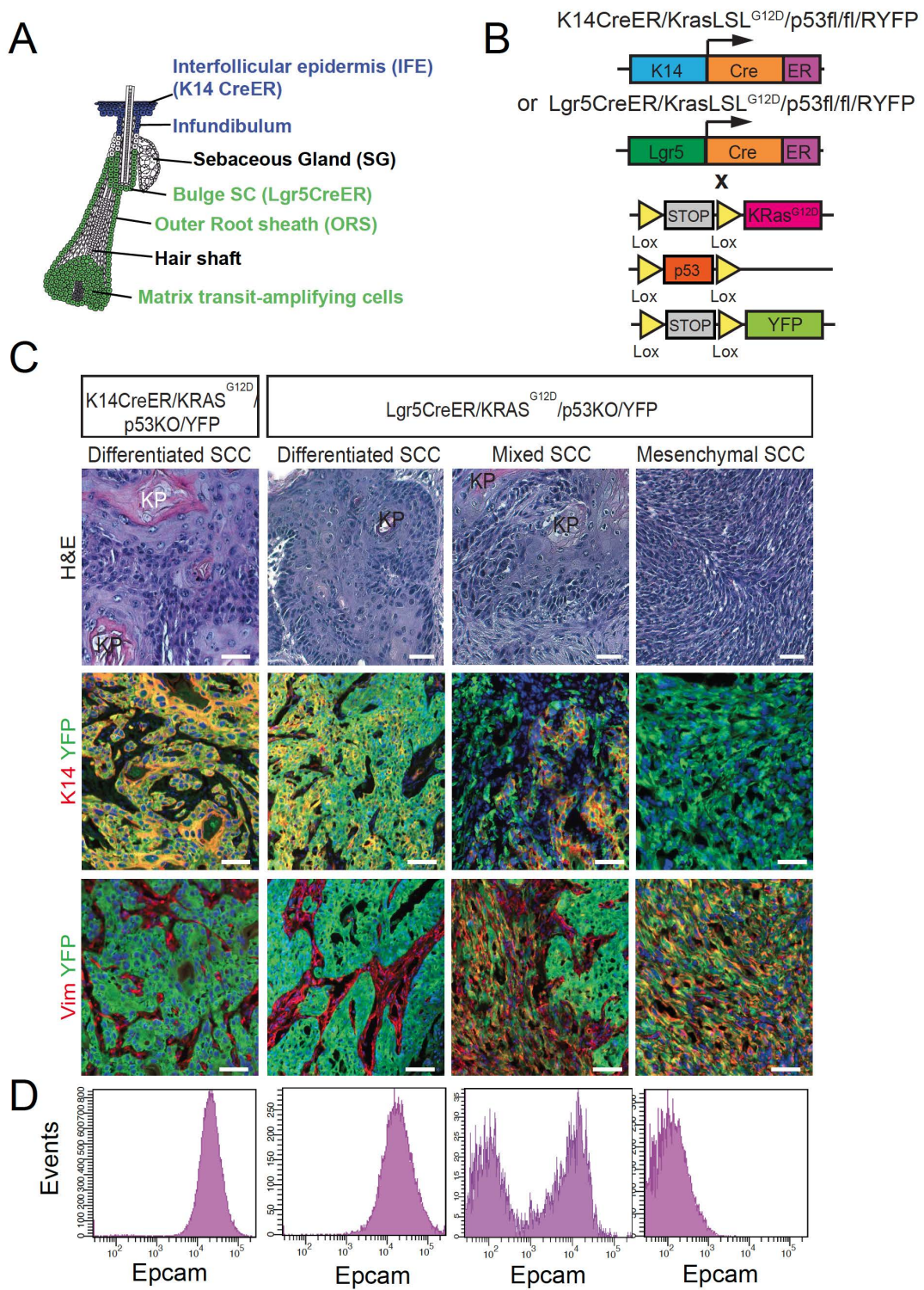
Figure 4.1: The cellular origin of skin SCC influences EMT related tumor heterogeneity. (continued in the next page)
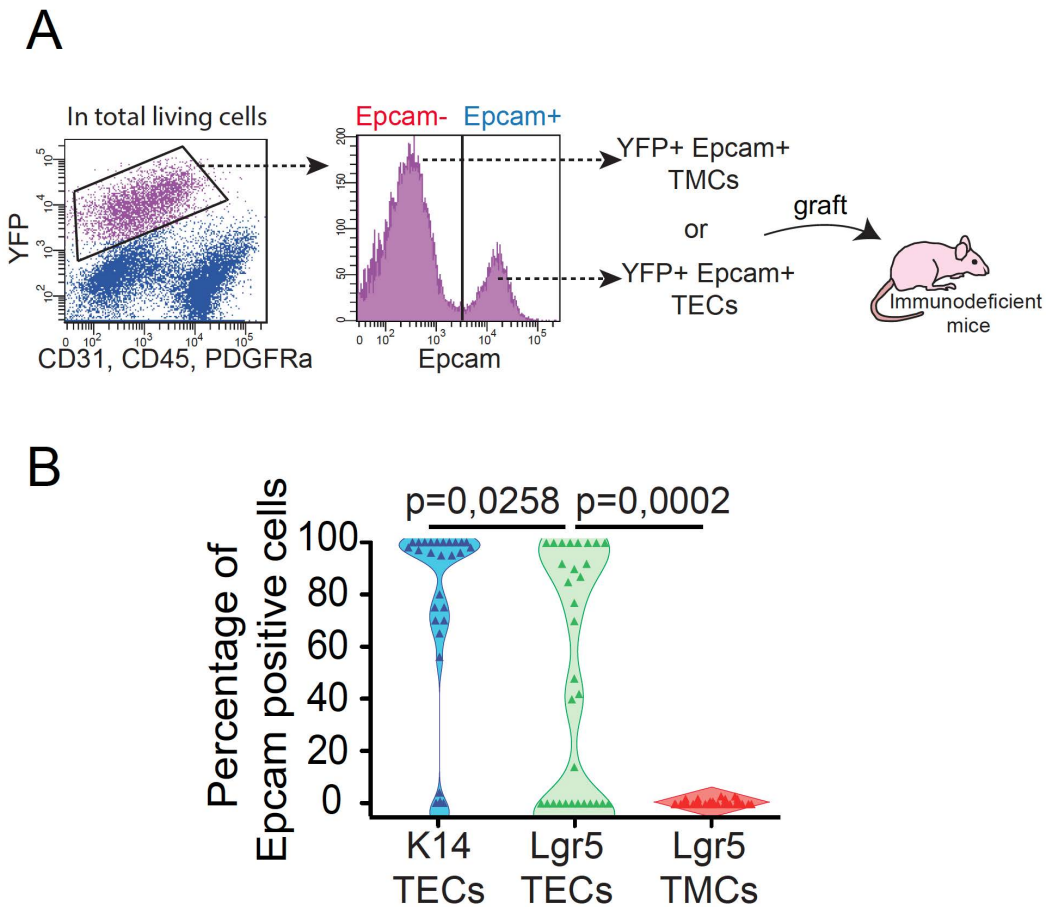
Figure 4.2 HF derived TECs are intrinsically biased to give rise to EMT tumors. (A) Scheme of testing cancer stemness of YFP+ Epcam+ TECs and YFP+ Epcam- TMCs. (B) Graph showing the proportion of Epcam+ cells in secondary tumor arising from the transplantation of K14 and Lgr5 YFP+ Epcam+ TECs and Lgr5 YFP+ Epcam- TMCs (n=11, n=12, n=10 respectively). All data from Latil and Nassar et al., submitted

Figure 4.1 The cellular origin of skin SCC influences EMT related tumor heterogeneity. (A) Scheme of the skin epidermis and its different lineage. (B) Genetically engineered mouse models of skin SCC induction using K14CreER or Lgr5CreER. (C) Phenotypic analysis of K14 and Lgr5-induced tumors. H&E and IF of YFP and K14 or Vimentin in the different SCC subtypes arising from the IFE or HF. (D), FACS analysis of Epcam expression in the different SCCs. Epcam is expressed homogeneously in YFP+ TECs from differentiated SCCs, while mixed SCCs are characterized by two distinct YFP+ tumor cell populations (Epcam+ and Epcam-) and mesenchymal SCCs do not express Epcam. All data from Latil and Nassar et al., submitted.

tumor cells, which allows to define the open chromatin regions and the TFs associated with these regions during tumorigenesis and EMT (Fig. 4.3). I first mapped the global cell state transition by hierarchical clustering and principle component analysis (PCA). Both clustering and PCA revealed that the major cell states during tumorigenesis take action from 4 weeks since tumor induction to Epcam+ tumor (Fig 4.3A-B). Non-induced, normal tissue display similar chromatin landscape as 4 weeks induced tissue, indicating nucleosome remodeling is not dominantly active at early stage of tumor formation. The HF shows slightly more significant chromatin remodeling than IFE at this stage. A second observation is that Epcam+ tumors derived from HFSCs and IFE are extremely similar (Fig 4.3B). Pair-wise differential occupancy analysis revealed that only 1% of ATAC-peaks shows significant change of accessibility between HF Epcam+ tumors and IFE Epcam+ tumors, which are at least 20-fold less than comparing between normal and tumor samples. These data suggests that if chromatin remodeling is involved in driving EMT-related tumor heterogeneity, it is probably a subtle difference, at least not a global phenomenon.

To define the chromatin signature associated with tumorigenesis, I filtered the chromatin regions (ATAC-seq peaks) that are changed by more than 3 fold, between the cell of origin and the respective tumor populations. I identified 3544 peaks corresponding to 2222 genes that gained accessibility during tumorigenesis in the different tumor cell populations analyzed (K14 TECs, Lgr5 TECs, Lgr5 TMCs) whereas 1088 peaks corresponding to 840 genes were decreased. Many of the changes in the chromatin landscape were associated with change in gene expression by microarray profiling, and represent the common epigenetic and transcriptional changes associated with tumor initiation irrespective of the cell lineages from which these tumors arise and the tumor
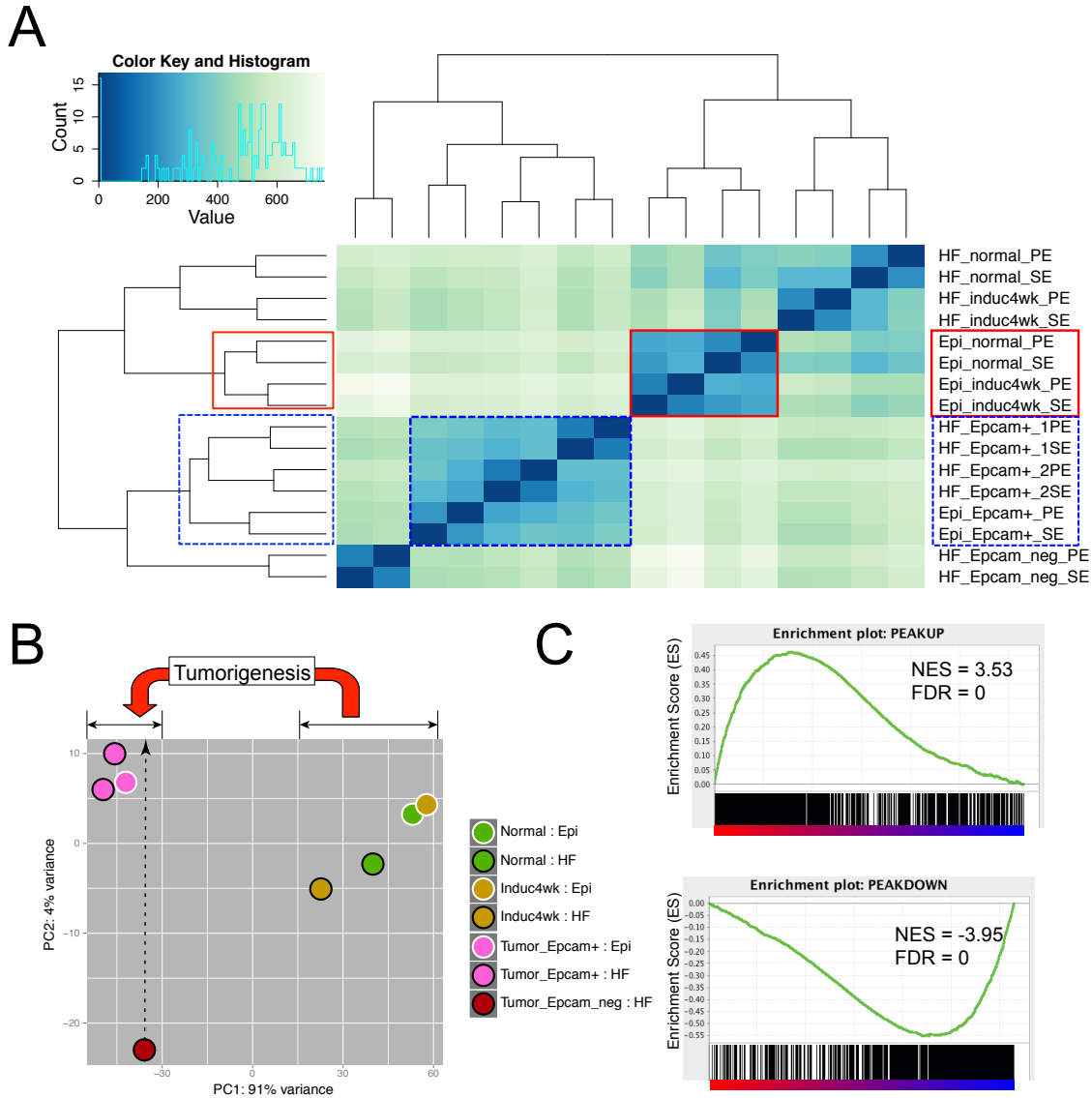
Figure 4.3 ATAC-seq of tumor induction model reveals cell state transition. (A) Heatmap of distance matrix among all ATAC-seq samples. Note that the normal epidermis samples cluster together (red) as well as all Epcam+ tumor samples include HF and IFE origin (blue). (B) PCA plot of all ATAC-seq samples. Three major cell state: normal, Epcam+ tumor and Epcam (-) tumor can be distinguished whereas two tumor state can be grouped to based on the major PC, which leads to the cell state transition in tumorigenesis (red arrow). (C) GSEA of the most differentially changed peak for a given gene between Lg5-TEC and Lg5-TMC, indicating gaining chromatin accessibility correlates with gene activation.

phenotypes. This common set of genes irrespective of EMT included ligands of the EGFR pathway (Areg), CD44, secreted proteins (Ccl2, CXCL1-3, Vegfa/vegfc) and transcription factors known to relay the activation of the Ras/MAPK pathway (Fos/Fosb, Fosl1, Nfe2l2, Ets1) and other TF promoting tumor stemness and invasion (Twist1, Hmga2)(Fig 4.4A). Some of these genes such as Fosl1, Ereg, Nfe2l2, Twist1 were already upregulated transcriptionally 4 weeks after oncogenic transformation, suggesting they represent one of the most common earliest events during tumorigenesis (Fig 4.4B). Motif enrichment analysis of the chromatin regions gaining accessibility during tumorigenesis revealed a strong enrichment for the binding site of TFs such as Jun/AP1 (65 %), BHLH proteins (33%), Ets1 (37%), Runx (29%), Nf-kb sites (22%) (Fig 4.4C).

To determine the changes in the chromatin landscape that occurred during EMT, I assessed the ATAC-seq peaks that were upregulated or downregulated by more than 3 fold between Lgr5 Epcam+ and Epcam- cells. During EMT, 10126 chromatin peaks corresponding to 3824 genes were upregulated by more than 3 fold. About 50% of the genes upregulated during EMT are associated with an increase in the opening of the chromatin region surrounding them (Fig. 4.4D). These genes included TFs such as Zeb1/2, Twist1, Tcf4, Tcf7l1, cell adhesion (eg: Cdh2, Cdh11), ECM (Col1a1, Col3a1, Cola2, Fn1, Lox/Loxl1), members of the YAP pathway (eg: YAP, Amotl2, Ctgf), signaling molecules (eg: Ltbp1/2, Flsl1, Grem1, Serpin2/h1, Cxcl12). In contrast, the chromatin regions that are closed during EMT are strongly associated with the genes upregulated in TEC (58%) (Fig 4.4G), consistent with the chromatin regions associated with the maintenance of the epithelial fate including TFs (eg: Grhl1/3, Ovol1/2, Klf5, Maf, P63, Trp73, Cebpa,), cytoskeleton (Krt5, Krt14) and adhesion molecules (Epcam, Cdh1, Dsg2, Dsc3, Fig 4.4F).
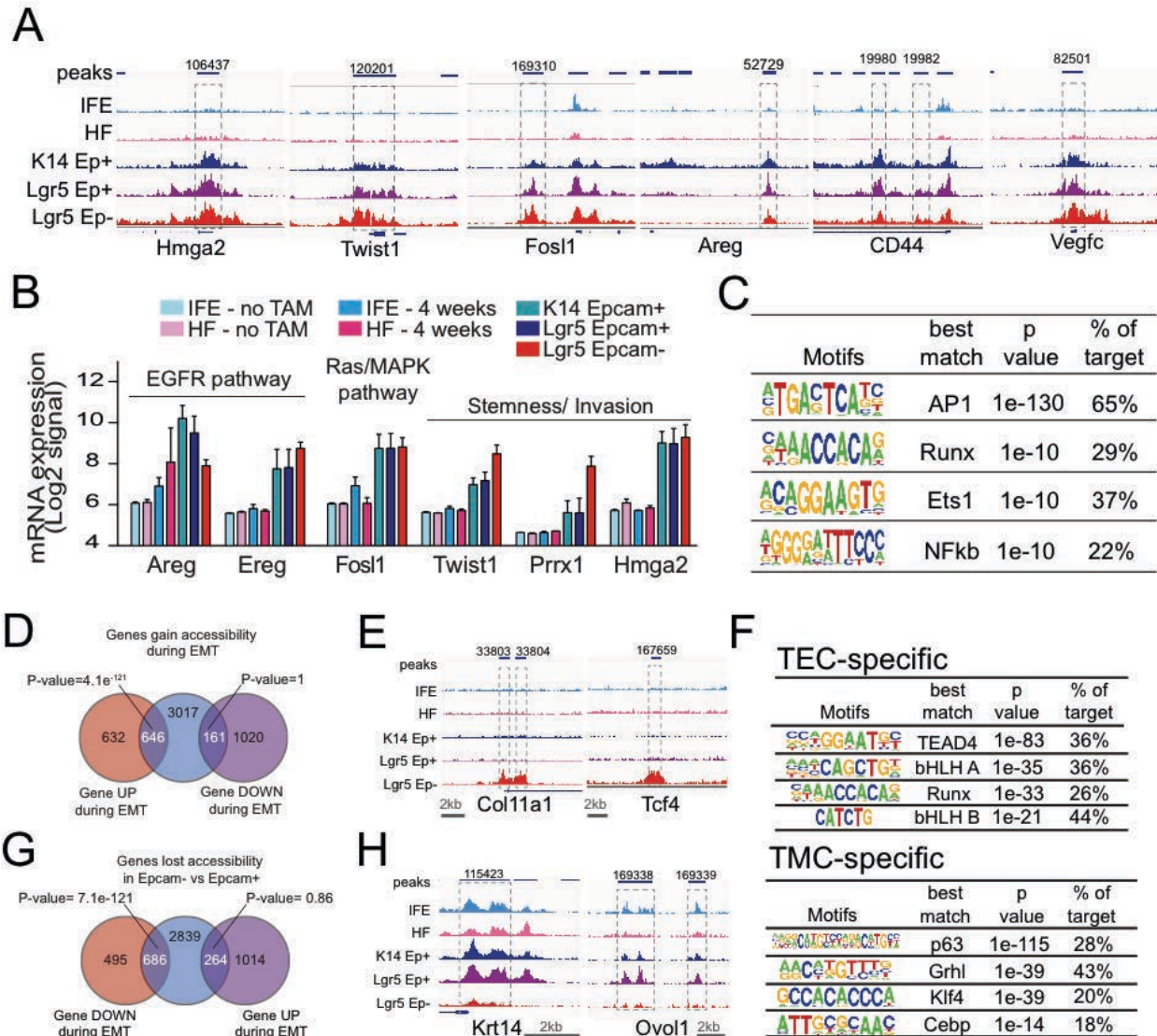
Figure 4.4 Chromatin landscape during tumorigenesis and EMT. (A, E, F) Illustrations of ATAC-seq profiles of IFE, HF, K14 Epcam+ TECs, Lgr5 Epcam+ TECs and Epcam- TMCs showing gain of open chromatin regions during tumorigenesis (A) and EMT (E) or lost of accessible chromatin during EMT (H). (B) Enriched TF motifs found in the common ATAC-seq peaks up-regulated during tumorigenesis. (C) Gene expression from microarray showing the dynamic of gene expression during tumorigenesis of genes involved in EGFR pathway, Ras/MAPK pathway and in transcription factors involved in tumor progression and stemness. (D, G)Venn diagram showing the overlap of the genes associated with ATAC-seq peaks gain opening (D) or lose opening (G) in during EMT and the genes up-regulated and during EMT, i.e. comparing Lgr5 Epcam TMC with Epcam+ TEC. (F) Enriched TF motifs found in the ATAC-seq peaks specific for TEC (Top) or TMC (bottom).

The most frequently recurrent motifs found with very high statistical significance in peaks that are upregulated during EMT and positively associated with gene expression was Jun/AP1(40,%), Tead4 (36%) , NF1 (44%), Ets1 (29%), bHLH TFs (36%), Runx (26%, Fig. 4.4F) and some of these TFs (eg: Runx1, Nfatc1, Twist1, Tcf4) were also upregulated in Epcam- TMCs (Fig. 3k). Motif analysis of peaks losing accessibility in Epcam- TMC compared toEpcam+ TECs were highly enriched for binding sites of the same core set of TFs including Jun/AP1 (47%), NF1 (50%), Ets1 (31%), and TFs that are specific for the epithelial tumor fate including P63 (28 %), Grhl (43%), Klf4 (20%), Cepba (18%) (Fig. 4.4F), suggesting that epithelial TFs contribute to the repression of EMT genes in TECs. Altogether, this global analysis allows defining the chromatin dynamics during EMT and allowing reconstructing for each gene regulated during EMT, the corresponding chromatin regions that positively and negatively regulate its expression.

4.3   Epigenetic priming of EMT genes in HF lineages

To get further insights into the lineage-specific gene regulatory network differs IFE and HF in the outcome of tumor formation, I assessed whether the chromatin landscape are differentially regulated between the two lineages. Because both lineage have tumor progression from normal cells to Epcam+, TECs, I first compared the chromatin landscape dynamics during this process, between IFE and HFs. Interestingly, peaks gained accessibility during tumorigenesis in HF lineages also tend to gain accessibility in IFE lineages and vise versa. Similar result was observed for peaks lost accessibility in tumorigenesis in HF or IFE (Fig 4.5A). This observation is consistent with precious cell state mapping (Fig 4.3), where K14-derived TEC and Lg5-derived TEC display significant

similarity, which suggests that tumorigenesis in IFE and HF are likely undergoing similar epigenetic dynamics, but whether mediated by similar groups of regulators is not guaranteed. To clarify this uncertainty, I searched for transcription factors that potentially distinguish the IFE and HF lineage using differential motif enrichment analysis. I searched motifs preferentially enriched in HF lineage and depleted in IFE lineage, or vice versa. Surprisingly, motifs that differentially enriched between IFE and HF-derived Epcam+ tumors are p63, Cebp, both of which are preferentially gained in HF tumor and Nfatc1 and Lhx2, which are preferentially lost in HF tumor (Fig 4.5B). Both p63 and Cebp are master regulators of epithelial fate (Mills et al., 1999; Yang et al., 1999; Lopez et al., 2009) and therefore it is counterintuitive to expect them enriched in hypothetically EMT-prone Lgr5 Epcam+ cells. Moreover, Nfatc1 and Lhx2 are both HF stem cell-specific transcription factors promoting stem cell quiescence (Horsley et al., 2008 and Folgueras et al., 2013) and gaining potential binding sites of quiescence regulators should consequently associated with decrease of metabolic activity, which is also untenable to be enriched in Lg5-TECs. However, since p63 and Cebp motifs are among the most depleted motifs in TMCs compared to Lgr5 TECs, an alternative interpretation for Lgr5 TECs gaining these motifs is that they are the predecessor population for EMT, instead, Lgr5 TECs are differentiated tumors that specifically avoid EMT fate and reinforced the epithelial fate, which leads to dramatic similarity with K14-TECs.

I next defined whether the chromatin landscape of HF cells primed them to give rise to EMT tumors. To this end, I defined the chromatin regions that gained or lost accessibility by at least 3 fold in ATAC-seq in both Lgr5 Epcam- tumors vs Lgr5 Epcam+ tumors and normal HF vs normal IFE, and examined gene expression associated with these peaks

87
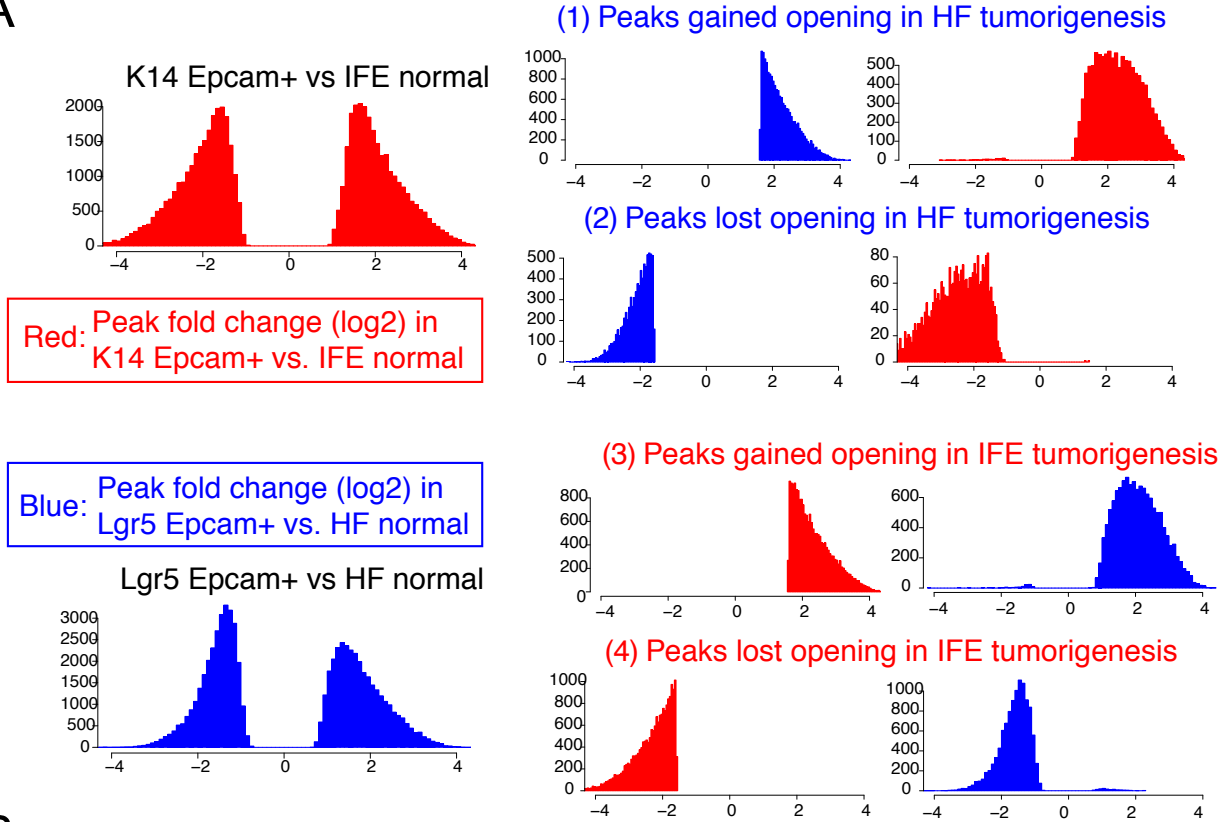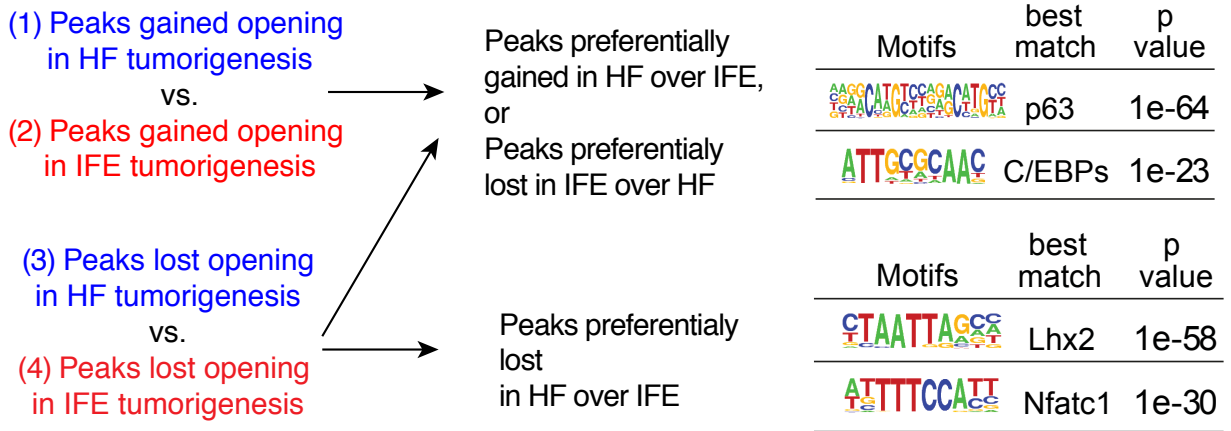
Figure 4.5 Differential motif analysis in HF and IFE tumorigenesis. (A) All peaks with significant changes in opening during IFE tumorigenesis (red, left) or HF tumorigenesis (blue, left) are shown as reference. Among them, peaks gained (list 1) or lost (list 2) opening in HF tumorigenesis tends to also gain or lose opening in IFE tumor formation. Similarly, peaks gained (list 3) or lost (list 4) in IFE tumor behave similarly in HF tumor. (B) Top: differential motif discovery on list 1 over 2 (or reverse) identified p63 and C/EBPs motifs, which are over-represented in Up-peaks in HF tumor. Bottom: comparing list 3 over 4 (or reverse) identified Lhx2 and NFatc1 as preferentially lost motifs in HF tumor.

in the same processes (Fig 4.6). Very interestingly, 652 genes presented chromatin peaks gained accessibility in HF vs IFE and in Lgr5 Epcam- vs Epcam+, which is significantly high proportion of overlap between genes from the two comparisons (Fig. 4.6A). 139 of these genes were upregulated during EMT and 62 of these genes were already upregulated at the transcriptional level in normal HF (Fig. 4.6A). In contrast, these common genes that presented more open chromatin regions in HFs and Epcam- tumors cells did not significantly overlapped with genes downregulated during EMT. Among these common genes, 73 genes presented identical upregulated peaks in HF and Epcam- TMCs, 20 of these genes including Col5a2, Serpinh1, Ltpb1, Tcf7l1 were upregulated in TMCs (Fig. 4.6B). These comparisons suggest that genes upregulated together with gain chromatin accessibility during EMT tends to express at higher level in normal HF than IFE.

I then analysed whether the chromatin landscape primed the IFE cells to give rise to well differentiated tumors. Interestingly, 1413 genes presented chromatin peaks upregulated IFE vs HF and in Lgr5 Epcam+ vs Epcam-, corresponding to 86% of the genes that have ATAC-seq peaks upregulated between IFE and HF, which is also significantly high proportion of overlap between genes from the two comparisons (Fig. 4.6C). 253 of these genes were upregulated in TECs, among which 112 genes were already upregulated at the mRNA level in IFE (Fig. 4.6C). These common genes that presented more open chromatin regions in IFE and Epcam+ tumors cells did not significantly overlapped with genes upregulated during EMT. Among these common genes, 531 peaks were identically upregulated in IFE and Epcam+ TECs, 98 of these genes including TFs regulating epidermal differentiation such as Grhl1/3, Cepba, or Klf5 were upregulated in Epcam+ TECs (Fig 4.6D). These commonly upregulated chromatin regions between IFE and Epcam+ TECs,
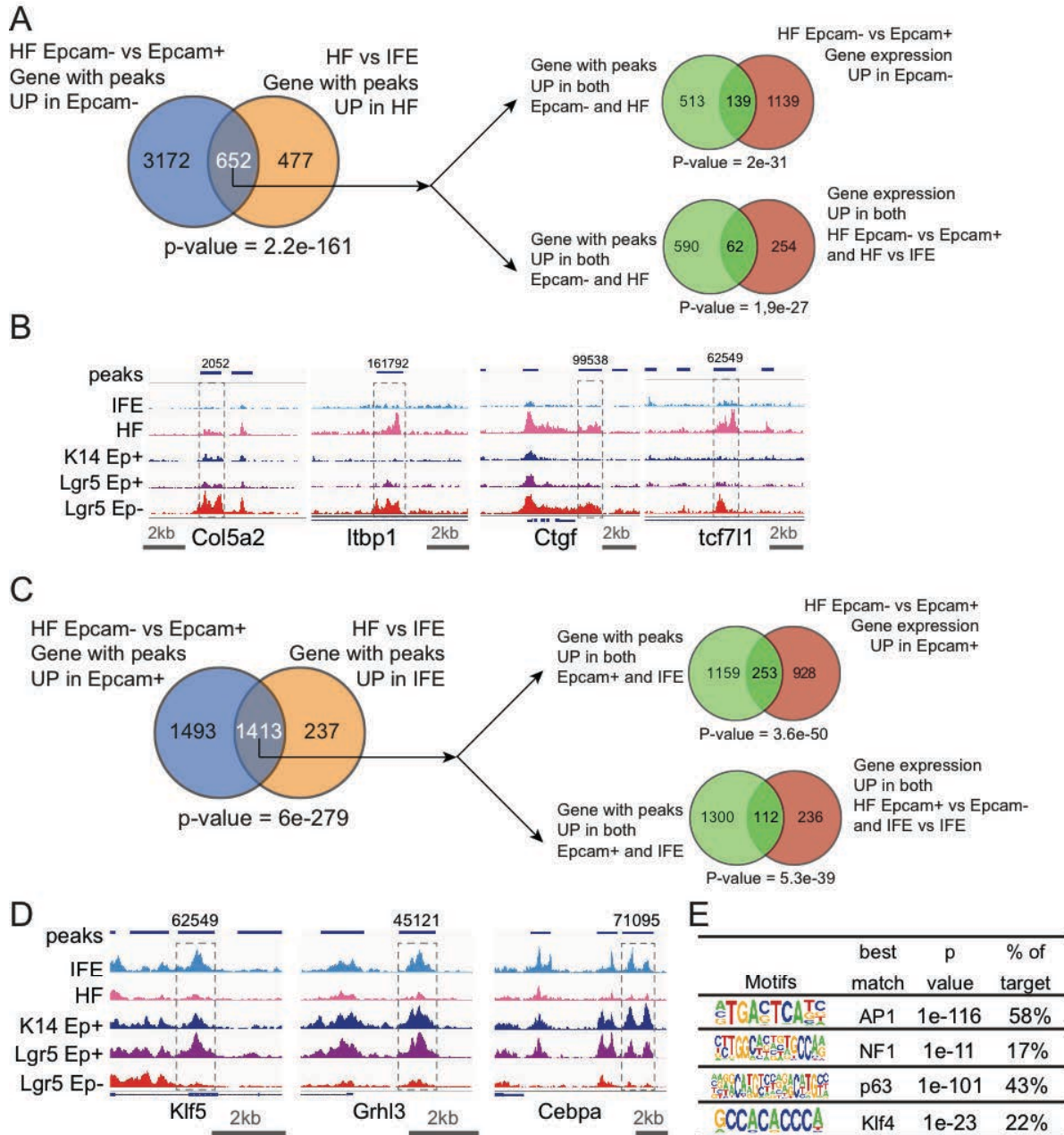
Figure 4.6 Epigenetic priming of EMT genes in HF lineages. (A) Venn diagram showing the overlap of the genes associated with ATAC-seq peaks upregulated in Lgr5 Epcam- TMCs vs TECs and upregulated in HF vs IFE. The intersected genes (a total of 652 genes, green circle) are further overlapped with genes upregulated in TMCs or both TMC and HF. (B) ATAC-seq snapshot of chromatin regions upregulated in HF and Epcam- TMCs. (C) Venn diagram of genes with ATAC peak upregulated in Lgr5 TEC and upregulated in IFE vs HF. The intersected genes (a total of 1413 genes, green circle) are further overlapped with genes upregulated in TMCs or both TEC and IFE. (D) ATAC-seq snapshot of chromatin regions that are upregulated in both IFE and K14 Epcam+ TECs. (E) Enriched TF motifs found in the ATAC-seq peaks up-regulated in Epcam+ TECs and IFE and also associated with gene upregulation in Epcam+ TECs.

90

were highly enriched for binding sites of TFs including Jun/AP1, p63 and Klf (Fig. 4.5E), suggesting that these TFs contribute to the epigenetic priming of IFE cells to progress into well-differentiated tumors. Together, these data indicate a significant portion of similarity in transcriptome and chromatin landscape of normal HF and Epcam- TMCs, and suggest that these primed genes may facilitate EMT in HF derived tumors.

4.4   Discussion

Our study demonstrates that the cellular origin of skin cancers controls EMT-related tumor heterogeneity. HFSC and their progeny are more primed to initiate EMT derived tumors whereas IFE/infundibulum cells are intrinsically primed to induce the formation of well-differentiated SCCs. Chromatin and transcriptional profiling demonstrate that HF cells are epigenetically and transcriptionally primed by common and distinct open chromatin regions within the genes that are upregulated during EMT, including TFs associated with HF stemness and differentiation (Runx1, Nfatc1, Tcf7l1, Tcf, Tbx1), soluble proteins promoting Tgf-b signaling (Ltbp1, 3) or inhibiting BMP signaling (Grem1, Flstl1), ECM proteins (Col, Postn, Lox) (Chen et al., 2012; Horsley et al., 2008; Morris et al., 2004; Nguyen et al., 2006; Tumbar et al., 2004). In contrast, the IFE cells present many chromatin regions that characterized Epcam+ TECs, resulting in a shared gene expression program that primed IFE cells toward differentiated tumors. These genes included many epithelial cytoskeleton and adhesion proteins (Cdh1, Itga2, Lama1), and many key TFs, known to promote epidermal differentiation such as p63, Grhl1/3, Cepba, Ovol1, Klf5 (Kenchegowda et al., 2011; Lopez et al., 2009; Mills et al., 1999; Sur et al., 2006; Yang et al., 1999).

Bioinformatic analysis of the chromatin regions that are gain or lost during EMT uncovered the combination of specific TFs that are associated with the chromatin remodeling during tumorigenesis and EMT. I found that the chromatin remodeling during tumorigenesis is associated with the combination of AP1, NF1, Ets1 motifs that were very significantly overrepresented in the chromatin regions that are gained during tumorigenesis, indicating that oncogenic that KRas signaling pathway converges to some common signaling program irrespective of the cell of origin and tumor phenotypes that involved the activation of Fos/Jun, NF1 and Ets TF family members, many of which are upregulated transcriptionally during tumorigenesis. In both Epcam+ and Epcam- cells, the chromatin regions that are dynamically changed during EMT are also very enriched for AP1, NF1, Ets1 motifs, but also contained other specific TF associated motifs that are associated with open chromatin regions in Epcam- (Nfat, bHLH, Tead, Runx, Smad2) or Epcam+ (P63, Klf, Grhl, Cepba). These results indicated that in addition to Jun/Fos, Ets1, and NF1, YAP pathway, which has been recently shown to regulate KRas mediated proliferation and EMT in cultured cancer cell lines in vitro58, in combination with HF TFs (Nfatc1, Runx, Smad2) are important for the chromatin remodeling that accompanied EMT and may contribute to the priming of HF lineages to EMT tumors. The enrichment of Smad2 binding site in chromatin region open during EMT is in good agreement with the recent data showing that TGF-b responsive cells mediated cell invasion in skin SCCs17. The specific open chromatin region that defined Epcam+ TECs presented many interesting similarities with IFE cells are highly enriched for the common tumorigenic AP1, NF1, Ets1 motifs motifs of Jun/Fos together with Grhl, p63, Klf, Cepba, which sustain the epithelial program of TECs by regulating the expression of the other epithelial genes such as keratins,

Cadherins and Desmogleins. Surprisingly, several chromatin regions surrounding EMT genes such as Zeb1 and Zeb2 are already highly opened in Epcam+ TECs, although their expression is repressed, suggesting that while EMT TFs are epigenetically primed in TECs, other chromatin regions restrict their expression in TECs, possibly related to the other opened chromatin regions in these genes that are lost during EMT.

Although K14 derived tumors are intrinsically primed to give rise to well differentiated tumors, our collaborators further performed co-transplantation experiments with heterologous stromal cells shows showed that Lgr5 CAFs presents more potent EMT inducing activity as compared to K14 CAFs, suggesting that the type of stromal cells recruited in the tumors or their specialization depend on the cancer cell of origin, and influences EMT in a non-cell autonomous manner. In classical tumor transplantation assay together with matrigel or with homologous mesenchyme, only Epcam+ TECs and not Epcam- TMCs presents cancer SC characteristics consisting of multi-lineage differentiation potential and the ability to reform the heterogeneity of the primary tumors. However, single cell tumor transplantation show that not all TECs present this multi-lineage differentiation potential, and TECs can be already committed to epithelial, mixed or mesenchymal fate only. In conclusion, our results provide novel insights about how the cancer cell of origin controls EMT related tumor heterogeneity and open new avenues to further define the intrinsic signals that regulates EMT in primary tumor cells.

4.5 Materials and Methods

4.5.1 ATAC-seq Library construction

Assay for transposase accessible chromatin (ATAC) followed by sequencing was performed as following: 100,000 sorted cells were collected in 1mL of PBS+3%FBS at 4°C. Cells were centrifuged, then cell pellets were resuspended in 50uL of lysis buffer (Tris HCl 10mM, NaCl 10mM, MgCl2 3mM, Igepal 0,1%) and centrifuged (500g) for 25 min at 4°C. Supernatant was discarded and nuclei were resuspended in 50uL of reaction buffer (Tn5 transposase 2.5uL, TD buffer 22.5uL and 25uL H2O – Nextera DNA sample preparation kit, Illumina). The reaction was performed for 30 min at 37°C and then blocked by addition of 5uL of clean up buffer (NaCl 900mM, EDTA 300mM). DNA was purified using the LinElute purification kit (Qiagen). DNA libraries were PCR amplified (Nextera DNA Sample Preparation Kit, Illumina), and size selected for 200 to 800 bp (BluePippin, Sage Sciences), following manufacturers' protocols.

4.5.2 Alignment and Peak-calling

ATAC-seq reads (single-end or paired-end) were aligned to mouse genome (NCBI37/mm10) using Bowtie2 (version 2.2.3) (Langmead and Salzberg, 2012) using option of "–local" for single-end and "--local --very-sensitive-local --dovetail --dovetail -X 1000" for paired-end. Mitochondrial reads were excluded from dowonstream analysis and duplicate reads were removed by Picard tools (http://broadinstitute.github.io/picard/). Alignment data tracks were visualized by Integrative Genomics Viewer (IGV) (Robinson et al., 2011).

Peak calling was performed on each individual sample by HOMER (Heinz et al., 2010) with parameters setting of "-L 0 -C 3 -size 1000 -minDist 1000 -tbp 3 -o auto". Peaks from different ATAC-seq samples were merged for downstream analysis.

4.5.3 Differential Peak analysis

Pair-wise comparisons of ATAC peaks between two conditions were performed by R package DESeq2 (Love et al., 2014), with reads count of each peak calculated by HTSeq (Anders et al., 2015). Significance is defined as adjusted p value smaller than 0.001 and fold change more than 3. Peaks were assigned to the nearest Refseq annotated genes with 500kb range.

4.5.4 Motif analysis

De novo motif search was performed using program of findMotifsGenome.pl in the HOMER package (Heinz et al., 2010) with parameters setting of "-size -250,250 -S 15 -len 6,8,10,12,16". Incidences of specific motif was examined by the program of annotatePeaks.pl in the HOMER package with size parameter "-size 500".

4.5.5 GSEA analysis

GSEA analysis was performed using ranked fold change values (Epcam- over Epcam+) of ATAC peaks for the displayed dataset. For the upregulated or downregulated genes, the highest fold change peak was selected to represent the gene and enrichment score was calculated following GSEA documentation.

CHAPTER 5   GENOME-WIDE MAPS OF POLYADENYLATION REVEAL DYNAMIC mRNA 3'END FORMATION IN MAMMALIAN CELL LINEAGES

Part of the work described in this chapter is included in Wang L., Dowell RD., and Yi R. (2013). Genome-wide maps of polyadenylation reveal dynamic mRNA 3'-end formation in mammalian cell lineages. RNA.

Post-transcriptional regulation, often mediated by miRNAs and RNA-binding proteins at the 3' untranslated regions (UTRs) of mRNAs, is implicated important roles in the output of transcriptome. To decipher this layer of gene regulation, it is essential to quantitatively measure global mRNA expression in a 3'UTR specific manner. In this chapter, I will describe an experimental and bioinformatic pipeline that simultaneously determines 3'end formation by leveraging local nucleotide composition and quantitatively measures mRNA expression by sequencing polyadenylated transcripts. When applied to purified mouse embryonic skin stem cells and their daughter lineages, I identified 18,060 3'UTRs representing 12,739 distinct mRNAs that are abundantly expressed in the skin. I also determined that ~78% of UTRs are formed by using canonical A[A/U]UAAA polyadenylation signals whereas ~22% of UTRs utilize alternative signals. By comparing to relative and absolute mRNA abundance determined by qPCR, our RNA-Seq approach can precisely measure mRNA fold-change and accurately determine the expression of mRNAs over four orders of magnitude. Surprisingly, only 829 out of 12,739 genes show differential 3'end usage between embryonic skin stem cells and their immediate daughter cells whereas the numbers increase to 933 genes when comparing embryonic skin stem cells to

the more remotely related hair follicle cells. It suggests an evolving diversity instead of switch-like dynamics in 3'end formation during development. Together, my quantitative analysis reveals a dynamic picture of mRNA 3'end formation in tissue stem cell lineages *in vivo*.

5.1   Introduction

It is widely recognized that 3'UTR of mRNAs harbor numerous regulatory elements that are implicated in important functions for mRNA metabolism (Di Giammartino et al., 2011). Alternative polyadenylation (APA) is of special interest because it generates 3'UTRs with different lengths, which provides a mechanism to control transcript stability, localization and translational efficiency by modulating cis-elements on the 3'UTR (Di Giammartino et al., 2011). APA has been implicated in modulating proliferation and transformation in cancer cells (Sandberg et al., 2008; Mayr & Bartel, 2009; Lin et al., 2012). Recent studies further suggest that APA is widely present with cell type-specific and tissue-specific patterns (Fu et al., 2011; Derti et al., 2012; Smibert et al., 2012; Ulitsky et al., 2012). However, most studies have focused on differences between *in vitro* cultured cells derived from different tissues, or whole organ-scale comparison. To date there is little information about whether APA is involved in developmental transitions among closely related cell lineages *in vivo*, which is critical to understand how APA is initiated and controlled in biological processes. In this regard, mouse embryonic skin affords an ideal system to examine how APA is controlled in somatic cell lineages. First, embryonic skin stem cells and their daughter cells are abundantly available and functionally distinct (Blanpain & Fuchs, 2009). Second, embryonic skin stem cells give rise to their differentiated daughter cells by

an asymmetric cell division mechanism (Lechler & Fuchs, 2005). Upon the cell division, the differentiating daughter cells exit the cell cycle and embark on the epidermal differentiation program. Thus, the stem cell populations are separated from the differentiated cells by a single cell division (Fig 5.1). Taken together, profiling mRNA 3'end formation in mouse embryonic skin provides an opportunity to examine APA during a developmental transition *in vivo* by global transcriptome analyses.

To determine the complexity of mRNA 3'end formation and quantify their expression, many RNA-Seq techniques have been recently developed (Ozsolak et al., 2009; Beck et al., 2010; Fu et al., 2011; Jan et al., 2011; Derti et al., 2012). Virtually all quantitative techniques for RNA molecules, including qPCR and microarray, are based on oligo(dT) or random priming followed by signal amplification. Therefore, oligo(dT) priming-based 3Seq has its unique advantage in mRNA quantification. I reasoned that if I can computationally distinguish authentic mRNA 3'end signals from internal priming events I can develop a tool to accurately and quantitatively measure mRNA 3'ends by 3Seq.

5.2   Distinct nucleotide composition patterns at mRNA 3'end regions

Among the three current 3'end sequencing approaches, direct sequencing and 3P-Seq experimentally distinguish authentic 3'end formation (Ozsolak et al., 2010; Jan et al., 2011). Interestingly, local nucleotide composition surrounding mRNA 3'end regions determined by 3P-Seq in *C.elegans* showed a distinct pattern (Jan et al., 2011). Importantly, this nucleotide composition pattern was universally observed regardless of whether canonical or alternative polyadenylation signals (PAS) were used or whether proximal or distal 3'ends were analyzed (Jan et al., 2011). These observations suggest a requirement for
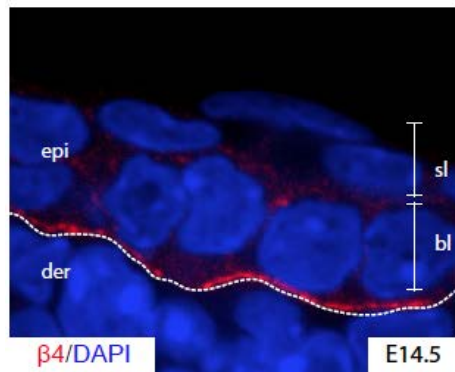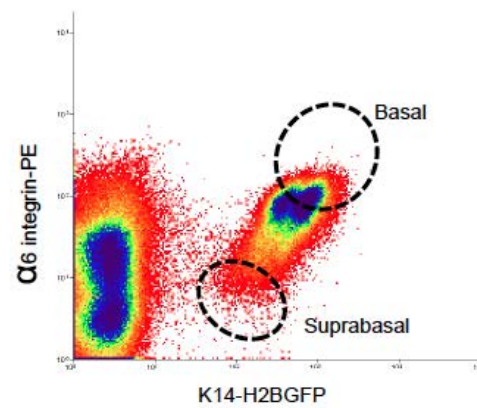
Figure 5.1 Isolation embryonic epidermal lineages. (A) Morphology of E14.5 (denoted as E14 for convenience) basal and suprabasal lineages. Dotted lines mark epidermal- dermal boundaries. bl, basal layer cells; sl: suprabasal layer cells; epi, epidermis ;der, dermis. (B) Cell sorting strategy for isolating basal and suprabasal cells.

multiple motifs with a specific positional distribution for 3'end formation in *C.elegans*. To determine whether this pattern is conserved in mammals, I analyzed a previously published human dataset generated by direct RNA sequencing (Ozsolak et al., 2010). As expected, the distinct pattern of nucleotide composition was observed (Fig. 5.2A). Specifically, I observed that: 1) A is significantly enriched at +1 position; 2) a ~20nt A-rich region with a peak at -18 position is identified at -10~-30 position, matching the position of the PAS; 3) a short U-rich and a short A-rich regions are observed at -1 ~-10 position; 4) a ~30nt U-rich region is detected downstream of the cleavage site. In contrast, when I analyzed a previously published human dataset generated by 3Seq (Jenal et al., 2012), the distinct pattern was not observed. Instead, the sequences downstream of the cleavage site were dominated by A-rich sequences, indicating widespread internal priming events in unfiltered 3Seq results (Fig. 5.2B). Together, these results suggested that authentic 3'end formation requires a distinct nucleotide composition pattern, and I could develop a bioinformatic approach to leverage this pattern to computationally distinguish authentic mRNA 3'end formation from the large number of internal priming events generated by 3Seq.

To test this possibility, I performed 3Seq and developed our bioinformatic pipeline with mouse embryonic skin lineages including basal stem cells and suprabasal cells (Fig 5.1). I uniquely mapped 7.9 million and 7.4 million 100mer reads for each library. Because of my optimization in the library construction, the reads were highly enriched around the cleavage site (nucleotide position 0) with a gradual slope from 5'end and a sharp drop in 3'end (Fig. 5.2C). To determine the accuracy of the cleavage site identification, I selected peaks whose cleavage site is followed by C, T or G (non-A) in the genomic sequences, and

plotted reads density of these trimmed ends (Fig. 5.2D). Indeed, trimmed ends were significantly enriched at the defined 0 position. Thus, our optimized protocol has generated a very high percentage of reads with the untemplated As that allows accurate determination of the cleavage site.

Because polyadenylation is a post-transcriptional processing, the poly(A) site should be located within actively transcribed region of RNA Pol II and close to the 3' end of actively transcribed region of RNA Pol II (Lin et al., 2012). To provide insights into the link between 3' end formation and RNA Pol II transcription, I performed ChIP-seq of histone H3K4me3 and H3K36me3 in the basal stem cells and defined actively transcribed gene as H3K4me3 present at promoter and H3K36me3 present at gene body (double positive). Metagene analysis of H3K36me3 for the most distal peak revealed that the H3K36me3 signals are enriched upstream of the cleavage site and are weakened rapidly downstream of the cleavage site for all distal peaks (Fig. 5.2E). This indicated that transcription termination takes place shortly after the distal 3'end cleavage site. In contrast, the drop-off of the H3K36me3 signals for proximal peaks was considerably milder, compared to that of the distal peaks (Fig. 5.2E). Collectively, these analyses strongly support that our optimized technique successfully defines 3' end of mRNAs and provide new insights for the relationship between Pol II transcription, as detected by H3K36me3, and 3'end formation.

5.3 Distinguish authentic 3' end signals from internal priming signals

To systematically identify authentic 3'end signals, I designed a bioinformatic pipeline to eliminate internal priming events by using a combination of motifs upstream and downstream of the defined cleavage site. I divided all 3Seq peaks into four mutually
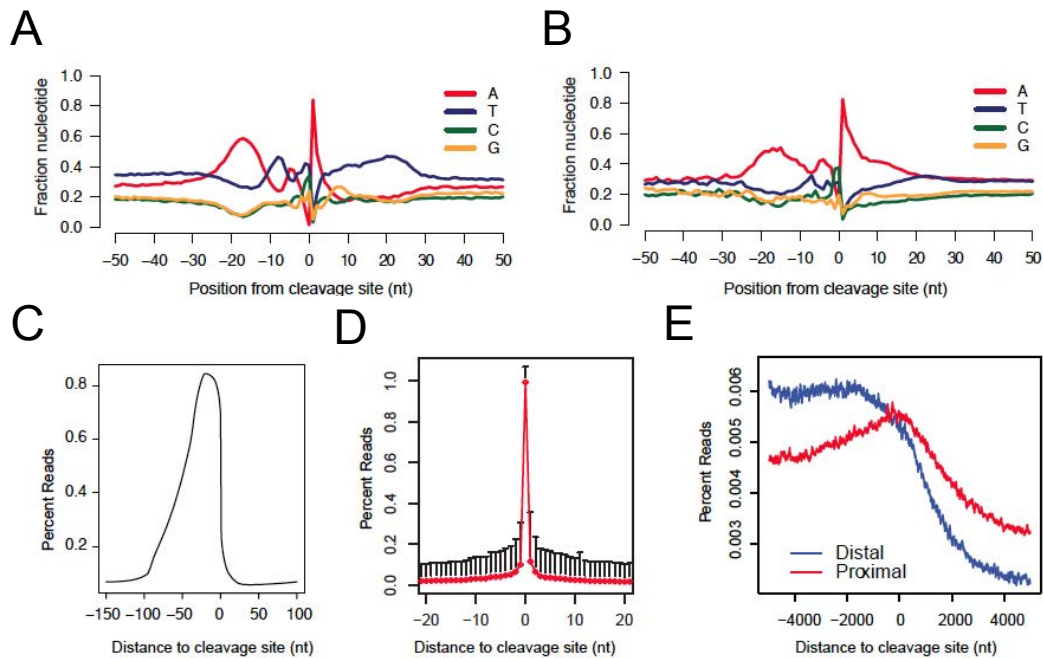
Figure 5.2 Distinct nucleotide composition patterns at mRNA 3'end regions.
Nucleotide composition of polyA site for Direct RNA sequencing data and 3Seq.
(C) 3Seq metagene analysis of -150~+100 window centered at the cleavage site.
(D) Percent reads plot of 3'end of trimmed reads centered at the cleavage site
which +1 position is T,G or C. (E) H3K36me3 metagene plot centered at the
cleavage site. Distal peaks (blue) are defined as the most 3' peak assigned to a
given gene. Proximal (red) are peaks mapped to RefSeq and Ensembl annotated
that is proximal in location to the distal peaks.

exclusive categories based on the existence of the canonical PAS and the downstream A-rich sequences (Fig. 5.3A). Peaks in category 1 contained canonical PAS (A[A/U]UAAA) and showed no A-rich genomic sequences downstream of the cleavage site. Thus, they represented authentic 3'end signals. Peaks in category 2 contained both canonical PAS and downstream A-rich sequences. Thus, some of them were derived from mRNA 3'ends and the others were likely derived from internal priming. Peaks in category 3 were defined by the lack of both canonical PAS and downstream A-rich sequences. These peaks were likely derived from 3' end formation using alternative PAS or sequencing artifacts as judged by the low reads number. Peaks in category 4 contained only downstream A-rich sequences and lacked canonical PAS upstream of the cleavage site. They should represent mostly internal priming peaks and a few true 3' end peaks that use alternative PAS.

To remove peaks that likely arise from sequencing noise, I first eliminated peaks from all four categories that have less that 10 reads covering the cleavage sites. Then, I analyzed nucleotide composition around the defined cleavage sites for peaks from all four categories. As expected, the nucleotide distribution among four categories was different from each other (Fig. 5.3). Notably, category 1 peaks showed the exact pattern for authentic 3'end as observed in direct sequencing data (compare Fig. 5.2A and Fig. 5.3B). Therefore, I concluded that peaks in category 1 are derived from authentic 3' end signals.

To analyze peaks in category 2 that contain both canonical PAS and downstream A-rich genomic sequences, I applied additional criteria to define high-confidence 3'end signals. These filters were: 1) positional distribution of PAS e.g. at least one canonical PAS should be localized in position -10 to -30 upstream to the defined cleavage site; 2) no more than three canonical PAS should exist in the 50nt window upstream of the cleavage site

103

because multiple occurrence of PAS is likely due to interspersed non-A nucleotide in an A-rich sequence segment; 3) peaks should be mapped to either annotated 3'UTR region (Refseq and Ensembl, with 10kb extension in the 3'end) or intergenic regions. With these filters, Iidentified 2,626 peaks in category 2 as genuine 3'end signals and eliminated 7,483 peaks. Strikingly, the nucleotide distribution of the filtered peaks closely resembled the pattern of true 3'ends as observed for category 1 peaks whereas the eliminated peaks showed strong A-rich signals and lacked any specific PAS signals (Fig. 5.3C, left panel). This indicated that the eliminated peaks are characteristic of internal priming events.

I next analyzed peaks in category 3 with a focus on alternative PAS. I searched for enriched 10-nucleotide motifs using MEME (Beiley et al. 2009) from position -1 ~ -50 sequences of category 3 peaks. Strikingly, the 10nt motif search yielded results dominated by 6nt motifs (Fig. 5.4A). I further identified all enriched 6nt motifs with $p < 1e^{-4}$ and selected 6nt motifs that significantly enriched at -10 to -30 position (Fig. 5.4A), consistent with the notion that the alternative PAS is also recognized and processed by the same polyadenylation machinery as the canonical PAS (Di Giammartino et al., 2011). Altogether, I identified 14 6mer motifs as candidates for alternative PAS (Fig. 5.4B), majority of which are consistent with previous studies (Beaudoing et al. 2000, Derti et al., 2012). I then used the presence of these 14 alternative PAS as a primary filter to screen for authentic 3'end signals from category 3 (p<0.01). I extracted 2,336 true 3'end peaks and eliminated 33,231 peaks from category 3. Although the eliminated peaks accounted for the majority of category 3 peaks, they showed significant enrichment for peaks with very low reads count (less than 10 reads mapped to a peak), which is characteristic of sequencing/mapping noise (see below for detailed analysis). Importantly, when I plotted the nucleotide
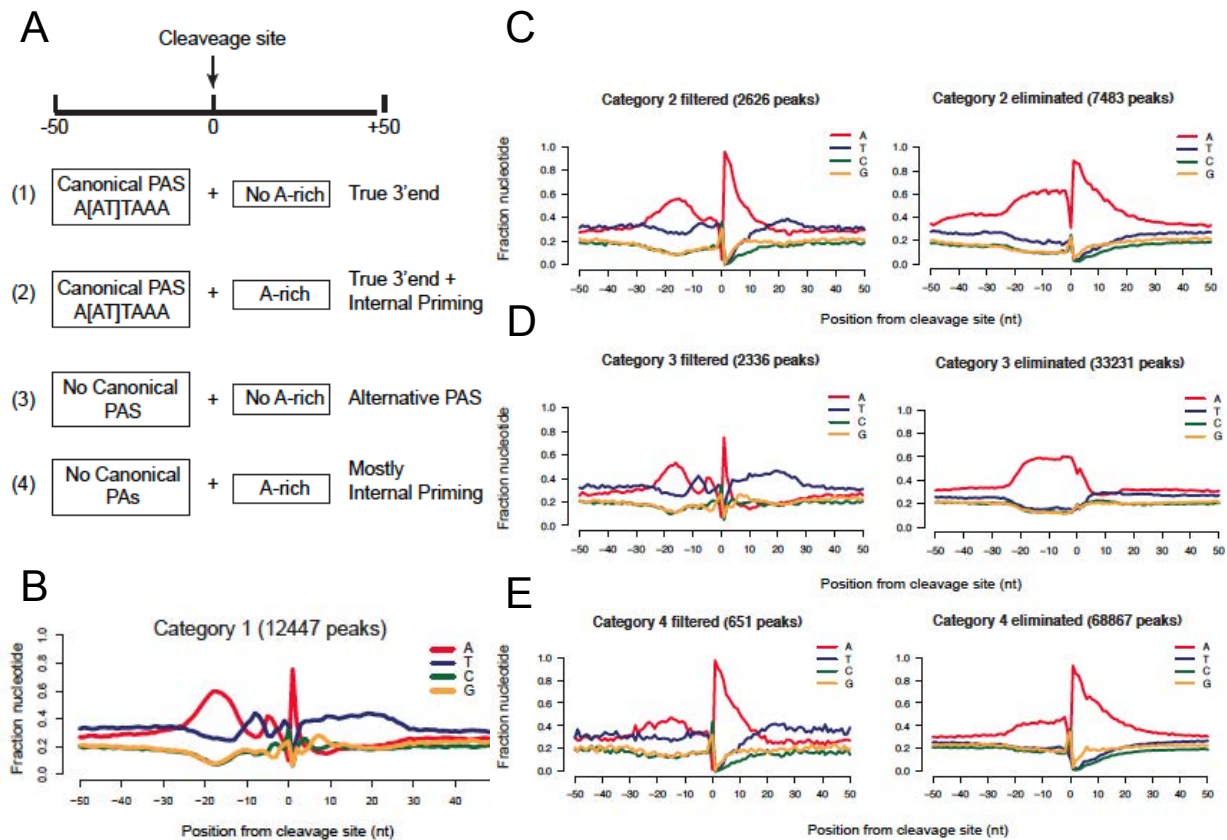
Figure 5.3 A bioinformatic pipeline distinguishes authentic 3'end signals from internal priming events. (A) Classification of 3Seq data is based on the presence of polyadenylation signal in position -50 to 0 and downstream A-rich sequence in position 0 to +50. Interpretation of the four categories is noted on the right and discussed in detail in the text. (B)-(E) Nucleotide composition plot of category 1 in (B), which is composed of true 3'end signals and categories 2 in (C), category 3 in (D) and (4) in (E). Fraction of nucleotide in -50 to +50 window centered at the defined cleavage site is plotted. In (C)-(E), each category is shown as a "filtered" panel, including all peaks pass the filter, which is classified as bona fide 3'end; and an "eliminated" panel, including all peaks fail to pass the filter, which is classified as false signals.

distribution of the filtered or eliminated peaks from category 3, only the pattern of the filtered peaks showed strong resemblance to that of category 1 whereas the eliminated peaks showed a simple A-rich pattern without any other specific signals (Fig. 5.3D). These results further validated the robustness of our identification of the peaks that are derived from alternative PAS. Since category 4 was composed of a large number of peaks with only A-rich sequence downstream of the cleavage site without canonical PAS upstream, I applied more stringent filters based on the presence of downstream U/GU-rich motifs for 3'end identification (p < 0.01). In this manner I identified 651 out of 68,841 peaks that were likely derived from authentic 3' end. Importantly, when I plotted the nucleotide distribution surrounding the cleavage site for these 651 peaks, I again observed a pattern that strongly resembles that of category 1 peaks (Fig. 5.3E). As expected, the pattern of the eliminated peaks was consistent with internal priming events including prominent A-rich sequences downstream of the defined cleavage site and lacking any other specific signals (Fig. 5.3E). Together, these results validated the effectiveness of our bioinformatic analysis to successfully distinguish authentic 3'end signals from a large number of sequencing artifacts. Overall, I defined 18,060 3Seq peaks as mRNA 3'ends, corresponding to 12,739 genes. Although the eliminated peaks significantly outnumbered the retained peaks (86.4% of peaks are eliminated), the eliminated peaks only accounted for 57.5% of total mappable reads.

To further evaluate the performance of the bioinformatic pipeline, I applied the analysis pipeline to a previously published 3Seq dataset generated from human U2OS cells (Jenal et al., 2012) and compared the results to the direct RNA sequencing (DRS) annotated 3'end generated from human liver (Fig 5.4, Ozsolak et al., 2010). Because these two dataset

were not from the same RNA source, I only compared genes that were detected by both 3Seq (before filtering) and DRS. When a 3Seq peak intersected with a DRS annotated 3'end region, it was called a match. I defined successful classification as when the filtered peaks match the DRS annotation and the eliminated peaks do not match the DRS annotation, which were denoted as true positive signals and true negative signals, respectively. Among a total of 95,332 3Seq peaks, 10,891 peaks (11.4%) were true positive signals whereas 52,101 peaks (77.0%) were true negative signals. This result indicated a 89% successful rate for our pipeline. 3Seq peaks that passed our filter but did not match the DRS annotation were classified as false positives, which constitutes 3.6% of of the total peaks. However, since these two dataset were derived from different cell types, false positives could also be derived from U2OS cell type-specific gene expression. Finally, 7,579 3Seq peaks (8.4%) were eliminated but had the DRS match, and were therefore considered as false negatives. Despite a relative high proportion, however, 71.6% of the false negative peaks had less than 10 reads covering the cleavage site, which resulted in the exclusion of these peaks by our analysis (Table 5.1). I expected that these peaks should be correctly classified with a deeper sequencing depth. Among the remaining false negatives (28.4% of the false negatives e.g. 2.4% of the all peaks), the leading causes for the false classification were: 1) failure to detect alternative PAS (15.8% of the total false negatives) and 2) failure to define the cleavage site (12.6% of the total false negatives). Thus, with the sequencing depth and stringency of filtering, only 2.4% of 3Seq peaks were mis-classified by our analysis.

5.4   Genome-wide analysis of polyadenylation signals and downstream sequence motifs in epidermal lineages

(A) Benchmarking Result

Pass filter and match DRS	True Positive	12.7%	⎫
Eliminated and not match DRS	True Negative	76.1%	⎬ 88.8% successful
Eliminated but match DRS	False Negative	7.3%	⎭
Passed filter but not match DRS	False Positive	3.9%	

(B) Composition of False Negative signals

False Negative	% of false negative	% of total peaks
match DRS, reads count at cleavage site ≤ 10	71.6%	5.7%
match DRS, reads count at cleavage site > 10	28.4%	2.3%
no PAS found	15.8%	1.3%
Cleavage site is not correct defined	12.6%	1.0%

Table 5.1 Benchmarking 3seq peak filtering with DRS annotation

After successfully classifying 3Seq peaks and distinguishing authentic 3' ends of mRNAs from sequencing artifacts, Iperformed genome-wide analyses of the utilization of PAS and other prominent motifs for 3'end formation in the skin lineages. Canonical PAS, A[A/U]UAAA, were dominantly utilized e.g. AAUAAA peaks count for 58.4% and AUUAAA peaks count for 19.2% of the total peaks and only 22.4% of mRNAs use the alternative PAS for 3' end formation (Fig. 5.4B). Taken together, these results indicate that the utilization of PAS is strongly biased towards canonical PAS in mouse.

Next I analyzed the sequences that are downstream of the cleavage sites, where U/GU-rich sequences have been reported as Cstf binding site (MacDonald et al., 1994; Takagaki and Manley, 1997; Beyer et al., 1997). These U/GU-rich motifs have been proposed to be present within 50nt downstream of the cleavage site, and their positional enrichment is less well defined compared to the PAS. I searched for 10nt motifs that are enriched in position 0~+100 using true 3'end peaks in category 1 and 3. I identified an U-rich motif and a GU-rich motif as shown in Figure 3C. Furthermore, the U/GU-rich motifs also showed significant positional enrichment at 0~+30nt immediately downstream of the cleavage site (Fig. 5.4C). Overall, I observed a strong enrichment of these motifs in 3'end formation when examining the presence of these motifs in category 1-3 peaks. For example, 90.8% of category 1 peaks, 80.4% of filtered category 2 peaks and 91.4% of filtered category 3 peaks contained at least one U/GU-rich motif at 0~+100 position downstream of the cleavage site (p < 0.01). These results suggest that for most genes both upstream PAS and downstream U/GU-rich motifs are utilized to direct 3'end formation.

Because for many genes more than one 3'UTR are formed, I next asked whether there is a preference for the canonical or alternative PAS at proximal and distal ends.
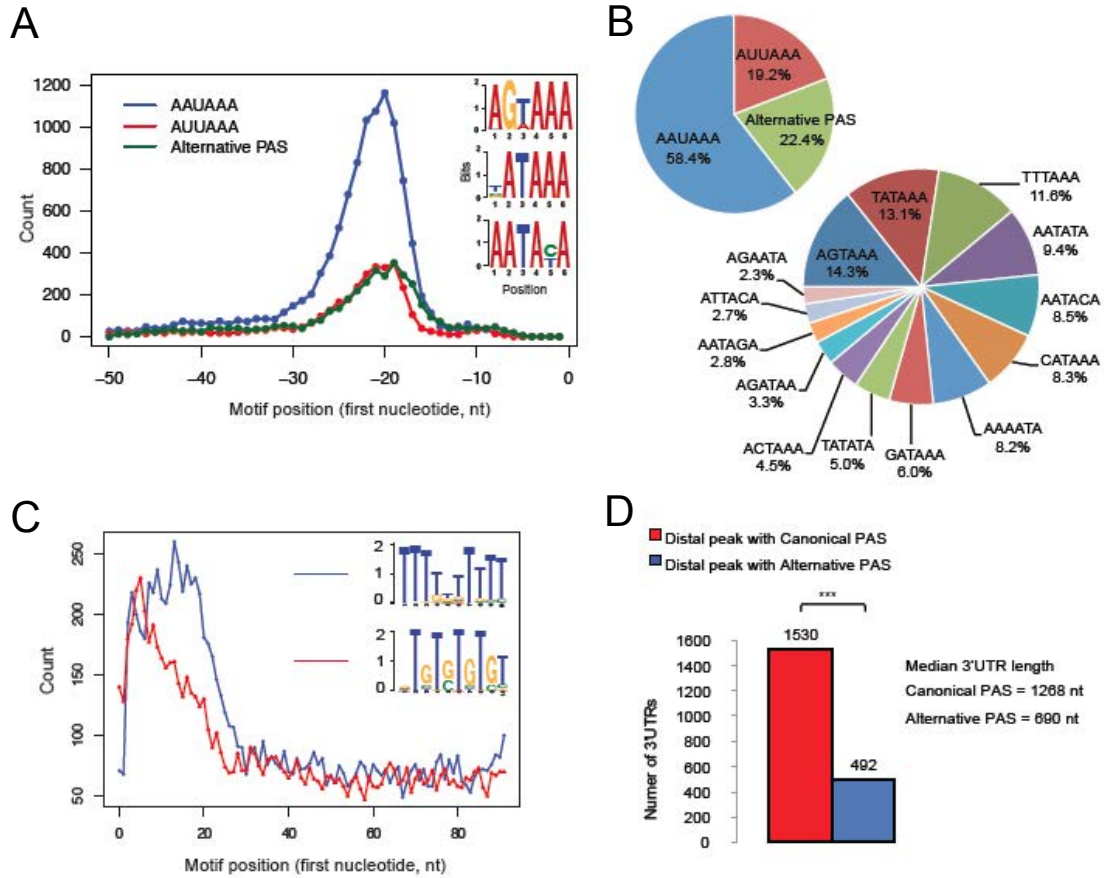
Figure 5.4. Characterization of sequence elements surrounding the polyadenylation site. (A) Positions of PAS are plotted in the -50 to 0 window upstream of the cleavage site. Three enriched 6mer motifs identified in the -50 to 0 window upstream of the cleavage site are shown in the right. T is shown instead U. (B) Genome-wide usage of PAS. All alternative PAS are displayed in details in a subset of the pie chart. (C) Positions of the U-rich (blue line) and GU-rich motifs (red line) are identified in the 0 to +100 window downstream of the cleavage site, T is shown instead U as in (A). (D) Preference for canonical and alternative PAS by distal and proximal 3'UTRs. ***: p<0.001, two sample Z-test.

I identified 2,022 genes using both canonical and alternative PAS. Interestingly, significantly more canonical PAS were used to generate longer 3'UTR than alternative PAS. For example, 1,530 distal ends (75.7%) were formed by using canonical PAS whereas 492 distal ends (24.3%) used alternative PAS (Fig. 5.4D). The median length for 3'UTRs formed by the canonical PAS was 1,268nt whereas it was 690nt for the 3'UTRs formed by alternative PAS. Taken together, these results show a strong bias towards canonical PAS by the distal peaks.

5.5   Genome-wide quantification of the transcriptome

I next benchmarked the quantitative performance of 3Seq. I first chose a group of genes with known differential expression between E14 basal stem cells and suprabasal differentiating cells including basal cell markers e.g. transcription factors (Lef1 and Trp63), extracellular basement membrane gene (α6 integrin), basal structural genes (Krt14 and Krt5) as well as a Wnt receptor (Fzd10) and suprabasal markers e.g. structural genes (Loricrin, Krt1 and Krt10) and cell cycle inhibitor (Cdkn1a also known as p21). I quantified their expression by counting reads mapped to each gene and normalizing to the total mappable reads from each library. Remarkably, the fold-change data determined by 3Seq measurement showed very robust correlation with the fold-change results obtained by qPCR for all of these genes (Pearson correlation r = 0.996, Fig. 5.5A). This result indicates that the 3Seq quantification is suitable for differential gene expression studies, and for individual genes its performance is comparable to RT-qPCR quantification.

In the genome-wide analysis, I noticed that the highest reads number for a single gene in each library was over 30,000. This observation suggested that with the overall

sequencing depth at 8 million uniquely mappable reads, 3Seq could provide a quantitative dynamic range of up to 10^4. To test this hypothesis, I first examined normalized reads count per million mappable reads (RPM) for 12,739 genes that showed detectable expression in the skin. The result showed that 3Seq can detect gene RPM over four orders of magnitude in both basal and suprabasal cells (Fig. 5.5B).

I then evaluated the impact of sequencing depth on the sensitivity of 3Seq. To this end, I performed sub-sampling from our E14 basal 3Seq data (11.2 million uniquely mappable reads). By sub-sampling from 0.1 million to 10 million reads following the exact filtering analysis, I determined that with 5 million uniquely mappable reads it was sufficient to detect 90% of 3'end events and the detection of authentic 3'ends reached a plateau with 8-10 million reads (Fig. 5.5C). I estimated that with 8-10 million mapped reads our experimental and bioinformatic pipeline represents a robust tool for genome-wide mRNA quantification. With the output of Illumina HiSeq routinely approaching 150-200M reads per lane (2012 data), I could profile a minimum of 8-10 samples with a single lane for the cost of ~$1,000.

To evaluate the accuracy of absolute copy number quantification by 3Seq, I used *in vitro* cultured keratinocytes, which provided us an ample amount of total RNA, to construct 3Seq libraries and cDNAs for extensive quantification. With these two different keratinocytes samples, I uniquely mapped 1.6 million 50mer 3Seq reads, respectively. Despite the low sequencing depth of these experiments, I still observed the dynamics range of ~10^4. I selected 16 genes whose expression spanned the entire dynamic range for quantitative analysis. To assure accurate copy number quantification by qPCR, I characterized amplification efficiency and determined the dynamic range of linear
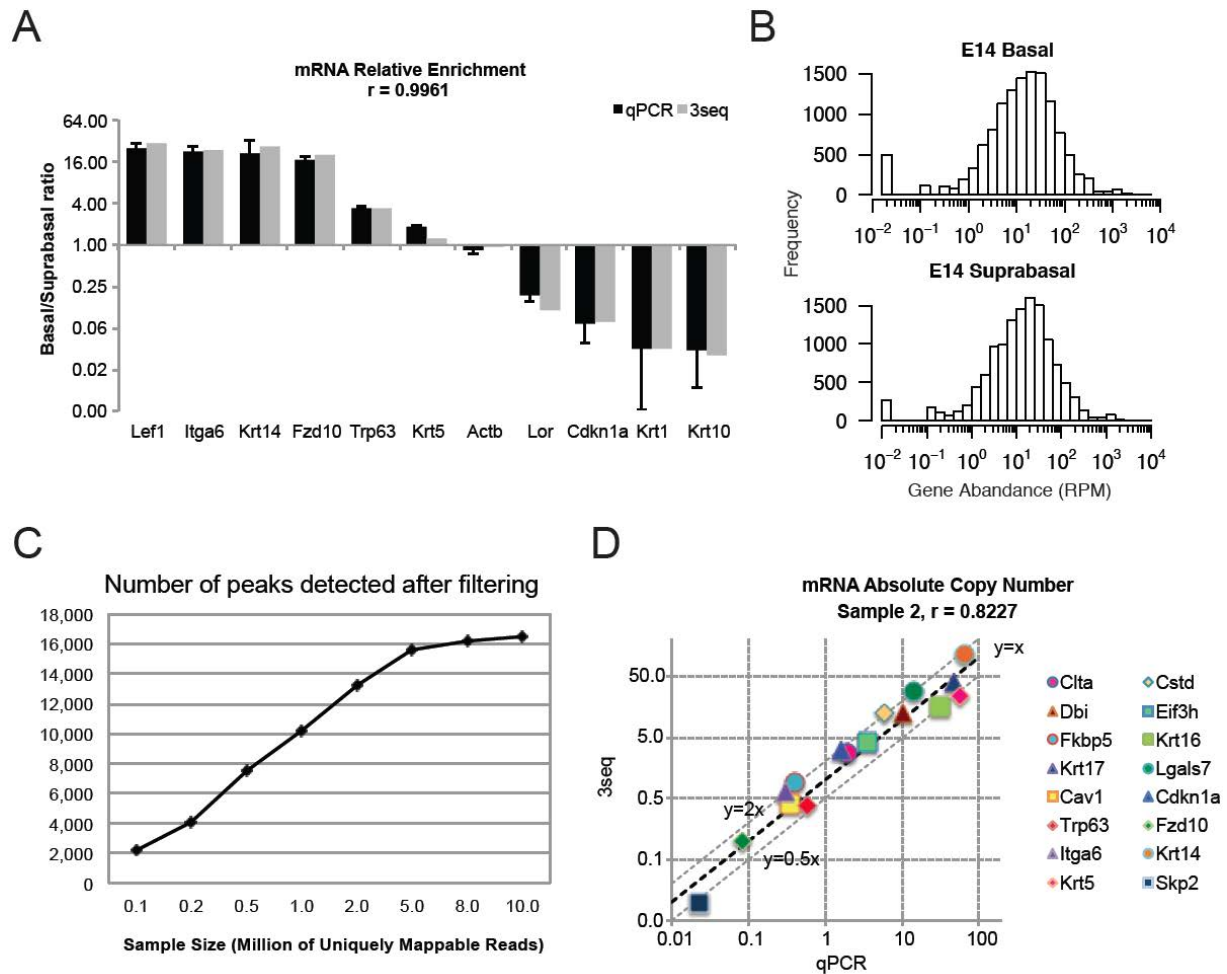
Figure 5.5 3Seq provides accurate measurement for both relative and absolute mRNA quantification. (*A*) Measurement of mRNA relative abundance between E14 basal and suprabasal samples. (*B*) Dynamic range of 3Seq gene quantification in E14 basal and suprabasal samples. (C) Impact of sequencing depth on 3'end detection and quantification. Sub-sampling of different size of mappable reads is used to perform 3seq peak filtering (solid line).
(*D*) Log 10 ratio of absolute copy number for 16 genes as measured by qPCR is compared to the measurement by 3Seq (normalized to Hprt).

amplification for each qPCR primer set before calculating the copy number for each gene. Strikingly, when I compared the normalized 3Seq quantification for each gene to their absolute copy number, it showed a remarkable consistency between these two measurements, with Pearson's correlation coefficient 0.8227 (Fig. 5.5D). In general, I also observed an average of 1.48 fold overestimation by 3Seq quantification, with a standard deviation of 0.79. This is probably due to the sequencing depth. Only 1.6 million mappable reads and the short reads length (50mer) for these experiments. I expect that with a depth of ~20-30M total reads per library and longer reads e.g. 100mer instead of 50mer, I could observe more robust absolute quantification by 3Seq. Nonetheless, these results demonstrated that 3Seq enables highly accurate quantification for mRNA throughout the entire dynamic range of mRNA expression.

5.6   Differential 3' UTR usage in the epidermal lineages

Transcriptome quantification by 3Seq enables the detection of differential 3'UTR usage in a quantitative manner. For example, if gene A forms two 3'UTR isoforms with a proximal and distal PAS, respectively, it would be important to determine if gene A prefers the proximal isoform than the distal isoform in one cell lineage vs the other. These insights may uncover a molecular basis for dynamic regulation of mRNA e.g. stability, localization and translation through differential 3'UTR expression in different cell lineages. Since 3Seq was able to quantify abundance of each 3'end isoform independently, I directly examined the data to identify 3'UTR switching events. I analyzed 3Seq data for differential 3'end formation between the E14 basal and suprabasal cells. Overall, 39.1% of genes contain more than one 3'end (Fig. 5.6A). I first asked whether differential PAS utilization is

114

associated with 3'UTR switching. I studied the PAS distribution in splicing-independent, differential 3' end formation and observed a strong preference for the alternative PAS in these 3'UTR switching events, compared to the genome-wide pattern (p<0.0001) (Fig. 5.6B). Interestingly, increased preference for the alternative PAS was mainly balanced by the decreased usage of AAUAAA, but not AUUAAA (Fig. 5.6B). On the other hand, the usage of alternative PAS at the distal 3'ends for these 3'UTR switching events is drastically reduced (7.1%) in contrast to the genome-wide usage rate (20.3%). In order to systematically identify 3'UTR switching events among these genes, I focused on the proportion of each 3'end isoform for a single transcript. Overall, I identified a total of 1,228 3'end isoforms from splicing-independent, authentic alternative polyadenylation that are differentially enriched between the basal and suprabasal lineages (p < 0.05, Benjiani & Hochberg correction). These 3' ends belong to 829 genes.

To validate these 3'UTR switching events experimentally, I selected 6 genes (Ddx46, Dicer, Eif1ad, Trappc5, Pcmtd2 and Opa1), which showed differential usage of 3' ends. I performed qPCR to validate the detection by 3Seq. After validating qPCR primers and expression variation between cell sorting experiments 6 out of 7 genes showed the same differential usage of the distal 3'ends in the qPCR analysis, consistent with the results obtained by the 3Seq analysis (Fig. 5.6c-D). This result further validated the ability of 3Seq in quantitative analysis for individual 3'ends.

To extend my study to more remotely related cell types, I compared the E14 embryonic skin lineages to the postnatal day 4 (P4) hair follicle lineages, which are originated from the E14 basal stem cells Blanpain & Fuchs, 2009), for 3'UTR switching events. I identified 1,394 3'ends that switched independent of splicing between these two
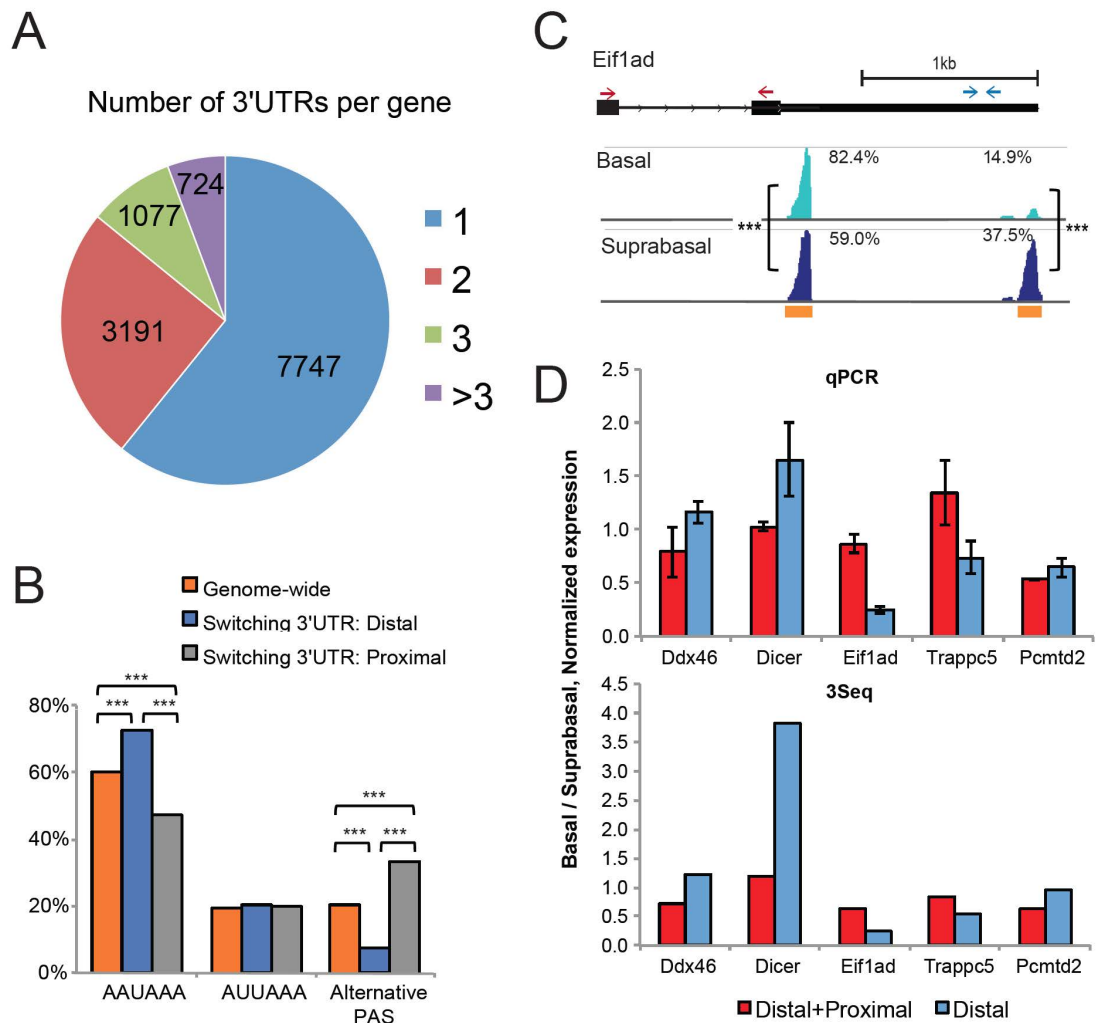
Figure 5.6 Differential 3'UTR formation is identified in embryonic skin stem cell lineages. (A) Number of 3'UTRs per gene. (B) Positional preference of PAS in splicing-independent 3'UTR switching, plotted as the proportions of 3'UTRs contains AAUAAA, AUUAAA or alternative PAS. (C) Examples of splicing-independent 3'UTR switching of Eif1ad gene. 3Seq coverage with RefSeq annotation is plotted on the left. Position of qPCR primers for detecting both proximal and distal 3'end expression (gray) and distal 3'end only (black) are shown (arrows). Gray bars below the coverage track indicates defined 3Seq peak region. The proportions of each 3'end are labeled. qPCR validation and comparison with 3Seq measurements are shown on the right. ***: $p<0.001$, **: $p<0.01$. (D) qPCR validation of switching 3'UTRs. A universal primer (black arrows in (C)) is used to detect both proximal and distal isoforms while another pair of primer only measures the distal isoforms (gray arrow in (C)). Error bar is standard deviation from biological replicates.

lineages, corresponding to 933 genes. Therefore, 3'UTR switching was slightly more widespread between the E14 and P4 lineages than that between the E14 basal and suprabasal lineages. I applied GO term analysis by using Gene Set Enrichment Analysis (Subramanian et al., 2005) for the genes with switched 3'UTRs. However, I did not observe any strong enrichment for any particular functional cluster even with a generous p-value cutoff (1e-4). This result suggests that differential 3'end formation takes place in many genes with diverse functions without any particular functional enrichment.

5.7   Discussion

I have optimized the 3Seq technique with a bioinformatic pipeline to accurately and quantitatively identify the 3' end of mRNA. By leveraging local sequence composition characteristic of mRNA 3'end formation I can accurately distinguish 'false' peaks derived from internal priming from authentic poly(A) sites. The accurate annotation of 3'end formation for specific cell populations will provide precise 3'UTR information for studies focusing on the regulatory function of 3'UTR e.g. cell type-specific identification of miRNA targets and characterization of cell type-specific binding motifs for RNA-binding proteins.

My results indicate that the quantification of mRNA levels by global 3Seq analysis has comparable accuracy as qPCR for differential expression of individual genes. In addition, 3Seq quantification shows robust correlation with the absolute copy number of individual genes over the dynamic range of $\sim 10^4$. This accuracy was obtained by less than 3 million uniquely mappable reads (~ 10 million raw reads). It should be noted that the proportion of mappable reads could be further increased by sequencing longer fragments or using a paired-end strategy. Considering the current yield for Illumina HiSeq platform is

117

approaching 250 million reads per lane, our 3Seq protocol would enable multiplexing of more that 10 samples per lane while retaining quantification accuracy. Practically, this significantly lowers the cost of transcriptome profiling. In particular, 3Seq is suitable for transcriptome profiling in gene functional studies where the main goal is usually global measurement of transcripts abundance and detection of full-length transcripts is not required.

The simultaneous detection of 3'end and quantification of mRNA expression gives us a unique opportunity to examine the dynamics of 3'end formation in a quantitative manner and reveal several hitherto unappreciated insights. For example, I showed that mRNAs using canonical PAS, especially AAUAAA, tend to express at a considerably higher level than mRNAs using alternative PAS (Fig. 4C). Furthermore, in 3'UTR switching events, the distal 3'ends are strongly biased towards AAUAAA whereas the proximal 3'ends are biased towards alternative PAS (Fig. 5B). These observations suggest an intrinsic relationship between polyadenylation and gene expression. Because of the simplicity and cost-effectiveness of 3Seq, I anticipate that 3Seq, when combining with robust bioinformatic analysis, can be widely applied in numerous studies for gene expression.

The quantitative profiling of polyadenylated RNAs in skin stem cell lineages yields important insights into the dynamics of mRNA 3'end formation in closely related somatic cells *in vivo*. Among 12,739 genes that are detected in our analysis, 4,992 of them (39.1%) contain more than one 3'end. Interestingly, this percentage is comparable to the proportion of mRNAs with alternative 3'ends that has been detected in *C. elegans* (30%) by 3P-Seq (Jan et al., 2011). Of note, my study examined the alternative 3'UTRs in a single mammalian lineage at a specific time point whereas the *C. elegans* study used the whole organism at

different developmental stages. This result indicates that a large number of genes employ different 3'UTRs to control their mRNAs e.g. stability, location as well as translation. I was particularly interested in detecting mRNA transcripts that differentially use distinct 3'UTR isoforms between the basal stem cells and the suprabasal differentiating cells. It could provide critical insights for the role of differential 3'UTR usage during developmental transitions. Surprisingly, I only detected 829 genes that show differential 3'end usage between these two lineages. It indicates that the impact of mRNA 3'end switch could be limited during developmental transitions. However, I could not rule out the possibility that a few master regulators, which employ differential 3'end formation, could have a significant impact for the process. For example, the 3'UTR of Dicer shows one of the most dramatic shortening once the stem cells embark on the differentiation program. In contrast to our findings, several previous studies suggest that differential 3'end formation might be a widespread mechanism for cell fate specification (Tian et al., 2005; Sandberg et al., 2008; Mayr & Bartel, 2009). However, these studies used very different samples for such comparison e.g. 1) a cohort of studies from a large number of cell lines (Tian et al., 2005); 2) *in vitro* cultured tumor cells (Mayr & Bartel, 2009; Fu et al., 2011; Shepard et al., 2011); or 3) a mixed cell populations isolated from a large tissue or organ (Ozsolak et al., 2010). As a result, those samples were either heterogeneous or highly transformed, which may not be able to accurately reflect transcriptome diversity in a single lineage *in vivo*. Unlike these cells, the basal stem cells and suprabasal differentiating cells are spatiotemporally well-defined and closely linked with a much shorter distance in their developmental timeline. To examine if increased developmental gap could lead to a more dramatic change in 3'end usage, I also compared embryonic skin stem cells to neonatal hair follicle lineages, which

are originated from the stem cells. I found that 933 genes show 3'end switching among these two cell types. Thus, our results indicate that differential usage of alternative polyadenylation could be either a specific event for a certain type of developmental transition and cell transformation or a process that gradually takes place through *in vivo* differentiation.

5.8   Materials and Methods

5.8.1   Flow cytometry

Mouse embryonic epidermis is dissected from E14 embryo, cut to small pieces (2mm x 2mm) and trypsinized for 10min in 37°C. Digested sample is resuspended with PBS and 3% chalexed serum and filtered to get single cell suspension. To isolated basal and suprabasal population from embryonic epidermis, transgenic mouse line K14H2BGFP is used and single cell suspension is further stained with Rat anti-human CD49f PE antibody (BD Pharmingen) and then undergo flow cytometry.

5.8.2   Cell culture, RNA extraction and cDNA construction and quantitative PCR

Primary keratinocytes are isolated and cultured as previously described (Yi et al 2008). RNA from *in vitro* cultured cell or *in vivo* sorted cells are extracted by TriZOL and precipitated with isopropanol. cDNA construction is performed using Invitrogen SuperScript III cDNA Construction Kit. Quantitative PCR is performed using BioRAD SYBR Green system. For absolute copy number qPCR, PCR amplicon of candidate genes are gel purified and concentration is measured. Standard curve of each amplicon is determined by

performing series dilution of amplicon and calculate linear relationship of Ct values and copy number input per reaction.

5.8.3 Chromatin Immunoprecipitation

ChIP experiments are performed using ChIP-IT™ Express kit from Active Motif (Catalog #53008), following manufacturer suggested procedure. H3K4me3 antibody is purchased from Active Motif (Catalog #39159) and H3K36me3 antibody is purchased from Abcam (Catalog #ab9050).

5.8.4 Sequencing library preparation

For 3Seq experiment, $0.5\sim2\mu g$ total RNA are polyA selected twice using Invitrogen Dynabeads mRNA DIRECT Kit. The polyA RNA are fragmented for 5min using Ambion RNA Fragmentation Kit. Reverse transcription is performed using fragmented RNA and anchored oligo dT_{20} probe, and second strands are further synthesized using RNase H and DNA polymerase I. Double strand DNA library is constructed following Illumina standard dsDNA library protocol. For ChIP-Seq experiment, ChIP-ed DNA is purified and followed by Illumina standard dsDNA library protocol.

5.8.5 Read mapping

Read mapping is performed using the short reads aligner Bowtie (version 0.12.7, Langmead et al. 2009). I observed that over 50% of reads contains concatenated A, a subset of which also contains an incomplete 3'adapter sequence. The adapter sequences are very degenerate and mismatch-prone, due to the effect of base-calling on concatenated A and

therefore they have low quality scores. All reads containing the lowest quality score are trimmed to remove the low quality sequences from the 3'end. Then, all reads are aligned against the mouse genome (mm9), allowing a maximum of two mismatches (bowtie parameter –v 2 –m 1 --nomaqround -l 20).

Since reads contains untemplated adenosine and 3' adapter sequences will fail to align as the full-length, I first trim the adapter sequence using Andy Chu's algorithm (publically available at http://www.bcgsc.ca/platform/bioinfo/software/adapter-trimming-for-small-rna-sequencing) and then trim off all concatenate adenosines from the 3'end. These reads are aligned using the same Bowtie setting against mouse genome (mm9). Finally reads failed to aligned are re-aligned using Tophat (v1.2.0, Trapell et all., 2009) to capture reads spanning exon-exon junctions. All reads uniquely aligned are pooled to generate reads coverage tracks.

5.8.6 Peak calling

Aligned reads cluster to form peaks. To define the reads enriched region, I used the peak calling algorithm MACS (version 1.4.0, Zhang et al. 2008). First the read alignment data is converted to read coverage data, (bedgraph format), which has single nucleotide resolution. Peaks are called using reads coverage data as input for MACS, using settings "--bw 120 --slocal 1000 --keep-dup auto --nomodel --shiftsize 20". Peaks are further processed using PeakSplitter (version: 0.1, settings "-f -v 0.2"), which will separate nearby peaks called as a single peak by MACS.

5.8.7 Determination of the cleavage site

Cleavage site of a given peak is determined using all trimmed reads, i.e. reads has been trimmed from 3'end due to untemplated concatenated A and adapter sequence. Briefly, after collecting all trimmed reads, positions of 3' terminal nucleotide of all trimmed reads are intersected with 3Seq peak regions. The nucleotide position that has the maximum count of 3' terminal of trimmed reads is defined as the cleavage site. If this position is not unique, which means multiple position has the same maximum 3' terminal count, only peaks with at most three maximum and maximum value is greater than 1 will be further analyzed.

5.8.8 Motif combination analysis

a. Canonical PAS motif search

Canonical PAS are well-established 6mer motif AATAAA and ATTAAA. I searched for these two 6mers directly from sequences 50nt upstream of cleavage site and positional distribution is analyzed.

b. Select training set for A-rich region motif search

To define the A-rich motif that necessary for internal priming, I first selected 3seq peaks that have canonical PAS and map to -50 to +50 window of 3'end annotated Refseq genes. Then 3seq peaks that don't have AATAAA or ATTAAA but map to the same gene are extracted. Finally reads count cutoff (>50) is applied to extract high confidence priming events. Peaks that pass all these criteria are considered enriched of internal priming events. I used MEME with default setting to search for 10mer motif on 0 to +50 sequences of the selected peaks. After the first round, I identified that A-rich motif is most informative

at 0 to +15 region of the cleavage site. Therefore, I performed a second round of MEME on sequences only from 0 to +15 region of cleavage site, which is used in the following analysis.

c. Alternative PAS search

3seq peaks in category 3 have no canonical PAS or A-rich sequence, therefore may be derived from 3'end using non-canonical PAS or sequencing noise. I applied a simple reads cutoff filter to remove peaks that have less than 50 reads to enrich high-confidence priming events. Then Iran MEME with default setting search for 10mer motifs from sequences of -50 to 0 of the cleavage site. As expected, the 10mer motifs found showed that information is concentrated at 6mer. I then extracted the 6mer motif and analyze the positional enrichment around the cleavage site. I used two-sample Z-test to define statistical significance of 6mer position at -10 to -30 region and then nucleotide distribution is inspected to further remove motifs that likely from internal priming, which has extremely A-rich sequence immediately upstream of the cleavage site.

d. U/GU-rich motif search

3seq peaks in category 1 and 2 are used to search for U/GU-rich motifs and positional preference analysis because these peaks lack A-rich region and as a result U/GU-rich position can be more reliably defined.

e. Definition of existence of PAS

Existence of a motif is determined using FIMO in the MEME package. Because of different information enrichment of different motifs (for example, A[AT]TAAA is not information-rich), p-value cutoff used to call a given motif exists varies for different motifs. A-rich motif: e<0.01, e is a corrected value of p value, Alternative PAS: p<0.01, U/GU-rich: p<0.01

5.8.9 mRNA quantification and switching 3'UTR detection

Peaks classified as authentic 3'ends are used for transcript quantification. I used Refseq and Ensembl annotation to assign peaks to gene symbols (database downloaded on Apr. 4th, 2012). Peaks within 10kb downstream of an annotated 3'end and not overlap with any annotated gene body are assigned as an extended 3'UTR. Expression level for a given gene is calculated by summing the reads count of all authentic 3'ends peaks mapped to the gene normalized to total number of mappable reads (millions). For absolute quantification comparison with qPCR, 3Seq quantification is further normalized to internal control Hprt. Peak lengths are not normalized. Lengths of 3'UTR are re-annotated using the distance between the stop codon and cleavage site of 3'seq peaks, with intron size subtraction. Genes with more than two major 3'UTR (quantity is larger than 10% of total gene quantity) are candidates for switching 3'UTR detection. P-value is calculated using two-sample Z-test with multiple comparison correction (Benjamini & Hochberg). Switching 3'UTRs are defined as at least 1.5-fold difference of proportion with $p < 0.05$.

5.8.10 Data Access

3Seq raw read sequences and processed data of E14 basal, E14 suprabasal, P4 hair follicle and *in vitro* cultured keratinocytes samples are available at NCBI/GEO (study GSE37641).

REFERENCE

Amin, S., and Bobola, N. (2014). Chromatin Immunoprecipitation and Chromatin Immunoprecipitation with Massively Parallel Sequencing on Mouse Embryonic Tissue. In Methods in Molecular Biology, (New York, NY: Springer New York), pp. 231–239.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Arai, F., Hirao, A., Ohmura, M., Sato, H., Matsuoka, S., Takubo, K., Ito, K., Koh, G.Y., and Suda, T. (2004). Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche. Cell *118*, 149–161.

Barker, N., van Es, J.H., Jaks, V., Kasper, M., Snippert, H., Toftgard, R., and Clevers, H. (2008). Very long-term self-renewal of small intestine, colon, and hair follicles from cycling Lgr5+ve stem cells. Cold Spring Harb. Symp. Quant. Biol. *73*, 351–356.

Baumann, K. (2014). Stem cells: keeping alert. Nat Rev Mol Cell Biol *15*, 428–429.

Beck, B., and Blanpain, C. (2013). Unravelling cancer stem cell potential. Nat Rev Cancer *13*, 727–738.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315–326.

Blanpain, C., and Fuchs, E. (2014). Plasticity of epithelial stem cells in tissue regeneration. Science *344*, 1242281–1242281.

Blanpain, C. (2013). Tracing the cellular origin of cancer. Nature Cell Biology *15*, 126–134.

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

Blanpain, C., Lowry, W.E., Geoghegan, A., Polak, L., and Fuchs, E. (2004). Self-renewal, multipotency, and the existence of two cell populations within an epithelial stem cell niche. Cell *118*, 635–648.

Botchkarev, V.A., Botchkareva, N.V., Nakamura, M., Huber, O., Funa, K., Lauster, R., Paus, R., and Gilchrest, B.A. (2001). Noggin is required for induction of the hair follicle growth phase in postnatal skin. Faseb J. *15*, 2205–2214.

Botchkarev, V.A., Botchkareva, N.V., Roth, W., Nakamura, M., Chen, L.H., Herzog,

W., Lindner, G., McMahon, J.A., Peters, C., Lauster, R., et al. (1999). Noggin is a mesenchymally derived stimulator of hair-follicle induction. Nature Cell Biology *1*, 158–164.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, dnA-binding proteins and nucleosome position. Nat Methods 1–8.

Burrell, R.A., McGranahan, N., Bartek, J., and Swanton, C. (2013). The causes and consequences of genetic heterogeneity in cancer evolution. Nature *501*, 338–345.

Chen, C.-C., and Chuong, C.-M. (2012). Multi-layered environmental regulation on the homeostasis of stem cells: the saga of hair growth and alopecia. J. Dermatol. Sci. *66*, 3–11.

Chen, T., Heller, E., Beronja, S., Oshimori, N., Stokes, N., and Fuchs, E. (2012). An RNA interference screen uncovers a new molecule in stem cell self-renewal and long-term regeneration. Nature *485*, 104–108.

Cheng, T., Rodrigues, N., Shen, H., Yang, Y., Dombkowski, D., Sykes, M., and Scadden, D.T. (2000). Hematopoietic stem cell quiescence maintained by p21cip1/waf1. Science *287*, 1804–1808.

Cheung, B.B., Koach, J., Tan, O., Kim, P., Bell, J.L., D'andreti, C., Sutton, S., Malyukova, A., Sekyere, E., Norris, M., et al. (2012). The retinoid signalling molecule, TRIM16, is repressed during squamous cell carcinoma skin carcinogenesis in vivo and reduces skin cancer cell migration in vitro. J. Pathol. *226*, 451–462.

Cheung, T.H., and Rando, T.A. (2013). Molecular regulation of stem cell quiescence. Nat Rev Mol Cell Biol *14*, 329–340.

Claudinot, S., Nicolas, M., Oshima, H., Rochat, A., and Barrandon, Y. (2005). Long-term renewal of hair follicles from clonogenic multipotent stem cells. Proc Natl Acad Sci USA *102*, 14677–14682.

Clavel, C., Grisanti, L., Zemla, R., Rezza, A., Barros, R., Sennett, R., Mazloom, A.R., Chung, C.-Y., Cai, X., Cai, C.-L., et al. (2012). Sox2 in the Dermal Papilla Niche Controls Hair Growth by Fine-Tuning BMP Signaling in Differentiating Hair Shaft Progenitors. Developmental Cell *23*, 981–994.

Cotsarelis, G., Sun, T.T., and Lavker, R.M. (1990). 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.

Craven, A.J., Ormandy, C.J., Robertson, F.G., Wilkins, R.J., Kelly, P.A., Nixon, A.J., and Pearson, A.J. (2001). Prolactin signaling influences the timing mechanism of

the hair follicle: analysis of hair growth cycles in prolactin receptor knockout mice. Endocrinology *142*, 2533–2539.

DasGupta, R., and Fuchs, E. (1999). Multiple roles for activated LEF/TCF transcription complexes during hair follicle development and differentiation. Development *126*, 4557–4568.

Ding, L., and Morrison, S.J. (2013). Haematopoietic stem cells and early lymphoid progenitors occupy distinct bone marrow niches. Nature *495*, 231–235.

Driskell, R.R., Giangreco, A., Jensen, K.B., Mulder, K.W., and Watt, F.M. (2009). Sox2-positive dermal papilla cells specify hair follicle type in mammalian epidermis. Development *136*, 2815–2823.

Festa, E., Fretz, J., Berry, R., Schmidt, B., Rodeheffer, M., Horowitz, M., and Horsley, V. (2011). Adipocyte lineage cells contribute to the skin stem cell niche to drive hair cycling. Cell *146*, 761–771.

Folgueras, A.R., Guo, X., Pasolli, H.A., Stokes, N., Polak, L., Zheng, D., and Fuchs, E. (2013). Architectural Niche Organization by LHX2 Is Linked to Hair Follicle Stem Cell Function. Cell Stem Cell *13*, 314–327.

Fuchs, E. (2008). Skin stem cells: rising to the surface. The Journal of Cell Biology *180*, 273–284.

Fuchs, E. (2007). Scratching the surface of skin development. Nature *445*, 834–842.

Fuchs, E. (2009). The Tortoise and the Hair: Slow-Cycling Cells in the Stem Cell Race. Cell *137*, 811–819.

Gat, U., DasGupta, R., Degenstein, L., and Fuchs, E. (1998). De Novo hair follicle morphogenesis and hair tumors in mice expressing a truncated beta-catenin in skin. Cell *95*, 605–614.

Genander, M., Cook, P.J., Ramsköld, D., Keyes, B.E., Mertz, A.F., Sandberg, R., and Fuchs, E. (2014). BMP Signaling and Its pSMAD1/5 Target Genes Differentially Regulate Hair Follicle Stem Cell Lineages. Cell Stem Cell *15*, 619–633.

Goldstein, J., Fletcher, S., Roth, E., Wu, C., Chun, A., and Horsley, V. (2014). Calcineurin/Nfatc1 signaling links skin stem cell quiescence to hormonal signaling during pregnancy and lactation. Genes & Development *28*, 983–994.

Greco, V., Chen, T., Rendl, M., Schober, M., Pasolli, H.A., Stokes, N., Cruz-Racelis, dela, J., and Fuchs, E. (2009). A Two-Step Mechanism for Stem Cell Activation during Hair Regeneration. Cell Stem Cell *4*, 155–169.

Hardy, M.H. (1992). The secret life of the hair follicle. Trends Genet. *8*, 55–61.

Hata, K., Takashima, R., Amano, K., Ono, K., Nakanishi, M., Yoshida, M., Wakabayashi, M., Matsuda, A., Maeda, Y., Suzuki, Y., et al. (2013). Arid5b facilitates chondrogenesis by recruiting the histone demethylase Phf2 to Sox9-regulated genes. Nature Communications *4*, 2850.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576–589.

Hinde, E., Haslam, I.S., Schneider, M.R., Langan, E.A., Kloepper, J.E., Schramm, C., Zouboulis, C.C., and Paus, R. (2013). A practical guide for the study of human and murine sebaceous glands in situ. Exp Dermatol *22*, 631–637.

Hobbs, R.M., and Pandolfi, P.P. (2010). Shape-shifting and tumor suppression by PLZF. Oncotarget *1*, 3–5.

Horsley, V., Aliprantis, A.O., Polak, L., Glimcher, L.H., and Fuchs, E. (2008). NFATc1 Balances Quiescence and Proliferation of Skin Stem Cells. Cell *132*, 299–310.

Horsley, V., O'Carroll, D., Tooze, R., Ohinata, Y., Saitou, M., Obukhanych, T., Nussenzweig, M., Tarakhovsky, A., and Fuchs, E. (2006). Blimp1 Defines a Progenitor Population that Governs Cellular Input to the Sebaceous Gland. Cell *126*, 597–609.

Hsu, Y.-C., Li, L., and Fuchs, E. (2014). Transit-Amplifying Cells Orchestrate Stem Cell Activity and Tissue Regeneration. Cell *157*, 935–949.

Hsu, Y.-C., Pasolli, H.A., and Fuchs, E. (2011). Dynamics between Stem Cells, Niche, and Progeny in the Hair Follicle. Cell *144*, 92–105.

Jaks, V., Barker, N., Kasper, M., van Es, J.H., Snippert, H.J., Clevers, H., and Toftgård, R. (2008). Lgr5 marks cycling, yet long-lived, hair follicle stem cells. Nat Genet *40*, 1291–1299.

Junttila, M.R., and de Sauvage, F.J. (2013). Influence of tumour micro-environment heterogeneity on therapeutic response. Nature *501*, 346–354.

Kadaja, M., Keyes, B.E., Lin, M., Pasolli, H.A., Genander, M., Polak, L., Stokes, N., Zheng, D., and Fuchs, E. (2014). SOX9: a stem cell transcriptional regulator of secreted niche signaling factors. Genes & Development *28*, 328–341.

Kandyba, E., Leung, Y., Chen, Y.-B., Widelitz, R., Chuong, C.-M., and Kobielak, K. (2013). Competitive balance of intrabulge BMP/Wnt signaling reveals a robust gene network ruling stem cell homeostasis and cyclic activation. Proc Natl Acad Sci USA

110, 1351–1356.

Kenchegowda, D., Swamynathan, S., Gupta, D., Wan, H., Whitsett, J., and Swamynathan, S.K. (2011). Conditional disruption of mouse Klf5 results in defective eyelids with malformed meibomian glands, abnormal cornea and loss of conjunctival goblet cells. Developmental Biology *356*, 5–18.

Keyes, B.E., Segal, J.P., Heller, E., Lien, W.-H., Chang, C.-Y., Guo, X., Oristian, D.S., Zheng, D., and Fuchs, E. (2013). Nfatc1 orchestrates aging in hair follicle stem cells. Proc Natl Acad Sci USA.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol *14*, R36.

Kimura-Ueki, M., Oda, Y., Oki, J., Komi-Kuramochi, A., Honda, E., Asada, M., Suzuki, M., and Imamura, T. (2012). Hair Cycle Resting Phase Is Regulated by Cyclic Epithelial FGF18 Signaling. J Invest Dermatol *132*, 1338–1345.

Kobielak, K., Stokes, N., la Cruz, de, J., Polak, L., and Fuchs, E. (2007). Loss of a quiescent niche but not follicle stem cells in the absence of bone morphogenetic protein signaling. Proc Natl Acad Sci USA *104*, 10063–10068.

Kume, T., Deng, K.Y., Winfrey, V., Gould, D.B., Walter, M.A., and Hogan, B.L. (1998). The forkhead/winged helix gene Mf1 is disrupted in the pleiotropic mouse mutation congenital hydrocephalus. Cell *93*, 985–996.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357–359.

Lapouge, G., Youssef, K.K., Vokaer, B., Achouri, Y., Michaux, C., Sotiropoulou, P.A., and Blanpain, C. (2011). Identifying the cellular origin of squamous skin tumors. Proc Natl Acad Sci USA *108*, 7431–7436.

Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretsky, I., Jaitin, D.A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S., et al. (2014). Chromatin state dynamics during blood formation. Science *345*, 943–949.

Leishman, E., Howard, J.M., Garcia, G.E., Miao, Q., Ku, A.T., Dekker, J.D., Tucker, H., and Nguyen, H. (2013). Foxp1 maintains hair follicle stem cell quiescence through regulation of Fgf18. Development *140*, 3809–3818.

Leushacke, M., and Barker, N. (2011). Lgr5 and Lgr6 as markers to study adult stem cell roles in self-renewal and cancer. Oncogene *31*, 3009–3022.

Li, L., and Clevers, H. (2010). Coexistence of Quiescent and Active Adult Stem Cells

in Mammals. Science *327*, 542–545.

Lien, W.-H., Guo, X., Polak, L., Lawton, L.N., Young, R.A., Zheng, D., and Fuchs, E. (2011). Genome-wide maps of histone modifications unwind in vivo chromatin states of the hair follicle lineage. Cell Stem Cell *9*, 219–232.

Lien, W.-H., Polak, L., Lin, M., Lay, K., Zheng, D., and Fuchs, E. (2014). In vivo transcriptional governance of hair follicle stem cells by canonical Wnt regulators. Nature Cell Biology *16*, 1–15.

Lopez, R.G., Garcia-Silva, S., Moore, S.J., Bereshchenko, O., Martinez-Cruz, A.B., Ermakova, O., Kurz, E., Paramio, J.M., and Nerlov, C. (2009). C/EBPalpha and beta couple interfollicular keratinocyte proliferation arrest to commitment and terminal differentiation. Nature Cell Biology *11*, 1181–1190.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 31.

Lowry, W.E., Blanpain, C., Nowak, J.A., Guasch, G., Lewis, L., and Fuchs, E. (2005). Defining the impact of beta-catenin/Tcf transactivation on epithelial stem cells. Genes & Development *19*, 1596–1611.

Lu, B., Rothnagel, J.A., Longley, M.A., Tsai, S.Y., and Roop, D.R. (1994). Differentiation-specific expression of human keratin 1 is mediated by a composite AP-1/steroid hormone element. J. Biol. Chem. *269*, 7443–7449.

Marusyk, A., Almendro, V., and Polyak, K. (2012). Intra-tumour heterogeneity: a looking glass for cancer? Nat Rev Cancer *12*, 323–334.

McKenzie, J.L., Gan, O.I., Doedens, M., Wang, J.C.Y., and Dick, J.E. (2006). Individual stem cells with highly variable proliferation and self-renewal properties comprise the human hematopoietic stem cell compartment. Nat. Immunol. *7*, 1225–1233.

Meacham, C.E., and Morrison, S.J. (2013). Tumour heterogeneity and cancer cell plasticity. Nature *501*, 328–337.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Mills, A.A., Zheng, B., Wang, X.J., Vogel, H., Roop, D.R., and Bradley, A. (1999). p63 is a p53 homologue required for limb and epidermal morphogenesis. Nature *398*, 708–713.

Milner, Y., Sudnik, J., Filippi, M., Kizoulis, M., Kashgarian, M., and Stenn, K. (2002). Exogen, shedding phase of the hair growth cycle: characterization of a mouse model. J Invest Dermatol *119*, 639–644.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631–642.

Morris, R.J., Liu, Y., Marles, L., Yang, Z., Trempus, C., Li, S., Lin, J.S., Sawicki, J.A., and Cotsarelis, G. (2004). Capturing and profiling adult hair follicle stem cells. Nature Biotechnology *22*, 411–417.

Müller-Röver, S., Handjiski, B., van der Veen, C., Eichmüller, S., Foitzik, K., McKay, I.A., Stenn, K.S., and Paus, R. (2001). A comprehensive guide for the accurate classification of murine hair follicles in distinct hair cycle stages. J Invest Dermatol *117*, 3–15.

Nguyen, H., Rendl, M., and Fuchs, E. (2006). Tcf3 governs stem cell features and represses cell fate determination in skin. Cell *127*, 171–183.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell *95*, 379–391.

Nowak, J.A., Polak, L., Pasolli, H.A., and Fuchs, E. (2008). Hair follicle stem cells are specified and function in early skin morphogenesis. Cell Stem Cell *3*, 33–43.

Omatsu, Y., Seike, M., Sugiyama, T., Kume, T., and Nagasawa, T. (2014). Foxc1 is a critical regulator of haematopoietic stem/progenitor cell niche formation. Nature *508*, 536–540.

Oshima, H., Rochat, A., Kedzia, C., Kobayashi, K., and Barrandon, Y. (2001). Morphogenesis and renewal of hair follicles from adult multipotent stem cells. Cell *104*, 233–245.

Oshimori, N., and Fuchs, E. (2012). Paracrine TGF-b Signaling Counterbalances BMP-Mediated Repressionin Hair Follicle Stem Cell Activation. Cell Stem Cell *10*, 63–75.

Paus, R., and Cotsarelis, G. (1999). The biology of hair follicles. N. Engl. J. Med. *341*, 491–497.

Plikus, M.V., and Chuong, C.M. (2014). Macroenvironmental Regulation of Hair Cycling and Collective Regenerative Behavior. Cold Spring Harbor Perspectives in

Medicine *4*, a015198–a015198.

Plikus, M.V., Baker, R.E., Chen, C.C., Fare, C., la Cruz, de, D., Andl, T., Maini, P.K., Millar, S.E., Widelitz, R., and Chuong, C.M. (2011). Self-Organizing and Stochastic Behaviors During the Regeneration of Hair Stem Cells. Science *332*, 586–589.

Plikus, M.V., and Chuong, C.-M. (2008). Complex Hair Cycle Domain Patterns and Regenerative Hair Waves in Living Rodents. J Invest Dermatol *128*, 1071–1080.

Plikus, M.V., Mayer, J.A., la Cruz, de, D., Baker, R.E., Maini, P.K., Maxson, R., and Chuong, C.-M. (2008). Cyclic dermal BMP signalling regulates stem cell activation during hair regeneration. Nature *451*, 340–344.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., et al. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. Cell *122*, 517–527.

Potten, C.S., Booth, C., and Pritchard, D.M. (1997). The intestinal epithelial stem cell: the mucosal governor. Int J Exp Pathol *78*, 219–243.

Potten, C.S., Hume, W.J., Reid, P., and Cairns, J. (1978). The segregation of DNA in epithelial stem cells. Cell *15*, 899–906.

Rhee, H., Polak, L., and Fuchs, E. (2006). Lhx2 maintains stem cell character in hair follicles. Science *312*, 1946–1949.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nature Biotechnology *29*, 24–26.

Rodgers, J.T., King, K.Y., Brett, J.O., Cromie, M.J., Charville, G.W., Maguire, K.K., Brunson, C., Mastey, N., Liu, L., Tsai, C.-R., et al. (2014). mTORC1 controls the adaptive transition of quiescent stem cells from G0 to G(Alert). Nature *510*, 393–396.

Rompolas, P., Deschene, E.R., Zito, G., Gonzalez, D.G., Saotome, I., Haberman, A.M., and Greco, V. (2012). Live imaging of stem cell and progeny behaviour in physiological hair-follicle regeneration. Nature *487*, 496–499.

Rompolas, P., Mesa, K.R., and Greco, V. (2014). Spatial organization within a niche as a determinant of stem-cell fate. Nature *502*, 513–518.

Schepeler, T., Page, M.E., and Jensen, K.B. (2014). Heterogeneity and plasticity of epidermal stem cells. Development *141*, 2559–2567.

Siegenthaler, J.A., Ashique, A.M., Zarbalis, K., Patterson, K.P., Hecht, J.H., Kane,

M.A., Folias, A.E., Choe, Y., May, S.R., Kume, T., et al. (2009). Retinoic Acid from the MeningesRegulates Cortical Neuron Generation. Cell *139*, 597–609.

Silva-Vargas, V., Celso, Lo, C., Giangreco, A., Ofstad, T., Prowse, D.M., Braun, K.M., and Watt, F.M. (2005). Beta-catenin and Hedgehog signal strength can specify number and location of hair follicles in adult epidermis without recruitment of bulge stem cells. Developmental Cell *9*, 121–131.

Slominski, A., and Paus, R. (1993). Melanogenesis is coupled to murine anagen: toward new concepts for the role of melanocytes and the regulation of melanogenesis in hair growth. J Invest Dermatol *101*, 90S–97S.

Snippert, H.J., Haegebarth, A., Kasper, M., Jaks, V., van Es, J.H., Barker, N., van de Wetering, M., van den Born, M., Begthel, H., Vries, R.G., et al. (2010). Lgr6 Marks Stem Cells in the Hair Follicle That Generate All Cell Lineages of the Skin. Science *327*, 1385–1389.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA *102*, 15545–15550.

Sur, I., Rozell, B., Jaks, V., Bergström, A., and Toftgård, R. (2006). Epidermal and craniofacial defects in mice overexpressing Klf5 in the basal layer of the epidermis. J. Cell. Sci. *119*, 3593–3601.

Tumbar, T., Guasch, G., Greco, V., Blanpain, C., Lowry, W.E., Rendl, M., and Fuchs, E. (2004). Defining the epithelial stem cell niche in skin. Science *303*, 359–363.

Vakoc, C.R., Sachdeva, M.M., Wang, H., and Blobel, G.A. (2006). Profile of histone lysine methylation across transcribed mammalian chromatin. Mol Cell Biol *26*, 9185–9195.

Van Mater, D., Kolligs, F.T., Dlugosz, A.A., and Fearon, E.R. (2003). Transient activation of beta -catenin signaling in cutaneous keratinocytes is sufficient to trigger the active growth phase of the hair cycle in mice. Genes & Development *17*, 1219–1224.

Venables, W.N., and Ripley, B.D. (2013). Modern Applied Statistics with S-PLUS (Springer Science & Business Media).

Vidal, V.P.I., Chaboissier, M.-C., Lützkendorf, S., Cotsarelis, G., Mill, P., Hui, C.-C., Ortonne, N., Ortonne, J.-P., and Schedl, A. (2005). Sox9 is essential for outer root sheath differentiation and the formation of the hair stem cell compartment. Current Biology *15*, 1340–1351.

Visvader, J.E. (2011). Cells of origin in cancer. Nature *469*, 314–322.

Waghmare, S.K., Bansal, R., Lee, J., Zhang, Y.V., McDermitt, D.J., and Tumbar, T. (2008). Quantitative proliferation dynamics and random chromosome segregation of hair follicle stem cells. Embo J. *27*, 1309–1320.

Wang, L., Dowell, R.D., and Yi, R. (2013). Genome-wide maps of polyadenylation reveal dynamic mRNA 3'-end formation in mammalian cell lineages. RNA.

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

Yang, A., Schweitzer, R., Sun, D., Kaghad, M., Walker, N., Bronson, R.T., Tabin, C., Sharpe, A., Caput, D., Crum, C., et al. (1999). p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. Nature *398*, 714–718.

Yi, R., O'Carroll, D., Pasolli, H.A., Zhang, Z., Dietrich, F.S., Tarakhovsky, A., and Fuchs, E. (2006). Morphogenesis in skin is governed by discrete sets of differentially expressed microRNAs. Nat Genet *38*, 356–362.

Zarbalis, K., Siegenthaler, J.A., Choe, Y., May, S.R., Peterson, A.S., and Pleasure, S.J. (2007). Cortical dysplasia and skull defects in mice with a Foxc1 allele reveal the role of meningeal differentiation in regulating cortical development. Proc Natl Acad Sci USA *104*, 14002–14007.

Zhang, Y.V., Cheong, J., Ciapurin, N., McDermitt, D.J., and Tumbar, T. (2009). Distinct Self-Renewal and Differentiation Phases in the Niche of Infrequently Dividing Hair Follicle Stem Cells. Cell Stem Cell *5*, 267–278.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.