# Genetic basis of human-specific skin characteristics in comparison with other primates

荒川　那海

博士（理学）

総合研究大学院大学
先導科学研究科
生命共生体進化学専攻

平成３０（２０１８）年度

**PhD Thesis**

# Genetic basis of human-specific skin characteristics in comparison with other primates

**Nami Arakawa**

**Doctor of Philosophy**

**Department of Evolutionary Studies of Biosystems**
**School of Advanced Sciences**
**SOKENDAI (The Graduate University for Advanced Studies)**
**Japan**

**2019**

## Summary

Skin is the largest organ of a body and located at interface between the inside and outside of an organism. It protects the inside of the body from external stresses, such as physical, chemical, and microbial insults. Skin phenotypes have evolved to protect the body and to allow the species to adapt to their habitat environments. Actually, human skin is morphologically and physiologically different from the skin of other primates. The reduced amount of hair and the high number of sweat glands are well-known examples of human-specific skin characteristics. It has been proposed that these human-specific characteristics allowed for efficient thermoregulation and adaptation to the savannah environment after our human ancestors abandoned the forest.

There are many phenotypic characteristics that are unique to the human lineage, including large brain size, bipedalism, and language development as well as skin characteristics. However, genetic causes that underlie human-specific characteristics remain poorly understood. Because skin phenotypes have evolved to protect the inside of the body and to adapt the species to their habitat environments, human-specific skin characteristics are likely to have significant roles in human evolution. In my PhD thesis, I therefore focused on human-specific skin characteristics and studied the genetic basis that underlies these characteristics.

Chapter 1 is a general introduction. I briefly explain the function of skin to protect the inside of the body from external environments. I also describe the main structure of mammalian skin including the epidermis, dermis, subcutaneous tissue, and epidermal basement membrane (BM) zone. Especially, I focus on the epidermal BM zone, which forms adhesion between the epidermis and dermis, for better understanding of the subsequent parts in this thesis. In addition,

I provide overviews of representative cases of the association between genetic elements and human-specific characteristics.

In chapter 2, I quantitatively distinguished histological skin differences between humans and other primates to investigate human-specific characteristics in skin structure. I found that the epidermis and dermis in human skin were significantly thicker than those in the three Old World monkey species examined. I also indicated that the epidermal BM zone topography in humans was undulating, which is known as a rete ridge, while that in the three Old World monkey species was flat. These results, together with previous qualitative studies, suggest that the thicker epidermis and rete ridge may be human-specific skin characteristics, although additional quantitative histological comparison between human and great ape skin is required.

In chapter 3, I then comprehensively compared gene expression levels between human and great ape skin using next-generation cDNA sequencing (RNA-seq) to investigate genes associated with human-specific skin characteristics. I found that the expression levels of four structural protein genes, *biglycan* (*BGN*), *collagen type XVIII alpha 1 chain* (*COL18A1*), *CD151 molecule* (*CD151*), and *laminin subunit beta 2* (*LAMB2*), in skin were significantly higher in humans than in great apes. *COL18A1*, *LAMB2*, and *CD151* are genes that encode proteins structurally associated with the epidermal BM zone. BGN regulates the formation of elastin, which is one of the components of elastic fibers in the dermis. According to previous studies of qualitative histological comparison between human and other primate skin, an abundance of elastic fibers seems to be human-specific skin characteristics. The human-specific expression patterns of the four structural protein genes identified may contribute to the rete ridge formation and rich elastic fibers in human skin.

Humans have a low amount of hair on their body compared with other primates, which gives humans a high level of thermoregulation. However, it is believed that human skin has lost the ability to protect the internal tissues from external physical stresses by hair. Compared to flat topography of the epidermal BM zone, a rete ridge increases the area where the epidermis and dermis connect, which may make strong adhesion between these two layers. The rete ridge, thick epidermis, and rich elastic fibers might contribute to physical strength of human skin. Although additional quantitative histological comparison between human and great ape skin is required to clarify human-specific skin characteristics, the human-specific expression patterns found in this chapter may contribute to adaptive skin characteristics specific to humans with less hair.

In chapter 4, I inferred substitutions responsible for the human-specific expression patterns of the four structural protein genes (*COL18A1*, *LAMB2*, *CD151*, and *BGN*) in their transcriptional regulatory regions. I first estimated transcriptional regulatory regions for each gene by identifying conserved noncoding regions around the genes with taking histone modifications for active regulatory regions in skin cells into consideration. The human-specific substitutions in putative transcriptional regulatory regions were estimated to be candidate substitutions responsible for the human-specific expression patterns in the genes of interest, resulting in two to ten candidate substitutions for each of the genes. These candidate substitutions, especially those located in the expected binding sites of transcription factors functioning in skin, may give humans adaptive skin characteristics through human-specific gene expression patterns.

Chapter 5 is a general discussion. I suggest that the candidate substitutions in the putative transcriptional regulatory regions inferred may cause the human-specific gene expression patterns that possibly lead to adaptive skin characteristics specific to humans with less hair. In the

near future, I am planning to conduct a promoter assay in cultured skin cells to examine whether these candidate substitutions are responsible for the expression differences between humans and great apes. In addition, for the candidate substitutions with expression changes in promoter assay, I will investigate whether these substitutions influence the expression of the genes of interest, but not other genes, in cultured skin cells by genome editing technique using CRISPR/Cas9 system. Identifying substitutions that may give humans adaptive skin characteristics through human-specific gene expression patterns will contribute to the understanding of how human-specific characteristics have been genetically acquired. Finally, I hope that my PhD study will provide further insight into the human evolution.

## Contents

# Chapter 1

General introduction

Skin is the largest organ of a body and located at interface between the inside and outside of an organism. The primary role of the skin is to form an effective barrier between the inside of a body and external environments. We can divide skin barrier function into two types: the outside-in barrier and inside-out barrier (Niehues, et al. 2018). From the outside-in, the skin protects the body against physical stress, chemical and ultraviolet (UV) exposure, and invasion of pathogens. From the inside-out, penetration of excessive water and electrolytes is prevented to protect the body from dehydration.

Human skin is covered with diverse communities of microorganisms that consist of bacteria, fungi, mites, and viruses (Weyrich, et al. 2015). Recent studies reported that the commensal skin microorganisms function as a barrier that protects against more pathogenic and harmful organisms (Niehues, et al. 2018; Scharschmidt and Fischbach 2013; Weyrich, et al. 2015). Disruptions of skin microbial community balance have been linked to skin diseases (Grice and Segre 2011; Weyrich, et al. 2015). For example, Firmicute species and Actinobacteria were significantly overrepresented and underrepresented in psoriatic skin, respectively, compared to normal skin from both healthy persons and the patients with psoriasis, suggesting that psoriasis is associated with changes of cutaneous bacterial composition (Gao, et al. 2008).

Skin phenotypes have evolved to protect the inside of a body and to adapt the species to their habitat environments. For example, fish skin is equipped with slimy, slippery mucus throughout the body. The mucus contains numerous antibacterial factors and protects fish from invading pathogens (Benhamed, et al. 2014). Fish are constantly in contact with their aquatic environment, which includes a multitude of diverse pathogenic organisms (Dash, et al. 2018). Therefore, the skin mucus plays an essential defensive role for fish to survive in their aquatic

environment (Benhamed, et al. 2014; Dash, et al. 2018). In other examples, humans show adaptive skin pigmentation polymorphisms that highly correlate with levels of UV radiation across the latitude; human populations near the equator generally have dark skin to protect against UV exposure, whereas those at high latitude have lighter skin to maintain cutaneous photosynthesis of vitamin D, a substance with manifold biological significance, under low-sunlight environments (Deng and Xu 2018; Jablonski and Chaplin 2010).

Mammalian skin is generally composed of three primary layers: the epidermis, dermis, and subcutaneous tissue (Smoller 2009; Watt and Fujiwara 2011). The epidermal basement membrane (BM) zone forms adhesion between the epidermal underside and the dermal upside (Has and Nystrom 2015). In addition, skin-associated structures, including nails, hair follicles, sweat glands, and sebaceous glands, are known as skin appendages. They have physically and physiologically important functions such as protection, perspiration, and thermoregulation. Here, I provide overviews of the three skin layers and epidermal BM zone.

*The epidermis*

The epidermis is a stratified squamous epithelium that is located at the most superficial layer of the skin. The epidermis primarily (greater than 90 %) consists of keratinocytes (Smoller 2009), and is divided into four cell layers from the bottom to the top: the basal layer, spinous layer, granular layer, and stratum corneum (Figure 1-1) (Fuchs and Raghavan 2002). The monolayered basal layer contains proliferative stem cells of keratinocyte, and progeny of those withdraws from the cell cycle after a limited number of divisions (Watt 1998). Keratinocytes then leave the basal layer to undergo terminal differentiation as they migrate towards the skin surface. Cells that reach

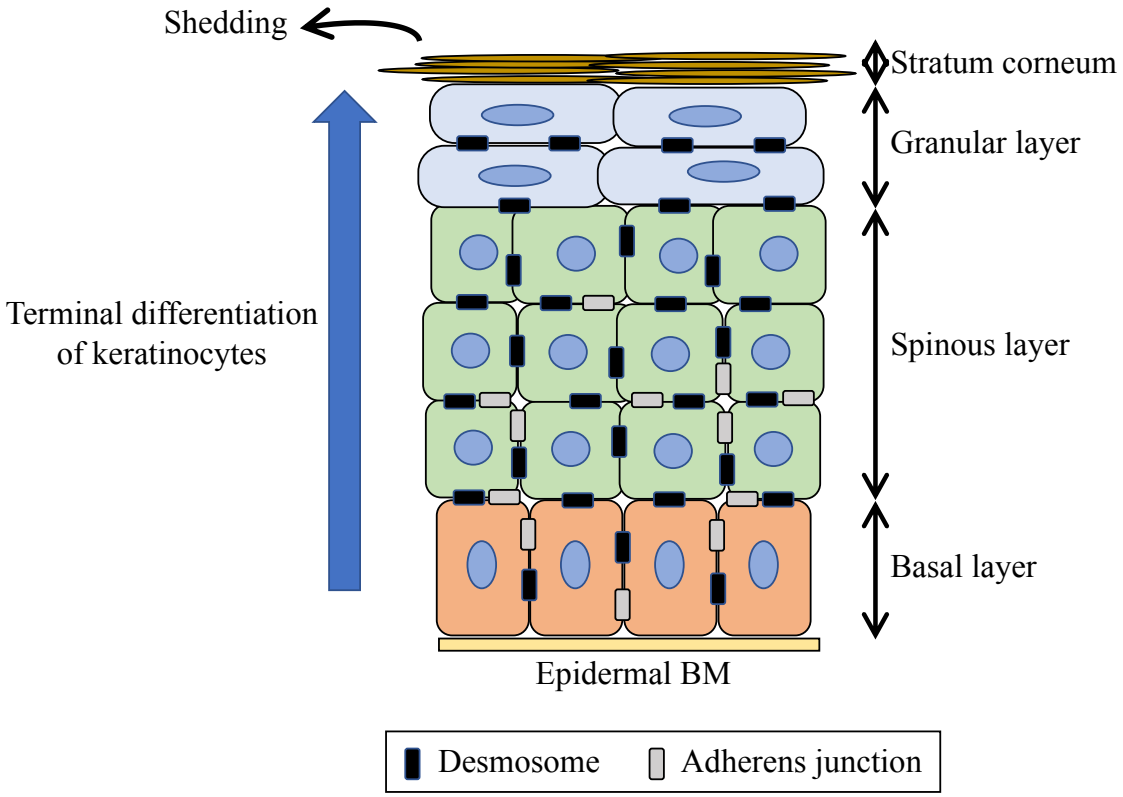


**Figure 1-1** A schematic representation of the epidermal layers and terminal differentiation of keratinocytes.

the stratum corneum are now called corneocytes, which have become dead, enucleated, and flattened, and subsequently shed from the skin surface. Intercellular adhesion between keratinocytes within nucleated epidermal layers is formed by junction structures called the desmosome and adherens junction (Fuchs and Raghavan 2002). The architecture of tightly adhering cells in the epidermis contributes to its protective function against external physical stresses (Fuchs and Raghavan 2002; Proksch, et al. 2008).

*The dermis*

The dermis is a supportive connective tissue enriched with extracellular matrix including collagens and elastic fibers. Collagens are the most abundant structural constituent of extracellular matrix in the dermis (Smoller 2009) and provide tissues with tensile strength (Rozario and DeSimone 2010). Elastic fibers give skin elasticity (Kielty, et al. 2002). Structural components of these two kinds of extracellular matrix are produced by fibroblasts, which are one of the cell types existing in the dermis (Smoller 2009). The dermis with extracellular matrix contributes to the mechanical strength of the skin (Has and Nystrom 2015; Wong, et al. 2016).

*The subcutaneous tissue*

The subcutaneous tissue is an inmost layer of the skin mainly composed of adipocytes (Ali, et al. 2013). This tissue protects the inside of the body from physical stresses and exposure to heat and cold, and also plays important roles in energy metabolism and storage (Baroni, et al. 2012).

*The epidermal BM zone*

In the interface between the epidermis and the dermis there is the epidermal BM, which is formed of extracellular matrix such as laminins and collagens IV and XVIII (Has and Nystrom 2015). It provides adhesion between the two layers by collaborating with other anchoring structures (Figure 1-2). The hemidesmosomes are supramolecular assemblies that form adhesion between the basal keratinocytes and the epidermal BM. They are composed of two plaque structures, an inner and an outer plaque (Darling, et al. 2013). The hemidesmosomal outer plaques are placed on inside surface of the basolateral cell membrane of basal keratinocytes, and contain the transmembrane proteins α6β4 integrin, collagen XVII, and CD151 (Goletz, et al. 2017; Walko, et al. 2015). The cytoplasmic domains of α6β4 integrin and collagen XVII are linked to two inner plaque proteins, plectin and BP230, which in turn connect with the keratin intermediate filament network of basal keratinocytes (Walko, et al. 2015). α6β4 integrin is heterodimeric cell surface adhesion receptor that preferentially binds to laminin-332 by interacting with CD151 molecule (Goletz, et al. 2017; Has and Nystrom 2015). Laminin-332 is a major component of the epidermal BM (Sugawara, et al. 2008) and consists of three distinct polypeptide chains of laminin subunit α3, β3, and γ2 (encoded by *LAMA3*, *LAMB3*, and *LAMC2* genes, respectively) (Yao 2017). Collagen XVII inserts into the epidermal BM with its extracellular domain and, together with α6β4 integrin-laminin-332-CD151 anchoring, mediates basal keratinocyte adhesion (Masunaga, et al. 1997). In addition, from the epidermal BM, anchoring fibrils extend into the upper part of the dermis (Bruckner-Tuderman and Has 2014). The distal ends of anchoring fibrils are integrated into extracellular matrix networks of the dermis or kink back to the epidermal BM, which attaches the epidermal BM to the dermis. Anchoring fibrils are mainly composed of collagen VII (Sakai,

et al. 1986). A highly specialized orchestrated continuum that forms adhesion between the epidermis and the dermis, namely, the hemidesmosomes for linking basal keratinocytes to the epidermal BM, the epidermal BM, and the anchoring fibrils extending from the epidermal BM into the upper dermis, is called the epidermal BM zone (Bruckner-Tuderman and Has 2014).

Collagens are one of the essential structural components for the epidermal BM zone. The collagen family is comprised of 28 members in vertebrates numbered with Roman numerals I-XXVIII (Ricard-Blum 2011). A collagen molecule contains three polypeptide α chains, and for the majority of the collagen family members, those are identical chains (Kadler, et al. 2007). For example, collagens VII, XVII, and XVIII are homotrimers of their respective α1 chains (encoded by *COL7A1*, *COL17A1*, and *COL18A1* genes, respectively). On the other hand, collagen IV is heterotrimer of six genetically different α chains (Ricard-Blum 2011). In the epidermal BM, collagen IV is formed by two α1 chains and one α2 chain (encoded by *COL4A1* and *COL4A2* genes, respectively) with some additional distribution of collagen IV consisting of two α5 chains and one α6 chain (encoded by *COL4A5* and *COL4A6* genes, respectively) (Has and Nystrom 2015; Hasegawa, et al. 2007).

Humans and their closest relatives, great apes (chimpanzees, gorillas, and orangutans), diverged from Old World monkey species including the macaques approximately 30.5-32.5 million years ago (Mya). The divergence of the orangutan, gorilla, and chimpanzee lineages then occurred about 14.5-18.3, 8.3-10.6, and 4.1-7.7 Mya, respectively (Das, et al. 2014; Hobolth, et al. 2007; Pozzi, et al. 2014; Steiper and Young 2006). After the divergence of the human and chimpanzee lineages, humans have evolved morphologically and physiologically unique skin

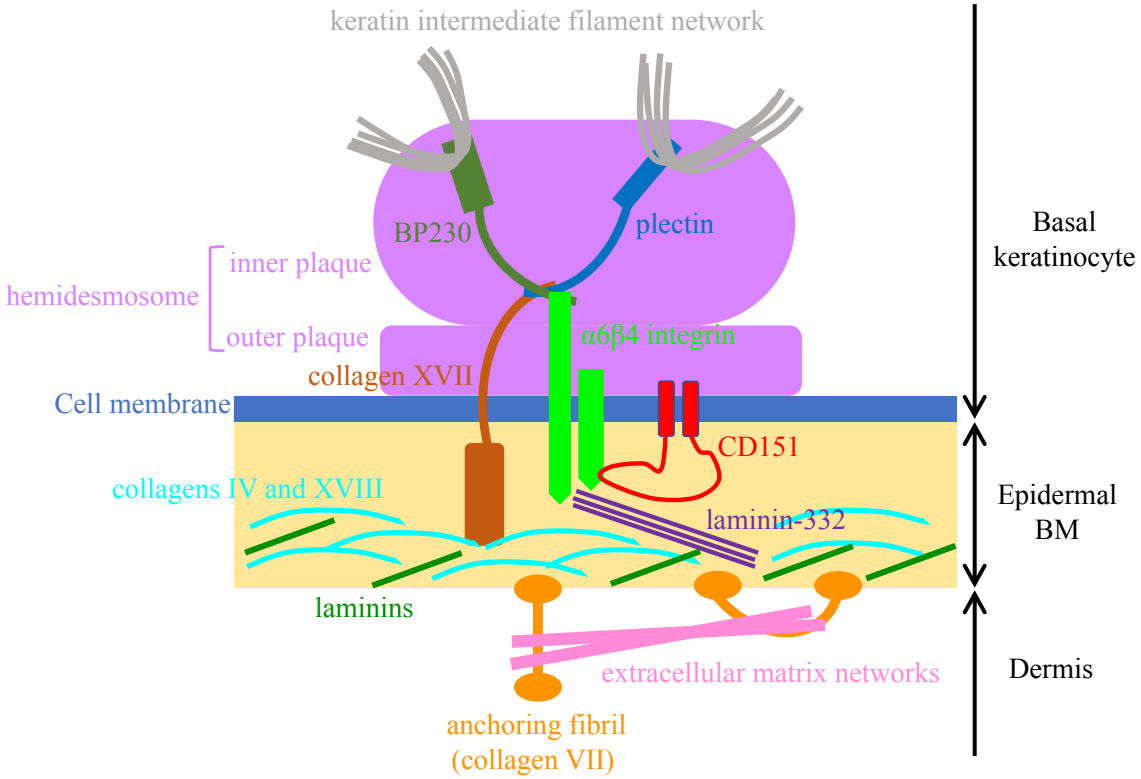


**Figure 1-2** A schematic representation of the epidermal BM zone.

characteristics that are not found in other primates. The reduced amount of hair is one of the most striking human-specific skin characteristics (Dávid-Barrett and Dunbar 2016). Although the number and density of hair follicles in humans are not significantly different from those of great apes, most of human hairs are extremely fine, short vellus except for hairs on the scalp, axilla, and pubic regions. Such miniscule hairs have lost the ability of protecting the skin from physical insults and UV radiation, which is one of the important functions of mammalian hairs as well as thermal insulation, visual communication, and camouflage (Lieberman 2015; Rantala 2007). The high number of eccrine sweat glands is also well-known human-specific skin characteristic (Folk and Semken 1991). The body is cooled through heat loss when water secreted by eccrine sweat glands vaporizes from the skin surface. In combination with the reduced amount of hair, which promotes water evaporation on the skin surface via increasing air convection, human skin exhibits efficient thermoregulation (Lieberman 2015). It has been generally thought that these human-specific skin characteristics evolved to allow for adaptation to the hot savannah environment after our human ancestors abandoned the shady forest (Folk and Semken 1991).

There are many phenotypic characteristics that are unique to the human lineage, including large brain size, bipedalism, and language development (Carroll 2003) as well as skin characteristics described above. However, the genetic causes that underlie human-specific characteristics remain poorly understood, except a few known cases. Here, I provide overviews of representative cases of the association between genetic elements and human-specific characteristics.

*Forkhead box P2 (FOXP2) and language development*

The transcription factor *FOXP2* is a gene implicated in the human ability of language. A patient with a mutant *FOXP2* gene exhibits a speech and language disorder (Lai, et al. 2001). The human *FOXP2* amino acid sequence is different at two residues from the identical homologous sequences of the chimpanzee and gorilla (Enard, et al. 2002), and these two replacements alter FOXP2 function to lead to human-specific transcriptional regulation of genes that may be involved in language development (Konopka, et al. 2009). *FOXP2* is therefore thought to have a functionally important role for establishing language in humans (Mitchell and Silver 2018).

*Human-accelerated conserved noncoding sequence 1 (HACNS1) and limb patterning*

The HACNS1 is a sequence element that may have contributed to the evolution of human limb characters (Douglas and Hill 2014). Although the orthologous regions in vertebrates are highly conserved, the human sequence shows extremely rapid evolution with 13 nucleotide substitutions in a short 81-bp region of HACNS1, exceeding neutral evolutionary rate. In transgenic mouse embryo experiments, these nucleotide changes were shown to provide the human-specific enhancer activity during limb development that was not detected in the orthologous sequences from the chimpanzee and rhesus macaque (Prabhakar, et al. 2008). Subsequent studies found that human HACNS1 lacking the 81-bp with the clustered substitutions retained enhancer activity similar to the full-length human HACNS1. This result suggests that the 13 nucleotide substitutions in the human lineage inactivate repressor activity of the 81-bp element observed in the other primates examined, rather than bring novel enhancer activity to the sequence

element (Sumiyama and Saitou 2011). It has been proposed that loss of function in the HACNS1 81-bp region may have led to human-specific digit and limb patterning, such as the opposable thumb, through altering the expression of genes associated with limb development (Prabhakar, et al. 2008; Sumiyama and Saitou 2011).

*Myosin heavy chain 16 (MYH16) and masticatory muscle size*

The *MYH16* is also one of the genes likely involved in the acquisition of human-specific characteristics. MYH16 is present in masticatory muscles in non-human primates, whereas human *MYH16* is pseudogenized by a two-nucleotide frameshifting deletion and does not produce functional protein (Stedman, et al. 2004). Since humans exhibit considerably smaller masticatory muscles compared with most other primates, the loss of function in human *MYH16* has been proposed to be associated with the striking difference in masticatory muscle size. In addition, it has been also suggested that the size reduction of masticatory muscles might have allowed the cranium to expand for encephalization in humans (Stedman, et al. 2004; Vallender, et al. 2008). Further studies, however, indicated that pseudogenization of the *MYH16* gene occurred approximately 5.3 Mya, which is significantly older than the first emergence of masticatory muscle reduction and large brain size during hominid evolution (Perry, et al. 2004). Although the loss of function in the *MYH16* gene might not have directly led to these two human-specific characteristics, it is a notable case that shows the genetic association with phenotypic characteristics unique to humans (Oh, et al. 2015).

In my PhD thesis, I focused on human-specific skin characteristics and studied the genetic basis that underlies those phenotypes. In chapter 2, I quantitatively distinguished histological skin differences between humans and other primates to investigate human-specific characteristics in skin structure. In chapter 3, I then comprehensively compared gene expression levels between human and great ape skin using next-generation cDNA sequencing (RNA-seq). I identified genes with human-specific expression patterns that may be related to human-specific characteristics in skin structure. In chapter 4, I identified possible transcriptional regulatory regions and DNA sequence substitutions likely responsible for the human-specific expression patterns of the genes. I hope that my PhD study will provide insight into the evolution of human-specific skin characteristics.

# Chapter 2

## Histological differences between human and other primate skin

## Introduction

Skin is an important organ that is constantly exposed to external environments. It protects the inside of a body from external stresses, such as physical, chemical, and microbial insults. It is likely that skin phenotypes have evolved to protect the inside of a body in species adapted to external environments such as terrestrial amniotes, including humans.

Human skin is morphologically and physiologically different from the skin of other primates. For example, the reduced amount of hair (Dávid-Barrett and Dunbar 2016) and the high number of sweat glands (Folk and Semken 1991) are examples of human-specific skin characteristics that are not found in other primates such as chimpanzees, gorillas, and orangutans. It has been proposed that these human-specific characteristics allowed for efficient thermoregulation and adaptation to the savannah environment after our human ancestors abandoned the forest (Folk and Semken 1991).

Skin is generally composed of three layers: the epidermis, dermis, and subcutaneous tissue (Smoller 2009). The epidermal BM zone forms adhesion between the epidermal underside and the dermal upside through anchoring structures (Has and Nystrom 2015). Humans seem likely to have unique characteristics also for skin structures. Until the early 1980s, the histological differences between human and other primate skin were reported, but solely from a qualitative perspective. For example, it was described that the human epidermis is thicker than that of other primates, although without quantitative analysis (Montagna 1982; Montagna 1985).

In addition, these qualitative studies also focused on the epidermal BM zone. The epidermal underside and dermal upside (the position of the epidermal BM zone) in the furred skin

of most non-human primates are described as flat (Montagna 1982). On the other hand, those in human skin, including those in the hairy skin of the scalp, are strongly sculptured and penetrate each other (Montagna 1982; Montagna 1985), resulting in undulating topography of the epidermal BM zone known as a rete ridge. In the furred skin of chimpanzees and gorillas, the epidermal underside has been reported with inconsistent descriptions; a degree of sculpturing in chimpanzees and gorillas (Montagna 1982), discrete and moderate sculpturing in chimpanzees (Montagna and Yun 1963), and a nearly flat topography in gorillas (Ellis and Montagna 1962).

Another striking difference between human and other primate skin described is the amount of elastic fibers, which give skin elasticity (Kielty, et al. 2002). Elastic fibers are rich in the human dermis but those in most other primates are not as numerous as humans (Montagna 1982; Montagna 1985). Also in this case, the amount of elastic fibers in the furred skin of chimpanzees and gorillas has not been reported consistently; while it was reported to be similar to the content in humans (Montagna 1982; Montagna 1985), the elastic fibers in the chimpanzee dermis were also described as nowhere numerous (Montagna and Yun 1963).

According to these qualitative studies, a thick epidermis, extensive rete ridge formation, and an abundance of elastic fibers seem to be human-specific skin characteristics. However, to the best of our knowledge, no recent study has quantitatively compared the characteristics of human and other primate skin. In this chapter, I therefore measured epidermis and dermis thickness and statistically compared them between humans and three Old World monkey species. I also provide skin section photographs to show the difference in epidermal BM zone topography between humans and the three Old World monkey species. These two analyses will help to reveal human-specific characteristics in skin structure.

## Materials and Methods

### Skin specimens

The use of human skin tissues was authorized by the Ethics Committee of the University of the Ryukyus for Medical and Health Research Involving Human Subjects (#18-1295). The research using Old World monkey [anubis baboons (*Papio anubis*), Sykes' monkeys (*Cercopithecus albogularis*), vervet monkeys (*Chlorocebus pygerythrus*)] skin tissues was approved by the Institutional Review committee of the Institute of Primate Research, National Museum of Kenya (No. IRC/05/14). The dorsal skin tissues of the three species were collected by Dr. Akiko Matsumoto-Oda under research permission (No. NCST/RRI/12/1/BS/240), and transferred to Dr. Kenzo Takahashi and Dr. Daisuke Utsumi (the University of the Ryukyus) under the regulation of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES: No. 0830732) (see acknowledgments).

### Measurement of skin thickness

Digital photographs of dissected skins of anubis baboons (n=6), Sykes' monkeys (n=6), vervet monkeys (n=6), and humans (n=4) were provided by the Department of Dermatology, Graduate School of Medicine, University of the Ryukyus. Epidermis and dermis thickness were measured at ten sites with regular intervals on dissected skin photographs for each individual using iViewer version 5.5.7 (http://www.claro.jp). Measurements that were unreliable due to skin condition were discarded. The average values of the ten measurements for epidermis and dermis thickness were calculated for each individual. Those values were then used to calculate the average thickness in the species. I compared the average thickness between humans and each of

the three Old World monkey species using a t-test with Bonferroni correction.

## Results

To clarify the human-specific characteristics in skin structure, I measured the thickness of the epidermis and dermis in humans and three Old World monkey species, anubis baboons, Sykes' monkeys, and vervet monkeys. I also observed epidermal BM zone topography in the skin of humans and the three Old World monkey species. The results are as follows.

*Epidermis and dermis thickness*

The average thickness of the epidermis was 39.15 ± 6.2, 23.23 ± 5.7, 27.31 ± 2.2, and 105.14 ± 14.6 μm in anubis baboons, Sykes' monkeys, vervet monkeys, and humans, respectively (Figure 2-1a). The epidermis was significantly thicker in humans than in the three Old World monkey species ($p < 0.05$, t-test with Bonferroni correction).

The average thickness of the dermis was 1098.23 ± 322.5, 760.18 ± 78.4, 669.79 ± 79.2, and 4131.87 ± 1155.1 μm in anubis baboons, Sykes' monkeys, vervet monkeys, and humans, respectively (Figure 2-1b). The dermis was significantly thicker in humans than in the three Old World monkey species ($p < 0.05$, t-test with Bonferroni correction).

*Epidermal BM zone topography*

I observed that the epidermal BM zone topography in human skin was undulating (i.e., showed a rete ridge) (Figure 2-1c). On the other hand, that in the three Old World monkey species was flat (Figure 2-1d-f).

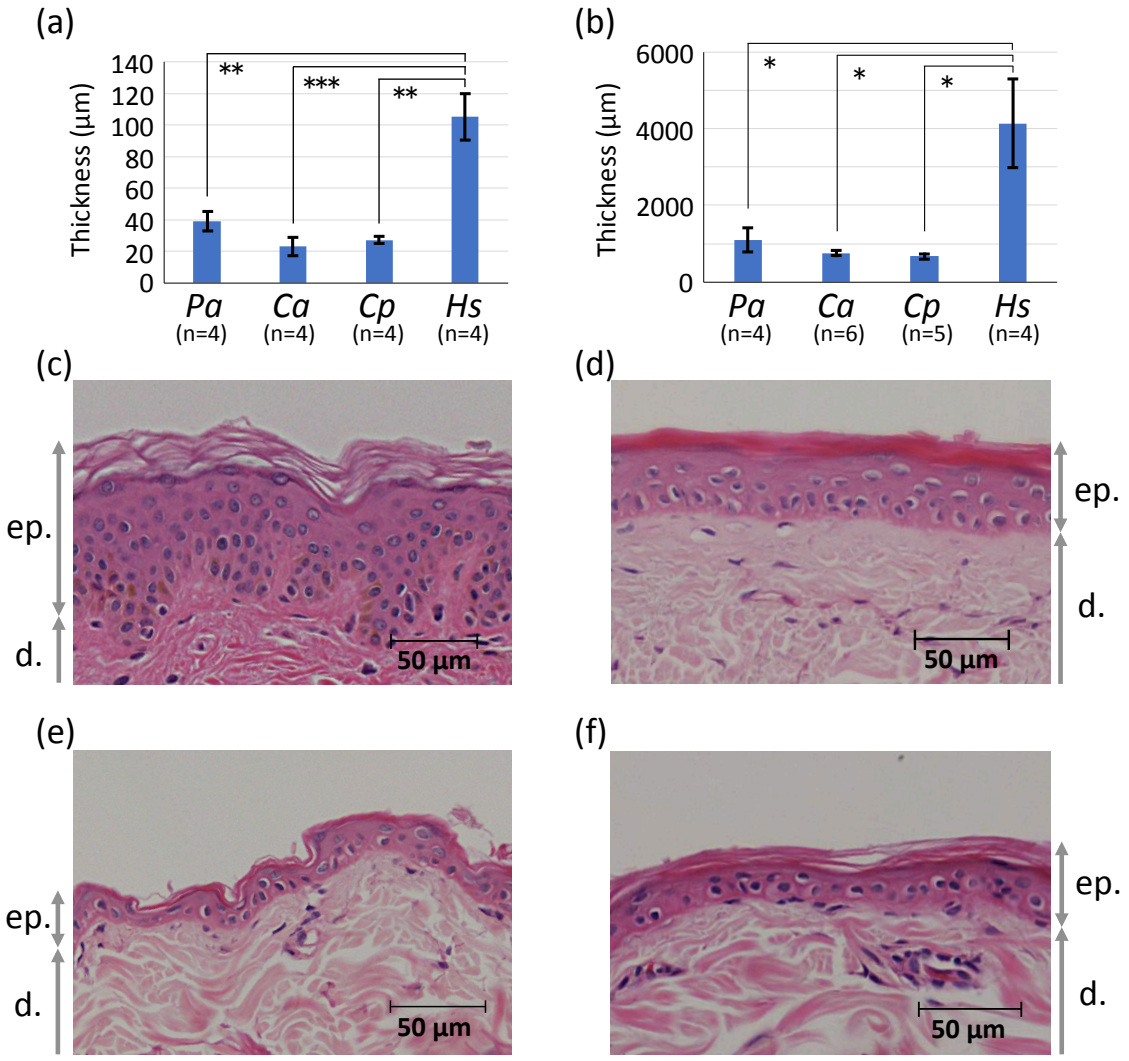


**Figure 2-1** The thickness and epidermal BM zone topography of skin in humans and three Old World monkey species.

The comparison of thickness of the epidermis (a) and dermis (b). *Pa*, *Ca*, *Cp*, and *Hs* indicate *Papio anubis*, *Cercopithecus albogularis*, *Chlorocebus pygerythrus*, and *Homo sapiens*, respectively. The numbers of individuals used for measurements are shown under each species name abbreviation. Photographs of hematoxylin–eosin stained histologic skin sections are from *H. sapiens* (c), *P. anubis* (d), *C. albogularis* (e), and *C. pygerythrus* (f). Scale bars are shown in each panel. ep. and d. indicate the epidermis and dermis, respectively. $^{*}$: $p < 0.05$, $^{**}$: $p < 0.01$, $^{***}$: $p < 0.001$, t-test with Bonferroni correction.

## Discussion

More than three decades ago, histological differences between human and other primate skin were qualitatively described (Montagna 1982; Montagna 1985). However, to the best of our knowledge, there is no report that quantifies the differences in skin structure between humans and other primates. In this study, I quantified two of the primary skin differences between humans and three Old World monkey species. One of the primary skin differences is that the epidermis and dermis in human skin are significantly thicker than those of the three Old World monkey species investigated. The other difference is that the epidermal BM zone topography shows a rete ridge in humans but is flat in the three Old World monkey species.

To definitively clarify the human-specific characteristics in skin structure, it would be logical to compare human skin with closely related species. However, I could not investigate the histological traits in great ape skin due to the absence of dissected skin digital photographs. Previous studies report that the epidermis in humans is thicker than that in other primates including great apes, although without quantitative analysis (Montagna 1982; Montagna 1985). In most non-human primates, the epidermal underside (the position of the epidermal BM zone) of the furred skin is described as flat (Montagna 1982). Those in chimpanzees and gorillas are reported with inconsistent descriptions; a degree of sculpturing (Montagna 1982), discrete and moderate sculpturing (Montagna and Yun 1963), and a nearly flat topography (Ellis and Montagna 1962). Thus, it is assumed that the thicker epidermis and strongly sculptured epidermal underside (i.e., rete ridge) may be human-specific skin characteristics. To fully clarify these points, additional quantitative comparisons between human and great ape skin are required.

In the near future, skin specimens of great apes will be available thanks to an offer

from my research collaborators. I am planning to measure thickness of the epidermis and dermis and observe epidermal BM zone topography in those great ape specimens. Histological comparison between human and great ape skin, together with the results shown in this chapter, will definitively clarify the human-specific characteristics in skin structure.

# Chapter 3

## Gene expression differences between human and great ape skin

## Introduction

The genetic causes that underlie human-specific characteristics remain poorly understood, except a few known cases (Enard, et al. 2002; Prabhakar, et al. 2008; Stedman, et al. 2004). Because skin phenotypes have evolved to protect the inside of a body and to adapt the species to their habitat environments, human-specific skin characteristics including those discussed in chapter 2 likely have significant roles in human evolution. Therefore, I focused on human-specific skin characteristics and investigated the genetic causes responsible for these characteristics.

It is widely accepted that most of the phenotypic differences observed between closely related species are a result of different quantitative and spatiotemporal expression patterns in functionally relevant genes, rather than amino acid differences in the protein-coding regions of those genes (Carroll 2008; Wray 2007). One of the famous papers that support this hypothesis was published by King and Wilson in 1975 (King and Wilson 1975). They showed that human protein sequences are nearly identical to homologous sequences in chimpanzees and argued that the small degree of these protein changes cannot account for the substantial phenotypic differences between the two species. They therefore proposed that changes in gene expression regulation are more likely to be responsible for the majority of biological differences between these closely related species.

In fact, human brain exhibited differential gene expression patterns compared with other primates, and approximately 90 % of the differentially expressed genes were up-regulated in humans (Cáceres, et al. 2003). Many of the genes highly expressed in human brain were related to a variety of brain functions, such as neuronal activity and physiological processes. It has been

suggested that the elevated gene expression patterns could provide the basis for extensive changes of brain function in humans, including human-specific high ability of cognition (Cáceres, et al. 2003; Naumova, et al. 2013).

RNA-seq is a recently developed method to simultaneously measure gene expression levels on a genome-wide scale using next-generation sequencing technologies (Wang, et al. 2009). The results of RNA-seq are highly reproducible (Marioni, et al. 2008) and enable us to compare expression levels for each gene between multiple RNA samples. Using this technique, we can comprehensively clarify gene expression differences between species that may contribute to their phenotypic differences (Romero, et al. 2012).

The typical workflow of an RNA-seq experiment is as follows (Figure 3-1) (Wang, et al. 2009). First, the mRNAs from tissue or cell samples are fragmented and converted into cDNAs. cDNAs are then amplified, and populations of the generated products are called libraries. cDNA libraries are subjected to next-generation sequencing to determine short sequences of one end or both ends of each cDNA molecule. The resulting sequences, called reads, are aligned (i.e., mapped) to corresponding exonic regions of genes in a reference genome, which provides quantification of expression levels for each gene in the sample.

In this chapter, to reveal genes with human-specific increased or decreased expression patterns, I comprehensively compared gene expression levels between human and great ape skin by RNA-seq analyses. I then discussed association between the human-specific gene expression patterns identified and human-specific skin characteristics.

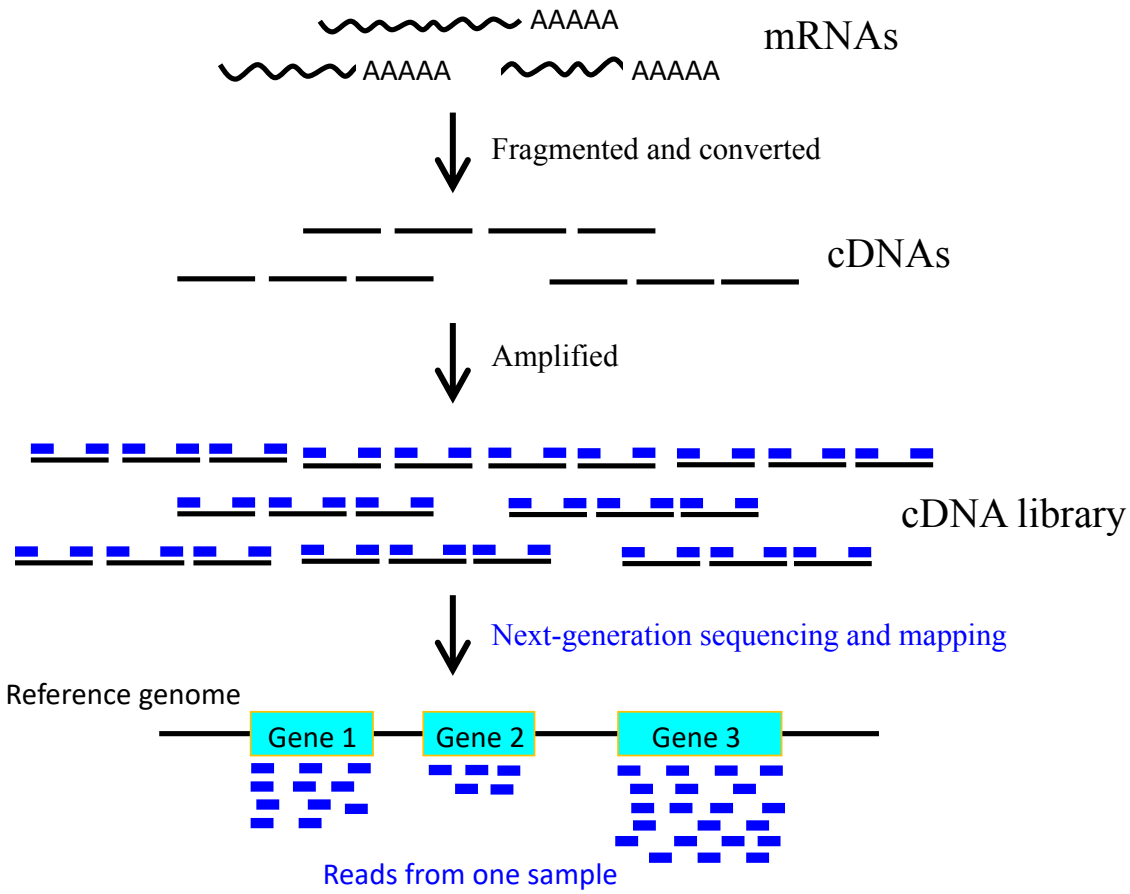


**Figure3-1** Typical workflow of an RNA-seq experiment.

## Materials and Methods

### Skin specimens

The skin tissue specimens from chimpanzees (*Pan troglodytes verus*), gorillas (*Gorilla gorilla gorilla*), and orangutans (*Pongo pygmaeus*) (n=3 for each species) (Table 3-1) were collected by the Primate Research Institute, Kyoto University via the Great Ape Information Network (GAIN) from zoos and the Kumamoto sanctuary, Wildlife Research Institute, Kyoto University.

### RNA extraction and sequencing

Total RNA was extracted from skin tissue samples of chimpanzees, gorillas, and orangutans using TRIzol reagent (Thermo Fisher Scientific, Waltham, MA, USA). Human skin total RNA of five individuals was obtained from commercial sources, and these individuals were not the same as ones used for measurement of skin thickness (Total RNA: BioChain, Newark, CA, USA; MVP Total RNA, Human Skin: Agilent Technologies, Santa Clara, CA, USA; Table 3-1). Skin total RNA was used to construct libraries for high-throughput next-generation sequencing using the NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA). Short cDNA sequences were determined from the libraries using the Illumina HiSeq2000 (paired-end, 100 bp) or HiSeq2500 (paired-end, 125 bp) platform.

### Comparison of RNA expression in skin

The procedure to compare skin RNA expression patterns between humans and great apes is shown in Figure 3-2. Sequenced reads from all libraries were mapped to each of the four

**Table 3-1** Human skin total RNA samples and great ape skin tissue samples used for RNA-seq

| Sample ID | Species | Age (years) | Sex* | Ethnicity | Body part |
|---|---|---|---|---|---|
| Human 1[a)] | *Homo sapiens* | 83 | F | Caucasian | Chest |
| Human 2[a)] | *Homo sapiens* | 46 | F | African American | Chest |
| Human 3[a)] | *Homo sapiens* | 29 | M | Asian | NA |
| Human 4[b)] | *Homo sapiens* | 64 | M | NA** | NA |
| Human 5[a)] | *Homo sapiens* | 75 | F | Caucasian | Chest |
| Chimpanzee 1 | *Pan troglodytes verus* | 28 | M | - | Abdomen |
| Chimpanzee 2 | *Pan troglodytes verus* | 35 | M | - | Abdomen |
| Chimpanzee 3 | *Pan troglodytes verus* | 24 | M | - | NA |
| Gorilla 1 | *Gorilla gorilla gorilla* | 46 | M | - | Abdomen |
| Gorilla 2 | *Gorilla gorilla gorilla* | 33 | M | - | NA |
| Gorilla 3 | *Gorilla gorilla gorilla* | 38 | M | - | NA |
| Orangutan 1 | *Pongo pygmaeus* | 32 | M | - | Abdomen |
| Orangutan 2 | *Pongo pygmaeus* | 24 | F | - | Dorsal |
| Orangutan 3 | *Pongo pygmaeus* | 45 | F | - | NA |

a) Manufacturer: BioChain, Product name: Total RNA.

b) Manufacturer: Agilent Technologies, Product name: MVP Total RNA, Human Skin.

*M: male, F: female.

**NA: not available.

**Table 3-2** Reference genomes used for RNA-seq mapping and differences in gene names between genomes

| | Human | Chimpanzee | Gorilla | Orangutan |
|---|---|---|---|---|
| Reference genome | GRCh37/hg19 | CGSC 2.1.3/ panTro3 | gorGor4.1/ gorGor4 | WUGSC 2.0.2/ ponAbe2 |
| Annotated genes | 57,773 genes | 39,699 genes | 32,212 genes | 34,087 genes |
| | Corresponding gene names* | | | |
| | CDHR1 | CDHR1 | CDHR1 | LOC100456512 |
| | CYP1B1 | LOC459163 | LOC101134278 | CYP1B1 |
| | LCE2A | LOC736270 | LOC109029453 | LCE2A |
| | HLA-DPA1 | PATR-DPA1 | LOC101141395 | LOC100438181 |
| | HLA-DPB1 | PATR-DPB1 | LOC101141025 | LOC100444477 |
| | HLA-DQB1 | LOC100616520 | LOC101133052 | LOC100457574 |
| | HLA-DQB2 | PATR-DQB2 | LOC101134815 | LOC100442586 |
| | HLA-DRA | PATR-DRA | LOC101128237 | LOC100460187 |

*The annotation names of these eight differentially expressed genes differ between the four reference genomes.

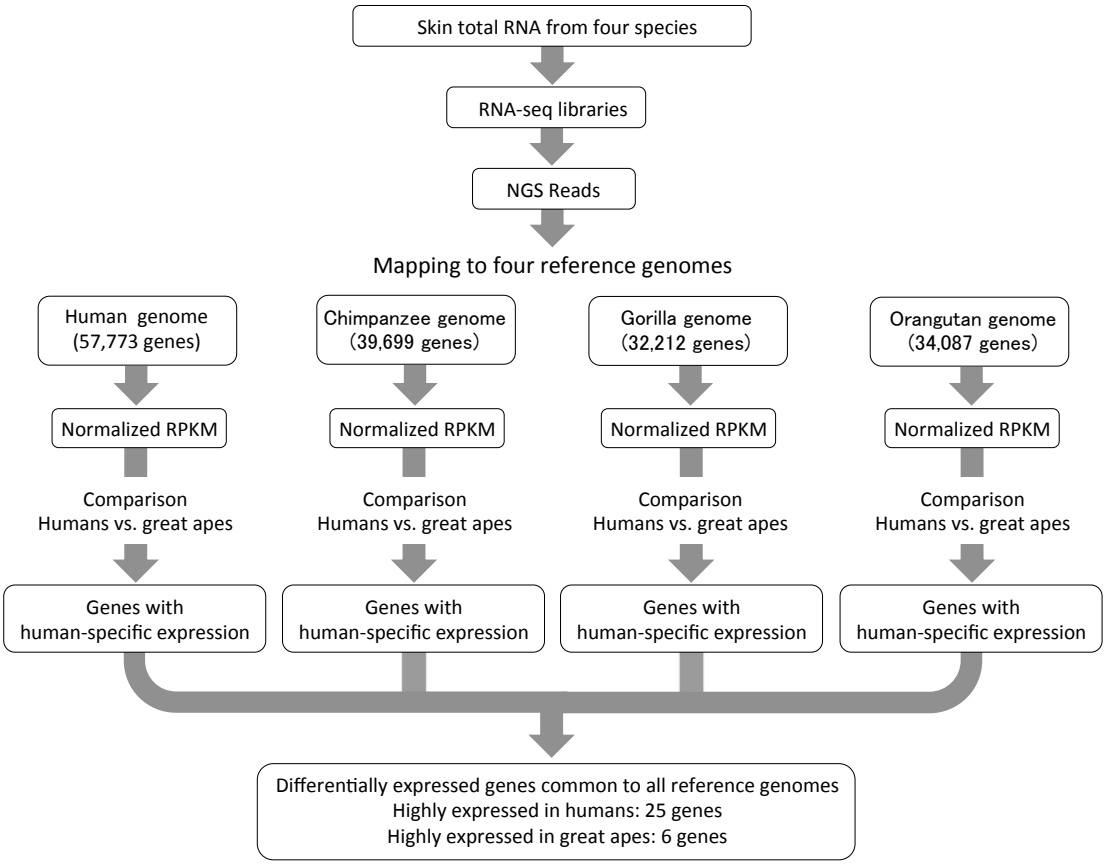


**Figure 3-2** Flow chart of the identification of differentially expressed genes between humans and great apes.

reference genome sequences of human, chimpanzee, gorilla, and orangutan (Figure 3-2, Table 3-2). In each of four mapping results, the expression values, RPKM (Reads Per Kilobase of an exon model per Million mapped reads) values, were calculated for each gene in each sample. I focused on genes with average RPKM values for humans or great apes ≥ 1 in each mapping result. I normalized the expression values by Quantile normalization (Bolstad, et al. 2003). The normalized expression data were checked by boxplot. The normalized expression values of five human individuals were compared with those of nine great apes by Baggerley's test (Baggerly, et al. 2003). The genes showing statistically significant differences [$p < 0.05$, with FDR p-value correction (Benjamini and Hochberg 1995)] in their average normalized RPKM values between humans (n=5) and great apes (n=9) were extracted in each of four mapping results. The mapping and comparison of normalized RPKM values were conducted using CLC Genomics Workbench (https://www.qiagenbioinformatics.com/). Then, the genes that were common to each of the extracted results were selected as differentially expressed genes between humans and great apes.

## Results

To investigate genes associated with human-specific skin characteristics, I identified differentially expressed genes between human and great ape skin using RNA-seq (Figure 3-2). I used total RNA from the skin of five human and nine great ape individuals (three individuals each from chimpanzees, gorillas, and orangutans) and sequenced their cDNA transcripts using the Illumina HiSeq platforms. The 25-45 million reads from each sample were mapped to the human reference genome. To avoid a mapping bias caused by genetic divergences between the human reference genome and the mapped reads, I also mapped reads from each sample to the chimpanzee, gorilla, and orangutan reference genomes. Details of mapped read depth for each sample were shown in Table 3-3. For each of the four mapping results, the expression values (i.e., RPKM values) were calculated for each gene of each sample and subsequently normalized (Figure 3-3). The average normalized RPKM values of five humans were compared with those of nine great apes for each gene. The genes that showed statistically significant differences in their average normalized RPKM values ($p < 0.05$, Baggerley's test with FDR p-value correction) were extracted for each of the four mapping results. Finally, I selected the genes that were common to each of the extracted results as differentially expressed genes.

As a result, I extracted 487, 126, 165, and 166 genes (including unannotated genes and pseudogenes) with differential expression from the mapping results using human, chimpanzee, gorilla, and orangutan reference genomes, respectively. Among these genes, 30 genes were common to all mapping results (Figure 3-4a-d). Differential expression of *COL18A1* was only detected in mapping to the human reference genome (Figure 3-4a); however, the annotations of *COL18A1* in the three great ape reference genomes were incomplete, and so the majority of reads

could not be mapped to these genomes. Therefore, I selected *COL18A1* as a differentially expressed gene between humans and great apes. In total, 31 genes were assigned as differentially expressed genes (Table 3-4). 25 and 6 genes showed higher and lower expression in humans than great apes, respectively. In this thesis, I focus on structural differences between human and other primate skin. Therefore, I further analyzed structural protein genes in my differential expression results, namely, *biglycan* (*BGN*), *collagen type XVIII alpha 1 chain* (*COL18A1*), *CD151 molecule* (*CD151*), and *laminin subunit beta 2* (*LAMB2*).

Both COL18A1 and BGN co-localize with other collagen proteins; therefore, I also analyzed the expression of the other collagen genes using the mapping result in the human reference genome. COL18A1 forms collagen XVIII (Marneros and Olsen 2005), which is a structural component of the epidermal BM (Has and Nystrom 2015). The epidermal BM and the other anchoring structures include collagens IV, VII, and XVII as well as collagen XVIII in the epidermal BM zone (Has and Nystrom 2015). I focused on α chain genes for collagens in the epidermal BM zone that were relatively highly expressed (average normalized RPKM values for humans or great apes ≥ 10). All of such genes (*COL4A1*, *COL4A2*, *COL7A1*, *COL17A1*, and *COL18A1*) showed higher expression in humans than in great apes (Table 3-5). Among them, the expression differences in the two genes, *COL17A1* and *COL18A1*, were statistically significant ($p < 0.05$, t-test with Bonferroni correction).

BGN binds to collagens I, II, III, VI, and IX (Chen and Birk 2013), and regulates collagen fibrillogenesis in skin (Halper 2014). I focused on α chain genes for BGN-binding collagens that were relatively highly expressed (average normalized RPKM values for humans or great apes ≥ 10). All of such genes (*COL1A1*, *COL1A2*, *COL3A1*, *COL6A1*, *COL6A2*, and

**Table 3-3** Read depth values for each sample in each mapping result

| | | Reference genome | | | |
|---|---|---|---|---|---|
| | Sample ID | Human | Chimpanzee | Gorilla | Orangutan |
| Mapped reads (million reads)* | Human 1 | 37.14 | 29.61 | 29.61 | 27.67 |
| | Human 2 | 24.84 | 20.23 | 20.05 | 18.40 |
| | Human 3 | 27.41 | 22.32 | 21.70 | 19.83 |
| | Human 4 | 34.51 | 26.24 | 25.96 | 24.36 |
| | Human 5 | 29.17 | 21.62 | 21.48 | 20.04 |
| | Chimpanzee 1 | 26.26 | 22.36 | 21.97 | 20.47 |
| | Chimpanzee 2 | 27.21 | 25.52 | 23.48 | 20.84 |
| | Chimpanzee 3 | 27.72 | 23.68 | 23.17 | 21.75 |
| | Gorilla 1 | 24.16 | 20.30 | 21.68 | 18.78 |
| | Gorilla 2 | 21.77 | 16.25 | 17.32 | 13.68 |
| | Gorilla 3 | 25.69 | 20.22 | 20.99 | 19.53 |
| | Orangutan 1 | 24.16 | 20.35 | 20.71 | 22.61 |
| | Orangutan 2 | 27.86 | 21.93 | 23.01 | 22.16 |
| | Orangutan 3 | 39.06 | 29.83 | 30.43 | 32.91 |

*Excluding repetitive sequences.

**Table 3-4** Differentially expressed genes between human and great ape skin

| | | Average normalized RPKM[1)] | | |
|---|---|---|---|---|
| | | Humans | Great apes | FD[2)] |
| **Higher expression in humans** | | | | |
| *DHCR24* | *24-dehydrocholesterol reductase* | 137.0 | 32.1 | 4.3*** |
| ***BGN*** | ***Biglycan*** | **138.3** | **41.8** | **3.3**** |
| *CDHR1* | *Cadherin related family member 1* | 20.9 | 0.7 | 30.8*** |
| ***CD151*** | ***CD151 molecule*** | **61.3** | **27.3** | **2.2***** |
| *CD207* | *CD207 molecule* | 29.3 | 1.9 | 15.5*** |
| *CD74* | *CD74 molecule* | 547.6 | 112.8 | 4.9*** |
| ***COL18A1*** | ***Collagen type XVIII alpha 1 chain*** | **41.7** | **18.3** | **2.3***** |
| *CFD* | *Complement factor D* | 986.0 | 130.0 | 7.6*** |
| *FAM57A* | *Family with sequence similarity 57 member A* | 34.3 | 12.5 | 2.7*** |
| *GEMIN4* | *Gem nuclear organelle associated protein 4* | 7.0 | 2.8 | 2.5* |
| *GRINA* | *Glutamate ionotropic receptor NMDA type subunit associated protein 1* | 42.8 | 19.1 | 2.2** |
| ***LAMB2*** | ***Laminin subunit beta 2*** | **83.1** | **27.3** | **3.0***** |
| *LCE2A* | *Late cornified envelope 2A* | 75.2 | 2.0 | 36.9* |
| *LCE6A* | *Late cornified envelope 6A* | 56.2 | 6.4 | 8.8* |
| *HLA-DPA1* | *Major histocompatibility complex, class II, DP alpha 1* | 30.8 | 7.0 | 4.4*** |
| *HLA-DPB1* | *Major histocompatibility complex, class II, DP beta 1* | 64.8 | 10.4 | 6.2*** |
| *HLA-DQB1* | *Major histocompatibility complex, class II, DQ beta 1* | 20.5 | 9.3 | 2.2* |
| *HLA-DQB2* | *Major histocompatibility complex, class II, DQ beta 2* | 40.1 | 2.4 | 17.0*** |
| *HLA-DRA* | *Major histocompatibility complex, class II, DR alpha* | 361.7 | 103.5 | 3.5*** |
| *NFE2L1* | *Nuclear factor, erythroid 2 like 1* | 80.0 | 42.1 | 1.9*** |
| *SCRN2* | *Secernin 2* | 12.0 | 5.0 | 2.4** |
| *SYT8* | *Synaptotagmin 8* | 14.3 | 3.2 | 4.4*** |
| *TREX1* | *Three prime repair exonuclease 1* | 14.0 | 6.4 | 2.2* |
| *TSR3* | *TSR3, acp transferase ribosome maturation factor* | 47.2 | 23.3 | 2.0** |
| *WFDC5* | *WAP four-disulfide core domain 5* | 118.6 | 45.5 | 2.6** |
| **Lower expression in humans** | | | | |
| *BNIP3* | *BCL2 interacting protein 3* | 14.7 | 39.3 | 2.7* |

| | | | | |
|---|---|---|---|---|
| *CYP1B1* | *Cytochrome P450 family 1 subfamily B member 1* | 4.8 | 22.9 | 4.7** |
| *HMGB2* | *High mobility group box 2* | 19.6 | 45.6 | 2.3* |
| *ID3* | *Inhibitor of DNA binding 3, HLH protein* | 41.1 | 74.2 | 1.8** |
| *NPM3* | *Nucleophosmin/nucleoplasmin 3* | 19.7 | 47.0 | 2.4*** |
| *SULT1C4* | *Sulfotransferase family 1C member 4* | 0.4 | 7.7 | 18.6*** |

1) In mapping to the human reference genome.

2) FD: fold difference.

Bold letters: structural protein genes.

*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, Baggerley's test with FDR p-value correction.

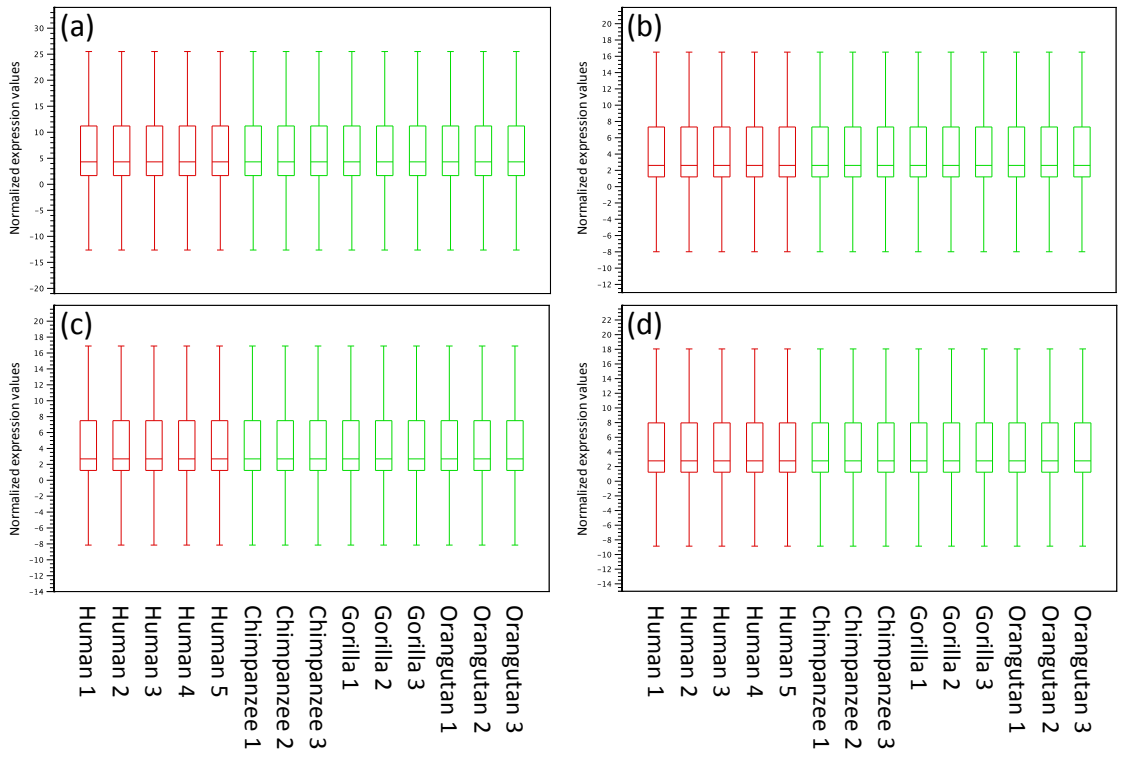


**Figure 3-3** Box plots for the normalized expression values.

Box plots of the normalized expression values in five human and nine great ape skin specimens based on the mapping to the reference genome of (a) human, (b) chimpanzee, (c) gorilla, and (d) orangutan.

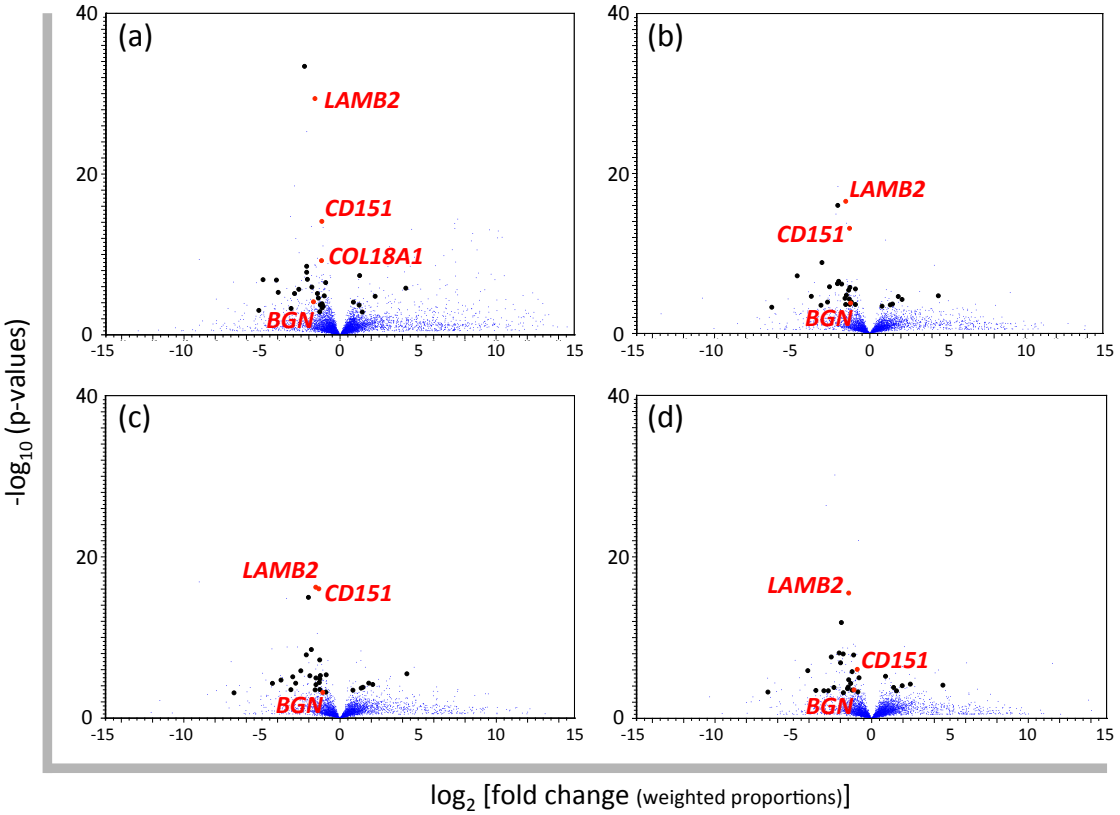


**Figure 3-4** Volcano plots for the gene expression differences between humans and great apes. Each blue dot represents the gene expressed in the skin. The normalized RPKM values are based on the mapping to the reference genome of (a) human, (b) chimpanzee, (c) gorilla, and (d) orangutan. The log2 fold changes of average normalized RPKM values of great apes compared to those of humans and the -log10 p-values resulting from Baggerley's test comparing average normalized RPKM values between humans and great apes for each gene are plotted on the x- and y-axis, respectively. Differentially expressed genes between humans and great apes selected in this study are shown by black and red dots. The gene names colored in red indicate structural protein genes.

**Table 3-5** Collagen genes with relatively high expression in skin

| | | Average normalized RPKM[1)] | | |
|---|---|---|---|---|
| | | Humans | Great apes | Fold difference |
| **(a) Genes encoding α chains for collagens in the epidermal BM zone** | | | | |
| *COL4A1* | *Collagen type IV alpha 1 chain* | 54.6 | 21.5 | 2.5 |
| *COL4A2* | *Collagen type IV alpha 2 chain* | 77.5 | 32.9 | 2.4 |
| *COL7A1* | *Collagen type VII alpha 1 chain* | 32.9 | 16.8 | 2.0 |
| *COL17A1* | *Collagen type XVII alpha 1 chain* | 181.7 | 76.4 | 2.4* |
| *COL18A1* | *Collagen type XVIII alpha 1 chain* | 41.7 | 18.3 | 2.3*** |
| **(b) Genes encoding α chains for BGN-binding collagens** | | | | |
| *COL1A1* | *Collagen type I alpha 1 chain* | 444.4 | 198.5 | 2.2 |
| *COL1A2* | *Collagen type I alpha 2 chain* | 311.3 | 134.9 | 2.3 |
| *COL3A1* | *Collagen type III alpha 1 chain* | 404.7 | 148.9 | 2.7 |
| *COL6A1* | *Collagen type VI alpha 1 chain* | 276.3 | 106.2 | 2.6 |
| *COL6A2* | *Collagen type VI alpha 2 chain* | 459.3 | 185.1 | 2.5** |
| *COL6A3* | *Collagen type VI alpha 3 chain* | 45.6 | 27.3 | 1.7 |

1) In mapping to the human reference genome.

*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, t-test with Bonferroni correction.

*COL6A3*) showed higher expression in humans than in great apes (Table 3-5). Among them the expression difference in the gene *COL6A2* was statistically significant ($p < 0.05$, t-test with Bonferroni correction).

In the same manner as the gene expression comparison between human and great ape skin, I identified differentially expressed genes in the skin of each of great ape species. The results are summarized in Table 3-6. These analyses for each great ape species would provide a first step to investigate differential gene expression patterns that might be responsible for their respective skin characteristics.

**Table 3-6** Differentially expressed genes in great ape skin

| Differentially expressed genes between chimpanzees and HGO[1)] | | Average normalized RPKM[2)] | | |
|---|---|---|---|---|
| | | Chimpanzees | HGO | FD[3)] |
| **Higher expression in chimpanzees** | | | | |
| *CA2* | *Carbonic anhydrase 2* | 90.5 | 17.0 | 5.3* |
| **Lower expression in chimpanzees** | | | | |
| *ADRA2A* | *Adrenoceptor alpha 2A* | 4.9 | 16.4 | 3.4* |
| *APEX1* | *Apurinic/apyrimidinic endodeoxyribonuclease 1* | 39.1 | 58.6 | 1.5** |
| *ARL2* | *ADP ribosylation factor like GTPase 2* | 9.1 | 21.2 | 2.3** |
| *C3* | *Complement C3* | 15.6 | 83.4 | 5.4*** |
| *CCL19* | *C-C motif chemokine ligand 19* | 2.4 | 24.6 | 10.4*** |
| *CLDN5* | *Claudin 5* | 14.0 | 75.7 | 5.4* |
| *DLL4* | *Delta like canonical Notch ligand 4* | 2.4 | 8.5 | 3.5* |
| *DUSP23* | *Dual specificity phosphatase 23* | 23.6 | 79.1 | 3.4** |
| *FAM110D* | *Family with sequence similarity 110 member D* | 1.3 | 7.6 | 5.7* |
| *GHDC* | *GH3 domain containing* | 1.9 | 6.8 | 3.5* |
| *INHBB* | *Inhibin subunit beta B* | 7.2 | 25.4 | 3.5* |
| *NT5C* | *5′, 3′-nucleotidase, cytosolic* | 10.7 | 20.2 | 1.9* |
| *PRDX5* | *Peroxiredoxin 5* | 66.4 | 173.3 | 2.6*** |
| *TGM2* | *Transglutaminase 2* | 0.8 | 8.0 | 10.2*** |
| *TNFRSF14* | *TNF receptor superfamily member 14* | 2.8 | 10.5 | 3.7* |
| *TSR3* | *TSR3, acp transferase ribosome maturation factor* | 16.8 | 35.9 | 2.1* |
| *UAP1L1* | *UDP-N-acetylglucosamine pyrophosphorylase 1 like 1* | 1.3 | 6.9 | 5.2** |
| *VAT1* | *Vesicle amine transport 1* | 88.0 | 146.2 | 1.7* |
| Differentially expressed genes between gorillas and HCO[1)] | | Average normalized RPKM[2)] | | |
| | | Gorillas | HCO | FD[3)] |
| **Higher expression in gorillas** | | | | |
| None | | | | |
| **Lower expression in gorillas** | | | | |
| *B2M* | *Beta-2-microglobulin* | 224.3 | 496.0 | 2.2* |
| *GSN* | *Gelsolin* | 70.7 | 330.2 | 4.7** |

| | | | | |
|---|---|---|---|---|
| *HS3ST6* | *Heparan sulfate-glucosamine 3-sulfotransferase 6* | 3.5 | 33.7 | 9.6** |
| *IRF2BP2* | *Interferon regulatory factor 2 binding protein 2* | 18.0 | 33.3 | 1.9* |
| *KLF4* | *Kruppel like factor 4* | 58.0 | 197.6 | 3.4** |
| *PARP3* | *Poly(ADP-ribose) polymerase family member 3* | 2.0 | 11.1 | 5.7* |
| *TNFRSF18* | *TNF receptor superfamily member 18* | 1.7 | 13.2 | 7.7*** |
| *TPPP3* | *Tubulin polymerization promoting protein family member 3* | 12.7 | 62.2 | 4.9* |

| | | Average normalized RPKM[2)] | | |
|---|---|---|---|---|
| Differentially expressed genes between orangutans and HCG[1)] | | Orangutans | HCG | FD[3)] |
| **Higher expression in orangutans** | | | | |
| *ANGPTL4* | *Angiopoietin like 4* | 176.1 | 39.5 | 4.5*** |
| *BEX4* | *Brain expressed X-linked 4* | 53.8 | 23.1 | 2.3*** |
| *BLOC1S4* | *Biogenesis of lysosomal organelles complex 1 subunit 4* | 14.7 | 6.7 | 2.2* |
| *C2* | *Complement C2* | 14.0 | 1.7 | 8.4*** |
| *CERS2* | *Ceramide synthase 2* | 21.6 | 9.9 | 2.2* |
| *FAM32A* | *Family with sequence similarity 32 member A* | 70.2 | 35.6 | 2.0* |
| *MAP1LC3A* | *Microtubule associated protein 1 light chain 3 alpha* | 48.5 | 27.8 | 1.7** |
| *NPC2* | *NPC intracellular cholesterol transporter 2* | 161.9 | 73.9 | 2.2*** |
| *RHOU* | *Ras homolog family member U* | 22.3 | 5.5 | 4.0*** |
| *SNRPD2* | *Small nuclear ribonucleoprotein D2 polypeptide* | 152.8 | 88.0 | 1.7** |
| *VMO1* | *Vitelline membrane outer layer 1 homolog* | 10.5 | 2.1 | 4.9** |
| **Lower expression in orangutans** | | | | |
| *ANKRD65* | *Ankyrin repeat domain 65* | 0.1 | 3.3 | 30.7*** |
| *APOL3* | *Apolipoprotein L3* | 0.0 | 9.4 | 209.7*** |
| *BGN* | *Biglycan* | 24.2 | 90.5 | 3.7** |
| *CCND1* | *Cyclin D1* | 5.5 | 27.7 | 5.0*** |
| *CELSR2* | *Cadherin EGF LAG seven-pass G-type receptor 2* | 2.5 | 14.6 | 5.9*** |
| *CFD* | *Complement factor D* | 13.2 | 550.9 | 41.7* |
| *CHP2* | *Calcineurin like EF-hand protein 2* | 1.4 | 35.7 | 25.5*** |

| | | | | |
|---|---|---|---|---|
| *CRABP2* | *Cellular retinoic acid binding protein 2* | 44.8 | 121.9 | 2.7*** |
| *DEGS2* | *Delta 4-desaturase, sphingolipid 2* | 5.6 | 24.8 | 4.4** |
| *DES* | *Desmin* | 31.7 | 276.7 | 8.7** |
| *DNAJB2* | *DnaJ heat shock protein family (Hsp40) member B2* | 14.5 | 33.5 | 2.3*** |
| *DSG1* | *Desmoglein 1* | 2.0 | 18.7 | 9.3** |
| *DSP* | *Desmoplakin* | 70.1 | 222.6 | 3.2*** |
| *EFS* | *Embryonal Fyn-associated substrate* | 12.8 | 26.7 | 2.1** |
| *FAM83H* | *Family with sequence similarity 83 member H* | 6.9 | 25.6 | 3.7** |
| *GAL3ST4* | *Galactose-3-O-sulfotransferase 4* | 2.4 | 10.2 | 4.2*** |
| *GPAA1* | *Glycosylphosphatidylinositol anchor attachment 1* | 18.2 | 34.7 | 1.9*** |
| *GPX2* | *Glutathione peroxidase 2* | 0.0 | 2.1 | 122.2* |
| *HDHD3* | *Haloacid dehalogenase like hydrolase domain containing 3* | 2.0 | 12.0 | 6.0*** |
| *HSPB7* | *Heat shock protein family B (small) member 7* | 1.9 | 11.3 | 6.1** |
| *IFI27* | *Interferon alpha inducible protein 27* | 1.2 | 15.0 | 12.7** |
| *IL36RN* | *Interleukin 36 receptor antagonist* | 1.5 | 12.6 | 8.4* |
| *IRF7* | *Interferon regulatory factor 7* | 1.1 | 8.9 | 8.3* |
| *JAG1* | *Jagged 1* | 10.2 | 27.8 | 2.7*** |
| *KLHDC8B* | *Kelch domain containing 8B* | 8.0 | 20.8 | 2.6* |
| *KLK1* | *Kallikrein 1* | 0.5 | 14.8 | 28.1** |
| *LAD1* | *Ladinin 1* | 21.9 | 93.6 | 4.3*** |
| *LAMB2* | *Laminin subunit beta 2* | 22.2 | 54.0 | 2.4* |
| *MCAM* | *Melanoma cell adhesion molecule* | 10.0 | 29.1 | 2.9** |
| *MPEG1* | *Macrophage expressed 1* | 0.2 | 4.4 | 22.2* |
| *MYCL* | *MYCL proto-oncogene, bHLH transcription factor* | 2.3 | 8.8 | 3.8* |
| *NOTCH3* | *Notch 3* | 11.9 | 44.2 | 3.7*** |
| *PHLDA3* | *Pleckstrin homology like domain family A member 3* | 27.2 | 56.7 | 2.1*** |
| *PKP1* | *Plakophilin 1* | 53.4 | 240.1 | 4.5*** |
| *PKP3* | *Plakophilin 3* | 30.6 | 90.0 | 2.9** |

| | | | | |
|---|---|---|---|---|
| *POLR2L* | *RNA polymerase II subunit L* | 42.2 | 90.7 | 2.1** |
| *PSMB8* | *Proteasome subunit beta 8* | 12.1 | 28.8 | 2.4*** |
| *RAB25* | *RAB25, member RAS oncogene family* | 26.9 | 64.4 | 2.4*** |
| *SERPINA12* | *Serpin family A member 12* | 10.9 | 78.7 | 7.2** |
| *SH3BGRL3* | *SH3 domain binding glutamate rich protein like 3* | 80.3 | 156.8 | 2.0*** |
| *SLC27A3* | *Solute carrier family 27 member 3* | 4.3 | 9.8 | 2.3* |
| *SPTBN2* | *Spectrin beta, non-erythrocytic 2* | 6.4 | 23.2 | 3.6** |
| *SUSD2* | *Sushi domain containing 2* | 1.8 | 14.1 | 7.8** |
| *SYT8* | *Synaptotagmin 8* | 1.8 | 8.6 | 4.7* |
| *SYTL1* | *Synaptotagmin like 1* | 10.6 | 24.0 | 2.3* |
| *TAP1* | *Transporter 1, ATP binding cassette subfamily B member* | 4.8 | 13.7 | 2.9* |
| *THEM5* | *Thioesterase superfamily member 5* | 0.2 | 49.3 | 261.0*** |
| *TP53AIP1* | *Tumor protein p53 regulated apoptosis inducing protein 1* | 0.0 | 6.5 | 5579.9*** |
| *TREX1* | *Three prime repair exonuclease 1* | 3.9 | 10.6 | 2.7** |
| *TUBA4A* | *Tubulin alpha 4a* | 22.2 | 64.1 | 2.9** |
| *UBL4A* | *Ubiquitin like 4A* | 10.0 | 17.8 | 1.8* |
| *WFDC5* | *WAP four-disulfide core domain 5* | 19.7 | 85.8 | 4.4** |
| *ZBTB22* | *Zinc finger and BTB domain containing 22* | 7.0 | 13.0 | 1.9* |
| *ZNF768* | *Zinc finger protein 768* | 10.3 | 19.5 | 1.9* |

1) Each abbreviation represents individuals of humans (H), chimpanzees (C), gorillas (G), and orangutans (O).

2) In mapping to the human reference genome.

3) FD: fold difference.

*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, Baggerley's test with FDR p-value correction.

## Discussion

Based on my RNA-seq analyses, expression levels of 25 and 6 genes in skin were found to be significantly higher and lower in humans than in great apes, respectively. Four of them (*COL18A1*, *LAMB2*, *CD151*, and *BGN*) encode structural proteins and showed higher expression in humans, suggesting the possibility that the expression changes of these genes influence human skin structure. As for the other differentially expressed genes, the comparison of nonstructural traits between human and other primate skin could reveal the correlation between the differential expression of these genes and human-specific characteristics. As promising genes, *late cornified envelope* (*LCE*) genes might be related to human-specific skin characteristics because previous studies reported the correlation of this gene family with skin function. I first provide discussion about association between the human-specific expression patterns identified in the structural protein genes and *LCE* genes and human-specific skin characteristics.

*Differential expression in the structural protein genes*

*COL18A1*, *LAMB2*, and *CD151* are genes that encode proteins structurally associated with the epidermal BM zone. COL18A1 forms collagen XVIII (Marneros and Olsen 2005), which comprises the epidermal BM (Has and Nystrom 2015). Since one of the collagens in the epidermal BM zone was highly expressed in humans, I focused on expression of α chain genes for collagens in the epidermal BM zone. Collagens IV, VII, and XVII as well as collagen XVIII function as structural components in the epidermal BM zone (Has and Nystrom 2015). The gene encoding α chain for collagen XVII (*COL17A1*) showed significantly higher expression levels in humans. Collagen XVII forms anchoring structure linking basal keratinocytes to the epidermal BM

(Masunaga, et al. 1997). Mice lacking *COL18A1* have a broadened epidermal BM (Utriainen, et al. 2004), and a patient with a known mutant *COL17A1* gene exhibits junctional epidermolysis bullosa (McGrath, et al. 1996). Both of these observations indicate a structural role of these genes in epidermal BM zone integrity.

LAMB2 is a component of the network structure of laminins in the epidermal BM (Has and Nystrom 2015). CD151 functions as an adhesion protein between the epidermis and epidermal BM (Has and Nystrom 2015). Therefore, it is plausible that the higher expression of *COL18A1*, *COL17A1*, *LAMB2*, and *CD151* may make the architecture of the epidermal BM zone in human skin different from that in great ape skin. If the rete ridge is specific to human skin, the increased expression of the two genes for structural components, *COL18A1* and *LAMB2*, may perhaps lead to the undulating epidermal BM topography. The elevated expression of the two genes for anchoring structures, *COL17A1* and *CD151*, may produce the strong adhesion between the epidermis and epidermal BM in human skin, or would correlate with the increased adhesive area of the rete ridge in human skin.

BGN is localized to both the epidermis and dermis (Li, et al. 2013). In the epidermis, BGN is on the cell surface of differentiating keratinocytes of the spinous layer (Bianco, et al. 1990). Although the function of this protein in epidermis is unknown, the higher expression of *BGN* may somehow correlate with the human-specific thicker epidermis.

BGN is a component of the extracellular matrix in the dermis (Li, et al. 2013). This protein interacts with collagens and regulates collagen fibrillogenesis to make the tensile strength of skin (Halper 2014). Since BGN interacts with collagen I, II, III, VI, and IX (Chen and Birk 2013), it was predicted that the expression of the α chain genes for the BGN-binding collagens

would also be higher in human skin. The expression of the gene encoding α chain for collagen VI (*COL6A2*) was significantly higher in humans. Thus, the increased expression of *BGN* and *COL6A2* may produce a stronger tensile strength in human skin.

Elastic fibers are another component of the extracellular matrix in the dermis (Smoller 2009) and give skin elasticity (Kielty, et al. 2002). Elastin is one of the components of elastic fibers (Kielty, et al. 2002), and BGN regulates elastin formation (Reinboth, et al. 2002). It is known that the human dermis is more enriched with elastic fibers compared to most other primates (Montagna 1982; Montagna 1985). The amount of elastic fibers in the furred skin of chimpanzees and gorillas is reported with inconsistent descriptions; similar to the content in humans (Montagna 1982; Montagna 1985) and nowhere numerous (Montagna and Yun 1963). Although the amount of elastic fibers in great apes is ambiguous, it is possible that the higher expression of *BGN* might contribute to the richness of elastic fibers in human skin.

Humans have a low amount of hair on their body compared with other primates, which gives humans a high level of thermoregulation (Folk and Semken 1991). However, it is believed that human skin has lost the ability to protect the internal tissues from external physical stresses by hair. Rete ridge increases the area where the epidermis and dermis connect compared to flat topography of the epidermal BM zone, which may make strong adhesion between these two layers. The rete ridge, thick epidermis, and rich elastic fibers in skin might contribute to the physical strength of human skin. Actually, it has been proposed that human skin has developed adaptive structural changes that give it strength and resilience (Montagna 1982). Although additional quantitative histological comparison between human and great ape skin is required to examine the human-specific skin characteristics, the human-specific expression patterns found in this

study may contribute to adaptive skin characteristics specific to humans with a low amount of hair on their body.

*Differential expression in the LCE genes*

The *LCE* gene family consists of 18 members subdivided into six subgroups, *LCE1* to *LCE6*, based on similarities of amino acid sequences, genomic organization, and patterns of expression (Bergboer, et al. 2011; Jackson, et al. 2005). Among the differentially expressed genes, *LCE2A* and *LCE6A* genes, which showed higher expression in humans, are the members of this gene family. *LCE6A* sequences were intact in humans and great apes (human: NM_001128600.1, chimpanzee: XM_003308479.1, gorilla: XM_004026698.1, and orangutan: XM_002810185.1), whereas *LCE2A* sequence was intact in humans (NM_178428.3) but those were independently pseudogenized by a premature stop codon in chimpanzees (LOC736270), by a frame shift in gorillas (LOC109029453), and by a large deletion in the coding region in orangutans (LOC103891408). The functions of LCE2A and LCE6A proteins are unknown so far.

In humans, *LCE2A* is expressed in normal healthy skin (Bergboer, et al. 2011), as observed in my RNA-seq analyses. Upregulation of this gene was induced by high concentration of extracellular calcium, UV irradiation (Jackson, et al. 2005), and Th17 cytokine stimulation (Niehues, et al. 2017), indicating functional roles of this gene in skin. Thus, only humans retain the functional LCE2A, which might be related to human-specific skin characteristics.

In the *LCE* gene family, the function of a few members was reported. A deletion of *LCE3B* and *LCE3C* is strongly associated with psoriasis (De Cid, et al. 2009), and the deletion is traced back to the ancient *Homo* lineage (Lin, et al. 2015; Pajic, et al. 2016). Recently, the

antimicrobial activity against a variety of bacterial taxa was shown in LCE3A, LCE3B, and LCE3C proteins (Niehues, et al. 2017). If the LCE2A and LCE6A proteins also have antimicrobial defensive roles, the increased expression of these genes might influence cutaneous bacterial composition in humans. Recent studies reported that the commensal skin microorganisms function as a barrier that protects against pathogenic and harmful organisms (Niehues, et al. 2018; Scharschmidt and Fischbach 2013; Weyrich, et al. 2015). Covered with less hair, human skin might have defensive characteristics that are different from great ape skin. I am planning to examine antimicrobial activity of the LCE2A and LCE6A proteins in the near future.

Since humans have a much less hair than great apes, it was predicted that the expression of the genes associated with the components of hair would be lower in humans. However, no such gene showed significantly lower expression in humans than in great apes based on the present RNA-seq analyses. Actually, the average normalized RPKM values of hair keratin genes were tens to thousands of fold higher in great apes than in humans. However, several great ape individuals showed the RPKM values similar to human individuals in those genes, resulting in insignificant differences in expression levels between humans and great apes (examples are shown in Table 3-7). This variation in RPKM values did not depend on differences in sex, age, or body part between great ape individuals.

It is conceivable that the genes with human-specific expression patterns found in this study cooperate with other differentially expressed genes to change human skin structure. For example, although significant differences in gene expression were detected only in a few genes, the expression of all the relatively highly expressed genes encoding α chains for collagens

**Table 3-7** Normalized RPKM values[1)] in *KRT* genes

| | *KRT31* | *KRT37* | *KRT81* | *KRT86* |
|---|---|---|---|---|
| Human 1 | 29.25 | 0.00 | 0.52 | 0.89 |
| Human 2 | 8.23 | 0.00 | 1.13 | 0.31 |
| Human 3 | 29.09 | 0.02 | 2.88 | 5.65 |
| Human 4 | 17.10 | 0.08 | 1.11 | 1.71 |
| Human 5 | 13.68 | 0.00 | 1.30 | 1.63 |
| Human mean | 19.47 | 0.02 | 1.39 | 2.04 |
| Chimpanzee 1 | 155.07 | 8.54 | 22.30 | 47.58 |
| Chimpanzee 2 | 1010.14 | 1.07 | 69.04 | 133.36 |
| Chimpanzee 3 | 2662.10 | 46.40 | 367.63 | 834.71 |
| Gorilla 1 | 3386.97 | 351.36 | 834.71 | 1882.13 |
| Gorilla 2 | 1.53 | 0.00 | 0.29 | 0.95 |
| Gorilla 3 | 7.99 | 1.24 | 0.00 | 0.10 |
| Orangutan 1 | 1.84 | 0.00 | 0.99 | 1.96 |
| Orangutan 2 | 20.72 | 0.25 | 3.27 | 6.97 |
| Orangutan 3 | 106.16 | 0.30 | 23.74 | 35.99 |
| Great ape mean | 816.95 | 45.46 | 146.88 | 327.08 |
| Fold difference | 41.96 | 2083.58 | 105.82 | 160.69 |
| p-value[2)] | 0.36 | 0.62 | 0.47 | 0.48 |

1) In mapping to the human reference genome.

2) Baggerley's test with FDR p-value correction.

associated with the epidermal BM zone and BGN was higher in humans than in great apes. These genes might also contribute to the rete ridge formation and stronger tensile strength in human skin. In this study, I used skin specimens from individuals of different sex and age and from different body parts for RNA-seq analyses. Therefore, the identified genes were consistently differentially expressed in human skin compared with great ape skin, regardless of these differences. In the future, increasing the number of skin specimens and comparing gene expression levels between humans and great apes using the skin specimens from individuals of the same condition (e.g., the same sex, age, and body part) would identify other differentially expressed genes specific to that condition. This approach would allow us to further reveal the genetic causes of human-specific skin characteristics.

# Chapter 4

## Inference of substitutions responsible for the human-specific gene expression patterns

## Introduction

In the previous chapters, I found that the expression levels of four structural protein genes (*COL18A1*, *LAMB2*, *CD151*, and *BGN*) in skin were significantly higher in humans than in great apes. These human-specific gene expression patterns may contribute to the rete ridge formation and rich elastic fibers, possible adaptive skin characteristics specific to humans with less hair. Therefore, I next focused on genetic causes responsible for the human-specific expression patterns of these structural protein genes in skin.

Certain noncoding regions within the genome (e.g., promoters and enhancers) are important for the regulation of gene expression (Lindblad-Toh, et al. 2011). Such transcriptional regulatory regions generally harbor a multitude of binding sites for sequence-specific transcription factors (TFs) that modulate gene expression (Carroll, et al. 2013). From an adaptive standpoint, these regions tend to evolve under functional constraint and are thus more conserved between species than the surrounding nonfunctional noncoding regions (He, et al. 2011; Pennacchio, et al. 2006). Transcriptional regulatory regions of a certain gene of interest can separately reside at various positions: immediately 5' of the transcription start site, in the adjacent intergenic regions, in the introns of the gene itself or neighboring genes, or/and even in the noncoding regions at considerable distances from the gene (Kleinjan and van Heyningen 2005). Mutations in transcriptional regulatory regions can change the expression level of the target gene by altering TF-binding affinities (Wittkopp and Kalay 2012), which plays important roles in phenotypic diversity for morphology, physiology, and behavior between species (Wray 2007).

Transcription is also regulated by histone modification. In eukaryotic cells, genomic DNA is folded into chromatin, which is composed of nucleosomes and linker DNA (Luger, et al.

2012). Histones comprise core histones (H2A, H2B, H3, and H4) and are wrapped by approximately 147 bp of DNA in nucleosomes (Luger, et al. 1997). Core histones are subjected to post-translational modifications on various amino acid residues, mostly in their N-terminal tails that extrude from the nucleosomes (Kimura 2013). Histone modifications, including methylation and acetylation, regulate the structure of chromatin, resulting in changing accessibility of TFs to potential transcriptional binding sites (Kouzarides 2007). Chromatin state is different between diverse cell types in a multicellular organism and contributes to cell type-specific gene expression patterns in the presence of an essentially identical genome in each cell (Heintzman, et al. 2009). The active promoters and enhancers of transcribed genes generally possess some specific modifications: mono-, di-, and tri-methylation of histone H3 lysine 4 (represented by H3K4m1, H3K4m2, and H3K4m3, respectively) and acetylation of histone H3 lysine 9 and histone H3 lysine 27 (H3K9ac and H3K27ac, respectively) (Barski, et al. 2007; Karmodiya, et al. 2012; Kimura 2013; Wang, et al. 2008). Chromatin immunoprecipitation followed by sequencing (ChIP-seq) is used to determine the genome-wide distribution of each histone modification in multiple cell types (Consortium 2012; Zhou, et al. 2011). The output from ChIP-seq allows us to estimate transcriptional regulatory regions in the genome, and ChIP-seq data are widely available in public databases [e.g., (Kent, et al. 2002)].

In this chapter, I inferred substitutions located in transcriptional regulatory regions responsible for the human-specific expression patterns of the four structural protein genes of interest. I estimated transcriptional regulatory regions for each gene by identifying conserved noncoding regions around the genes and focusing on histone modifications for active regulatory regions in skin cells. The human-specific substitutions in putative regulatory regions were

estimated to be the candidate substitutions responsible for the human-specific expression patterns in the genes of interest. These candidate substitutions may give humans adaptive skin characteristics through human-specific gene expression patterns.

## Materials and Methods

### Inference of substitutions responsible for the human-specific expression patterns

Noncoding regions that were conserved in non-human lineages (Figure 4-1, gray lines) were identified to estimate transcriptional regulatory regions for the four focused genes. The analyzed genomic regions were set to include noncoding regions at both sides of the genes of interest and were 372 kb, 100 kb, 100 kb, and 78 kb in length for the *COL18A1*, *LAMB2*, *CD151*, and *BGN* genes, respectively, in the human genome (GRCh38). Each of the four genes of interest was located in the center of their respective regions. The genomic sequence alignments of human, chimpanzee, gorilla, orangutan, and rhesus macaque were obtained from Ensembl (https://asia.ensembl.org/index.html). Alignment sites that showed one or more gaps in at least one of the five species were removed.

To identify conserved domains throughout the analyzed genomic regions, a sliding-window analysis was performed using a 120-bp window size and a 4-bp step size (Figure 4-2). For each window, pair-wise nucleotide differences between the sequences of the species were estimated using the Jukes-Cantor model implemented in DnaSP 5.0 (Rozas, et al. 2003). Then, the numbers of substitutions in non-human lineages (Figure 4-1, gray lines) for each window were calculated using the Fitch-Margoliash algorithm (Fitch and Margoliash 1967). For each analyzed genomic sequence alignment, the pair-wise nucleotide divergences between species excluding exonic and unaligned regions and their standard errors were calculated with the Jukes-Cantor model and a bootstrap method (1000 replicates), respectively, using MEGA 7 (Kumar, et al. 2016). Using the same algorithm, the average expected number of substitutions in non-human lineages for a 120-bp region in each analyzed genomic region was calculated using these pair-wise

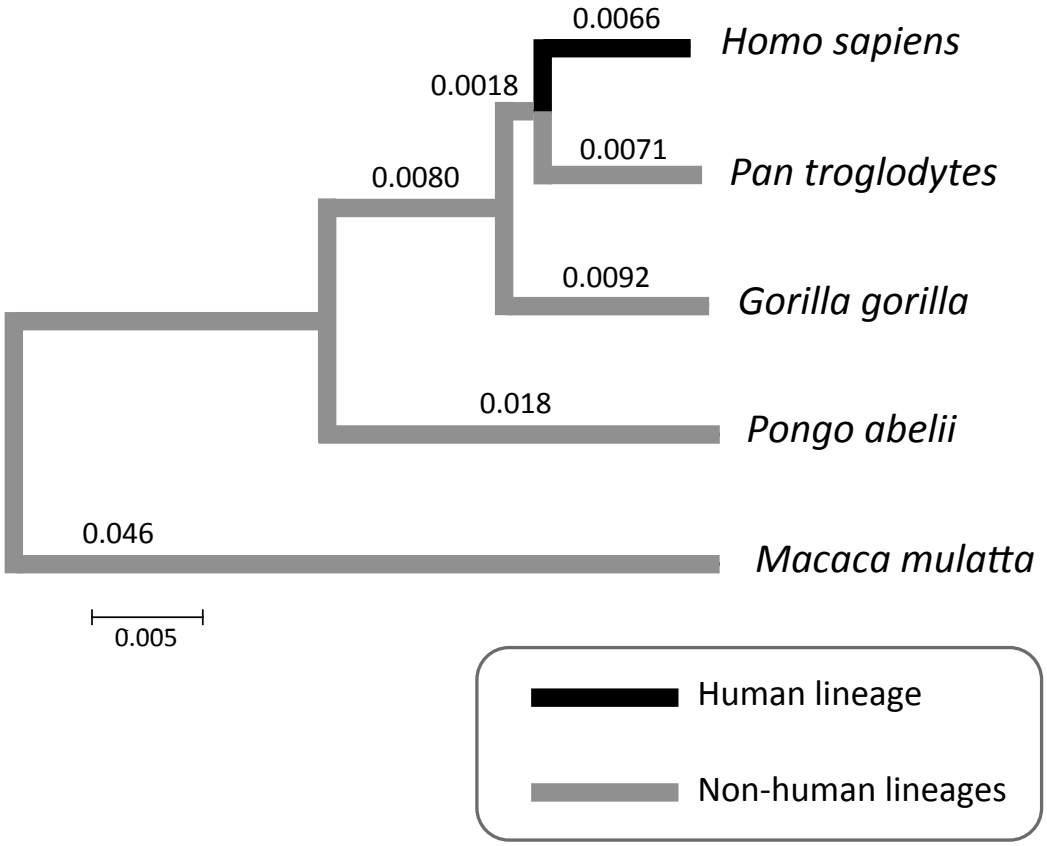


**Figure 4-1** Phylogenetic relationships between human, great apes, and rhesus macaque.

A phylogenetic tree was constructed using the neighbor-joining method and the pair-wise nucleotide divergence of whole genome sequences (Scally, et al. 2012). The scale bar represents 0.005 substitutions per site. The distance on each branch was calculated by the Fitch-Margoliash algorithm (Fitch and Margoliash 1967) using the pair-wise nucleotide divergence.

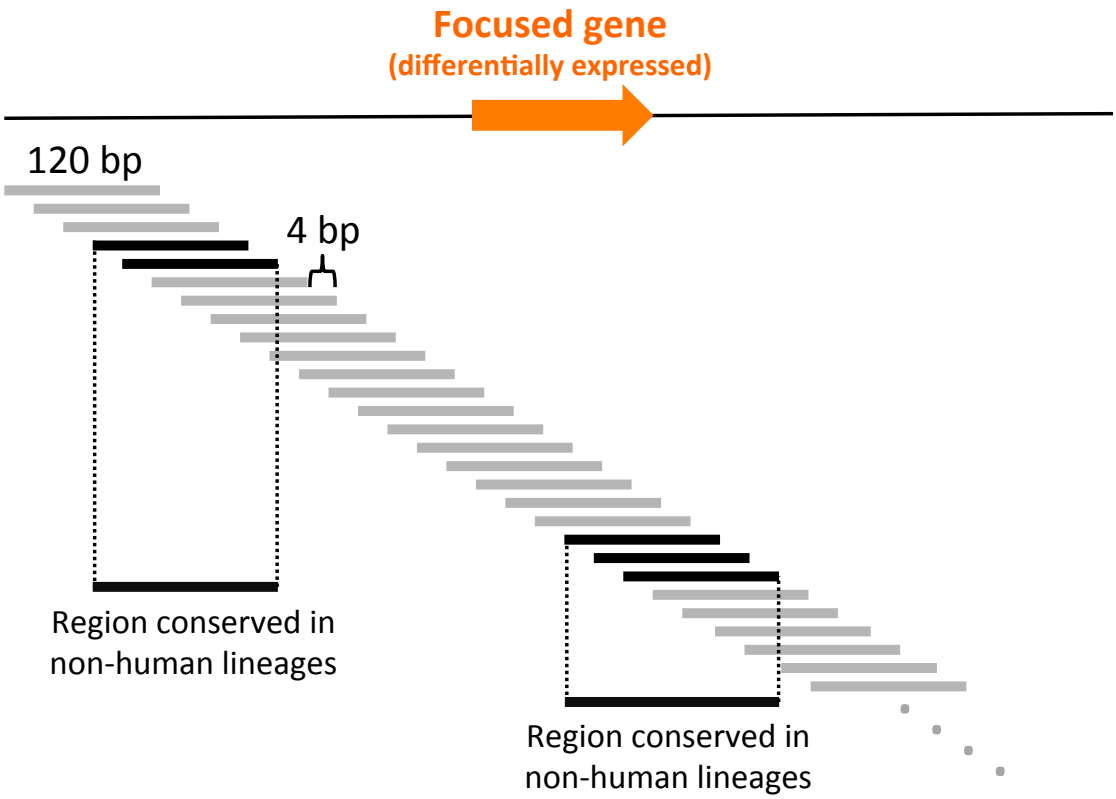


**Figure 4-2** A schematic representation of sliding-window analysis.

The gene of interest is shown by an orange arrow. Gray and black short horizontal lines indicate the positions of windows. The colors of gray and black represent non-conserved and conserved regions in non-human lineages, respectively. When the multiple conserved regions were continuous, these regions were concatenated into a single conserved region.

nucleotide divergence values for noncoding sequences. The 120-bp regions with the significantly smaller numbers of substitutions in non-human lineages than expected under a Poisson distribution ($p < 0.05$) were identified as the conserved regions. When multiple conserved regions were continuous, these regions were concatenated into a single conserved region (Figure 4-2).

Subsequently, conserved regions completely overlapping with exonic regions were eliminated by referring to exonic positions in the human genome [UCSC Genome Browser (https://genome.ucsc.edu), Human GRCh38/hg38]. I then extracted conserved regions harboring human-specific substitutions in noncoding regions.

Next, I selected regions with histone modifications (H3K4m1, H3K4m2, H3K4m3, H3K9ac, H3K27ac) for active regulatory regions from the conserved regions with human-specific substitutions identified above. ChIP-seq data for two skin cell strains, the NHEK (normal human epidermal keratinocytes) and NHDF-Ad (normal human dermal fibroblasts from adult skin), in the UCSC Genome Browser was used. Each gene of interest is expressed in either or both of these skin cell strains (Has and Nystrom 2015; Iivanainen, et al. 1995; Li, et al. 2013; Saarela, et al. 1998). The histone modifications around the transcription start sites of neighboring genes are expected to regulate the transcription of those genes, but not the genes of interest. Therefore, conserved regions with such histone modifications were not selected. The human-specific substitutions in the selected regions were estimated to be the candidate substitutions responsible for the human-specific expression patterns in the four genes of interest. The ancestral allele frequencies at the candidate substitution loci in human populations were investigated using the 1000 Genomes Project data (phase 3) in Ensembl (https://asia.ensembl.org/index.html).

**Evolutionary analyses and TF-binding site search for the candidate substitutions**

I assumed that the candidate substitutions most likely to change the expression levels of the genes of interest would be (1) in highly conserved 120-bp regions; or (2) in conserved 120-bp regions with the larger numbers of substitutions in the human lineage (Figure 4-1, black line). Among the most conserved 120-bp regions for each candidate substitution, regions with the significantly smaller numbers of substitutions in non-human lineages than expected ($p < 0.01$, Poisson distribution) were regarded as matches to condition (1), above. Next, I focused on the conserved 120-bp regions with the largest numbers of substitutions in the human lineage for each candidate substitution. The expected numbers of substitutions in the human lineage in each region were calculated from the numbers of substitutions in non-human lineages in the same 120-bp regions, according to the ratio of the numbers of substitutions in non-human and the human lineage(s) in each analyzed genomic region. The 120-bp regions with the significantly larger numbers of substitutions in the human lineage than expected ($p < 0.05$, Poisson distribution) were regarded as matches to condition (2), above.

The 51-bp sequence regions in which each candidate substitution locus was located at the center were screened for TF-binding sites using the JASPAR 2016 database (Mathelier, et al. 2015). I screened two sequences that differed at one base pair in the candidate substitution locus: (1) the human sequence with the human-specific allele; and (2) the human sequence with the ancestral allele. Relative scores in the JASPAR database were used to show the similarity with the consensus sequences of TF-binding sites.

## Results

I inferred substitutions in transcriptional regulatory regions responsible for the expression differences between human and great ape skin in the four structural protein genes of interest (i.e., *COL18A1*, *LAMB2*, *CD151*, and *BGN*). Transcriptional regulatory regions are expected to be conserved noncoding regions due to functional constraint (He, et al. 2011; Pennacchio, et al. 2006). Substitutions responsible for human-specific gene expression patterns are expected to be human-specific among the four primate species. Therefore, I identified regions that 1) were noncoding and conserved in non-human lineages (Figure 4-1, gray lines) and 2) harbored human-specific substitutions.

The genomic sequence alignments of human, chimpanzee, gorilla, and orangutan were used for this analysis. I hypothesized that the expression patterns of the four genes of interest in the skin of one Old World monkey species, the rhesus macaque (*M. mulatta*), would be similar to those of the three great ape species, and included the genomic sequence of *M. mulatta* in the multiple primate sequence alignments to improve detection of conserved regions. I intended to infer transcriptional regulatory regions located at short distance from each gene of interest. In general, transcriptional regulatory regions located at long distance from a target gene of interest are difficult to infer accurately due to the increased possibility that the inferred region is part of the regulatory network of a non-target neighboring gene. The analyzed genomic regions were set to include intergenic regions adjacent to the genes of interest and to locate target genes in the center of the analyzed region. When adjacent genes were close to the genes of interest (*CD151* and *LAMB2*), I set the analyzed regions as 100 kb to increase the lengths of regions under analysis. As a result, the size of my analyzed genomic regions was 372 kb, 100 kb, 100 kb, and 78 kb in

length for *COL18A1*, *LAMB2*, *CD151*, and *BGN*, respectively.

In my comparative analysis, genetic distances of noncoding regions within the analyzed genomic regions between species were similar to the average divergence based on whole genome sequences (Scally, et al. 2012) (Table 4-1). I designated regions as conserved if they showed the significantly smaller numbers of substitutions compared to the divergence of the analyzed genomic region ($p < 0.05$, Poisson distribution, $\mu$=11.31, 11.02, 11.25, and 11.16 for *COL18A1*, *LAMB2*, *CD151*, and *BGN*, respectively). I identified such conserved regions with a 120-bp sliding-window analysis throughout the analyzed genomic regions. Subsequently I eliminated conserved regions completely overlapping with exonic regions from the analysis. I then extracted regions harboring human-specific substitutions in noncoding regions from the conserved regions. As a result, the numbers of extracted regions finally obtained were 49, 39, 10, and 32 for *COL18A1*, *LAMB2*, *CD151*, and *BGN*, respectively (Figure 4-3, black and orange vertical lines).

The activity of a transcriptional regulatory region differs among diverse cell types in a multicellular organism (Heintzman, et al. 2009). Active promoters and enhancers generally possess specific histone modifications (H3K4m1, H3K4m2, H3K4m3, H3K9ac, and H3K27ac). From my conserved regions with human-specific substitutions, I selected regions with these histone modifications in human skin cells. As a result, the numbers of selected regions were one, six, two, and seven for *COL18A1*, *LAMB2*, *CD151*, and *BGN*, respectively (Figure 4-3, Figure 4-4, orange vertical lines). In addition to these regions, I also selected two regions each for *COL18A1* and *LAMB2*. Region I and II for *COL18A1* were near (~2 kb proximity) the histone modifications around the transcription start site of this target gene, making it likely that they

**Table 4-1** Genetic distances of the analyzed genomic regions and the whole genomes

| | Human | Chimpanzee | Gorilla | Orangutan |
|---|---|---|---|---|
| **Analyzed region**[1)] | | | | |
| ***BGN*** | | | | |
| Chimpanzee | 0.0137±0.0004 | | | |
| Gorilla | 0.0148±0.0004 | 0.0190±0.0004 | | |
| Orangutan | 0.0330±0.0006 | 0.0372±0.0008 | 0.0339±0.0006 | |
| Macaque | 0.0636±0.0008 | 0.0677±0.0007 | 0.0642±0.0007 | 0.0656±0.0007 |
| ***COL18A1*** | | | | |
| Chimpanzee | 0.0147±0.0002 | | | |
| Gorilla | 0.0189±0.0002 | 0.0185±0.0002 | | |
| Orangutan | 0.0363±0.0002 | 0.0359±0.0002 | 0.0366±0.0003 | |
| Macaque | 0.0660±0.0004 | 0.0657±0.0003 | 0.0667±0.0004 | 0.0676±0.0003 |
| ***CD151*** | | | | |
| Chimpanzee | 0.0138±0.0006 | | | |
| Gorilla | 0.0183±0.0007 | 0.0176±0.0004 | | |
| Orangutan | 0.0364±0.0008 | 0.0355±0.0009 | 0.0374±0.0007 | |
| Macaque | 0.0655±0.0013 | 0.0645±0.0013 | 0.0659±0.0013 | 0.0684±0.0012 |
| ***LAMB2*** | | | | |
| Chimpanzee | 0.0106±0.0005 | | | |
| Gorilla | 0.0182±0.0008 | 0.0172±0.0007 | | |
| Orangutan | 0.0327±0.0007 | 0.0320±0.0007 | 0.0355±0.0010 | |
| Macaque | 0.0645±0.0009 | 0.0638±0.0009 | 0.0661±0.0011 | 0.0677±0.0013 |
| **Whole genome**[2)] | | | | |
| Chimpanzee | 0.0137 | | | |
| Gorilla | 0.0175 | 0.0181 | | |
| Orangutan | 0.0340 | 0.0344 | 0.0350 | |
| Macaque | 0.0623 | 0.0627 | 0.0629 | 0.0635 |

1) Excluding exonic and unaligned regions. The numbers of substitutions per site between the analyzed sequences of the species and their standard errors are shown.

2) The numbers of substitutions per site between the whole genome sequences of the species (Scally, et al. 2012).

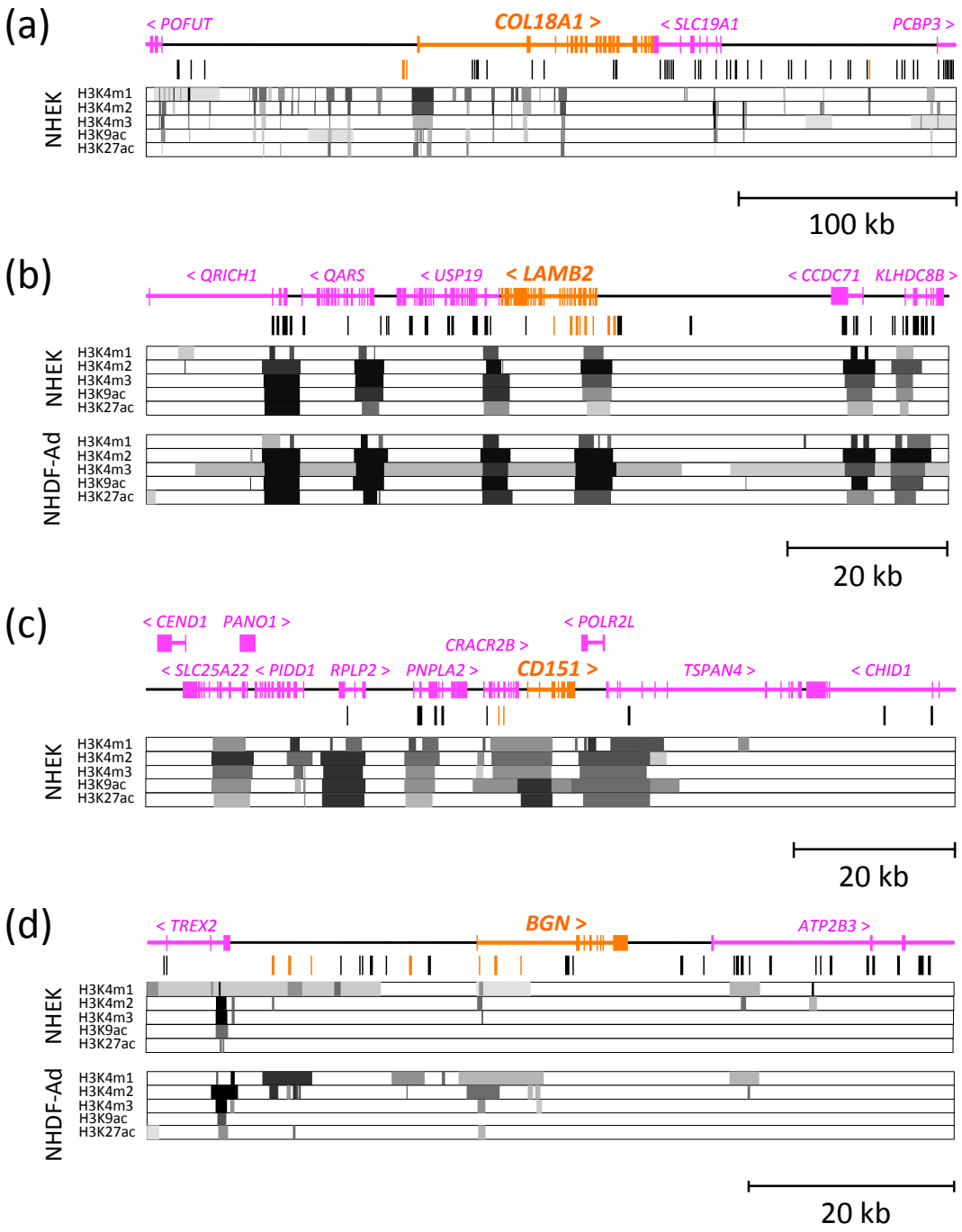


**Figure 4-3** The positions of the conserved regions with human-specific substitutions.

The exons of (a) *COL18A1*, (b) *LAMB2*, (c) *CD151*, and (d) *BGN* genes are shown in orange in each panel. The exons of genes (except the genes of interest) are shown in pink. The symbols "<" and ">" indicate the direction of the genes. Black vertical lines indicate the positions of the conserved regions with human-specific substitutions, and orange vertical lines indicate the

positions of those with the histone modifications. H3K4m1, H3K4m2, H3K4m3, H3K9ac, and H3K27ac indicate mono-methylation of histone H3 lysine 4, di-methylation of histone H3 lysine 4, tri-methylation of histone H3 lysine 4, acetylation of histone H3 lysine 9, and acetylation of histone H3 lysine 27, respectively. The positions and density of the gray scale bars indicate the positions and intensity of histone modifications shown at the side of the bars, respectively. The skin cell strains referred to for the histone modifications are shown in the left side of the abbreviations of the histone modification names.

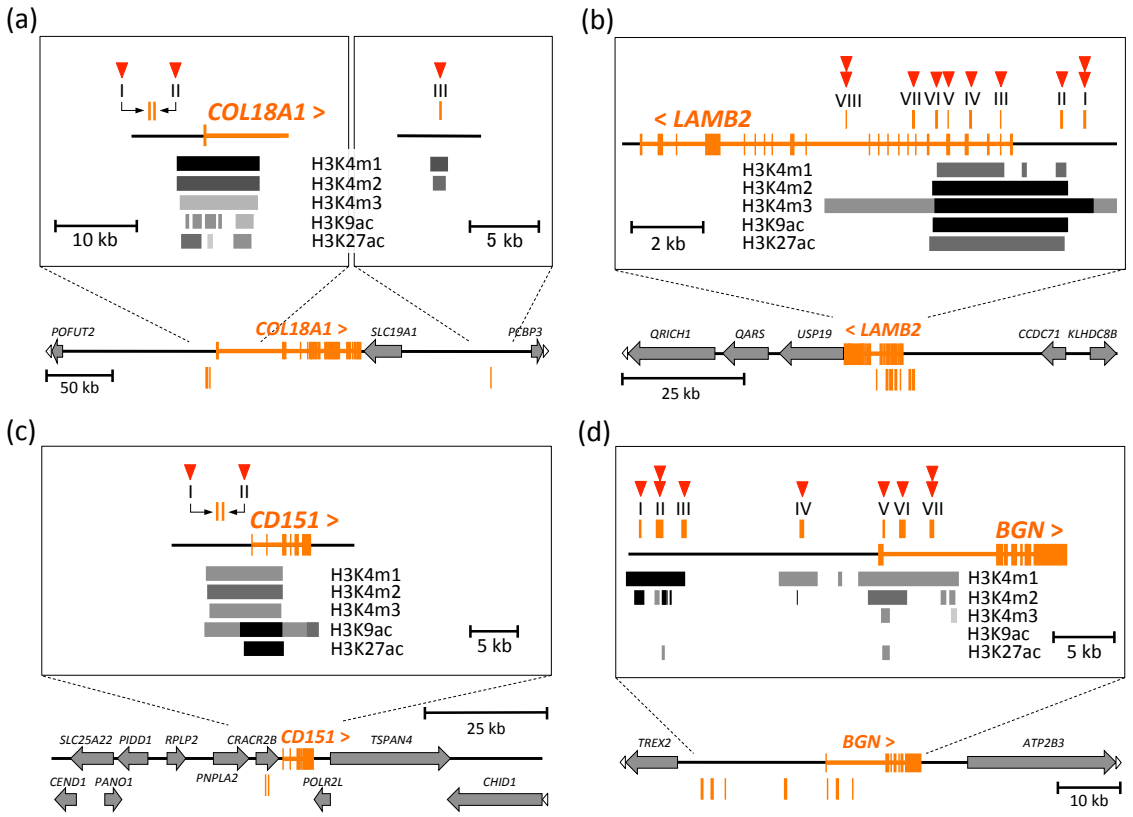


**Figure 4-4** The positions of the putative transcriptional regulatory regions with the candidate substitutions.

The exons of (a) *COL18A1*, (b) *LAMB2*, (c) *CD151*, and (d) *BGN* genes are shown in orange in each panel. The symbols "<" and ">" indicate the direction of the genes. All genes (except genes of interest) are shown by gray arrows. White arrowheads at the end of horizontal lines indicate that the genes are continuous beyond the schematic representation. Orange vertical lines indicate the position of the putative transcriptional regulatory regions numbered by Roman numerals. The zoomed-in view around these regions are shown in the rectangles of each panel. H3K4m1, H3K4m2, H3K4m3, H3K9ac, and H3K27ac indicate mono-methylation of histone H3 lysine 4, di-methylation of histone H3 lysine 4, tri-methylation of histone H3 lysine 4, acetylation of histone H3 lysine 9, and acetylation of histone H3 lysine 27, respectively. The positions and density of the gray scale bars indicate the positions and intensity of histone modifications shown at the side of the bars, respectively. The skin cell strains referred to for the histone modifications were as below: a: NHEK, b: NHDF-Ad, c: NHEK, and d: NHDF-Ad. For *LAMB2* and *BGN*, we also referred to the NHEK strain, in which the genes are also expressed (Figure 4-3b and 4-3d). The red arrowheads indicate the numbers of the candidate substitutions located in each putative

transcriptional regulatory region. A substitution in *COL18A1* region I was later removed from the candidate substitutions because of the high ancestral allele frequency in the locus.

regulate the expression of *COL18A1* (Figure 4-4a). Moreover, these regions showed di-methylation of histone H3 lysine 79 (H3K79m2), which is known to correlate with multiple roles including activating regulatory regions (Farooq, et al. 2016). Region VII and VIII for *LAMB2* were near (~2 kb proximity) four histone modifications (H3K4m1, H3K4m2, H3K9ac, and H3K27ac), and in a weak histone modification (H3K4m3) (Figure 4-4b). These two regions possessed mono-methylation of histone H3 lysine 9 (H3K9m1), which may be associated with activating regulatory regions (Barski, et al. 2007), as well as H3K79m2. The selected regions were assumed to be putative transcriptional regulatory regions for each of the genes of interest. Human-specific substitutions in these regions were regarded as candidate substitutions that would result in human-specific gene expression patterns. In total, the numbers of the candidate substitutions were three, ten, two, and nine for *COL18A1*, *LAMB2*, *CD151*, and *BGN*, respectively (Figure 4-4).

Among the 24 candidate substitution loci for the genes of interest, the ancestral alleles were found at one and two loci for *CD151* and *LAMB2*, respectively, with low frequencies (ancestral allele frequencies: 0.0126 for *CD151*; 0.0002 and 0.0016 for *LAMB2*) in human populations, based on data from the 1000 Genomes Project (Table 4-2). The human-specific alleles at these loci are not fixed completely. However, they could still have a possibility to be responsible for the observed human-specific expression patterns in the genes because the ancestral alleles could remain in low frequencies when the human-specific alleles make a small contribution to the human-specific expression patterns. Therefore, I included these almost-fixed mutations in the candidate substitutions putatively responsible for the human-specific expression patterns. On the other hand, an ancestral allele was found at the candidate substitution locus in

**Table 4-2** Chromosomal locations of the putative transcriptional regulatory regions and candidate substitutions

| Region | Length (bp) | Candidate substitution[1)] | Chromosomal location[2)] | |
|---|---|---|---|---|
| | | | Region | Candidate substitution |
| ***BGN*** | | | | |
| I | 140 | S1 | chrX:153,475,169-153,475,315 | 153,475,205 |
| II | 196 | S1<br>S2 | chrX:153,476,720-153,476,916 | 153,476,830<br>153,476,874 |
| III | 204 | S1 | chrX:153,478,882-153,479,085 | 153,479,000 |
| IV | 204 | S1 | chrX:153,488,428-153,488,631 | 153,488,559 |
| V | 120 | S1 | chrX:153,495,141-153,495,260 | 153,495,158 |
| VI | 240 | S1 | chrX:153,496,681-153,496,920 | 153,496,912 |
| VII | 120 | S1<br>S2 | chrX:153,499,138-153,499,257 | 153,499,169<br>153,499,170 |
| ***COL18A1*** | | | | |
| II | 144 | S1 | chr21:45,399,049-45,399,192 | 45,399,144 |
| III | 148 | S1 | chr21:45,612,482-45,612,629 | 45,612,532 |
| ***CD151*** | | | | |
| I | 212 | S1 | chr11:829,466-829,677 | 829,573[*1] |
| II | 160 | S1 | chr11:830,015-830,174 | 830,152 |
| ***LAMB2*** | | | | |
| I | 200 | S1<br>S2 | chr3:49,135,206-49,135,407 | 49,135,210<br>49,135,379 |
| II | 176 | S1 | chr3:49,134,445-49,134,624 | 49,134,566 |
| III | 212 | S1 | chr3:49,132,491-49,132,702 | 49,132,689 |
| IV | 268 | S1 | chr3:49,131,511-49,131,778 | 49,131,528 |
| V | 120 | S1 | chr3:49,130,814-49,130,933 | 49,130,878 |
| VI | 224 | S1 | chr3:49,130,584-49,130,809 | 49,130,667 |
| VII | 296 | S1 | chr3:49,129,774-49,130,069 | 49,130,023[*2] |
| VIII | 136 | S1<br>S2 | chr3:49,127,681-49,127,824 | 49,127,690<br>49,127,811[*3] |

1) S1 and S2 represent substitutions 1 and 2 in one region, respectively.

2) Location in human genome GRCh38/hg38.

[*]The ancestral alleles were found in human populations at low frequencies (ancestral allele frequencies: [*1]0.0126; [*2]0.0002; [*3]0.0016).

*COL18A1* region I with high frequency (ancestral allele frequency: 0.683). This mutation was thus removed from the list of candidate substitutions putatively responsible for the human-specific gene expression patterns. The expression changes of the genes of interest in the human lineage may be attributable to the independent or combined effects of these candidate substitutions. The positions of the putative transcriptional regulatory regions and the candidate substitutions within those regions in the human genome are shown in Table 4-2.

I further explored the possibility that these candidate substitutions change the gene expression levels by using two independent evolutionary analyses. I focused on conserved 120-bp regions ($p < 0.05$, Poisson distribution) where each candidate substitution was located. First, I assumed that regions conserved more in non-human lineages ($p < 0.01$, Poisson distribution) than the other regions ($0.01 \leq p < 0.05$, Poisson distribution) were likely to have a role in gene expression regulation. Among the most conserved 120-bp regions for each candidate substitution, three and two regions for *BGN* and *LAMB2*, respectively, matched with this condition. They contained four and two candidate substitutions for *BGN* and *LAMB2*, respectively (Table 4-3, bold letters). Second, I assumed that the conserved 120-bp regions with the significantly larger numbers of substitutions in the human lineage than expected from the numbers of substitutions in non-human lineages in the same 120-bp regions ($p < 0.05$, Poisson distribution) are likely to change their functions in gene expression regulation. Among the conserved 120-bp regions with the largest numbers of substitutions in the human lineage for each candidate substitution, two regions for each of *BGN* and *CD151* and three regions for *LAMB2* had the significantly larger numbers of substitutions (two or three substitutions) in the human lineage. They contained four, two, and four candidate substitutions for *BGN*, *CD151*, and *LAMB2*, respectively (Table 4-3, bold

**Table 4-3** The numbers of substitutions in conserved 120-bp regions with candidate substitutions

| Region | Candidate substitution[1)] | Substitutions[2)] in non-human lineages | Substitutions[3)] in the human lineage |
|---|---|---|---|
| ***BGN*** | | | |
| I | S1 | 3.79* | 1 |
| II | S1<br>S2 | **1.01**[4)]** | **2**[4)]** |
| III | S1 | **2.52**** | 1 |
| IV | S1 | **2.02**** | 1 |
| V | S1 | 4.05* | 1 |
| VI | S1 | 4.05* | 1 |
| VII | S1<br>S2 | 4.06[4)]* | **2**[4)]* |
| ***COL18A1*** | | | |
| II | S1 | 3.03* | 1 |
| III | S1 | 3.80* | 1 |
| ***CD151*** | | | |
| I | S1 | 3.03* | **2*** |
| II | S1 | 3.03* | **2*** |
| ***LAMB2*** | | | |
| I | S1 | 4.09* | 1 |
| | S2 | 4.06* | 1 |
| II | S1 | 3.04* | 1 |
| III | S1 | 3.04* | 1 |
| IV | S1 | 3.03* | **2*** |
| V | S1 | 4.06* | 1 |
| VI | S1 | **1.00**** | **2*** |
| VII | S1 | **2.02**** | 1 |
| VIII | S1<br>S2 | 4.25[4)]* | **3**[4)]** |

1) S1 and S2 represent substitutions 1 and 2 in one region, respectively.

2) The 120-bp regions with the smallest numbers of substitutions were selected for each candidate substitution.

3) The 120-bp regions with the largest numbers of substitutions were selected for each candidate substitution.

4) Two substitutions (S1, S2) were located in the same 120-bp region.

Bold letters: significantly highly conserved regions in non-human lineages ($p < 0.01$, Poisson distribution) or conserved regions with the significantly larger numbers of substitutions in the human lineage ($p < 0.05$, Poisson distribution).

*: $p < 0.05$, **: $p < 0.01$, Poisson distribution.

letters). I hypothesize that the candidate substitutions in the conserved 120-bp regions indicated by these two evolutionary analyses have a high likelihood of changing target gene expression.

In addition to the evolutionary analyses above, I searched for TF-binding sites that contain each candidate substitution locus. Candidate substitutions in TF-binding sites have a high likelihood of changing target gene expression levels. Searching for TF-binding sites showed that all of the candidate substitutions were located in TF-binding sites (Table 4-4). Especially, one, six, two, and five candidate substitutions for *COL18A1*, *LAMB2*, *CD151*, and *BGN*, respectively, were in binding sites of TFs that are reported to function in skin and were also expressed based on my RNA-seq analyses (average normalized RPKM values for humans and great apes $\geq 1$ in the mapping result of the human reference genome) (Table 4-5). The candidate substitutions from the ancestral alleles to the human-specific alleles changed the similarity with the consensus sequences of binding sites for such TFs associated with skin function (Table 4-5). However, without further experimentation it remains difficult to infer whether these differences in similarity values would affect the TF-binding affinities.

**Table 4-4** TF classes and the numbers of TF-binding sites that contain candidate substitution loci

| Candidate substitution[1)] | TF class | Ancestral allele | Human-specific allele |
|---|---|---|---|
| ***BGN*** | | | |
| I (S1) | C2CH THAP-type zinc finger factors | 1 | 1 |
| | C2H2 zinc finger factors | 2 | 3 |
| | Fork head/winged helix factors | 2 | 3 |
| | High-mobility group (HMG) domain factors | 0 | 3 |
| | Homeo domain factors | 2 | 1 |
| | SMAD/NF-1 DNA-binding domain factors | 2 | 2 |
| II (S1) | Basic helix-loop-helix factors (bHLH) | 0 | 6 |
| | C2H2 zinc finger factors | 1 | 1 |
| | Homeo domain factors | 15 | 14 |
| | SMAD/NF-1 DNA-binding domain factors | 1 | 1 |
| | Tryptophan cluster factors | 1 | 0 |
| II (S2) | Basic helix-loop-helix factors (bHLH) | 2 | 0 |
| | Basic leucine zipper factors (bZIP) | 1 | 2 |
| | Basic leucine zipper factors (bZIP)::Basic leucine zipper factors (bZIP) | 0 | 2 |
| | Homeo domain factors | 15 | 2 |
| | Paired box factors | 0 | 1 |
| III (S1) | Basic helix-loop-helix factors (bHLH) | 1 | 6 |
| | C2CH THAP-type zinc finger factors | 1 | 0 |
| | C2H2 zinc finger factors | 2 | 1 |
| | Homeo domain factors | 0 | 1 |
| | Tryptophan cluster factors | 0 | 8 |
| | Nuclear receptors with C4 zinc fingers | 1 | 0 |
| | SMAD/NF-1 DNA-binding domain factors | 1 | 0 |
| IV (S1) | Basic helix-loop-helix factors (bHLH) | 1 | 2 |
| | Basic leucine zipper factors (bZIP) | 3 | 0 |
| | Basic leucine zipper factors (bZIP)::Basic leucine zipper factors (bZIP) | 3 | 0 |
| | C2H2 zinc finger factors | 1 | 1 |
| | Homeo domain factors | 7 | 0 |
| | Nuclear receptors with C4 zinc fingers | 1 | 0 |
| | Paired box factors | 0 | 1 |
| | STAT domain factors | 1 | 1 |
| | TEA domain factors | 1 | 0 |
| V (S1) | Basic helix-loop-helix factors (bHLH) | 1 | 4 |
| | Basic helix-loop-helix factors (bHLH)::Basic helix-loop-helix factors (bHLH) | 1 | 0 |
| | Basic helix-span-helix factors (bHSH) | 9 | 10 |
| | Basic leucine zipper factors (bZIP)::Basic leucine zipper factors (bZIP) | 1 | 0 |
| | C2H2 zinc finger factors | 1 | 4 |
| | Nuclear receptors with C4 zinc fingers | 1 | 1 |
| | Nuclear receptors with C4 zinc fingers::Nuclear receptors with C4 zinc fingers | 1 | 1 |
| | Paired box factors | 1 | 1 |
| | Runt domain factors | 1 | 0 |
| | SMAD/NF-1 DNA-binding domain factors | 1 | 1 |
| | TEA domain factors | 1 | 0 |

| | | | |
|---|---|---|---|
| | Tryptophan cluster factors | 1 | 0 |
| VI (S1) | Basic helix-loop-helix factors (bHLH) | 0 | 2 |
| | Basic leucine zipper factors (bZIP) | 0 | 1 |
| | Basic leucine zipper factors (bZIP)::Basic leucine zipper factors (bZIP) | 0 | 1 |
| | C2H2 zinc finger factors | 13 | 5 |
| | Fork head/winged helix factors | 1 | 1 |
| | Homeo domain factors | 0 | 6 |
| | Nuclear receptors with C4 zinc fingers | 1 | 0 |
| | Rel homology region (RHR) factors | 1 | 0 |
| | T-Box factors | 0 | 1 |
| | TEA domain factors | 1 | 2 |
| | Tryptophan cluster factors | 0 | 1 |
| VII (S1, S2)* | Basic helix-loop-helix factors (bHLH) | 3 | 2 |
| | Basic helix-span-helix factors (bHSH) | 2 | 0 |
| | Basic leucine zipper factors (bZIP)::Basic leucine zipper factors (bZIP) | 1 | 0 |
| | C2H2 zinc finger factors | 6 | 1 |
| | Homeo domain factors | 6 | 7 |
| | MADS box factors | 1 | 1 |
| | Nuclear receptors with C4 zinc fingers | 1 | 0 |
| | SMAD/NF-1 DNA-binding domain factors | 0 | 1 |
| | Tryptophan cluster factors | 0 | 10 |
| ***COL18A1*** | | | |
| II (S1) | ARID domain factors | 0 | 1 |
| | Basic helix-loop-helix factors (bHLH) | 7 | 0 |
| | Basic leucine zipper factors (bZIP) | 1 | 3 |
| | Basic leucine zipper factors (bZIP)::Basic leucine zipper factors (bZIP) | 1 | 1 |
| | Fork head/winged helix factors | 1 | 2 |
| | High-mobility group (HMG) domain factors | 0 | 3 |
| | Homeo domain factors | 13 | 69 |
| | Nuclear receptors with C4 zinc fingers | 1 | 0 |
| | STAT domain factors | 1 | 1 |
| | TEA domain factors | 0 | 1 |
| III (S1) | Homeo domain factors | 3 | 2 |
| | MADS box factors | 1 | 1 |
| | Nuclear receptors with C4 zinc fingers::Nuclear receptors with C4 zinc fingers | 1 | 1 |
| | TATA-binding proteins | 0 | 1 |
| ***CD151*** | | | |
| I (S1) | Basic helix-loop-helix factors (bHLH) | 0 | 16 |
| | Basic helix-loop-helix factors (bHLH)::Basic helix-loop-helix factors (bHLH) | 1 | 0 |
| | Basic helix-span-helix factors (bHSH) | 1 | 3 |
| | Basic leucine zipper factors (bZIP) | 3 | 3 |
| | C2CH THAP-type zinc finger factors | 1 | 1 |
| | C2H2 zinc finger factors | 3 | 4 |
| | Homeo domain factors | 1 | 4 |
| | Paired box factors | 3 | 1 |
| | SMAD/NF-1 DNA-binding domain factors | 2 | 4 |
| | STAT domain factors | 0 | 1 |

| | | | |
|---|---|---|---|
| II (S1) | Basic helix-loop-helix factors (bHLH) | 0 | 2 |
| | Basic helix-loop-helix factors (bHLH)::Basic helix-loop-helix factors (bHLH) | 1 | 0 |
| | C2H2 zinc finger factors | 9 | 10 |
| | Fork head/winged helix factors | 2 | 2 |
| | GCM domain factors | 0 | 1 |
| | Heteromeric CCAAT-binding factors | 1 | 1 |
| | Homeo domain factors | 1 | 0 |
| | Runt domain factors | 1 | 0 |
| | TEA domain factors | 0 | 1 |
| ***LAMB2*** | | | |
| I (S1) | Basic helix-loop-helix factors (bHLH) | 35 | 3 |
| | Basic helix-loop-helix factors (bHLH)::Basic helix-loop-helix factors (bHLH) | 1 | 0 |
| | Basic leucine zipper factors (bZIP)::Basic leucine zipper factors (bZIP) | 1 | 0 |
| | C2H2 zinc finger factors | 3 | 2 |
| | Homeo domain factors | 2 | 2 |
| | Nuclear receptors with C4 zinc fingers | 1 | 0 |
| | Tryptophan cluster factors | 1 | 0 |
| I (S2) | Basic helix-loop-helix factors (bHLH)::Basic helix-loop-helix factors (bHLH) | 0 | 1 |
| | Basic leucine zipper factors (bZIP)::Basic leucine zipper factors (bZIP) | 0 | 1 |
| | Homeo domain factors | 0 | 1 |
| II (S1) | ARID domain factors | 0 | 1 |
| | Basic helix-loop-helix factors (bHLH) | 1 | 0 |
| | Basic leucine zipper factors (bZIP) | 2 | 0 |
| | C2H2 zinc finger factors | 3 | 0 |
| | DM-type intertwined zinc finger factors | 1 | 1 |
| | Homeo domain factors | 2 | 4 |
| | Nuclear receptors with C4 zinc fingers | 1 | 5 |
| | Paired box factors | 1 | 1 |
| | Tryptophan cluster factors | 1 | 0 |
| III (S1) | Basic leucine zipper factors (bZIP) | 5 | 3 |
| | C2H2 zinc finger factors | 0 | 1 |
| | Homeo domain factors | 1 | 0 |
| | Rel homology region (RHR) factors | 0 | 1 |
| | STAT domain factors | 0 | 1 |
| | Tryptophan cluster factors | 2 | 18 |
| IV (S1) | Basic helix-loop-helix factors (bHLH) | 1 | 1 |
| | Basic helix-loop-helix factors (bHLH)::Basic helix-loop-helix factors (bHLH) | 1 | 1 |
| | C2CH THAP-type zinc finger factors | 1 | 2 |
| | C2H2 zinc finger factors | 6 | 7 |
| | GCM domain factors | 0 | 1 |
| | Homeo domain factors | 1 | 3 |
| | Nuclear receptors with C4 zinc fingers | 0 | 1 |
| | SMAD/NF-1 DNA-binding domain factors | 3 | 3 |
| | T-Box factors | 3 | 0 |
| | Tryptophan cluster factors | 1 | 0 |
| V (S1) | Basic helix-loop-helix factors (bHLH) | 5 | 3 |

| | | | |
|---|---|---|---|
| | Basic helix-loop-helix factors (bHLH)::Basic helix-loop-helix factors (bHLH) | 1 | 1 |
| | C2CH THAP-type zinc finger factors | 1 | 0 |
| | C2H2 zinc finger factors | 2 | 2 |
| | Homeo domain factors | 1 | 1 |
| | SMAD/NF-1 DNA-binding domain factors | 1 | 0 |
| | SMAD/NF-1 DNA-binding domain factors::SMAD/NF-1 DNA-binding domain factors::SMAD/NF-1 DNA-binding domain factors | 1 | 0 |
| | TEA domain factors | 0 | 2 |
| | Tryptophan cluster factors | 1 | 1 |
| VI (S1) | Basic helix-loop-helix factors (bHLH) | 2 | 0 |
| | Basic leucine zipper factors (bZIP) | 1 | 1 |
| | C2H2 zinc finger factors | 1 | 0 |
| | Fork head/winged helix factors | 0 | 1 |
| | Grainyhead domain factors | 1 | 0 |
| | Heteromeric CCAAT-binding factors | 2 | 1 |
| | Homeo domain factors | 2 | 35 |
| VII (S1) | Basic helix-loop-helix factors (bHLH) | 4 | 0 |
| | Basic helix-span-helix factors (bHSH) | 3 | 9 |
| | Basic leucine zipper factors (bZIP) | 3 | 5 |
| | Basic leucine zipper factors (bZIP)::Basic leucine zipper factors (bZIP) | 0 | 1 |
| | C2CH THAP-type zinc finger factors | 1 | 1 |
| | C2H2 zinc finger factors | 2 | 0 |
| | Homeo domain factors | 20 | 3 |
| | MADS box factors | 1 | 1 |
| | Other C4 zinc finger-type factors::Basic helix-loop-helix factors (bHLH) | 1 | 0 |
| | Rel homology region (RHR) factors | 4 | 2 |
| | Runt domain factors | 0 | 1 |
| | SMAD/NF-1 DNA-binding domain factors | 6 | 1 |
| | STAT domain factors | 2 | 2 |
| VIII (S1) | Basic helix-loop-helix factors (bHLH) | 13 | 3 |
| | Basic leucine zipper factors (bZIP) | 2 | 1 |
| | Basic leucine zipper factors (bZIP)::Basic leucine zipper factors (bZIP) | 1 | 1 |
| | C2H2 zinc finger factors | 0 | 1 |
| | Nuclear receptors with C4 zinc fingers | 4 | 6 |
| | Paired box factors | 1 | 0 |
| VIII (S2) | C2H2 zinc finger factors | 0 | 2 |
| | Homeo domain factors | 2 | 2 |
| | MADS box factors | 1 | 1 |
| | Tryptophan cluster factors | 0 | 1 |

1) S1 and S2 represent substitutions 1 and 2 in one region, respectively.

*These two substitutions (S1 and S2) are located next to each other. The 25 bp sequences on both sides of these substitutions were retrieved for the TF search. The binding sites of one of the tryptophan cluster factors and homeo domain factors contain only the "S2" substitution locus of the human-specific allele.

**Table 4-5** Candidate substitutions in binding sites of TFs with a function in skin

| Region | Candidate substitution[1)] | TF | Binding site[2)] | Relative score[3)] | |
|---|---|---|---|---|---|
| | | | | Ancestral allele | Human-specific allele |
| ***BGN*** | | | | | |
| II | S2 | HOXB2 | 22-31 (f) | 0.819 | 0.713 |
| III | S1 | FLI1 | 22-31 (r) | 0.701 | 0.827 |
| VI | S1 | KLF4 | 20-29 (f) | 0.808 | 0.719 |
| VII | S1, S2[4)] | SPDEF | 20-30 (f) | 0.595 | 0.803 |
| | | FLI1 | 21-30 (f) | 0.614 | 0.825 |
| ***COL18A1*** | | | | | |
| II | S1 | HOXB2 | 23-32 (f) | 0.755 | 0.821 |
| | | MSX2 | 24-31 (f) | 0.686 | 0.810 |
| | | PRRX2 | 24-31 (f) | 0.757 | 0.858 |
| ***CD151*** | | | | | |
| I | S1 | FOSL2 | 26-36 (f) | 0.859 | 0.841 |
| | | SNAI2 | 23-31 (f) | 0.763 | 0.902 |
| II | S1 | KLF4 | 21-30 (f) | 0.809 | 0.814 |
| ***LAMB2*** | | | | | |
| I | S1 | SNAI2 | 20-28 (r) | 0.971 | 0.876 |
| II | S1 | KLF4 | 25-34 (f) | 0.818 | 0.727 |
| | | SPDEF | 19-29 (r) | 0.842 | 0.731 |
| III | S1 | FLI1 | 20-29 (r) | 0.736 | 0.866 |
| V | S1 | SNAI2 | 20-28 (f) | 0.779 | 0.929 |
| | | SNAI2 | 24-32 (f) | 0.937 | 0.842 |
| VI | S1 | SNAI2 | 22-30 (r) | 0.840 | 0.785 |
| | | MSX2 | 21-28 (f) | 0.713 | 0.810 |
| VII | S1 | HOXB2 | 21-30 (r) | 0.854 | 0.748 |

1) S1 and S2 represent substitutions 1 and 2 in one region, respectively.

2) The numbers represent the positions within the 51-bp sequences retrieved for the TF search. Each candidate substitution is located at position 26. “f” and “r” indicate TF-binding site on forward and reverse strands, respectively.

3) This score shows the similarity with the consensus sequence of TF-binding site in the JASPAR database. The score changes by the candidate substitutions from the ancestral alleles to the human-specific alleles are shown.

4) These two substitutions (S1 and S2) are located next to each other. The 25 bp sequences on both sides of these substitutions were retrieved for the TF search (candidate substitutions: positions 26 and 27).

## Discussion

Substitutions in transcriptional regulatory regions can change the expression of their target genes (Wittkopp and Kalay 2012). I hypothesized that the substitutions responsible for the human-specific expression patterns in the four structural protein genes of interest (*COL18A1*, *LAMB2*, *CD151*, *BGN*) were (1) in noncoding regions that were conserved in non-human lineages and (2) specific to humans. Conserved regions in great ape and rhesus macaque lineages are difficult to find through sequence alignment, as the species are closely related and the sequences are largely identical. Therefore, I utilized a sliding-window analysis to identify regions that exhibited the significantly smaller numbers of substitutions than expected from the divergences of the each analyzed genomic region between the species.

The conserved noncoding regions with human-specific substitutions I identified were taken for the next analysis. I suggested that the human-specific substitutions in those regions with histone modifications for active regulatory regions in skin cells could be the candidate substitutions responsible for the human-specific expression patterns in the genes of interest.

Substitutions changing the expression of target genes in transcriptional regulatory regions generally alter the binding affinities for TFs that modulate the gene expression (Wittkopp and Kalay 2012). According to the TF-binding site searches, all of the candidate substitution loci were expected to be located in TF-binding sites, suggesting a possibility that the regions containing the candidate substitution loci are associated with gene expression regulation. Expected TFs listed in Table 4-5 are reported to function in skin, and therefore they are likely to regulate the expression of the genes of interest. Comparison in skin injury repair between *MSX2* null mutant and wild type mice suggested that MSX2 regulates the cellular competence of

keratinocytes and fibroblasts (Yeh, et al. 2009). Deletion of the *PRRX2* gene in mice reduced fetal fibroblast proliferation during wound healing (White, et al. 2003). *FOSL2* mutant mice caused skin barrier defects due to reduced expression of epidermal differentiation genes, and ectopic expression of *FOSL2* induced expression of those genes (Wurm, et al. 2015). Microarray analysis showed that *HOXB2* and *SPDEF* genes were highly expressed in the regenerating skin during tissue expansion (Yang, et al. 2011). The other three TFs function in skin and are somewhat associated with the genes of interest. Homozygous *KLF4* deletion mutant mice lose the skin barrier function (Segre, et al. 1999), and KLF4 accelerated epidermal barrier acquisition (Patel, et al. 2006). KLF4 regulates the expression of some members of the laminin family (Ghaleb and Yang 2017). The reepithelialization component was reduced during wound healing in *SNAI2* null mice (Hudson, et al. 2009), and SNAI2 is intrinsically linked to CD151 (Yin, et al. 2014). The homozygous deletion of the C-terminal transcriptional activation domain of the *Fli1* gene upregulates expression levels of the α chain genes for collagens I and III in mouse skin (Asano, et al. 2009). These collagens are known to bind to BGN (Chen and Birk 2013). Their predicted binding sites suggest that KLF4, SNAI2, and FLI1 are likely to regulate the expression of *LAMB2*, *CD151*, and *BGN*, respectively. The candidate substitutions from the ancestral alleles to the human-specific alleles changed the similarity with the consensus sequences of binding sites for the TFs associated with skin function. This result suggests that these substitutions may change the binding affinities for the predicted TFs and may change the expression levels in the genes of interest. In addition, it is likely that the candidate substitutions located in highly conserved regions in non-human lineages and in conserved regions with the larger numbers of substitutions in the human lineage than expected would change the expression levels in the genes of interest.

I suggest that the candidate substitutions in the putative transcriptional regulatory regions may cause the human-specific gene expression patterns that possibly lead to the human-specific characteristics in skin structure. In the near future, to examine whether these candidate substitutions are responsible for the expression differences between humans and great apes, I will conduct a promoter assay in skin cells using the putative transcriptional regulatory regions with the ancestral alleles and the human-specific alleles located at the candidate substitution loci. Identifying substitutions that may give humans adaptive skin characteristics through human-specific gene expression patterns will contribute to the understanding of how human-specific characteristics have been genetically acquired.

# Chapter 5

## General discussion and Perspectives

In this thesis, I suggest that the candidate substitutions in the putative transcriptional regulatory regions inferred may have caused the expression differences of four structural protein genes, *COL18A1*, *LAMB2*, *CD151*, and *BGN*, between human and great ape skin. These human-specific gene expression changes may contribute to the rete ridge formation and rich elastic fibers in human skin, which are possible adaptive skin characteristics specific to humans with less hair.

In the near future, I am planning to conduct a promoter assay using cultured skin cells to examine whether these candidate substitutions are responsible for the expression differences between humans and great apes. In the promoter assay, each putative transcriptional regulatory region with one or two candidate substitution(s) is inserted into upstream of the luciferase gene in a plasmid vector. For one putative transcriptional regulatory region, two human sequences that differ at nucleotide(s) only in the candidate substitution site(s) [i.e. human sequence with human-specific allele(s) or ancestral allele(s)] are used. The vectors are then transfected into skin cells, and the expression level of the luciferase gene is measured though luminescence. When there is expression difference between the vectors with two candidate substitutions, I will convert one of the human-specific alleles into the ancestral allele and again compare the expression. By investigating the candidate substitutions one by one in this way, I can identify substitutions responsible for the expression differences between humans and great apes.

Promotor assay can detect the transcriptional regulatory activity of a certain candidate region, but the gene regulated by the transcriptional regulatory region is predicted only based on the distance from the region to the surrounding genes. Therefore, I will analyze the candidate substitutions by a different experimental strategy. For the candidate substitutions with expression changes in promoter assay, I will investigate whether these substitutions influence the expression

of the genes of interest, but not other genes in the genome, in skin cells. A human-specific allele at the candidate locus in the genome of cultured human skin cells is converted into the ancestral allele by genome editing technique using CRISPR/Cas9 system (Sander and Joung 2014). The expression of the structural protein gene of interest is determined by quantitative PCR and compared between skin cells with a human-specific allele and ancestral allele at the candidate substitution locus. Identifying substitutions that may give humans adaptive skin characteristics through human-specific gene expression patterns will contribute to the understanding of how human-specific characteristics have been genetically acquired.

Humans show much less hair on their body, which is one of the most striking skin differences between humans and other primates. Human skin is thus exposed more directly to external environments compared to other primates. Since the skin plays an important role as protection against external insults, this fact likely induces human skin to evolve differently from the furred skin of other primates. Actually, this study implied that the human-specific expression patterns in the structural protein genes may lead to the rete ridge and rich elastic fibers possibly unique to human skin. These characteristics might contribute to the physical strength of skin to protect the inside of the body against physical stresses under less hair. Further studies may find more association between human-specific skin characteristics and reduced amount of hair in humans.

There are many phenotypic characteristics that are unique to the human lineage. However, the genetic causes that underlie human-specific characteristics remain poorly understood. In my PhD study, I focused on human-specific skin characteristics and studied the genetic basis that underlies these phenotypes. I started with quantitative histological comparison

between human and other primate skin to know possible human-specific characteristics in skin structure. I then identified genes with human-specific expression patterns in skin that may lead to the human-specific characteristics observed in the histological skin comparison. For these genes, I further inferred substitutions responsible for the human-specific expression patterns. In this way, I integrated those three approaches to elucidate the genetic basis that underlie human-specific characteristics based on skin. I hope that my PhD study will provide further insight into the human evolution.

## Acknowledgments

I would like to express my thanks to Dr. Akiko Matsumoto-Oda, Dr. Atunga Nyachieo, Dr. Daniel Chai, and Dr. Ngalla Jillani for providing Old World monkey skin tissues. Without these skin samples, I could not discuss the histological differences between human and other primate skin at all. To know histological characteristics in human skin is essential for inferring phenotypic results of the human-specific gene expression patterns identified in my RNA-seq analyses.

I would like to extend my thanks to Dr. Kenzo Takahashi and Dr. Daisuke Utsumi for providing the digital photographs of dissected skin. Without their histological technique, I could not investigate skin characteristics in humans and other primates. They are dermatologists and also gave me much knowledge of skin including their experience as doctors, which was very helpful for me to study on skin characteristics.

I would also like to thank Dr. Hiroo Imai for managing and providing the skin samples of great apes. Without these skin samples, I could not investigate gene expression levels in great ape skin and identify the human-specific expression patterns compared with them. He also provided helpful discussion and gave me a good opportunity for presentation as one of the studies that focus on human-specific characteristics compared with other primates, in an annual meeting of Primate Society of Japan.

I am sincerely grateful to my supervisors, Dr. Yoko Satta and Dr. Yohey Terai, for their entire support to my study. Dr. Yoko Satta and Dr. Yohey Terai directed me with deep knowledge of evolutionary analyses and molecular biological experiments, respectively. I could identify the human-specific gene expression patterns by RNA-seq experiment and further infer the candidate

substitutions responsible for these expression patterns by methods including evolutionary analyses. I think combination of the two approaches made my study more significant for understanding genetic basis of human-specific characteristics. I would also like to thank the members of Satta lab and ESB, for their helpful comments on my study and an enjoyable time in the university.

This study was supported by an internal SOKENDAI grant to Dr. Yoko Satta and Dr. Yohey Terai and the Cooperative Research Program of Primate Research Institute, Kyoto University. This study was supported in part by SOKENDAI and the GAIN. This study was supported by JSPS KAKENHI Grant Number JP26650172 to Dr. Akiko Matsumoto-Oda. I thank the following facilities for providing great ape skin samples: Kumamoto sanctuary, Himeji Central Park, Fukuoka City Zoological Garden, Kamine Zoo, Kobe Oji Zoo, Kyoto City Zoo, Osaka-city Tennoji Zoo, and Sapporo Maruyama Zoo. These samples were provided through the GAIN. I thank the Institute of Primate Research, National Museum of Kenya for help with collection of Old World monkey skin samples.

## References

Ali AT, Hochfeld WE, Myburgh R, Pepper MS 2013. Adipocyte and adipogenesis. European journal of cell biology 92: 229-236.

Asano Y, et al. 2009. Transcription factor Fli1 regulates collagen fibrillogenesis in mouse skin. Mol Cell Biol 29: 425-434. doi: 10.1128/MCB.01278-08

Baggerly KA, Deng L, Morris JS, Aldaz CM 2003. Differential expression in SAGE: accounting for normal between-library variation. Bioinformatics 19: 1477-1483.

Baroni A, et al. 2012. Structure and function of the epidermis related to barrier properties. Clinics in dermatology 30: 257-262.

Barski A, et al. 2007. High-resolution profiling of histone methylations in the human genome. Cell 129: 823-837.

Benhamed S, Guardiola FA, Mars M, Esteban MÁ 2014. Pathogen bacteria adhesion to skin mucus of fishes. Veterinary microbiology 171: 1-12.

Benjamini Y, Hochberg Y 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the royal statistical society. Series B (Methodological): 289-300.

Bergboer JG, et al. 2011. Psoriasis risk genes of the late cornified envelope-3 group are distinctly expressed compared with genes of other LCE groups. The American journal of pathology 178: 1470-1477.

Bianco P, Fisher LW, Young MF, Termine JD, Robey PG 1990. Expression and localization of the two small proteoglycans biglycan and decorin in developing human skeletal and non-skeletal tissues. Journal of Histochemistry & Cytochemistry 38: 1549-1563.

Bolstad BM, Irizarry RA, Åstrand M, Speed TP 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19: 185-193.

Bruckner-Tuderman L, Has C 2014. Disorders of the cutaneous basement membrane zone—The paradigm of epidermolysis bullosa. Matrix Biology 33: 29-34.

Cáceres M, et al. 2003. Elevated gene expression levels distinguish human from non-human primate brains. Proceedings of the National Academy of Sciences 100: 13030-13035.

Carroll SB 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell 134: 25-36.

Carroll SB 2003. Genetics and the making of Homo sapiens. Nature 422: 849.

Carroll SB, Grenier JK, Weatherbee SD. 2013. From DNA to diversity: molecular genetics and the evolution of animal design: John Wiley & Sons.

Chen S, Birk DE 2013. The regulatory roles of small leucine-rich proteoglycans in extracellular matrix assembly. The FEBS journal 280: 2120-2137.

Consortium EP 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57.

Darling T, Sperling L, Turiansky CG 2013. The Basement Membrane Zone: Making the Connection.

Das R, et al. 2014. Complete mitochondrial genome sequence of the Eastern Gorilla (Gorilla beringei) and implications for African ape biogeography. Journal of Heredity 105: 846-855.

Dash S, Das S, Samal J, Thatoi H 2018. Epidermal mucus, a major determinant in fish health: a review. Iranian Journal of Veterinary Research 19: 72-81.

Dávid-Barrett T, Dunbar RI 2016. Bipedality and hair loss in human evolution revisited: The impact of altitude and activity scheduling. Journal of Human Evolution 94: 72-82.

De Cid R, et al. 2009. Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis. Nature genetics 41: 211.

Deng L, Xu S 2018. Adaptation of human skin color in various populations. Hereditas 155: 1.

Douglas AT, Hill RD 2014. Variation in vertebrate cis-regulatory elements in evolution and disease. Transcription 5: e28848.

Ellis RA, Montagna W 1962. The skin of primates. VI. The skin of the gorilla (Gorilla gorilla). American Journal of Physical Anthropology 20: 79-93.

Enard W, et al. 2002. Molecular evolution of FOXP2, a gene involved in speech and language. Nature 418: 869.

Farooq Z, Banday S, Pandita TK, Altaf M 2016. The many faces of histone H3K79 methylation. Mutation Research/Reviews in Mutation Research 768: 46-52.

Fitch WM, Margoliash E 1967. Construction of phylogenetic trees. Science 155: 279-284.

Folk GE, Semken A 1991. The evolution of sweat glands. International Journal of Biometeorology 35: 180-186.

Fuchs E, Raghavan S 2002. Getting under the skin of epidermal morphogenesis. Nature Reviews Genetics 3: 199.

Gao Z, Tseng C-h, Strober BE, Pei Z, Blaser MJ 2008. Substantial alterations of the cutaneous bacterial biota in psoriatic lesions. PLoS One 3: e2719.

Ghaleb AM, Yang VW 2017. Krüppel-like factor 4 (KLF4): What we currently know. Gene 611: 27-37.

Goletz S, Zillikens D, Schmidt E 2017. Structural proteins of the dermal-epidermal junction targeted by autoantibodies in pemphigoid diseases. Experimental dermatology 26: 1154-1162.

Grice EA, Segre JA 2011. The skin microbiome. Nature Reviews Microbiology 9: 244.

Halper J. 2014. Proteoglycans and diseases of soft tissues. In. Progress in Heritable Soft Connective Tissue Diseases: Springer. p. 49-58.

Has C, Nystrom A 2015. Epidermal Basement Membrane in Health and Disease. Curr Top Membr 76: 117-170. doi: 10.1016/bs.ctm.2015.05.003

Hasegawa H, et al. 2007. The distributions of type IV collagen α chains in basement membranes of human epidermis and skin appendages. Archives of histology and cytology 70: 255-265.

He Q, et al. 2011. High conservation of transcription factor binding and evidence for combinatorial regulation across six Drosophila species. Nature genetics 43: 414.

Heintzman ND, et al. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459: 108.

Hobolth A, Christensen OF, Mailund T, Schierup MH 2007. Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. PLoS genetics 3: e7.

Hudson LG, et al. 2009. Cutaneous wound reepithelialization is compromised in mice lacking functional Slug (Snai2). Journal of dermatological science 56: 19-26.

Iivanainen A, et al. 1995. The human laminin β2 chain (s-laminin): structure, expression in fetal tissues and chromosomal assignment of the LAMB2 gene. Matrix Biology 14: 489-497.

Jablonski NG, Chaplin G 2010. Human skin pigmentation as an adaptation to UV radiation. Proceedings of the National Academy of Sciences: 200914628.

Jackson B, et al. 2005. Late cornified envelope family in differentiating epithelia—response to calcium and ultraviolet irradiation. Journal of investigative dermatology 124: 1062-1070.

Kadler KE, Baldock C, Bella J, Boot-Handford RP 2007. Collagens at a glance. J Cell Sci 120: 1955-1958. doi: 10.1242/jcs.03453

Karmodiya K, Krebs AR, Oulad-Abdelghani M, Kimura H, Tora L 2012. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. BMC genomics 13: 424.

Kent WJ, et al. 2002. The human genome browser at UCSC. Genome research 12: 996-1006.

Kielty CM, Sherratt MJ, Shuttleworth CA 2002. Elastic fibres. Journal of cell science 115: 2817-2828.

Kimura H 2013. Histone modifications for human epigenome analysis. Journal of human genetics 58: 439.

King M-C, Wilson AC 1975. Evolution at two levels in humans and chimpanzees. Science 188: 107-116.

Kleinjan DA, van Heyningen V 2005. Long-range control of gene expression: emerging mechanisms and disruption in disease. The American Journal of Human Genetics 76: 8-32.

Konopka G, et al. 2009. Human-specific transcriptional regulation of CNS development genes by FOXP2. Nature 462: 213.

Kouzarides T 2007. Chromatin modifications and their function. Cell 128: 693-705.

Kumar S, Stecher G, Tamura K 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular biology and evolution 33: 1870-1874.

Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP 2001. A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 413: 519.

Li Y, et al. 2013. Age-dependent alterations of decorin glycosaminoglycans in human skin. Scientific reports 3: 2422.

Lieberman DE 2015. Human locomotion and heat loss: an evolutionary perspective. Compr Physiol 5: 99-117.

Lin Y-L, Pavlidis P, Karakoc E, Ajay J, Gokcumen O 2015. The evolution and functional impact of human deletion variants shared with archaic hominin genomes. Molecular biology and evolution 32: 1008-1019.

Lindblad-Toh K, et al. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. Nature 478: 476.

Luger K, Dechassa ML, Tremethick DJ 2012. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? Nature reviews Molecular cell biology 13: 436.

Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 389: 251.

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome research.

Marneros AG, Olsen BR 2005. Physiological role of collagen XVIII and endostatin. The FASEB Journal 19: 716-728.

Masunaga T, et al. 1997. The extracellular domain of BPAG2 localizes to anchoring filaments and its carboxyl terminus extends to the lamina densa of normal human epidermal basement membrane. Journal of investigative dermatology 109: 200-206.

Mathelier A, et al. 2015. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic acids research 44: D110-D115.

McGrath JA, et al. 1996. Compound heterozygosity for a dominant glycine substitution and a recessive internal duplication mutation in the type XVII collagen gene results in junctional epidermolysis bullosa and abnormal dentition. The American journal of pathology 148: 1787.

Mitchell C, Silver DL editors. Seminars in cell & developmental biology. 2018.

Montagna W editor.; 1982 Berlin, Heidelberg.

Montagna W 1985. The evolution of human skin (?). Journal of Human Evolution 14: 3-22.

Montagna W, Yun JS 1963. The skin of primates. XV. The skin of the chimpanzee (Pan satyrus). American Journal of Physical Anthropology 21: 189-203.

Naumova OY, Lee M, Rychkov SY, Vlasova NV, Grigorenko EL 2013. Gene expression in the human brain: the current state of the study of specificity and spatiotemporal dynamics. Child development 84: 76-88.

Niehues H, et al. 2018. 3D skin models for 3R research: The potential of 3D reconstructed skin models to study skin barrier function. Experimental dermatology.

Niehues H, et al. 2017. Psoriasis-associated late cornified envelope (LCE) proteins have antibacterial activity. Journal of investigative dermatology 137: 2380-2388.

Oh HJ, Choi D, Goh CJ, Hahn Y 2015. Loss of gene function and evolution of human phenotypes. BMB reports 48: 373.

Pajic P, Lin Y-L, Xu D, Gokcumen O 2016. The psoriasis-associated deletion of late cornified envelope genes LCE3B and LCE3C has been maintained under balancing selection since Human Denisovan divergence. BMC evolutionary biology 16: 265.

Patel S, Xi ZF, Seo EY, McGaughey D, Segre JA 2006. Klf4 and corticosteroids activate an overlapping set of transcriptional targets to accelerate in utero epidermal barrier acquisition. Proceedings of the National Academy of Sciences 103: 18668-18673.

Pennacchio LA, et al. 2006. In vivo enhancer analysis of human conserved non-coding sequences. Nature 444: 499.

Perry GH, Verrelli BC, Stone AC 2004. Comparative analyses reveal a complex history of molecular evolution for human MYH16. Molecular biology and evolution 22: 379-382.

Pozzi L, et al. 2014. Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. Molecular phylogenetics and evolution 75: 165-183.

Prabhakar S, et al. 2008. Human-specific gain of function in a developmental enhancer. Science 321: 1346-1350.

Proksch E, Brandner JM, Jensen JM 2008. The skin: an indispensable barrier. Experimental dermatology 17: 1063-1072.

Rantala MJ 2007. Evolution of nakedness in Homo sapiens. Journal of Zoology 273: 1-7.

Reinboth B, Hanssen E, Cleary EG, Gibson MA 2002. Molecular interactions of biglycan and decorin with elastic fiber components: biglycan forms a ternary complex with tropoelastin and microfibril-associated glycoprotein 1. Journal of Biological Chemistry 277: 3950-3957.

Ricard-Blum S 2011. The collagen family. Cold Spring Harbor perspectives in biology 3: a004978.

Romero IG, Ruvinsky I, Gilad Y 2012. Comparative studies of gene expression and the evolution of gene regulation. Nature Reviews Genetics 13: 505.

Rozario T, DeSimone DW 2010. The extracellular matrix in development and morphogenesis: a dynamic view. Developmental biology 341: 126-140.

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19: 2496-2497.

Saarela J, Rehn M, Oikarinen A, Autio-Harmainen H, Pihlajaniemi T 1998. The short and long forms of type XVIII collagen show clear tissue specificities in their expression and location in basement membrane zones in humans. The American journal of pathology 153: 611-626.

Sakai LY, Keene DR, Morris NP, Burgeson RE 1986. Type VII collagen is a major structural component of anchoring fibrils. The Journal of cell biology 103: 1577-1586.

Sander JD, Joung JK 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. Nature biotechnology 32: 347.

Scally A, et al. 2012. Insights into hominid evolution from the gorilla genome sequence. Nature 483: 169.

Scharschmidt TC, Fischbach MA 2013. What lives on our skin: ecology, genomics and therapeutic opportunities of the skin microbiome. Drug Discovery Today: Disease Mechanisms 10: e83-e89.

Segre JA, Bauer C, Fuchs E 1999. Klf4 is a transcription factor required for establishing the barrier function of the skin. Nature genetics 22: 356.

Smoller BR 2009. Lever's Histopathology of the Skin. Journal of Cutaneous Pathology 36: 605-605.

Stedman HH, et al. 2004. Myosin gene mutation correlates with anatomical changes in the human lineage. Nature 428: 415.

Steiper ME, Young NM 2006. Primate molecular divergence dates. Molecular phylogenetics and evolution 41: 384-394.

Sugawara K, Tsuruta D, Ishii M, Jones JC, Kobayashi H 2008. Laminin-332 and-511 in skin. Experimental dermatology 17: 473-480.

Sumiyama K, Saitou N 2011. Loss-of-function mutation in a repressor module of human-specifically activated enhancer HACNS1. Molecular biology and evolution 28: 3005-3007.

Utriainen A, et al. 2004. Structurally altered basement membranes and hydrocephalus in a type XVIII collagen deficient mouse line. Human molecular genetics 13: 2089-2099.

Vallender EJ, Mekel-Bobrov N, Lahn BT 2008. Genetic basis of human brain evolution. Trends in neurosciences 31: 637-644.

Walko G, Castañón MJ, Wiche G 2015. Molecular architecture and function of the hemidesmosome. Cell and tissue research 360: 529-544.

Wang Z, Gerstein M, Snyder M 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10: 57.

Wang Z, et al. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. Nature genetics 40: 897.

Watt FM 1998. Epidermal stem cells: markers, patterning and the control of stem cell fate. Philosophical Transactions of the Royal Society of London B: Biological Sciences 353: 831-837.

Watt FM, Fujiwara H 2011. Cell-extracellular matrix interactions in normal and diseased skin. Cold Spring Harbor perspectives in biology: a005124.

Weyrich LS, Dixit S, Farrer AG, Cooper AJ, Cooper AJ 2015. The skin microbiome: associations between altered microbial communities and disease. Australasian Journal of Dermatology 56: 268-274.

White P, et al. 2003. Deletion of the homeobox gene PRX-2 affects fetal but not adult fibroblast wound healing responses. Journal of investigative dermatology 120: 135-144.

Wittkopp PJ, Kalay G 2012. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nature Reviews Genetics 13: 59.

Wong R, Geyer S, Weninger W, Guimberteau JC, Wong JK 2016. The dynamic anatomy and patterning of skin. Experimental dermatology 25: 92-98.

Wray GA 2007. The evolutionary significance of cis-regulatory mutations. Nature Reviews Genetics 8: 206.

Wurm S, et al. 2015. Terminal epidermal differentiation is regulated by the interaction of Fra-2/AP-1 with Ezh2 and ERK1/2. Genes & development 29: 144-156.

Yang M, et al. 2011. A preliminary study of differentially expressed genes in expanded skin and normal skin: implications for adult skin regeneration. Archives of dermatological research 303: 125-133.

Yao Y 2017. Laminin: loss-of-function studies. Cellular and molecular life sciences 74: 1095-1115.

Yeh J, et al. 2009. Accelerated closure of skin wounds in mice deficient in the homeobox gene Msx2. Wound repair and regeneration 17: 639-648.

Yin Y, et al. 2014. CD151 represses mammary gland development by maintaining the niches of progenitor cells. Cell Cycle 13: 2707-2722.

Zhou VW, Goren A, Bernstein BE 2011. Charting histone modifications and the functional organization of mammalian genomes. Nature Reviews Genetics 12: 7.