Summer 8-15-2016

The Role of Epidermal Enhancer 923 in the Chromatin Architecture and Transcriptional Regulation of the Epidermal Differentiation Complex

Inez Oh
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Human and Statistical Genetics

Dissertation Examination Committee:
Cristina de Guzman Strong, Chair
Shiming Chen
John Edwards
Eugene Oltz
Gary Stormo

The Role of Epidermal Enhancer 923 in the Chromatin Architecture and Transcriptional Regulation of the Epidermal Differentiation Complex

by
Inez Ying Li Oh

A dissertation presented to the
Graduate School of Arts & Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2016
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank my mentor, Dr. Cristina de Guzman Strong, who has provided me with invaluable guidance, not only in the pursuit of my scientific goals, but also in dealing with all that life throws my way.

I would like to acknowledge all past and present members of the de Guzman Strong lab who have helped make the lab a supportive and friendly workplace. Thank you for making every work day enjoyable by interspersing intense scientific discussion with equally intense conversations about Harry Potter, food, musical theater, and everything else under the sun. I consider you to be some of my most valued friends. My graduate student experience would not have been the same without you.

I would like to thank my thesis committee for their encouragement and guidance throughout this journey, and for helping me believe that I have what it takes to succeed.

Finally, I would like to acknowledge my family and my friends who have provided advice and unwavering support in all my endeavors. Thank you for being there for me through all the highs and lows of life.

This work has been made possible by many sources of funding and with the help of many individuals.

For the work in Chapter 2, we thank members of the de Guzman Strong lab and Anne Bowcock for critical reading of the manuscript, Jun Cheng (NHGRI) for generating the transgenic mice, and the Genome Technology Access Center (GTAC, Genetics, Washington University School of Medicine) for Illumina sequencing and advice. This work was supported by NIH [R00AR055948 (C.S.), P30CA091842 and UL1RR024992 (GTAC)] and Washington University Faculty Diversity Scholar Award [C.S].

For the work in Chapter 3, we thank Renate Lewis at the Hope Center for the design and targeting of the gRNAs, Mia Wallace at the Mouse Core facility for the mice, and Toni Sinnwell and Eric Tycksen of the Washington University Genome Technology Access Center for the RNA-seq libraries and preliminary analyses. The work cited in this chapter was performed in a facility supported by NCRR grant C06RR015502, by the Genome Technology Access Center in the Department of Genetics for genomic analysis with partial support from NCI Caner Center Support Grant P30CA91842 (Siteman Cancer Center) and ICTS/CTSA ULTR000448 from National Center for Research Resources (NCRR) and by NIAMS R01AR065523 of the National Institutes of Health and NIH Roadmap for Medical Research.

For the work in Chapter 4, we thank Eugene Oltz for sharing the P5424 T cell line, Kinjal Majumder for discussions about experimental design, Toni Sinnwell and Seth Crosby of the Washington University Genome Technology Access Center for assistance with the design of custom sequencing primers and sequencing of the 4C-seq libraries. This work was funded by NIAMS R01AR065523 of the National Institutes of Health and NIH Roadmap for Medical Research.

Inez Oh

Washington University in St. Louis

August 2016

Dedicated to my father, Jackson,

who started me on the path of scientific discovery,

and to my very patient and loving husband, Zac,

expert in “The Care and Handling of a PhD Student”.

ABSTRACT OF THE DISSERTATION

The Role of Epidermal Enhancer 923 in the Chromatin Architecture and Transcriptional Regulation of the Epidermal Differentiation Complex

by

Inez Ying Li Oh

Doctor of Philosophy in Biology and Biomedical Sciences

Human and Statistical Genetics

Washington University in St. Louis, 2016

Cristina de Guzman Strong, PhD, Chair

The epidermis covers the surface of the skin and provides a functional barrier across the entire body. Epidermal cells or keratinocytes proliferate in the innermost basal layer and migrate upwards into the suprabasal spinous and granular layers as they differentiate, and finally into the terminally differentiated outermost stratum corneum. Keratinocytes undergoing terminal differentiation are marked by tissue-specific concomitant expression of genes encoded in the Epidermal Differentiation Complex (EDC) locus. The EDC genes are organized into four gene families - *S100, Sprr, Lce,* and *Flg*-like, which are coordinately expressed upon activation of the terminal differentiation program in keratinocytes. The molecular mechanisms that govern the activation of the EDC during epidermal differentiation are poorly understood. The synteny and colinearity of the locus across multiple mammalian species and the coordinate expression of EDC genes upon keratinocyte differentiation suggest molecular mechanisms operating at the chromatin level. I hypothesize coordinate activation of the EDC by an enhancer regulatory element. Enhancers are non-coding regulatory DNA sequences that upon binding specific transcription factors, are able to increase expression of a proximal or distal target gene. Previous

work in our lab identified an epidermal-specific enhancer, CNE 923, that was active in in cell-based luciferase assays and transgenic mice. Here, I examine the function of the 923 enhancer for epidermal differentiation. Using an independent transgenic mouse line, I identified spatiotemporal sensitivity of the 923 enhancer that correlated with the patterning of epidermal barrier formation during mouse embryonic development. To determine if 923 formed chromatin interactions with the EDC gene promoters, I performed chromosome conformation capture (3C) assays in proliferating and differentiated primary mouse keratinocytes. The 3C studies identified physiologically sensitive chromatin interactions between 923 and EDC gene promoters. The data supports a dynamic EDC chromatin topology during keratinocyte differentiation. A requirement for c-Jun/AP-1 in relation to 923-mediated EDC chromatin remodeling for normal EDC gene expression during keratinocyte differentiation was further determined by chromatin immunoprecipitation, 3C, and RNA-seq upon pharmacological inhibition of AP-1 binding. To further determine the function of 923 *in vivo*, I generated a series of mutation alleles using CRISPR/Cas9 genome editing in mice. Cas9 nuclease activity targeted to the flanking ends of the 923 enhancer in mouse zygotes by a pair of guide RNAs, coupled with homologous recombination-mediated *loxP* insertions, generated 1 floxed (923^{flox}), 2 independent deletions (923^{delA}, 923^{delB}), and 1 partial deletion (923^{pdel}) alleles for the 923 enhancer. My results from the 923 knockout mice identified decreased expression of nearby *Smcp*, *Lce6a*, and involucrin gene expression, decreased distal *Crnn* and *Lce* gene family members, and a correlative increase in expression of *Sprr* gene family members. To identify the chromatin interactions for the 923 enhancer on a genome-wide scale, I performed high-throughput circular chromosome conformation capture (4C-seq) assays with respect to the 923 enhancer and an additional *Flg* promoter viewpoint in proliferating and differentiated keratinocytes and P5424 T-cells. My

results revealed 923 enhancer-mediated chromatin interactions indicative of a topologically associated domain encompassing the EDC. However, an enrichment of 923 mediated chromatin interactions within the EDC, were identified in keratinocytes relative to the T-cells, specifically between the 923 enhancer and the *Sprr* and *Lce* gene families, and with non-coding regions in the gene desert between the *S100* and *Sprr* gene families. Of note was a 923 interaction with another putative enhancer near *Crct1*, enriched specifically in proliferating keratinocytes, and suggesting cross-talk between enhancers. Keratinocyte-specific trans-interactions identified by MACS and GREAT algorithms included genes important for epidermal function including *Trp63,* an important regulator of keratinocyte differentiation. Together, my 4C-seq identifies unique chromatin architectures of the EDC in keratinocytes and T cells, including keratinocyte-specific enhancer-enhancer crosstalk in cis and interactions between transcriptionally active loci in trans. My studies identify, for the first time, a link between the 923 enhancer and proximal (*Ivl*, *Smcp*, *Lce6a*) and distal genes (*Crnn,* distal *Lce* family), the loss of which coincides with upregulation of other epidermal differentiation genes (*Sprr* family) to maintain skin barrier function. Together, my work has identified 923 as an epidermal-specific enhancer that participates in a chromatin looping network to co-regulate expression of genes important for epidermal development, as a mechanism for maintaining skin barrier integrity.

Chapter 1

Introduction

1.1 Overview

Transcriptional regulation plays a major role in the development of the epidermis. The development and organization of the epidermis relies on the spatiotemporal-specific activation of genes in order to direct the fate of each cell. With each somatic cell containing an identical genome, how does the keratinocyte interpret the genome in a way that activates a transcriptional program specific to epidermal differentiation? Early studies of gene regulation focused on single gene promoters and their transcription factor binding sites (Visel et al. 2007). However, not all transcriptional activation is attributable to biochemical activity at the gene promoter, thus suggesting the contribution of other loci. The notion of a non-promoter regulatory element that could modulate the expression of a target gene located some distance away arose with the discovery of the SV40 tandem repeat sequence (Banerji et al. 1981; Benoist and Chambon 1981; Gruss et al. 1981; Moreau et al. 1981; Fromm and Berg 1983). The SV40 sequence "enhanced" transcription from distal genes in a position and independent manner, leading to coinage of the term, "enhancer". The complete genomes of humans and other model organisms in conjunction with Next-Generation sequencing and high-throughput methods have allowed us to identify, define, and ascertain the function, of regulatory elements such as enhancers on a genome-wide scale (Visel et al. 2007; Levo and Segal 2014; Shlyueva et al. 2014). In this introductory

chapter, I discuss a history of the conceptual advances in our understanding of the enhancer in epidermal development, specifically with regards to EDC gene expression.

1.2 What is an Enhancer?

The concept of an "enhancer" emerged in 1981. In that year, Pierre Chambon and George Khoury independently discovered a non-coding 72 bp tandem repeat sequence upstream of the SV40 early gene promoters that was required for transcription (Benoist and Chambon 1981; Gruss et al. 1981). A subsequent study from Walter Schaffner identified the ability of the SV40 DNA sequence to "enhance" the expression of rabbit β-globin in an expression vector even when the SV40 sequence was placed thousands of base pairs away from the β-globin gene promoter (Banerji et al. 1981). Chambon observed similar results using the gene for conalbumin (Moreau et al. 1981). Further experiments performed by Paul Berg and Michael Fromm showed that the SV40 sequence was able to "enhance" transcription independent of its location (upstream or downstream of its target gene) and orientation (forward or reverse) (Fromm and Berg 1983). This established the SV40 sequence as the prototype of a novel genetic element, an enhancer, and established the definition of a classical enhancer as a non-coding sequence that can modulate gene expression in a position- and orientation-independent manner.

The discovery of the SV40 enhancer paved the way for the identification of enhancers in other tissue types (Shlyueva et al. 2014). Often, searches for enhancers were prioritized and interrogated in the immediate vicinity of target genes, including upstream or downstream sequences (including 5' and 3' untranslated regions (UTRs)), introns, and intergenic regions (reviewed in (Kleinjan and van Heyningen 2005)).

In vivo studies of these putative enhancers identified the spatiotemporal specificity of these sequences to ensure biologically relevant cell- and tissue-specific gene expression (reviewed in (Levine 2010; Williamson et al. 2011)). Biochemical studies further identified clusters or arrays of transcription factor binding sites that act as "building blocks" of cis-regulatory modules that also show enhancer activity (reviewed in (Hardison and Taylor 2012; Levo and Segal 2014; Shlyueva et al. 2014)). The enrichment of multiple transcription factor binding sites within an enhancer facilitates cell-specific expression largely attributable to combinatorial and differential binding of transcription factor family members in the context of different microenvironments. Based on these later studies, we more loosely define an enhancer as a non-coding sequence containing clusters of transcription factor binding sites that drives cell-, tissue-, or developmental stage-specific gene expression.

Following completion of the human genome sequence, the National Human Genome Research Institute, recognizing the need to more fully understand the regulation of gene expression, launched the Encyclopedia of Non-Coding Elements (ENCODE), a collaborative public research project to identify and characterize the function of noncoding elements in the genome, and develop the tools and technology to achieve this goal (The ENCODE Project Consortium 2007). Next-generation sequencing was instrumental in generating these genome datasets in a cost-effective manner. The ENCODE studies as well as work by others have greatly facilitated our ability to identify enhancers on a genome-wide scale based on chromatin modifications that are unique to these regulatory elements such as DNaseI hypersensitivity (open chromatin) and histone modification epigenetic marks (H3K27Ac, H3K4me1), and transcription factor binding (p300, activating TFs) associated with functional enhancers (reviewed in (Ong and Corces 2011; Hardison and Taylor 2012; Shlyueva et al. 2014)). The high-throughput

chromatin immunoprecipitation of chromatin containing specific histone modifications enabled the discovery of enhancers and future downstream functional analyses (reviewed in (Shlyueva et al. 2014)).

1.3 Development of the Epidermis

The epidermis is located at the surface of the skin and whose architecture provides the key structure for the physical barrier of the skin. As well, the epidermis provides a tractable and spatially hierarchical model to investigate the development of committed cells. Epidermal cells or keratinocytes within the multiple stratified layers of the epidermis must strike a critical balance between self-renewal and differentiation in order to build a functional epidermal barrier across the entire body (Hsu et al. 2014). Keratinocyte self-renewal is marked by parallel cell division within the basal layer. Expression of Keratin 5 (*K5*) and Keratin 14 (*K14*) marks these basal proliferating keratinocytes.

During differentiation, the basal keratinocyte asymmetrically divides giving rise to a basal daughter keratinocyte and a suprabasal daughter keratinocyte that migrates outwards, entering the spinous layer (Lechler and Fuchs 2005; Blanpain and Fuchs 2009; Hsu et al. 2014). Here, the spinous and granular keratinocytes activate Keratin 1 (*K1*) and Keratin 10 (*K10)* expression concomitant with *K5*/*K14* downregulation. During late terminal differentiation, the keratinocytes coordinately express many Epidermal Differentiation Complex (EDC) genes, including the filaggrin (*FLG*) and *FLG*-like, late cornified envelope (*LCE*), small proline-rich region (*SPRR*), and *S100* genes (Mischke et al. 1996; Zhao and Elder 1997; Marshall et al. 2001; de Guzman Strong et al. 2010). As the keratinocytes reach the outermost stratum corneum, they enucleate and are surrounded by their own cornified envelope (the single most basic structural

unit of the skin barrier) formed by the cross-linking of scaffold proteins by transglutaminase-1, and sealed together by keratinocyte-extruded lipids to form a semi-permeable barrier (Lechler and Fuchs 2005). In mice, the pattern of functional barrier acquisition corresponds to maturation of the cornified envelopes, and proceeds from specific dorsal initiation sites at embryonic day (E)16, spreading to converge at the dorsal and ventral midline so that the whole embryo is impermeable at E17 (Hardman et al. 1998).

Epidermal differentiation can be recapitulated *in vitro* by exposing keratinocytes proliferating to high calcium concentrations (Yuspa et al. 1989; Pillai et al. 1990). This process, called calcium switching, stimulates the calcium receptor (CaR) and downstream phosphokinase C (PKC) signaling, thus activating the Fos/Jun family of transcription factors that play an important role in keratinocyte differentiation (Reviewed in (Bikle et al. 2012)). Fos and Jun proteins form homo- or heterodimers that comprise the AP-1 transcription factor complex (Shaulian and Karin 2001). In normal epidermis as well as in organotypic epidermal cultures, the expression pattern of AP-1 proteins is tightly regulated even within the differentiated layers (Mehic et al. 2005). Fos proteins are found in the nuclei of both basal and suprabasal keratinocytes. JunB and JunD are expressed in all layers of normal epidermis. Interestingly, c-Jun is expressed in the spinous layer, then disappears and reemerges in the outermost granular layer directly at the transition zone to the stratum corneum. Many of the genes expressed in keratinocytes, in either proliferative or differentiated layers of the epidermis have AP-1 binding sites. Together, this suggests that the binding of different combinations of AP-1 protein complexes to different enhancers drives the region-specific expression of the genes in the epidermis.

1.4 Epidermal Differentiation Complex

The EDC locus, on human chromosome 1q21 and mouse chromosome 3q, contains a dense cluster of genes which encode proteins that are the major molecular markers for terminal differentiation in the mammalian stratified epidermis.

Loricrin and involucrin are major protein components of the cornified envelope (CE) - a structural unit of the skin barrier (Rice and Green 1977; Simon and Green 1984). As early scaffolds for the CE, loricrin and involucrin were the first EDC genes to be discovered (Eckert and Green 1986; Mehrel et al. 1990; Hohl et al. 1991). The functional cloning of mRNAs in UV-treated and calcium-treated human keratinocytes led to the additional discovery of the *SPRR* and *S100* genes (Kartasova and van de Putte 1988; Marenholz et al. 2004). It was later determined that these gene families are physically linked together on human chromosome 1q21 by the hybridization of gene-specific probes on electrophoresed genome restriction fragments (Volz et al. 1993). In 1996, the Epidermal Differentiation Complex name was proposed upon higher resolution mapping (Mischke et al. 1996). Later, a search for molecular markers that coincided spatio-temporally with skin barrier formation in mice identified a set of Expressed Sequence Tags (ESTs) that shared sequence homology to *SPRR1* but were expressed at a later stage of epidermal differentiation, and therefore named late envelope proteins (*LEPs*) (Zhao and Elder 1997; Marshall et al. 2001; Wang et al. 2001). The nomenclature for *LEP*s was subsequently changed to late cornified envelopes (*LCE*s) to more accurately reflect the shared genomic organization, protein homology, and expression pattern (Jackson et al. 2005). Thus, with the inclusion of the *LCE*s, the human EDC comprises a cluster of 64 coding genes and 4 gene families: filaggrin and *FLG*-like, late cornified envelope (*LCE*s), small proline rich region (*SPRR*s), and *S100* genes (Volz et al. 1993; Rothnagel et al. 1994; Mischke et al. 1996; Song et

al. 1999; Jackson et al. 2005; de Guzman Strong et al. 2010). The *FLG*, *LCE*, and *SPRR* gene families encode structural components of the epidermal skin barrier, while many of the *S100* genes encode chemoattractant proteins that are expressed upon disruption of the barrier (Segre 2006). The *FLG*-like genes, including trichohyalin (*TCHH*), repetin (*RPTN*), hornerin (*HRNR*), and filaggrin-2 (*FLG-2),* represent evolved paralogous genes given the fusion of the consensus S100 domain (two Ca^{2+}-binding EF domains) to gene-specific unique central repeat and C-terminal domains (Henry et al. 2012). The clustering and number of the EDC genes, the shared homology at the N- and C-terminal domains, and the variability in the internal repeat sequences underscore the evolution and divergence of the EDC from a common ancestor (Backendorf and Hohl 1992; The Chimpanzee Sequencing and Analysis Consortium 2005).

Loricrin-deficient mice exhibited a delay in barrier formation suggesting the existence of a compensatory mechanism for the skin barrier (Koch et al. 2000). Although no phenotype was observed in the involucrin-deficient mice (Djian et al. 2000), triple knockout mice of involucrin, and two non-EDC genes, periplakin and envoplakin, that contribute to the early protein scaffold of the cornified envelope, led to defects in the epidermal barrier (Sevilla et al. 2007).

The role of filaggrin in the epidermis was gleaned from human genetic studies that identified semi-dominant stop-gain *FLG* mutations in patients with ichthyosis vulgaris (IV) resulting in a complete loss of profilaggrin (Smith et al. 2006). Moreover, the overlap between IV and atopic dermatitis (AD) led to the discovery of common loss-of-function *FLG* variants for AD in Europe (Palmer et al. 2006), and has been one of the most widely replicated genetic risk factors for a common disease to date (Rodríguez et al. 2009). '*Flaky tail*' mice with dry skin, orthokeratosis, and acanthosis (Presland et al. 2000) also exhibited a predisposition to AD that

was associated with a spontaneous frameshift deletion mutation in *Flg* resulting in loss-of-function and the *'matted'* allele of *Tmem79* (Fallon et al. 2009).

The shared biology of the EDC gene components led to the conceptual recognition of the EDC genes as a cluster. This in turn has been pivotal in driving current studies to understand the transcriptional regulation of this important locus in cutaneous biology.

1.5 Regulation of EDC gene expression

1.5.1 Understanding gene regulation: Pre-Human Genome Era

Pioneer studies to elucidate gene expression for epidermal differentiation focused on the expressions of involucrin (*IVL*) and loricrin (*LOR*). These two important marker genes are distinctively expressed in terminally differentiated keratinocytes, and encode structural proteins that are cross-linked with many of the other proteins encoded by the EDC genes to form the cornified envelope (Segre 2006). IVL is cross-linked early in the formation of the cornified envelope (Eckert et al. 2004) and LOR is in turn cross-linked to the existing scaffolding containing IVL (Nithya et al.). In the developing mouse embryo, *Ivl* and *Lor* transcripts are upregulated as early as E15.5 (Oh et al. 2014), and protein expression can be observed by E16.5, corresponding to the onset of skin barrier formation (Hardman et al. 1998; Marshall et al. 2001). The tight correlation of *IVL* and *LOR* expression with keratinocyte terminal differentiation renders these genes as ideal candidates for studying the mechanisms that underlie the switch from a proliferating to a differentiating program in keratinocytes.

Involucrin

Transcriptional activation of involucrin (*IVL*) regulation using a β-galactosidase reporter gene construct in transgenic mice identified keratinocyte-specific expression driven by a 3.7 kb upstream sequence of *IVL* (Carroll and Taichman 1992). Using a series of deletion constructs of the aforementioned reporter construct, optimal expression of *IVL* in transgenic mice was further attributed to two discrete *IVL* upstream regions, the distal- and proximal-regulatory regions (DRR and PRR) (Welter et al. 1995; Banks et al. 1998). Within the DRR, an AP-1 binding site was required for *IVL* expression above basal levels, while a synergistic adjacent SP1 binding site was necessary for optimal expression (Banks et al. 1999). The DRR AP-1 site was found to interact with Fra-1, JunB, JunD, and p300, a histone acetyltransferase often associated with enhancers (Ogryzko et al. 1996; The ENCODE Project Consortium 2007), while the DRR SP1 site was observed to interact with SP1, SP3, and KLF4 transcription factors (Welter et al. 1995; Banks et al. 1998; Balasubramanian et al. 2005; Crish and Eckert 2008; Chew et al. 2013). Because Fra-1 and KLF4 are known to interact with p300 (Kaczynski et al. 2003; Crish and Eckert 2008), this observation and others suggest a complex of transcription factors that forms on the DRR to drive *IVL* expression during keratinocyte differentiation.

Loricrin

Transcriptional activation of mouse loricrin expression was first localized to a 6.5 kb region spanning the loricrin gene (DiSepio et al. 1995). Transgenic reporter mice in which the *LOR* coding sequences were replaced by a β-galactosidase gene revealed that the remaining 1.5 kb of 5′-flanking sequence, a small noncoding exon, a 1.1 kb intron, a single coding exon, and 2.2 kb of 3′-flanking sequence from the mouse loricrin gene drove epidermal-specific, but not differentiation-specific expression. Minimal promoter activity, dependent on an AP1 site conserved between mouse and human, was mapped to a 60 bp upstream sequence of the

transcription start site. In the case of the human *LOR* gene, enhancers located within 1.5 kb of 5′-flanking sequence and 9 kb of 3′-sequence were responsible for the tissue- and differentiation-specific expression of the human *LOR* transgene in transgenic mice (Yoneda and Steinert 1993). As few as 154 bp of 5′-upstream sequence from the cap site directed expression specifically in cultured keratinocytes (NHEK and HaCaT), in a Sp1/c-Jun and p300/CREB-dependent manner (Jang and Steinert 2002). Differential occupation of the keratinocyte-specific repressor-1 (KSR-1) comprised Sp3, CREB-1/CREMα/ATF-1, Jun B, while an AP-2-like protein was lost upon Sp1/c-Jun/p300/CBP recruitment during differentiation at this *LOR* site, thus enabling *LOR* transcriptional resolution in stratified layers (Jang and Steinert 2002).

1.5.2 EDC loci in other mammals: Identifying Conserved Noncoding Elements by Comparative Genomics in the Post-Human Genome Era

Many of the genes in the EDC are coordinately expressed at the onset of mouse epidermal differentiation at embryonic (E)15.5 (de Guzman Strong *et al.*, 2010). The dorsal-to-ventral patterning of skin barrier formation is conserved between 4 mammalian species (mouse, rat, rabbit, and opossum) and is associated with concomitant EDC gene activation (Hardman *et al.*, 1998; de Guzman Strong *et al.*, 2010). These observations raised an interesting question. How are the EDC genes concomitantly activated? A likely explanation is the involvement of regulatory DNA elements such as enhancers to direct spatiotemporal expression of the EDC in the epidermis. The availability of complete genome sequences of the human, mouse, and other mammalian species greatly facilitated the timely identification of regulatory elements (Visel *et al.*, 2007). Potential regulatory elements can be identified by sequence conservation in the noncoding regions of phylogenetically distinct animal species. Comparative genomic sequence alignments of 7 orthologous mammalian EDC loci across eutherian (placental) and metatherian

(marsupial) highlighted the evolutionarily conserved colinearity (order of the genes) and synteny (located on the same chromosome) of the EDC (de Guzman Strong *et al.*, 2010). Moreover, 48 conserved noncoding elements (CNEs) from the 7-mammal alignment data set were identified in the EDC. The regulatory activity for each CNE was tested in transfected keratinocytes using the firefly luciferase reporter assays. Approximately 50% of the CNEs exhibited regulatory activity, either enhancing or repressing luciferase activity in either or both proliferating or differentiated keratinocytes. The results demonstrated the physiological plasticity of these CNEs relevant to gene transcription.

1.5.3 Coordinate regulation of EDC gene expression

The discovery of enhancers in the EDC enabled hypothesis-driven research towards the elucidation of a potential locus control region (LCR) in the EDC. An LCR is defined as a strong enhancer that is capable of directing tissue-specific expression in a position independent manner (Li *et al.*, 2002). CNE 923, located 923 kb away from the most 5' EDC gene, was hypothesized to be a LCR of the EDC since it exhibited the highest reporter and hence enhancer activity in the keratinocytes. The enhancer activity for 923 was further validated based on DNaseI hypersensitivity in primary human keratinocytes. Ectopic expression of β-galactosidase by 923 in transgenic mice further demonstrated the epidermal-specificity of the 923 enhancer (de Guzman Strong *et al.*, 2010) and recapitulated the spatio-temporal migration of epidermal barrier formation (Oh *et al.*, 2014).

While these studies supported an intriguing role for the 923 enhancer as an LCR in epidermal-specific transcriptional activation, the mechanism was less clear. My work, described in the following chapters, seeks to further elucidate the role of CNE 923 to regulate EDC expression. Chapter 2 details my study that further characterized the enhancer activity 923 in an

independent transgenic mouse line while also identifying the 923 enhancer/c-Jun/AP-1 transcription factor axis that linked chromatin state to gene expression of the EDC (Oh *et al.*, 2014). Chapter 3 discusses the function and necessity of the endogenous 923 using a mouse model generated via CRISPR-Cas9 genome editing (Oh et al. 2016). Finally, Chapter 4 details the genome-scale chromatin architectures of the EDC in keratinocytes and T cells.

1.6 The Role of Chromatin Architecture in the Control of Gene Expression

Recent cell biology studies in epidermal development have highlighted the importance of chromatin remodeling as a mechanism for efficient and coordinate regulation of gene clusters (Fessing et al. 2011; Mardaryev et al. 2013; Oh et al. 2014; Sethi et al. 2014) During epidermal development, the EDC locus relocates away from the nuclear periphery and towards the nuclear interior prior to the activation of EDC gene expression (Gdula et al. 2013). Ablation of either p63, a master regulator of epidermal development (Mills et al. 1999; Yang et al. 1999; Viganò and Mantovani 2014), or Satb1, a higher-order genome organizer that binds to the EDC in epidermal progenitor cells (Fessing et al. 2011), resulted in the loss of keratinocyte-specific and EDC gene expression associated with alterations in the chromatin conformation of the EDC. The observation that p63 directly regulates the expression of Satb1, designated Satb1 as an important downstream target of p63 required for the proper establishment of higher-order EDC chromatin structure and coordinated gene expression (Fessing et al. 2011). Similarly p63 and its direct target Brg1 are essential in remodeling the higher-order chromatin structure of the EDC and positioning the locus within the 3D chromatin landscape to allow efficient expression of EDC

genes in epidermal progenitor cells during skin development (Mardaryev et al. 2013; Sethi et al. 2014).

1.6.1 The formation and biology of enhancer-promoter chromatin loops

The idea of the formation of enhancer-promoter chromatin loops to drive gene activation has emerged as a major framework around which we can approach investigations of enhancer-promoter interactions. As mentioned before, the first and thus one of the most well-studied loci, is the evolutionarily conserved β-globin locus important for hematopoiesis (reviewed in (Kiefer et al. 2008)). The 5 globin genes (ε, Gγ, Aγ, δ and β) form a cluster and are expressed in a developmental-stage- and tissue-specific manner controlled by a locus control region (LCR). The functional relevance of enhancer-promoter chromatin interactions for gene activation at this locus was recently demonstrated in an elegant series of studies (Deng et al. 2012). Previously, it was demonstrated that the GATA-1 transcription factor and Ldb1 were required to form a chromatin interaction between the β-globin locus control region and the β-globin promoter for transcriptional activation in erythroid cells (Song et al. 2007; Tripic et al. 2009). The more convincing experiment demonstrated a requirement for the formation of the chromatin loop for gene transcription arose from the use of artificial zinc fingers (ZF) (Deng et al. 2012). Introduction of the artificial targeted ZF forced chromatin loop formation by tethering Ldb1 to the β-globin locus control region in GATA-1 null erythroblasts and was sufficient to activate β-globin gene expression. This work was the first to demonstrate the causality of chromatin spatial interactions in promoting gene transcription.

My own work has also identified a role for such chromatin looping interactions in epidermal development. As will be described in greater detail in Chapter 2, epidermal-specific enhancer 923, whose optimal activity requires AP-1 transcription factor binding, was found to

interact with several of the EDC gene promoters (Oh et al. 2014). The loss of AP-1 binding resulted in changes in 923-mediated interactions, and correlating to the loss of EDC gene expression during keratinocyte differentiation, demonstrating the importance of enhancer-promoter interactions in keratinocyte differentiation.

The identification of chromatin looping interactions

The distant arrangement between enhancers and their target genes in metazoans incited debates on how the enhancers were regulating their distant target genes (Krivega and Dean 2012). Chromatin looping and tracking were proposed as models to explain this phenomenon (Dean 2006). However, the first experimental confirmation of the close proximity between enhancers and target genes came with the chromosome conformation capture (3C) technique (Dekker et al. 2002). 3C was used to demonstrate that loop formation between the β-globin locus control region (LCR) enhancer and gene accompanied transcriptional activation (Carter et al. 2002; Tolhuis et al. 2002; Palstra et al. 2003) and established a paradigm that was later validated in numerous other loci, including the α-globin gene cluster, T_H2, IFNG, MHC class II and IgH loci (Kadauke and Blobel 2009). Transcription factor ChIP-chip studies also revealed that enhancers could be located even further from their target genes than previously thought, as far as 10-20 kbs to several Mbs away (Hong et al. 2008). Often, these proximal and distal enhancers interact to co-regulate a target gene.

Recent improvements to the chromatin conformation methods have allowed us to examine the chromatin interactions of genomic regions at varying levels of depth and resolution. 4C (circular chromosome conformation capture) detects all interacting sequences with a sequence of interest (bait such as an enhancer) (Zhao et al. 2006). 5C (chromosome

conformation capture carbon-copy) is designed to detect many known interactions with numerous baits and typically within a gene locus (Dostie et al. 2006) whereas Hi-C approach is aimed to detect all chromatin interactions (Lieberman-Aiden et al. 2009). Methods such as ChIA-PET (Chromatin interaction analysis with paired-end tag sequencing) combine the two to simultaneously identify genome-wide chromatin interactions and the proteins that are binding these interactions (Li et al. 2010). The evidence for epigenetic modifications, chromatin looping, and the interplay between the two are relatively recent advancements that have provided new insights to our understanding of the biochemical aspects of enhancer-mediated transcriptional regulation.

1.6.2 Higher-level chromatin architecture: Topologically Associated Domains (TAD) and Chromosome Territories (CT)

Topologically associated domains (TADs) were first identified by Hi-C (a variation of the 3C technique) (Dixon et al. 2012; Sexton et al. 2012). TADs represent distinct clusters of enhancer-promoter interactions (Symmons et al. 2014; Lupiáñez et al. 2015). At the highest order of chromosome organization, spatially proximal TADs compose a chromosome territory (CT), which is a compartment within the nucleus that is often segregated in a chromosome-specific manner (reviewed in (Fraser et al. 2015)). It has been noted that actively transcribed gene-rich loci that are in an open conformation are more likely to loop out of their CTs, suggesting that the space between CTs is important for genomic loci to access the transcription machinery (reviewed in (Fraser et al. 2015)).

The influence of CTs over long range enhancer-promoter interactions was demonstrated for the developing limb bud (Amano et al. 2009). The differential expression of the Sonic

hedgehog (*Shh)* gene across regions of the limb bud as shown in the developing mouse embryo was mediated by specific interactions between the *Shh* promoter and a long-range enhancer MFCS1. In the intermediate portion of the limb bud which lacked ectopic Shh expression, the long-range enhancer was spatially and linearly distant from the *Shh* coding region. In the anterior limb bud cells, the long-range enhancer interacted with the *Shh* coding region, thus representing a poised state within their CT as *Shh* was not expressed. However, in the cells of the zone of polarizing activity (ZPA) where *Shh* is actively expressed, the chromatin interactions between the *Shh* promoter and MFCS1were observed in a 3D-FISH experiment to relocate outside the CT. Thus, the gene regulatory effect of an enhancer-promoter interaction was shown to be affected by the location of the interacting regions relative to their CT.

1.6.3 Involvement of cohesin and CTCF in forming active chromatin hubs

Enhancer activity is often modulated by a different class of regulatory elements called insulators that function as physical barriers to the optimal enhancer-promoter formation for transcriptional activation. Here we discuss newly recognized attributes of enhancers and new direct roles for CCCTC-binding factor (CTCF) insulators in enhancer–promoter interactions and in broadly configuring the genome (Reviewed in (Krivega and Dean 2012)).

Cohesin is a complex of proteins that holds sister chromatids together during DNA replication until the sister chromatids separate at anaphase (reviewed in (Hagstrom and Meyer 2003)). The *Drosophila Nipped-B* gene is similar to cohesin regulatory factor Scc2 (Rollins et al. 1999). Mutations in *Nipped-B* diminished the ability of an enhancer to overcome an intervening insulator by interacting with a distal promoter, thereby identifying what was then considered a non-canonical role for cohesin to regulate enhancer-promoter interactions and gene expression (Rollins et al. 1999). Subsequently, it was discovered that mammalian cohesin complexes can be

recruited to DNase I hypersensitive sites and regulatory elements by the CTCF DNA binding protein (Parelho et al. 2008; Rubio et al. 2008; Wendt et al. 2008). CTCF binds at insulators and at boundary elements to demarcate active chromatin hubs and limit the effect of enhancers (Wallace and Felsenfeld 2007), and cohesin contributes to CTCF's enhancer blocking activity (Parelho et al. 2008; Rubio et al. 2008). Studies of the apolipoprotein gene cluster (Mishiro et al. 2009), the globin locus (Hou et al. 2010), and the T-cell receptor (*Tcra*) locus (Seitan et al. 2011) have since demonstrated the cooperation of CTCF and cohesin to mediate insulators corresponding to TAD boundaries, thereby maintaining proper chromatin loop formation and localization of transcriptional apparatus at the gene promoters to control gene expression.

A powerful example of the biological function of CTCF is that of chromosomal rearrangements of the conserved TAD-spanning *WNT6/IHH/EPHA4/PAX3* locus that disrupt a CTCF-associated boundary domain within a TAD and cause limb malformations in humans. Mice harboring the equivalent disease-relevant rearrangements of the locus that interrupted a CTCF-associated boundary domain were generated using CRISPR/Cas9 genome editing, and displayed ectopic limb expression of an uninvolved gene in the locus due to misplacement of a cluster of limb enhancers relative to TAD boundaries (Lupiáñez et al. 2015). This demonstrated the functional importance of TADs for orchestrating gene expression via genome architecture and suggests the potential for disease-associated large-scale chromosomal abnormalities to pinpoint TAD boundaries.

A more recent study identified a functional role for the directionality of a CTCF insulator to influence chromatin topology and enhancer-promoter function (Guo et al. 2015). Inversions of CTCF boundary elements in the P-cadherin enhancer using CRISPR/Cas9-genome editing altered the chromatin topology and gene expression. This study demonstrated a novel governing

principle for chromatin architecture for gene expression by linear DNA, and holds great potential for the more accurate prediction of enhancers in the skin and other tissues.

1.6.4 Super-enhancers and the role of Mediator

More recently, a new class of enhancers called "super-enhancers" was identified (Whyte et al. 2013). Super-enhancers are marked by high levels of the Mediator coactivator complex occupation as determined by ChIP-seq and span much larger distances than typical enhancers (8.7 kb versus 703 bp). Mediator is a major component of the transcription pre-initiation complex (PIC) machinery with RNA polymerase II (RNA pol II) and is required for activator-dependent transcription *in vitro* and *in vivo* (reviewed in (Poss et al.)). Reduced levels of Mediator specifically affected gene expression near the super-enhancers (Whyte et al. 2013). This was convincingly demonstrated with the loss of enhancer-promoter loops of select genes upon loss of Mediator (Sung et al. 2005; Wang et al. 2005). Mediator-occupied super-enhancers also exhibited enriched binding of transcription factors that are master regulators involved in cell-identity including ESCs, pro-B cells, T helper cells, myotubes, and macrophages.

The role of super-enhancers was recently identified in epidermal stem cells (Adam et al. 2015). The target genes associated with the epidermal stem cell-specific super-enhancers identified by H3K27Ac and Mediator ChIP-seq methods contained a high frequency of transcription factor binding motifs for Sox9, Lhx2, Nfatc1 and Nfib, previously reported to be important for maintaining the hair follicle stem cell niche (Chang et al. 2013; Folgueras et al. 2013; Keyes et al. 2013; Kadaja et al. 2014). As further evidence of their importance, these transcription factors were shown via ChIP-seq to bind at high frequency to super-enhancers relative to typical enhancers, in particular Sox9, corroborating its known role as a master transcription factor. Lineage tracing during epidermal development in the mouse, wound-healing

and in cell culture enabled the detection of the remodeling of super-enhancer regions and thereby supporting the idea that enhancers are activated or silenced in lineage-specific fashion. The ability of Mediator to identify key transcription factors and enhancer sequences in a variety of cell types and its sensitivity to changing conditions highlights its potential as a tool to pinpoint important regulatory sequences involved in cell and tissue homeostasis, even without prior knowledge of the transcription factors or genes involved.

1.7 References

Adam, RC, Yang, H, Rockowitz, S, *et al.* (2015). Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. *Nature* 521: 366–70.

Amano, T, Sagai, T, Tanabe, H, *et al.* (2009). Chromosomal Dynamics at the Shh Locus: Limb Bud-Specific Differential Regulation of Competence and Active Transcription. *Dev Cell* 16: 47–57.

Backendorf, C, Hohl, D (1992). A common origin for cornified envelope proteins? *Nat Genet* 2: 91–91.

Balasubramanian, S, Sturniolo, MT, Dubyak, GR, *et al.* (2005). Human epidermal keratinocytes undergo (-)-epigallocatechin-3-gallate-dependent differentiation but not apoptosis. *Carcinogenesis* 26: 1100–8.

Banerji, J, Rusconi, S, Schaffner, W (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27: 299–308.

Banks, EB, Crish, JF, Eckert, RL (1999). Transcription factor Sp1 activates involucrin promoter activity in non-epithelial cell types. *Biochem J* 337 (Pt 3: 507–12.

Banks, EB, Crish, JF, Welter, JF, *et al.* (1998). Characterization of human involucrin promoter distal regulatory region transcriptional activator elements-a role for Sp1 and AP1 binding sites. *Biochem J* 331 (Pt 1: 61–8.

Benoist, C, Chambon, P (1981). In vivo sequence requirements of the SV40 early promotor region. *Nature* 290: 304–10.

Bikle, DD, Xie, Z, Tu, C-L (2012). Calcium regulation of keratinocyte differentiation. *Expert Rev Endocrinol Metab*.

Blanpain, C, Fuchs, E (2009). Epidermal homeostasis: a balancing act of stem cells in the skin. *Nat Rev Mol Cell Biol* 10: 207–17.

Carroll, JM, Taichman, LB (1992). Characterization of the human involucrin promoter using a transient beta-galactosidase assay. *J Cell Sci* 103 (Pt 4: 925–30.

Carter, D, Chakalova, L, Osborne, CS, *et al.* (2002). Long-range chromatin regulatory interactions in vivo. *Nat Genet* 32: 623–6.

Chang, C-Y, Pasolli, HA, Giannopoulou, EG, *et al.* (2013). NFIB is a governor of epithelial-melanocyte stem cell behaviour in a shared niche. *Nature* 495: 98–102.

Chew, YC, Adhikary, G, Xu, W, *et al.* (2013). Protein kinase C δ increases Kruppel-like factor 4 protein, which drives involucrin gene transcription in differentiating keratinocytes. *J Biol Chem* 288: 17759–68.

Crish, JF, Eckert, RL (2008). Synergistic activation of human involucrin gene expression by Fra-

1 and p300--evidence for the presence of a multiprotein complex. *J Invest Dermatol* 128: 530–41.

Dean, A (2006). On a chromosome far, far away: LCRs and gene expression. *Trends Genet*.

Dekker, J, Rippe, K, Dekker, M, *et al.* (2002). Capturing chromosome conformation. *Science* 295: 1306–11.

Deng, W, Lee, J, Wang, H, *et al.* (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* 149: 1233–44.

DiSepio, D, Jones, A, Longley, MA, *et al.* (1995). The Proximal Promoter of the Mouse Loricrin Gene Contains a Functional AP-1 Element and Directs Keratinocyte-specific but Not Differentiation-specific Expression. *J Biol Chem* 270: 10792–9.

Dixon, JR, Selvaraj, S, Yue, F, *et al.* (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485: 376–80.

Djian, P, Easley, K, Green, H (2000). Targeted ablation of the murine involucrin gene. *J Cell Biol* 151: 381–8.

Dostie, J, Richmond, TA, Arnaout, RA, *et al.* (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res* 16: 1299–309.

Eckert, RL, Crish, JF, Efimova, T, *et al.* (2004). Regulation of involucrin gene expression. *J Invest Dermatol*.

Eckert, RL, Green, H (1986). Structure and evolution of the human involucrin gene. *Cell* 46: 583–9.

Fallon, PG, Sasaki, T, Sandilands, A, *et al.* (2009). A homozygous frameshift mutation in the mouse Flg gene facilitates enhanced percutaneous allergen priming. *Nat Genet* 41: 602–8.

Fessing, MY, Mardaryev, AN, Gdula, MR, *et al.* (2011). p63 regulates Satb1 to control tissue-specific chromatin remodeling during development of the epidermis. *J Cell Biol* 194: 825–39.

Folgueras, AR, Guo, X, Pasolli, HA, *et al.* (2013). Architectural niche organization by LHX2 is linked to hair follicle stem cell function. *Cell Stem Cell* 13: 314–27.

Fraser, J, Williamson, I, Bickmore, WA, *et al.* (2015). An Overview of Genome Organization and How We Got There: from FISH to Hi-C. *Microbiol Mol Biol Rev* 79: 347–72.

Fromm, M, Berg, P (1983). Simian virus 40 early- and late-region promoter functions are enhanced by the 72-base-pair repeat inserted at distant locations and inverted orientations. *Mol Cell Biol* 3: 991–9.

Gdula, MR, Poterlowicz, K, Mardaryev, AN, *et al.* (2013). Remodeling of three-dimensional organization of the nucleus during terminal keratinocyte differentiation in the epidermis. *J Invest Dermatol* 133: 2191–201.

Gruss, P, Dhar, R, Khoury, G (1981). Simian virus 40 tandem repeated sequences as an element of the early promoter. *Proc Natl Acad Sci U S A* 78: 943–7.

Guo, Y, Xu, Q, Canzio, D, *et al.* (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* 162: 900–10.

de Guzman Strong, C, Conlan, S, Deming, CB, *et al.* (2010). A milieu of regulatory elements in the epidermal differentiation complex syntenic block: implications for atopic dermatitis and psoriasis. *Hum Mol Genet* 19: 1453–60.

Hagstrom, KA, Meyer, BJ (2003). Condensin and cohesin: more than chromosome compactor and glue. *Nat Rev Genet* 4: 520–34.

Hardison, RC, Taylor, J (2012). Genomic approaches towards finding cis-regulatory modules in animals. *Nat Rev Genet*.

Hardman, M, Sisi, P, Banbury, D, *et al.* (1998). Patterned acquisition of skin barrier function during development. *Development* 125: 1541–52.

Henry, J, Toulza, E, Hsu, C-Y, *et al.* (2012). Update on the epidermal differentiation complex. *Front Biosci (Landmark Ed* 17: 1517–32.

Hohl, D, Mehrel, T, Lichti, U, *et al.* (1991). Characterization of human loricrin. Structure and function of a new class of epidermal cell envelope proteins. *J Biol Chem* 266: 6626–36.

Hong, J-W, Hendrix, DA, Levine, MS (2008). Shadow enhancers as a source of evolutionary novelty. *Science* 321: 1314.

Hou, C, Dale, R, Dean, A (2010). Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc Natl Acad Sci U S A* 107: 3651–6.

Hsu, Y-C, Li, L, Fuchs, E (2014). Emerging interactions between skin stem cells and their niches. *Nat Med* 20: 847–56.

Jackson, B, Tilli, CMLJ, Hardman, MJ, *et al.* (2005). Late cornified envelope family in differentiating epithelia - Response to calcium and ultraviolet irradiation. *J Invest Dermatol* 124: 1062–70.

Jang, S-I, Steinert, PM (2002). Loricrin expression in cultured human keratinocytes is controlled by a complex interplay between transcription factors of the Sp1, CREB, AP1, and AP2 families. *J Biol Chem* 277: 42268–79.

Kaczynski, J, Cook, T, Urrutia, R (2003). Sp1- and Krüppel-like transcription factors. *Genome Biol* 4: 206.

Kadaja, M, Keyes, BE, Lin, M, *et al.* (2014). SOX9: A stem cell transcriptional regulator of secreted niche signaling factors. *Genes Dev* 28: 328–41.

Kadauke, S, Blobel, GA (2009). Chromatin loops in gene regulation. *Biochim Biophys Acta - Gene Regul Mech*.

Kartasova, T, van de Putte, P (1988). Isolation, characterization, and UV-stimulated expression of two families of genes encoding polypeptides of related structure in human epidermal keratinocytes. *Mol Cell Biol* 8: 2195–203.

Keyes, BE, Segal, JP, Heller, E, *et al.* (2013). Nfatc1 orchestrates aging in hair follicle stem cells. *PNAS* 110: E4950–9.

Kiefer, CM, Hou, C, Little, JA, *et al.* (2008). Epigenetics of beta-globin gene regulation. *Mutat Res* 647: 68–76.

Kleinjan, DA, van Heyningen, V (2005). Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet* 76: 8–32.

Koch, PJ, de Viragh, PA, Scharer, E, *et al.* (2000). Lessons from Loricrin-Deficient Mice: Compensatory Mechanisms Maintaining Skin Barrier Function in the Absence of a Major Cornified Envelope Protein. *J Cell Biol* 151: 389–400.

Krivega, I, Dean, A (2012). Enhancer and promoter interactions-long distance calls. *Curr Opin Genet Dev* 22: 79–85.

Lechler, T, Fuchs, E (2005). Asymmetric cell divisions promote stratification and differentiation of mammalian skin. *Nature* 437: 275–80.

Levine, M (2010). Transcriptional enhancers in animal development and evolution. *Curr Biol*.

Levo, M, Segal, E (2014). In pursuit of design principles of regulatory sequences. *Nat Rev Genet* 15: 453–68.

Li, G, Fullwood, MJ, Xu, H, *et al.* (2010). ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biol* 11: R22.

Lieberman-Aiden, E, van Berkum, NL, Williams, L, *et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326: 289–93.

Lupiáñez, DG, Kraft, K, Heinrich, V, *et al.* (2015). Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* 161: 1012–25.

Mardaryev, AN, Gdula, MR, Yarker, JL, *et al.* (2013). p63 and Brg1 control developmentally regulated higher-order chromatin remodelling at the epidermal differentiation complex locus in epidermal progenitor cells. *Development* 141: 101–11.

Marenholz, I, Heizmann, CW, Fritz, G (2004). S100 proteins in mouse and man: From evolution to function and pathology (including an update of the nomenclature). *Biochem Biophys Res Commun*.

Marshall, D, Hardman, MJ, Nield, KM, *et al.* (2001). Differentially expressed late constituents of the epidermal cornified envelope. *Proc Natl Acad Sci U S A* 98: 13031–6.

Mehic, D, Bakiri, L, Ghannadan, M, *et al.* (2005). Fos and jun proteins are specifically expressed during differentiation of human keratinocytes. *J Invest Dermatol* 124: 212–20.

Mehrel, T, Hohl, D, Rothnagel, JA, *et al.* (1990). Identification of a major keratinocyte cell envelope protein, loricrin. *Cell* 61: 1103–12.

Mills, AA, Zheng, B, Wang, XJ, *et al.* (1999). p63 is a p53 homologue required for limb and epidermal morphogenesis. *Nature* 398: 708–13.

Mischke, D, Korge, BP, Marenholz, I, *et al.* (1996). Genes encoding structural proteins of epidermal cornification and S100 calcium-binding proteins form a gene complex ("epidermal differentiation complex") on human chromosome 1q21. *J Invest Dermatol* 106: 989–92.

Mishiro, T, Ishihara, K, Hino, S, *et al.* (2009). Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *EMBO J* 28: 1234–45.

Moreau, P, Hen, R, Wasylyk, B, *et al.* (1981). The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res* 9: 6047–68.

Nithya, S, Radhika, T, Jeddy, NLoricrin - an overview. *J Oral Maxillofac Pathol* 19: 64–8.

Ogryzko, V V., Schiltz, RL, Russanova, V, *et al.* (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* 87: 953–9.

Oh, IY, Albea, DM, Goodwin, ZA, *et al.* (2014). Regulation of the Dynamic Chromatin Architecture of the Epidermal Differentiation Complex Is Mediated by a c-Jun/AP-1-Modulated Enhancer. *J Invest Dermatol* 134: 2371–80.

Oh, IY, Quiggle, AM, de Guzman Strong, C (2016). Cis Regulation of EDC genes identified in mice with a deletion of EDC enhancer by CRISPR/Cas9 genome editing. *Unpubl Manuscr*.

Ong, C-T, Corces, VG (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet* 12: 283–93.

Palmer, CNA, Irvine, AD, Terron-Kwiatkowski, A, *et al.* (2006). Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat Genet* 38: 441–6.

Palstra, R-J, Tolhuis, B, Splinter, E, *et al.* (2003). The beta-globin nuclear compartment in development and erythroid differentiation. *Nat Genet* 35: 190–4.

Parelho, V, Hadjur, S, Spivakov, M, *et al.* (2008). Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms. *Cell* 132: 422–33.

Pillai, S, Bikle, DD, Mancianti, ML, *et al.* (1990). Calcium regulation of growth and differentiation of normal human keratinocytes: modulation of differentiation competence by stages of growth and extracellular calcium. *J Cell Physiol* 143: 294–302.

Poss, ZC, Ebmeier, CC, Taatjes, DJThe Mediator complex and transcription regulation. *Crit Rev Biochem Mol Biol* 48: 575–608.

Presland, RB, Boggess, D, Patrick Lewis, S, *et al.* (2000). Loss of Normal Profilaggrin and

Filaggrin in Flaky Tail (ft/ft) Mice: an Animal Model for the Filaggrin-Deficient Skin Disease Ichthyosis Vulgaris. *J Invest Dermatol* 115: 1072–81.

Rice, RH, Green, H (1977). The cornified envelope of terminally differentiated human epidermal keratinocytes consists of cross-linked protein. *Cell* 11: 417–22.

Rodríguez, E, Baurecht, HH, Herberich, E, *et al.* (2009). Meta-analysis of filaggrin polymorphisms in eczema and asthma: Robust risk factors in atopic disease. *J Allergy Clin Immunol* 123: 1361–70.e7.

Rollins, RA, Morcillo, P, Dorsett, D (1999). Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. *Genetics* 152: 577–93.

Rothnagel, JA, Longley, MA, Bundman, DS, *et al.* (1994). Characterization of the mouse loricrin gene - linkage with profilaggrin and the flaky tail and soft coat mutant loci on chromosome-3. *Genomics* 23: 450–6.

Rubio, ED, Reiss, DJ, Welcsh, PL, *et al.* (2008). CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci U S A* 105: 8309–14.

Segre, JA (2006). Epidermal barrier formation and recovery in skin disorders. *J Clin Invest* 116: 1150–8.

Seitan, VC, Hao, B, Tachibana-Konwalski, K, *et al.* (2011). A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. *Nature* 476: 467–71.

Sethi, I, Sinha, S, Buck, MJ (2014). Role of chromatin and transcriptional co-regulators in mediating p63-genome interactions in keratinocytes. *BMC Genomics* 15: 1042.

Sevilla, LM, Nachat, R, Groot, KR, *et al.* (2007). Mice deficient in involucrin, envoplakin, and periplakin have a defective epidermal barrier. *J Cell Biol* 179: 1599–612.

Sexton, T, Yaffe, E, Kenigsberg, E, *et al.* (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*.

Shaulian, E, Karin, M (2001). AP-1 in cell proliferation and survival. *Oncogene* 20: 2390–400.

Shlyueva, D, Stampfel, G, Stark, A (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet* 15: 272–86.

Simon, M, Green, H (1984). Participation of membrane-associated proteins in the formation of the cross-linked envelope of the keratinocyte. *Cell* 36: 827–34.

Smith, FJD, Irvine, AD, Terron-Kwiatkowski, A, *et al.* (2006). Loss-of-function mutations in the gene encoding filaggrin cause ichthyosis vulgaris. *Nat Genet* 38: 337–42.

Song, HJ, Poy, G, Darwiche, N, *et al.* (1999). Mouse Sprr2 genes: a clustered family of genes showing differential expression in epithelial tissues. *Genomics* 55: 28–42.

Song, SH, Hou, C, Dean, A (2007). A Positive Role for NLI/Ldb1 in Long-Range ??-Globin

Locus Control Region Function. *Mol Cell* 28: 810–22.

Sung, WP, Li, G, Lin, YP, *et al.* (2005). Thyroid hormone-induced juxtaposition of regulatory elements/factors and chromatin remodeling of Crabp1 dependent on MED1/TRAP220. *Mol Cell* 19: 643–53.

Symmons, O, Uslu, VV, Tsujimura, T, *et al.* (2014). Functional and topological characteristics of mammalian regulatory domains. *Genome Res* 24: 390–400.

The Chimpanzee Sequencing and Analysis Consortium (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437: 69–87.

The ENCODE Project Consortium (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447: 799–816.

Tolhuis, B, Palstra, RJ, Splinter, E, *et al.* (2002). Looping and interaction between hypersensitive sites in the active ??-globin locus. *Mol Cell* 10: 1453–65.

Tripic, T, Deng, W, Cheng, Y, *et al.* (2009). SCL and associated proteins distinguish active from repressive GATA transcription factor complexes. *Blood* 113: 2191–201.

Viganò, MA, Mantovani, R (2014). Hitting the numbers: The emerging network of p63 targets. *Cell Cycle* 6: 233–9.

Visel, A, Bristow, J, Pennacchio, LA (2007). Enhancer identification through comparative genomics. *Semin Cell Dev Biol*.

Volz, A, Korge, BP, Compton, JG, *et al.* (1993). Physical Mapping of a Functional Cluster of Epidermal Differentiation Genes on Chromosome 1q21. *Genomics* 18: 92–9.

Wallace, JA, Felsenfeld, G (2007). We gather together: insulators and genome organization. *Curr Opin Genet Dev*.

Wang, a, Johnson, DG, MacLeod, MC (2001). Molecular cloning and characterization of a novel mouse epidermal differentiation gene and its promoter. *Genomics* 73: 284–90.

Wang, Q, Carroll, JS, Brown, M (2005). Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Mol Cell* 19: 631–42.

Welter, JF, Crish, JF, Agarwal, C, *et al.* (1995). Fos-related Antigen (Fra-1), junB, and junD Activate Human Involucrin Promoter Transcription by Binding to Proximal and Distal AP1 Sites to Mediate Phorbol Ester Effects on Promoter Activity. *J Biol Chem* 270: 12614–22.

Wendt, KS, Yoshida, K, Itoh, T, *et al.* (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451: 796–801.

Whyte, WA, Orlando, DA, Hnisz, D, *et al.* (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153: 307–19.

Williamson, I, Hill, RE, Bickmore, WA (2011). Enhancers: from developmental genetics to the

genetics of common human disease. *Dev Cell* 21: 17–9.

Yang, A, Schweitzer, R, Sun, D, *et al.* (1999). p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. *Nature* 398: 714–8.

Yoneda, K, Steinert, PM (1993). Overexpression of human loricrin in transgenic mice produces a normal phenotype. *Proc Natl Acad Sci U S A* 90: 10754–8.

Yuspa, SH, Kilkenny, AE, Steinert, PM, *et al.* (1989). Expression of murine epidermal differentiation markers is tightly regulated by restricted extracellular calcium concentrations in vitro. *J Cell Biol* 109: 1207–17.

Zhao, XP, Elder, JT (1997). Positional cloning of novel skin-specific genes from the human epidermal differentiation complex. *Genomics* 45: 250–8.

Zhao, Z, Tavoosidana, G, Sjölinder, M, *et al.* (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet* 38: 1341–7.

Chapter 2

Regulation of the Dynamic Chromatin Architecture of the Epidermal Differentiation Complex is Mediated by a c-Jun/AP-1-Modulated Enhancer

by

Inez Y. Oh, Danielle M. Albea, Zane A. Goodwin,
Ashley M. Quiggle, Breeana P. Baker, Ann M. Guggisberg,
Jessica H. Geahlen, Grace M. Kroner,
and Cristina de Guzman Strong

Journal of Investigative Dermatology 134: 2371-80

Format adapted for dissertation

2.1 Abstract

The Epidermal Differentiation Complex (EDC) locus comprises a syntenic and linear cluster of genes whose concomitant expression is a hallmark feature of differentiation in the developing skin epidermis. Many of the EDC proteins are cross-linked together to form the cornified envelope, an essential and discrete unit of the mammalian skin barrier. The mechanism underlying coordinate transcriptional activation of the EDC is unknown. Within the human EDC, we identified an epidermal-specific regulatory enhancer, 923, that responded to the developmental and spatio-temporal cues at the onset of epidermal differentiation in the mouse embryo. Comparative chromosomal conformation capture (3C) assays in proliferating and differentiated primary mouse keratinocytes revealed multiple chromatin interactions that were physiologically sensitive between the 923 enhancer and EDC gene promoters and thus depict the dynamic, chromatin topology of the EDC. We elucidate a mechanistic link between c-Jun/AP-1 and 923, whereby AP-1 and 923-mediated EDC chromatin remodeling is required for functional EDC gene activation. Thus, we identify a critical enhancer/transcription factor axis governing the dynamic regulation of the EDC chromatin architecture and gene expression and provide a framework for future studies towards understanding gene regulation in cutaneous diseases.

2.2 Introduction

The epidermis lies at the surface of the skin and provides the first line of defense against the external environment (Koster and Roop 2007; Fuchs 2009; Kubo et al. 2012). Protecting against infection and inflammation, the epidermis comprises stratified layers of epidermal cells or keratinocytes that are individually surrounded by a cornified envelope and function as one of the essential core units of the skin barrier. To build the epidermal barrier akin to a “bricks-and-

mortar" architecture, a basal keratinocyte in the innermost layer of the epidermis receives an inductive cue to differentiate and orients its mitotic spindle perpendicularly to the basement membrane (Lechler and Fuchs 2005). In doing so, an asymmetric cell division gives rise to a basal cell and a suprabasal daughter cell that is committed to terminal differentiation. As the keratinocyte completes the differentiation process, it is pushed upward and sequentially through the spinous and granular layers and finally to the outermost stratum corneum.

A hallmark feature for the execution of the terminal epidermal differentiation program is the expression of genes encoded by the Epidermal Differentiation Complex (EDC) locus (Mischke et al. 1996; Zhao and Elder 1997; Marshall et al. 2001; de Guzman Strong et al. 2010). The EDC (located on human 1q21 and mouse 3q) consists of 4 gene families that are associated with skin barrier formation: Small Proline Rich Region (*SPRR*), Late Cornified Envelope (*LCE*), filaggrin (*FLG*) and filaggrin-like (*FLG*-like), and *S100* genes. Genes encoded in the EDC are coordinately activated during embryonic epidermal differentiation (de Guzman Strong et al. 2010). Exciting and recent studies in mice have identified a role for epigenetics in the regulation of the EDC during skin development (reviewed in (Botchkarev et al. 2012)). Epidermal-specific loss of Ezh2, an essential component of the Polycomb repressor complex for histone modification, resulted in early epidermal differentiation owing to precocious recruitment of AP-1 transcription factor to the EDC for gene expression (Ezhkova et al. 2009). Furthermore, Satb1, a higher-order genome organizer, was recently identified as a p63 target and binds to the EDC (Fessing et al. 2011). Satb1-/- mice exhibited alterations in the chromatin conformation of the EDC resulting in defects in keratinocyte-specific and EDC gene expression and hence abnormal epidermal morphology and further demonstrated a requirement for the proper establishment of higher order EDC chromatin structure and coordinated gene expression. This was further

supported by confocal microscopy and computational modeling that identified distinct and active remodeling of the nuclear architecture associated with gene expression specifically in the terminally differentiated keratinocyte (Gdula et al. 2013). Comprehensive studies of the β-globin locus control region and the X-inactivation center further support evidence for causality of chromatin folding and 3D genome organization with respect to gene regulation (Deng et al. 2012; Nora et al. 2012). However, despite these studies, the molecular mechanism(s) that underlie activation and coordinate regulation of the EDC genes at the nucleotide level are unknown. The synteny and linearity of the EDC across a wide range of mammalian species suggests a molecular mechanism originating at the proximal genomic level.

One plausible model is the activation of critical EDC expression by cis-regulatory elements during skin barrier formation. Comparative genomics and the Encyclopedia of DNA Elements (ENCODE) consortium that has annotated 80% of the genome attributed to function have greatly facilitated identification of regulatory elements (Dunham et al. 2012). We previously identified many conserved non-coding elements (CNEs) within the human EDC that could synergistically or independently coordinate EDC gene expression (de Guzman Strong et al. 2010). Approximately 50% of them exhibit regulatory activity. CNE 923 (approximately 923 kb from the transcriptional start site of *S100A10*, the most 5' EDC gene) displayed the strongest enhancer activity in proliferating and differentiated keratinocytes in our functional screen. This result corroborated with transgenic reporter mice that demonstrated epidermal-specific enhancer activity for CNE 923 *in vivo*. This led us to hypothesize a role for 923 in the coordinate transcriptional activation of the EDC.

Here, we tracked the activity of CNE 923 during development in transgenic mice and identified spatio-temporal sensitivity for 923 that coincides with the onset and patterning of

epidermal differentiation. Chromatin conformation capture (3C) studies were employed to determine the physical interactions between 923 and EDC gene promoters and revealed multiple chromatin spatial interactions surrounding 923. Comparative 3C analyses between proliferating and differentiated primary keratinocytes revealed a dynamic 923-centric EDC chromatin domain associated with concomitant EDC gene expression. Comparative genomics and genetic studies identified an AP-1 transcription factor binding site within 923 that was required for enhancer activity. We determine that the AP-1 binding site in 923 is functionally relevant, as pharmacological inhibition of AP-1 in calcium-induced keratinocytes repressed EDC gene expression and was associated with aberrant chromatin remodeling and loss of c-Jun/AP-1 binding to 923. Thus, our results provide a framework to examining molecular mechanisms that link DNA sequence to chromatin architecture and biological functions relevant to development and disease.

2.3 Results

2.3.1 923 is an epidermal-specific enhancer responsive to the spatial and temporal cues in the developing mouse epidermis

We previously observed that human CNE 923 exhibited epidermal-specific enhancer activity, driving lacZ expression in G0 transgenic mice (923-*hsp68*-lacZ) analyzed only at mouse embryonic day (E)16.5 (de Guzman Strong et al. 2010). However, the onset and the spatial and temporal patterning for 923 during mouse embryonic development were unclear. To address this, we generated additional 923-*hsp68*-lacZ transgenic mice. 923 enhancer activity (as measured by *lacZ* transcript levels) was detected as early as E15.5 in the developing mouse epidermis (Figure 2.1e), and coincided with the onset of early epidermal differentiation as

demonstrated by positive Keratin 1 (K1) expression (Figure 2.1d) and activation of involucrin (*Ivl*), *Flg*, and loricrin (*Lor*) expression (Figure 2.1e). X-galactose reactivity (blue) was not detectable at E15.5 in whole-mount or cross-sections of the epidermis (Figure 2.1a, 2.1b) owing to the lack of β-galactosidase protein expression. At E16.5 and E17.5, we observed expression of 923 enhancer activity correlating with the patterning of barrier acquisition (dorsal to ventral migration pattern (Hardman et al. 1998) (Figure 2.1a). 923 β-galactosidase enhancer activity was localized to the spinous to stratum corneum layers of the dorsal epidermis at E16.5 and E17.5 (Figure 2.1b). Failure to detect β-galactosidase activity on the dorsal epidermis of E17.5 whole-mount embryos is consistent with barrier acquisition that precludes substrate penetration to detect β-galactosidase activity. Together, the data supports the responsiveness of 923 enhancer activity to the spatio-temporal cues of the developing mouse epidermis.

2.3.2 The dynamic chromatin architecture of the EDC

We previously identified DNaseI hypersensitivity for 923 (de Guzman Strong et al. 2010) (de Guzman Strong *et al.*, 2010) and note ENCODE-annotated H3K4me1 histone modification mark in proliferating primary human keratinocytes that independently tags functional enhancers (Ernst et al. 2011) (Figure 2.3a). Enhancers are known to form long-range physical interactions with target gene promoters for activation (Tolhuis et al. 2002). Given these observations and the spatio-temporal sensitivity in the developing mouse epidermis, we hypothesized a role for 923 in mediating the chromatin conformation of the EDC.

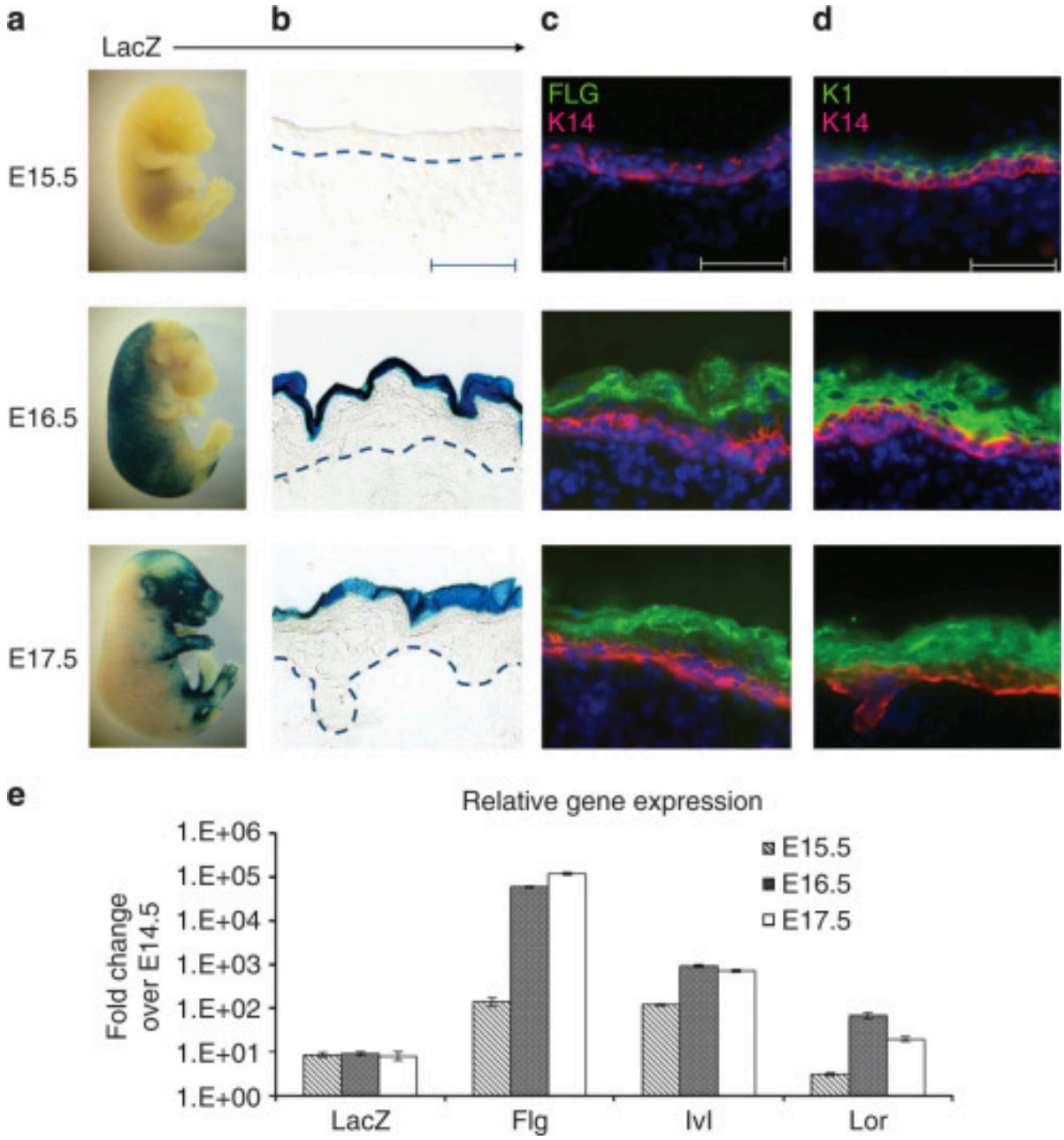


Figure 2.1. 923 is sensitive to spatio-temporal cues during mouse embryonic epidermal development. (**a**) Whole mount lacZ staining of 923-*hsp68-lacZ* mice demonstrates 923 activity (β-galactosidase/X-gal blue reactivity) following spatio-temporal patterns of epidermal barrier formation in the developing embryo by initial observation of activity at E16.5 (dorsal) that migrates ventrally by E17.5. (**b**) 923 activity localizes to granular and spinous layers of embryonic dorsal epidermis with corresponding (**c**) filaggrin (FLG) and (**d**) keratin1 (K1) immunofluorescent staining (green). Keratin14, K14 (red) marks basal keratinocytes, 20X. Dotted lines, basement membrane. Experiments observed in ≥ 2 independent mice. (**e**) 923 activity (qPCR, lacZ transcript) is noted at E15.5, E16.5, and E17.5 dorsal epidermis, concomitant with *Flg*, involucrin (*Ivl),* and loricrin *(Lor)* transcription relative to E14.5. Error bars represent mean+/-SD.

To test this hypothesis, 3C assays coupled with quantitative PCR (qPCR) (Hagège et al. 2007) were employed in proliferating and differentiated primary mouse keratinocytes to detect physical chromatin interactions at the sub-megabase level between the endogenous mouse 923 ortholog and the EDC genes. 923 formed multiple interactions with EDC gene promoters (9 out of 46 tested queries, *Sprr2a1*, *Sprr2d*, *Sprr2f*, *Sprr1b*, *Sprr3*, *Ivl*, *Lce1b*, *Lce1a2*, and *Crct1* [cysteine-rich C-terminal 1]) in proliferating keratinocytes despite the lack of EDC gene expression relative to the differentiated keratinocytes (Figure 2.2a, 2.2b). In differentiated keratinocytes, a reconfiguration of the EDC chromatin state was identified and was associated with eleven 923-mediated chromatin spatial interactions between a HindIII fragment 5' of *Lce3b and S100a6*, *Sprr2a1*, *Sprr2b*, *Sprr3*, *Sprr4*, *Ivl*, *Lce6a*, *Lce1b*, *Lce1e*, and *Crct1* gene promoters that was relatively consistent with their expression during terminal differentiation (Figure 2.2a, 2.2b). In comparison to the proliferating cells, the observed interactions in differentiated keratinocytes that were lost included *Sprr2d*, *Sprr2f*, and *Sprr1b*, and *Lce1a2* (within 250 kb of 923) as well as a gain with 5' of *Lce3b*, *S100a6*, *Sprr2b*, *Sprr4*, *Lce6a* and *Lce1e*. Notably, the gain of 923's interaction with *S100a6* was located 2Mb away across a gene desert and observed higher frequencies of interactions with *Sprr2a1* and *Sprr2b* (both >250kb away from 923). All of the above genes with the exception of *Sprr2f* are expressed by E15.5 as previously described in the newly differentiated dorsal epidermis of the developing mouse embryo (de Guzman Strong et al. 2010). The genes for which there were gains of interactions in differentiated keratinocytes had similar if not increased levels of expression at E16.5 relative to E15.5. These observations further identify a longer range EDC chromatin topology in differentiated cells (Figure 2.2a) and are underestimated given the modest coverage associated with 3C methodology and as not all keratinocytes completely differentiated. In sum, our results support both shared and unique

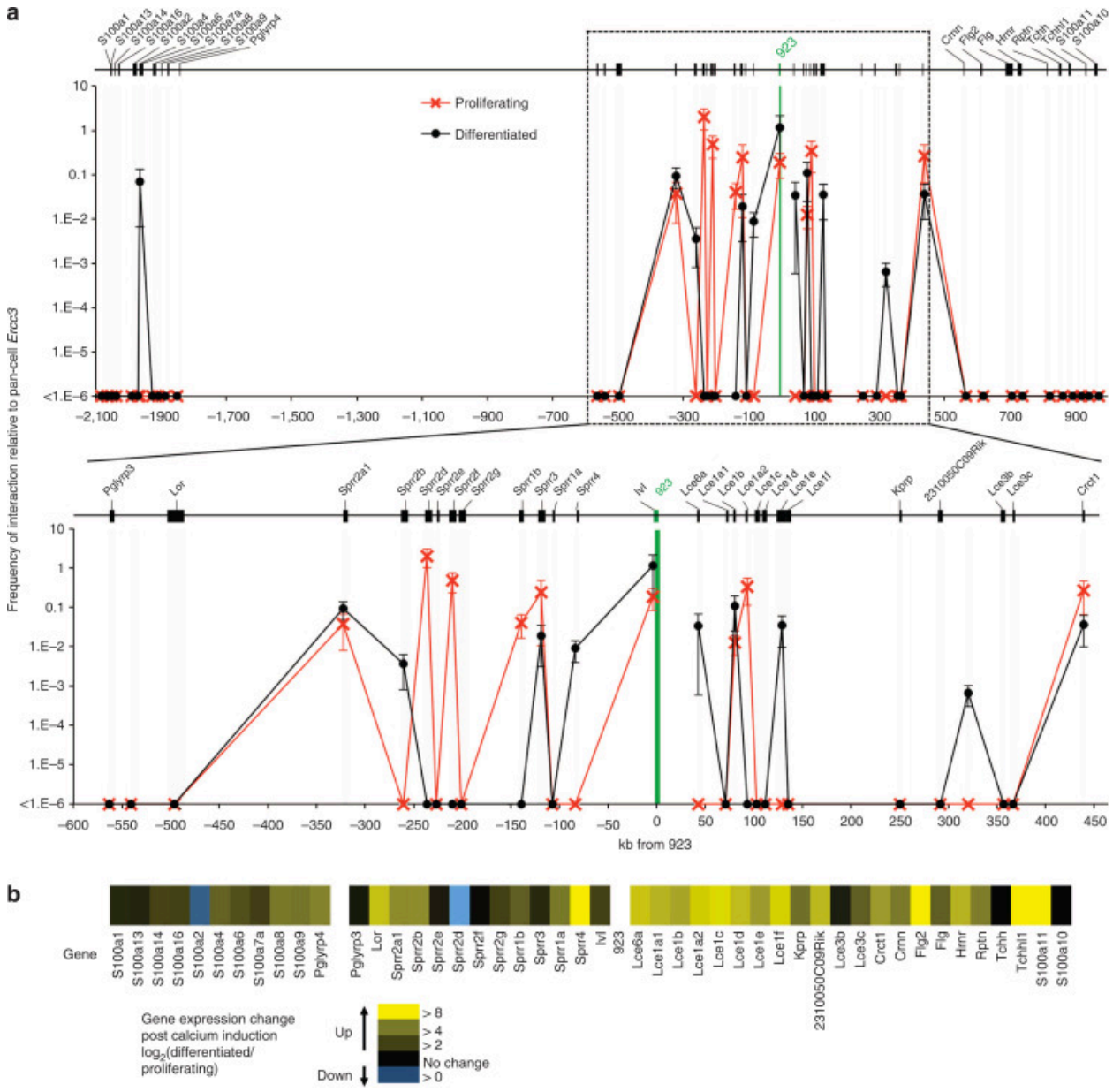


Figure 2.2. The Chromatin State of the Mouse EDC is Dynamic. (**a**) Semi-quantitative chromosomal conformation capture (3C-qPCR) assays were performed on primary mouse keratinocytes (proliferating and differentiated). Peaks = frequencies of physical interactions observed between HindIII restriction fragments containing 923 (green bar) and queried restriction fragments containing/neighboring EDC gene promoters (black bars + gray lines) or proximal sequences (gray lines) relative to a cell-ubiquitous *Ercc3* control. HindIII fragment (5' of *Lce3b*) represents a chromatin interaction that was not enhancer-promoter specific. Peaks, average of at least 2 biological replicates. Error bars represent the mean+/-SEM. (**b**) EDC gene expression heatmap (fold change) in differentiated vs. proliferating keratinocytes based on analysis of 3 pairwise RNA-seq libraries, green/yellow = upregulated, blue = downregulated, black = no change, (Table S3).

chromatin spatial interactions between the 923 epidermal-specific enhancer and EDC gene promoters in proliferating and differentiated primary mouse keratinocytes that represent the dynamic chromatin architecture of the EDC.

2.3.3 AP-1 transcription factor binding is required for 923 enhancer activity and EDC gene expression

To elucidate the molecular mechanism underlying 923 enhancer activity, we performed a bioinformatics search to identify core transcription factor binding sequences within 923 that are responsible for driving functional enhancer activity. Assessment of core enhancer functional activity in 4 PhastCons blocks were prioritized and represent highly conserved sequences between 28 vertebrate species and therefore likely to impart function (Siepel et al. 2005; Alexander et al. 2010) (Figure 2.3b). Deletions of blocks 1 and 4 at the 5' and 3' ends of 923 significantly decreased luciferase activity (Figure 2.3c) thus demonstrating a functional role for these blocks for 923 enhancer activity. As deletion of block 1 resulted in the greatest and more significant decrease in enhancer activity, we prioritized a search for transcription factor binding motifs within block 1. We identified an AP-1 transcription factor binding site (Figure 2.4a) and hypothesized that AP-1 is required for 923 enhancer activity. Deletion of the AP-1 binding site by site-directed mutagenesis led to a significant decrease in 923 enhancer activity under proliferating and differentiated conditions (Figure 2.4b) thus demonstrating a functional role for AP-1 to mediate 923 enhancer activity in both physiological states.

To examine a role for AP-1 activity with respect to 923 enhancer activity and EDC gene activation, AP-1 binding was inhibited by Tanshinone IIA (TanIIA) treatment in primary mouse keratinocytes induced to differentiate (calcium induction) (Ezhkova et al. 2009). Calcium-

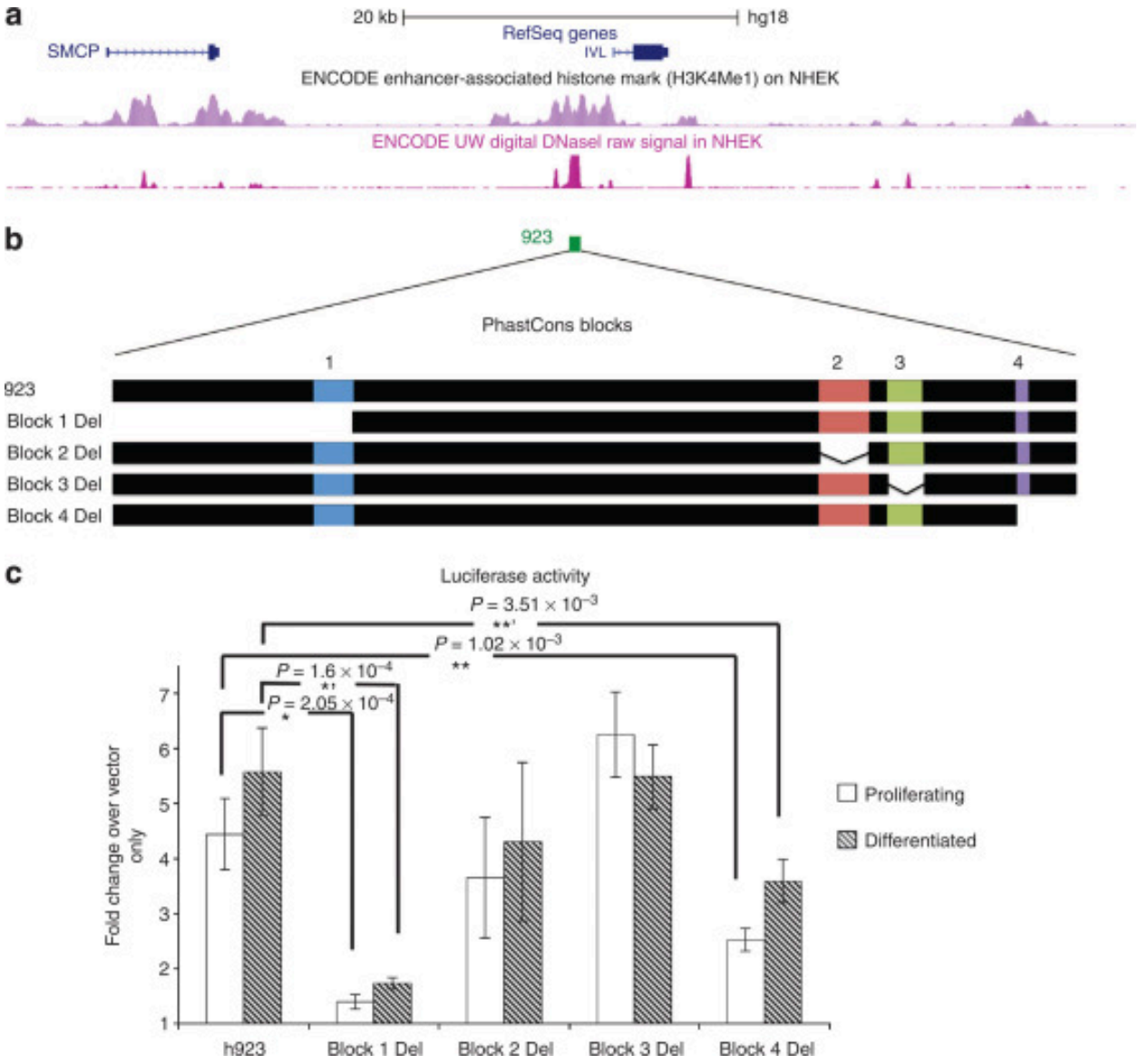


Figure 2.3. PhastCons (vertebrate conserved) blocks 1 and 4 are required for 923 enhancer activity. (**a**) 923 correlates with ENCODE-annotated strong enhancer (H3K4me1) and DNaseI hypersensitivity clusters in normal human epidermal keratinocytes (NHEK) and 4 PhastCons blocks (UCSC, hg18, block 1 (23 bp): chr1:151145182-151145204; block 2 (34 bp): chr1:151145525-151145558; block 3 (26 bp): chr1:151145572-151145597; block 4 (9 bp): chr1:151145661-151145669). (**b**) Individual deletion (del) of each PhastCons block reveals that c) blocks 1 and 4 are required for enhancer activity based on transient dual-luciferase reporter assays in proliferating and differentiating keratinocytes (n=2). *, P = 2.05 x 10^{-4} , *', 1.6 x 10^{-4}, **, P = 31.02 x 10^{-3}, **', 3.51 x 10^{-3}. P-values are based on a two-tailed t-test. Error bars represent mean+/-SE.

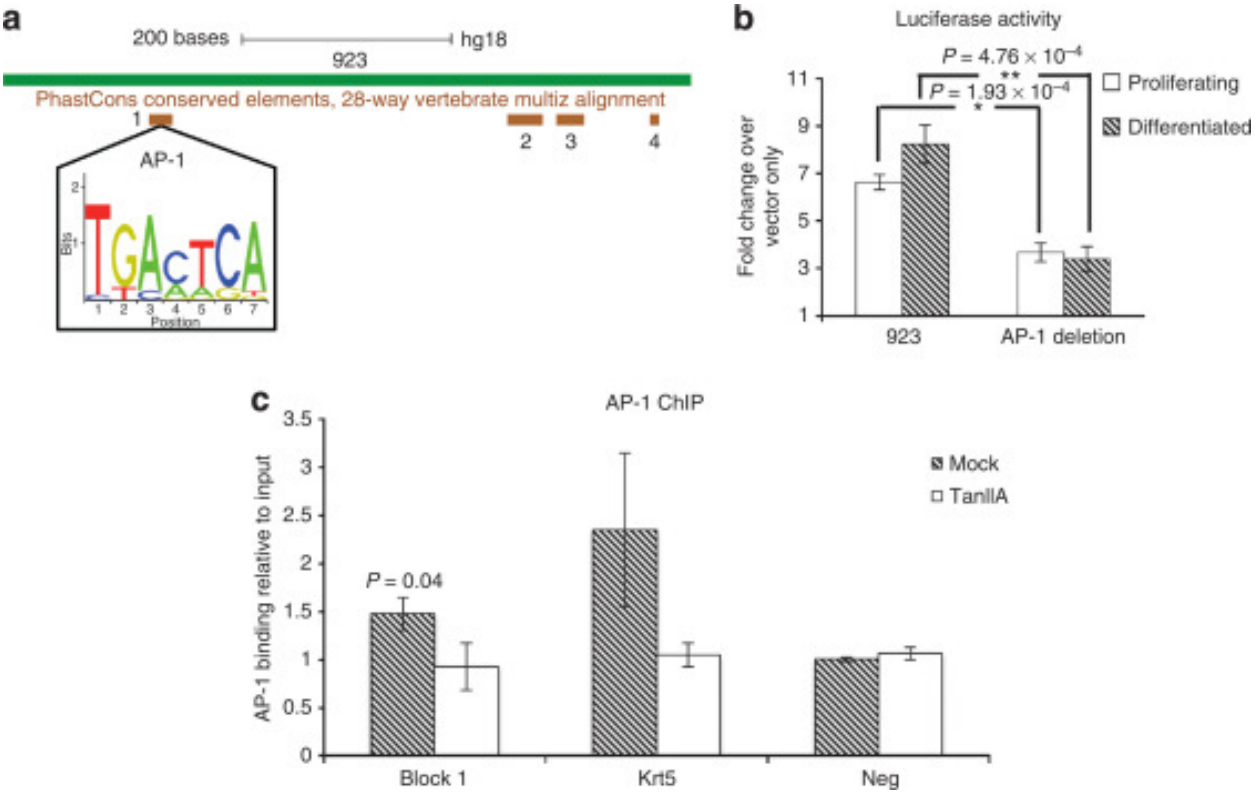


Figure 2.4. c-Jun/AP-1 transcription factor binding to PhastCons block 1 is required for 923 enhancer activity. (**a**) PhastCons block 1 contains a conserved AP-1 transcription factor binding sequence (UCSC, hg18, chr1:151145183-151145189). (**b**) Deletion of the AP-1 binding site significantly decreased luciferase and hence enhancer activity (n=3), * P=1.93 x 10^{-4}, ** P=4.76 x 10^{-4} (two-tailed t-test). (**c**) Chromatin immunoprecipitation on differentiated cells demonstrates 1.6-fold decrease in AP-1 binding to PhastCons block 1 in TanIIA-treated versus mock-treated cells (P=0.04). The positive control, an ENCODE-annotated site within Keratin5 (*Krt5*), displayed 2.3-fold decrease in AP-1 binding in TanIIA-treated versus mock-treated cells (P=0.06). A negative control (no AP-1 site) showed no difference between mock-treated and TanIIA-treated cells (n=2). P-values based on a one-tailed t-test.

induced keratinocytes treated with TanIIA exhibited repressed EDC gene expression (Figure 2.5b). Moreover, chromatin immunoprecipitation for AP-1 (c-Jun) revealed that the repression was associated with the loss of AP-1 binding to block 1 in 923 in TanIIA-treated, calcium-induced keratinocytes compared to mock controls (Figure 2.4c). Together, the data demonstrates a requirement for AP-1 in activating EDC gene expression, specifically associated with functional AP-1 binding within the 923 enhancer *in vivo*.

2.3.4 The c-Jun/AP-1/923 axis regulates the EDC transcriptome by modulating the chromatin architecture

To identify the mechanism by which AP-1 inhibition represses EDC gene expression, we returned to 3C assays to examine the chromatin conformation of the EDC with respect to 923 in the context of AP-1 pharmacological inhibition. The chromatin conformation was assessed at 48 hours post-TanIIA treatment to best ascertain the direct effects of AP-1 inhibition as opposed to secondary effects beyond 48 hours. Although there was no significant differences in the number of 923-mediated chromatin interactions in TanIIA-treated vs. mock-treated differentiated keratinocytes (11 vs. 12), only 6 interactions were shared (*Sprr2a1*, *Sprr3*, *Ivl*, *Lce1d*, *Lce1e*, and 5' of *Lce3b*) and are close by (within 325 kb) (Figure 2.5a). Moreover, in TanIIA-treated keratinocytes, there was a loss of 923 interactions with *Sprr2b*, *Sprr2d*, *Sprr1b*, *Lce6a*, and *2310050C09Rik* and a gain of spatial interactions with *Sprr1a*, *Lce1b*, and *Lce1c* and at the relatively extreme 5' and 3' ends of the EDC reaching as far away as >2 Mb and 866 kb in *S100a13* and *Tchh* (trichohyalin), respectively. Despite the gain of chromatin interactions, no appreciable differences in gene expression for *Sprr1a*, *Lce1b*, *Lce1c*, *S100a13* and *Tchh* were observed. It appears that AP-1 pharmacological inhibition resulting in decreased c-Jun/AP-1 binding at 923 (Figure 2.4c) was not sufficient to completely abrogate all chromatin spatial

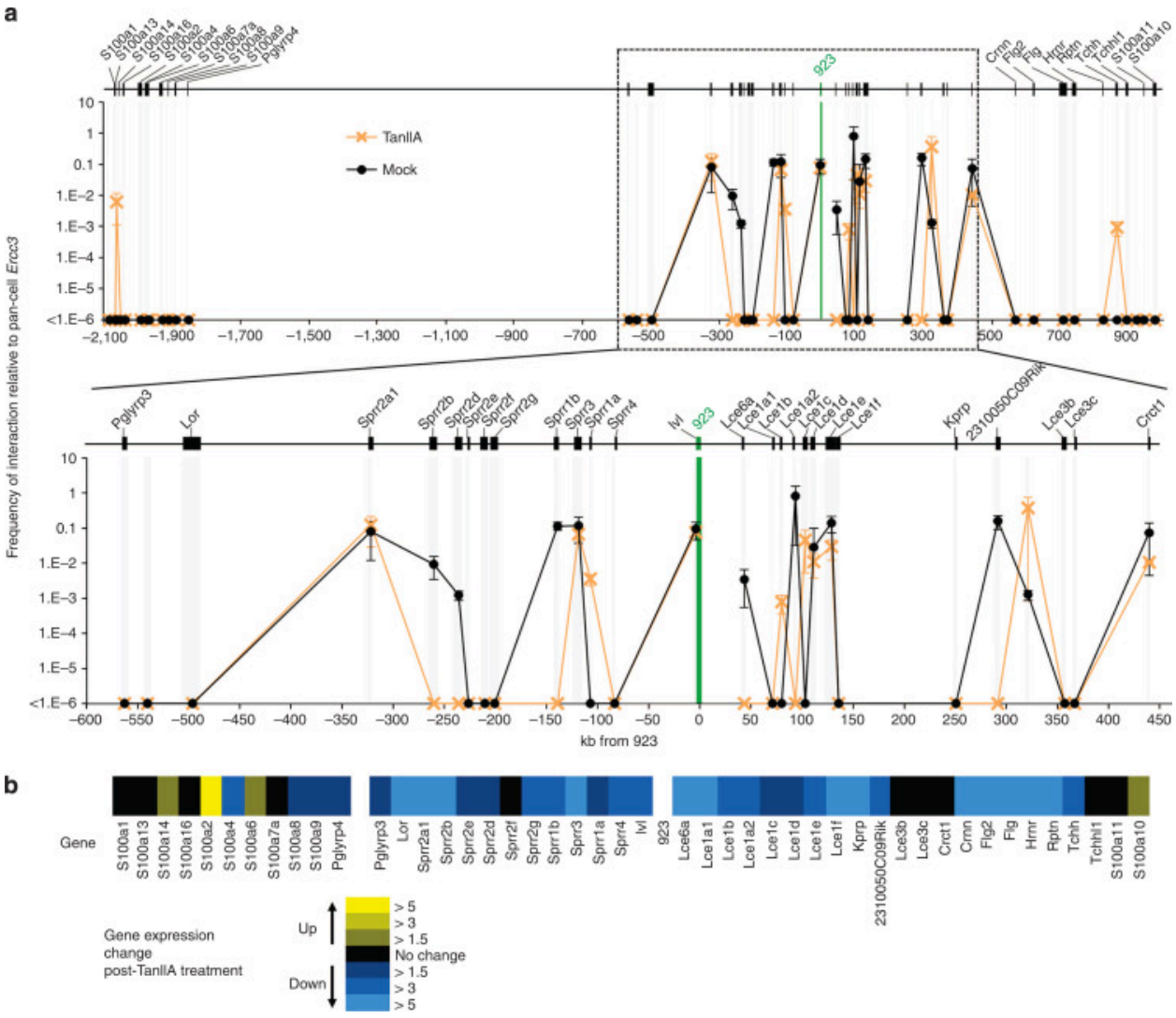


Figure 2.5. c-Jun/AP-1 activity is required for 923-mediated chromatin state remodeling to activate EDC gene expression. (**a**) Pharmacological inhibition of AP-1 binding using TanIIA (1.0 mg/L) modulates the chromatin interactions between 923 (green bar) and EDC gene promoters (gray bars) (n=2) in differentiated keratinocytes and (**b**) represses EDC gene expression based on heatmap depiction, yellow = upregulated, blue = downregulated in TanIIA-treated relative to mock.

interactions within the EDC that could be maintained by other transcription factors. However, our data supports a role for AP-1 in mediating proper 923-centric EDC chromatin conformation for EDC gene activation.

2.4 Discussion

Although 3C assays and recent high-throughput genomic studies have enhanced our understanding of chromatin architecture and gene regulatory modules (de Wit and de Laat, 2012), the mechanisms governing chromosomal spatial interactions are poorly understood. Our studies identify a molecular mechanism describing transcription factor/enhancer modulation of a cluster of genes, namely AP-1 in the EDC architecture required for epidermal differentiation. We translate a "linear" interpretation (de Wit and de Laat 2012) of the keratinocyte genome from our studies and ENCODE and prioritize functional studies on 923 to elucidate the 3D structure or chromatin interactions within the EDC. Our study demonstrates that 923 displays epidermal-specific enhancer activity that tracks with spatial and temporal patterns of epidermal differentiation and barrier formation during normal mouse development. We further elucidate an association of 923 with the coordinate activation of EDC genes during epidermal differentiation based on 3C assays that identified nearby chromatin spatial interactions between 923 and several EDC genes located as far as 2Mb away. Specifically, we observe a chromatin state of the EDC in proliferating keratinocytes that are marked by fewer cis-spatial interactions with 923 and do not express EDC genes. By contrast, the chromatin state of the EDC remodels in differentiated keratinocytes that express many EDC genes, as demonstrated by greater observed 923-mediated interactions with EDC gene promoters. Although *in vivo* knockout studies for 923 are beyond the scope of this study and would address the functional role of 923 as an intriguing locus control region (LCR) for the EDC during mouse development, our data nevertheless support a functional

role for 923 in mediating the chromatin spatial interactions of the EDC. In support of this model, a recent study evaluating a ZF-mediated Ldb1/β-globin LCR physical tethering to the β-globin promoter in GATA-1 deficient erythroid cells demonstrated causality of chromatin spatial interactions to gene transcription (Deng et al. 2012). The requirement of the AP-1 binding site for 923 enhancer activity in both proliferating and differentiated states and the repression of EDC expression by pharmacological inhibition of AP-1, suggest that the AP-1/923 axis is an important mechanism to coordinate the EDC transcriptome. A bioinformatic analysis of transcription factor binding sites in 923 suggests additional putative transcription factor binding sites including CREB within PhastCons Block 4 that could likely contribute to 923 enhancer function.

It is interesting to note that even in a proliferative state, the loss of the AP-1 binding site in 923 led to a significant decrease in enhancer activity and chromatin interactions were observed between EDC gene promoters and 923. A majority of AP-1 members are expressed in basal keratinocytes with a more restricted expression of specific AP-1 members in the suprabasal layers (Jochum et al. 2001). AP-1 is known to translate extracellular signals to a transcriptional response (Schonthaler et al. 2011). Together, these observations and our data suggest a role for AP-1 (c-Jun) in modulating 923 activity in basal keratinocytes by folding the EDC chromatin state, and for which activation of EDC transcription in the terminally differentiated keratinocyte is driven by the specificity of an AP-1 homo/heterodimeric partner. Although the epidermis with targeted loss of c-Jun (Zenz et al. 2003) and c-Jun/JunB (Guinea-Viniegra et al. 2009) exhibited normal skin morphology (that could be attributed to compensatory mechanisms to correct for skin barrier (Koch et al. 2000; Huebner et al. 2012)), the epidermal-specific c-Jun/JunB and JunB knockouts exhibited inflammatory defects owing to interleukin 6 (IL-6) and tumor necrosis

factor α (TNFα) expression respectively. Together, these observations lay the groundwork for investigations linking the role of chromatin architecture in skin barrier as well as inflammation.

A recent study revealed that chromatin architectural proteins may play a greater role than transcription factors in mediating promoter-enhancer interactions (Phillips-Cremins et al. 2013). Genome-wide analysis of chromatin interactions lost during the differentiation of embryonic stem cells (ESCs) to neural progenitor cells (NPCs) unveiled a strong colocalization of the architectural proteins Mediator and cohesin to the ESC-specific interactions, a partial colocalization of the Oct4/Sox2/Nanog (OSN) transcription factors with Mediator and cohesin, and far fewer interactions that were enriched for only transcription factors. The roles of Mediator and cohesin in mediating enhancer-promoter interactions were validated by the abrogation of an interaction between *Olig1* and a putative ESC-specific enhancer in Mediator and cohesin knockdown cells. This data suggests that a proportion of AP-1-mediated 923 interactions may in fact be dependent on the presence of chromatin architectural proteins that are able to maintain these interactions even when AP-1 activity is inhibited, while a smaller proportion of interactions are mediated solely by AP-1.

A genomic study has recently elucidated the chromatin topologies of the human and mouse genome that are marked by distinguishing structural topological domains (Dixon et al. 2012). These domains are stable, highly conserved, and often demarcated by boundaries enriched for CCCTC-binding factor (CTCF), housekeeping genes, tRNAs, and short interspersed nuclear element (SINE) transposons. That the EDC is also syntenic and linear across a wide range of metatherian genomes (de Guzman Strong et al. 2010) suggests a model for the EDC as a single distinct topological domain. Of note, ENCODE-annotated CTCF elements flank the gene families within and just outside the EDC, suggesting a role for CTCF as boundary elements for a

putative EDC topology (Ernst et al. 2011). More high-throughput and higher resolution characterization of the EDC chromatin conformation using 4C or 5C methodology would certainly address this hypothesis as well as to depict additional chromatin interactions in an unbiased manner.

The EDC has been implicated in atopic dermatitis and psoriasis (Giardina et al. 2006; Palmer et al. 2006; Sandilands et al. 2007; de Cid et al. 2009; Esparza-Gordillo et al. 2009; Hirota et al. 2012; Paternoster et al. 2012). Specifically, discovery of *FLG* mutations initially in ichthyosis vulgaris (Smith et al. 2006) and particularly in atopic dermatitis (AD) (Palmer et al. 2006) and other atopic diseases such as asthma and allergic rhinitis (Irvine et al. 2011), highlights the importance of how even one of the EDC components broadly affects prevalent allergic diseases. Even at the exclusion of common *FLG* mutations, genetic association to the EDC continues to persist in atopic dermatitis suggesting additional genetic variants within the EDC (Morar et al. 2007; Esparza-Gordillo et al. 2009). Our analysis provides a genomic framework for which we can begin to interrogate regulatory element variants as causative in these diseases. Although discovery of causative variants is prioritized in genomic regions in linkage disequilibrium (LD) with genome-wide association study (GWAS)-identified SNPs, our chromatin experiments suggest discovery of causative SNPs that are not in LD but are in "physical proximity" and *trans* to GWAS-identified SNPs.

2.5 Materials and Methods

2.5.1 Mice

h923-*hsp68-lacZ* reporter FVB/N mice were housed in pathogen-free, barrier facilities at NIH (Bethesda, MD) and Washington University School of Medicine (St. Louis, MO). All

animal procedures were approved by the NHGRI Animal Care and Use Committee and Washington University Division of Comparative Medicine Animal Studies Committee. All animal work was conducted in accordance with the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. Morning observation of a vaginal plug was designated as embryonic day (E) 0.5.

2.5.2 LacZ staining and Immunohistochemistry

Whole-mount embryos and frozen OCT sections were stained overnight for β-galactosidase activity as previously described (de Guzman Strong et al. 2010) and imaged on a Nikon SMZ 1500 Stereomicroscope and a Nikon Eclipse 80i brightfield microscope (Nikon, Tokyo, Japan), respectively. Primary antibodies used for immunofluorescence are rabbit K1 (17iKSCN, 1:500), rabbit FLG (5C-KSCN, 1:500) and chicken K14 (5560, 1:1000) (courtesy of J. Segre). Secondary antibodies used were goat anti-rabbit (Alexa Fluor 488, 1:500) and goat anti-chicken (Alexa Fluor 594, 1:1000) IgG antibodies (Life Technologies, Frederick, MD). Sections were fixed in 4% paraformaldehyde (Electron Microscopy Sciences, Hatfield, PA) prior to permeabilization with 0.1% Triton X-100 and subsequent antibody incubation. Sections were counterstained with SlowFade Gold antifade reagent with DAPI (Life Technologies, Frederick, MD) prior to fluorescent imaging on a Zeiss AxioImager Z1 and captured with AxioCam MRc and Axiovision software (Carl Zeiss, Stockholm, Sweden).

2.5.3 Chromosomal conformation capture (3C) assay

Primary keratinocytes were isolated from newborn mice as previously described (Lichti *et al.*, 2008) and plated under proliferating or differentiating (2.0 mM Ca^{2+}) conditions in custom keratinocyte media. 3C assays were performed as previously described (Hagège et al. 2007).

Briefly, approximately 10 million cells were harvested at 72 hours post-calcium treatment and cross-linked with 2% formaldehyde prior to overnight HindIII (New England Biolabs, Ipswich, MA) digestion. Each 3C library was assessed for efficient digestion efficiency by qPCR, and then further ligated with T4 DNA ligase (New England BioLabs, Ipswich, MA) overnight, decrosslinked using Proteinase K (IBI Scientific, Peosta, IA), and purified by phenol/chloroform (Life Technologies, Frederick, MD). Each putative physical interaction between 923 and an EDC gene promoter (detected by head-to-head [same strand] primer pair within 50-150 bp of a HindIII cut site and designed in NCBI37/mm9, Table S1) was detected by qPCR (Quantitect SYBR Green, Qiagen, Chatsworth, CA, ViiA7, Applied Biosystems, Foster City, CA) in triplicate in 3C libraries of equivalent concentrations. C_T values for each measured interaction were normalized against C_T values across an uncut region and a pan-cell *Ercc3* chromatin interaction (Hagège et al. 2007). Positive 3C interactions represent a minimum average of at least 1 replicate from 2 independent 3C libraries as a criterion to best exclude false positive and random events.

2.5.4 RNA isolation and analysis

RNA-seq RNA isolation and analysis is described in supplementary methods (Appendix A). RNA-seq data have been deposited in the NCBI SRA under accession number PRJNA210793.

qRT-PCR Real-time qPCR on cDNA (generated using SuperScript II reverse transcriptase (Life Technologies)) using SYBR Green was performed in triplicate (ABI ViiA7, Foster City, CA) and normalized to β_2-microglobulin. Only C_T values with single peaks on melt-curve analyses were included. Primers are listed in Table S2.

2.5.5 Luciferase assay

923 deletion constructs were cloned synthetically (IDT) or by PCR amplification. AP-1 site deletion was generated by site-directed mutagenesis (QuikChange Site-Directed Mutagenesis Kit, Agilent, Palo Alto, CA). All clones were verified by Sanger sequencing. Dual luciferase assays were performed in duplicate as previously described (de Guzman Strong et al. 2010).

2.5.6 Transcription factor binding prediction

Transcription factor binding sites were predicted by aligning each PhastCons block sequence (with a relative profile score threshold of 80%) against the JASPAR CORE database of transcription factor binding profiles (http://jaspar.genereg.net/) (Bryne et al. 2008).

2.5.7 AP-1 binding inhibition assay

Primary keratinocytes grown under differentiating conditions (2.0mM Ca^{2+}) were treated with either 1.0mg/L Tanshinone IIA (Biomol, Plymouth Meeting, PA) or DMSO (mock) during calcium shifting 24 hours after plating. Cells were harvested at 2 days post-calcium treatment for 3C assays and RNA isolation for gene expression analysis by real-time qPCR.

2.5.8 Chromatin Immunoprecipitation

Chromatin (approximately 5 x 10^6 cells) was sonicated (Bioruptor XL [Diagenode, Denville, NJ]) prior to immunoprecipitation with rabbit antibodies: c-Jun (AP-1) (Abcam, Cambridge, MA; Cat. # ab31419) and IgG (Millipore, Billerica, MA; Cat. # 12-370) antibodies bound to Dynabeads Protein A (Life Technologies). ChIP Primers are listed in Table S4.

2.6 Supplementary Material

Supplementary material may be viewed in Appendix A.

2.7 References

Alexander, RP, Fang, G, Rozowsky, J, *et al.* (2010). Annotating non-coding regions of the genome. *Nat Rev Genet* 11: 559–71.

Botchkarev, VA, Gdula, MR, Mardaryev, AN, *et al.* (2012). Epigenetic regulation of gene expression in keratinocytes. *J Invest Dermatol* 132: 2505–21.

Bryne, JC, Valen, E, Tang, M-HE, *et al.* (2008). JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res* 36: D102–6.

de Cid, R, Riveira-Munoz, E, Zeeuwen, PLJM, *et al.* (2009). Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis. *Nat Genet* 41: 211–5.

Deng, W, Lee, J, Wang, H, *et al.* (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* 149: 1233–44.

Dixon, JR, Selvaraj, S, Yue, F, *et al.* (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485: 376–80.

Dunham, I, Kundaje, A, Aldred, SF, *et al.* (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57–74.

Ernst, J, Kheradpour, P, Mikkelsen, TS, *et al.* (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473: 43–9.

Esparza-Gordillo, J, Weidinger, S, Fölster-Holst, R, *et al.* (2009). A common variant on chromosome 11q13 is associated with atopic dermatitis. *Nat Genet* 41: 596–601.

Ezhkova, E, Pasolli, HA, Parker, JS, *et al.* (2009). Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. *Cell* 136: 1122–35.

Fessing, MY, Mardaryev, AN, Gdula, MR, *et al.* (2011). p63 regulates Satb1 to control tissue-specific chromatin remodeling during development of the epidermis. *J Cell Biol* 194: 825–39.

Fuchs, E (2009). Finding One's Niche in the Skin. *Cell Stem Cell*.

Gdula, MR, Poterlowicz, K, Mardaryev, AN, *et al.* (2013). Remodeling of three-dimensional organization of the nucleus during terminal keratinocyte differentiation in the epidermis. *J Invest Dermatol* 133: 2191–201.

Giardina, E, Sinibaldi, C, Chini, L, *et al.* (2006). Co-localization of susceptibility loci for psoriasis (PSORS4) and atopic dermatitis (ATOD2) on human chromosome 1q21. *Hum Hered* 61: 229–36.

Guinea-Viniegra, J, Zenz, R, Scheuch, H, *et al.* (2009). TNF?? shedding and epidermal

inflammation are controlled by Jun proteins. *Genes Dev* 23: 2663–74.

de Guzman Strong, C, Conlan, S, Deming, CB, *et al.* (2010). A milieu of regulatory elements in the epidermal differentiation complex syntenic block: implications for atopic dermatitis and psoriasis. *Hum Mol Genet* 19: 1453–60.

Hagège, H, Klous, P, Braem, C, *et al.* (2007). Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* 2: 1722–33.

Hardman, M, Sisi, P, Banbury, D, *et al.* (1998). Patterned acquisition of skin barrier function during development. *Development* 125: 1541–52.

Hirota, T, Takahashi, A, Kubo, M, *et al.* (2012). Genome-wide association study identifies eight new susceptibility loci for atopic dermatitis in the Japanese population. *Nat Genet* 44: 1222–6.

Huebner, AJ, Dai, D, Morasso, M, *et al.* (2012). Amniotic Fluid Activates the Nrf2/Keap1 Pathway to Repair an Epidermal Barrier Defect In Utero. *Dev Cell* 23: 1238–46.

Irvine, AD, McLean, WHI, Leung, DYM (2011). Filaggrin Mutations Associated with Skin and Allergic Diseases. *N Engl J Med* 1315–27.

Jochum, W, Passegué, E, Wagner, EF (2001). AP-1 in mouse development and tumorigenesis. *Oncogene* 20: 2401–12.

Koch, PJ, de Viragh, PA, Scharer, E, *et al.* (2000). Lessons from Loricrin-Deficient Mice: Compensatory Mechanisms Maintaining Skin Barrier Function in the Absence of a Major Cornified Envelope Protein. *J Cell Biol* 151: 389–400.

Koster, MI, Roop, DR (2007). Mechanisms regulating epithelial stratification. *Annu Rev Cell Dev Biol* 23: 93–113.

Kubo, A, Nagao, K, Amagai, M (2012). Review series Epidermal barrier dysfunction and cutaneous sensitization in atopic diseases 122.

Lechler, T, Fuchs, E (2005). Asymmetric cell divisions promote stratification and differentiation of mammalian skin. *Nature* 437: 275–80.

Marshall, D, Hardman, MJ, Nield, KM, *et al.* (2001). Differentially expressed late constituents of the epidermal cornified envelope. *Proc Natl Acad Sci U S A* 98: 13031–6.

Mischke, D, Korge, BP, Marenholz, I, *et al.* (1996). Genes encoding structural proteins of epidermal cornification and S100 calcium-binding proteins form a gene complex ("epidermal differentiation complex") on human chromosome 1q21. *J Invest Dermatol* 106: 989–92.

Morar, N, Cookson, WOCM, Harper, JI, *et al.* (2007). Filaggrin mutations in children with severe atopic dermatitis. *J Invest Dermatol* 127: 1667–72.

Nora, EP, Lajoie, BR, Schulz, EG, *et al.* (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*.

Palmer, CNA, Irvine, AD, Terron-Kwiatkowski, A, *et al.* (2006). Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat Genet* 38: 441–6.

Paternoster, L, Standl, M, Chen, C-M, *et al.* (2012). Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. *Nat Genet* 44: 187–92.

Phillips-Cremins, JE, Sauria, MEG, Sanyal, A, *et al.* (2013). Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. *Cell* 153: 1281–95.

Sandilands, A, Terron-Kwiatkowski, A, Hull, PR, *et al.* (2007). Comprehensive analysis of the gene encoding filaggrin uncovers prevalent and rare mutations in ichthyosis vulgaris and atopic eczema. *Nat Genet* 39: 650–4.

Schonthaler, HB, Guinea-Viniegra, J, Wagner, EF (2011). Targeting inflammation by modulating the Jun/AP-1 pathway. *Ann Rheum Dis* 70 Suppl 1: i109–12.

Siepel, A, Bejerano, G, Pedersen, JS, *et al.* (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 15: 1034–50.

Smith, FJD, Irvine, AD, Terron-Kwiatkowski, A, *et al.* (2006). Loss-of-function mutations in the gene encoding filaggrin cause ichthyosis vulgaris. *Nat Genet* 38: 337–42.

Tolhuis, B, Palstra, RJ, Splinter, E, *et al.* (2002). Looping and interaction between hypersensitive sites in the active ??-globin locus. *Mol Cell* 10: 1453–65.

de Wit, E, de Laat, W (2012). A decade of 3C technologies: insights into nuclear organization. *Genes Dev* 26: 11–24.

Zenz, R, Scheuch, H, Martin, P, *et al.* (2003). c-Jun regulates eyelid closure and skin tumor development through EGFR signaling. *Dev Cell* 4: 879–89.

Zhao, XP, Elder, JT (1997). Positional cloning of novel skin-specific genes from the human epidermal differentiation complex. *Genomics* 45: 250–8.

Chapter 3

Proximal and distal regulation of EDC genes identified in mice with a deletion of an EDC enhancer by CRISPR/Cas9 genome editing

by

Inez Y. Oh, Ashley M. Quiggle,
and Cristina de Guzman Strong

Manuscript in preparation

Format adapted for dissertation

3.1 Abstract

Tissue-specific gene expression is driven largely by enhancers. Despite genome-wide discovery of enhancers in the skin, their function in skin biology is not fully known. Here we address the function of the epidermal-specific enhancer 923 in the Epidermal Differentiation Complex (EDC) locus via CRISPR/Cas9 genome editing in mice. Targeting of the 923 enhancer in mouse zygotes using a pair of flanking guide RNAs to direct Cas9 nuclease activity coupled with homologous recombination-mediated loxP insertions generated 1 floxed (923^{flox}), 2 independent deletions (923^{delA}, 923^{delB}), and 1 partial deletion (923^{pdel}) 923 enhancer alleles. Our results identify a significant dose-dependent requirement of the 923 enhancer for nearby *Ivl*, *Smcp*, and *Lce6a* gene expression, distal *Lce3* and *Crnn* gene expression, and a compensatory increase in members of the *Sprr* gene family (*Sprr4*, *Sprr2i*, *Sprr2g*, etc.) to reinforce the epidermal barrier. The results demonstrate a requirement of the 923 enhancer for directing proximal and distal gene expression.

3.2 Introduction

Enhancers are regulatory elements that drive tissue-specific gene expression and are often critical for cell fate decision and function. Despite genome-wide discovery of enhancers across a wide and diverse range of tissues and cell types using ChIP-seq studies, we know very little about how the enhancer regulates gene expression in vivo. The concomitant activation of the genes encoded by the Epidermal Differentiation Complex (EDC) specific to the stratified skin epidermis and the discovery of enhancers within the EDC provides a unique opportunity to investigate enhancer-driven events in EDC activation specific to the skin. We initially identified several candidate enhancers within the EDC locus based on sequence conservation across a wide

range of mammalian species (de Guzman Strong et al. 2010). Their enhancer activities were confirmed in cell-based reporter assays and were dynamic and physiologically dependent. The 923 enhancer in the human EDC, named after its kb distance away from the most 5'EDC gene, *S100A10*, demonstrated the highest activity in this assay as well as being DNaseI hypersensitive. Moreover, 923 exhibited epidermal-specific activity and was sensitive to the spatiotemporal cues during epidermal development (de Guzman Strong et al. 2010; Oh et al. 2014). Chromosome conformation capture studies further revealed remodeling of the mouse EDC surrounding the endogenous 923 enhancer upon epidermal differentiation and a requirement for the AP-1 transcription factor binding to 923 for inducing EDC expression. Given these results, we hypothesized that the 923 enhancer is required for EDC gene activation and the development of the epidermis. To test this hypothesis, we generated mice with a deletion of the 923 enhancer using the CRISPR/Cas9 genome editing tool. Loss of the 923 enhancer in two independent mouse models for the 923 deletion exhibited a dose-dependent decrease in proximal gene expression of *Ivl*, *Smcp*, and *Lce6a*, distal gene expression of *Lce3e*, *Lce3f*, and *Crnn,* and an observed compensatory increase in the *Sprr* family members to reinforce the epidermal barrier. Together, our data identifies a requirement for the 923 enhancer for proximal and distal gene expression in the EDC.

3.3 Results

3.3.1 Generation of an Allelic Series of the EDC 923 enhancer in mice using CRISPR/Cas9 Genome Editing

To generate a deletion for the 923 enhancer, we employed the CRISPR/Cas9 genome editing strategy in mice. Two short guide (sg)RNA were designed to target the flanking ends of

the orthologous mouse 923 sequence for cleavage by Cas9. Two single stranded oligodeoxynucleotides (ssODNs), each containing a *loxP* site and a restriction enzyme recognition site flanked by specific homology arms, were also introduced to generate a floxed allele. The strategy leverages endogenous DNA repair mechanisms in response to Cas9-induced double strand breaks, generating a floxed allele by homology-directed repair enabled by the ssODNs, as well as a deletion allele by non-homologous end joining (Figure 3.1a). Targeting specificity was verified using *in vitro* pilot studies prior to zygote injection. Of the 779 C57BL6/6XCBA hybrid zygotes injected, 80 F_0 newborns were recovered, of which 75 survived to weaning age. Our initial PCR screen identified 7 out of 80 mice whose 923 allele size deviated from the wild type allele (partial or full deletions) demonstrating 8.75% targeting efficiency. We also identified *loxP* insertion at either the 5' or 3' end at 13.75% efficiency collectively (5' only: 3/80; 3' only: 5/80; 5' and 3': 3/80). From these mice, we confirmed germline transmission for one floxed allele (923^{flox}) with both *loxP* inserting in cis, two deletion alleles (923^{delA}, 923^{delB}), and one partial deletion with 238 bp of 923 intact (923^{pdel}) (Figure 3.1b). Additionally, the 923^{delA} allele of 923 was flanked by both 5' and 3' *loxP* sites, while the 923^{delB} allele also contained a 3' *loxP* site.

3.3.2 923 deletion mice were viable and appeared normal

To directly identify the phenotype for the loss of the 923 enhancer, we first prioritized our studies on the homozygous 923^{delA}, 923^{delB}, and 923^{pdel} -/- mice. All 923^{delA}, 923^{delB}, and 923^{pdel} -/- mice appeared normal throughout gestation and into adulthood and were viable (Figure 3.2, 3.5). This was further supported by no deviations of expected versus observed genotypes for

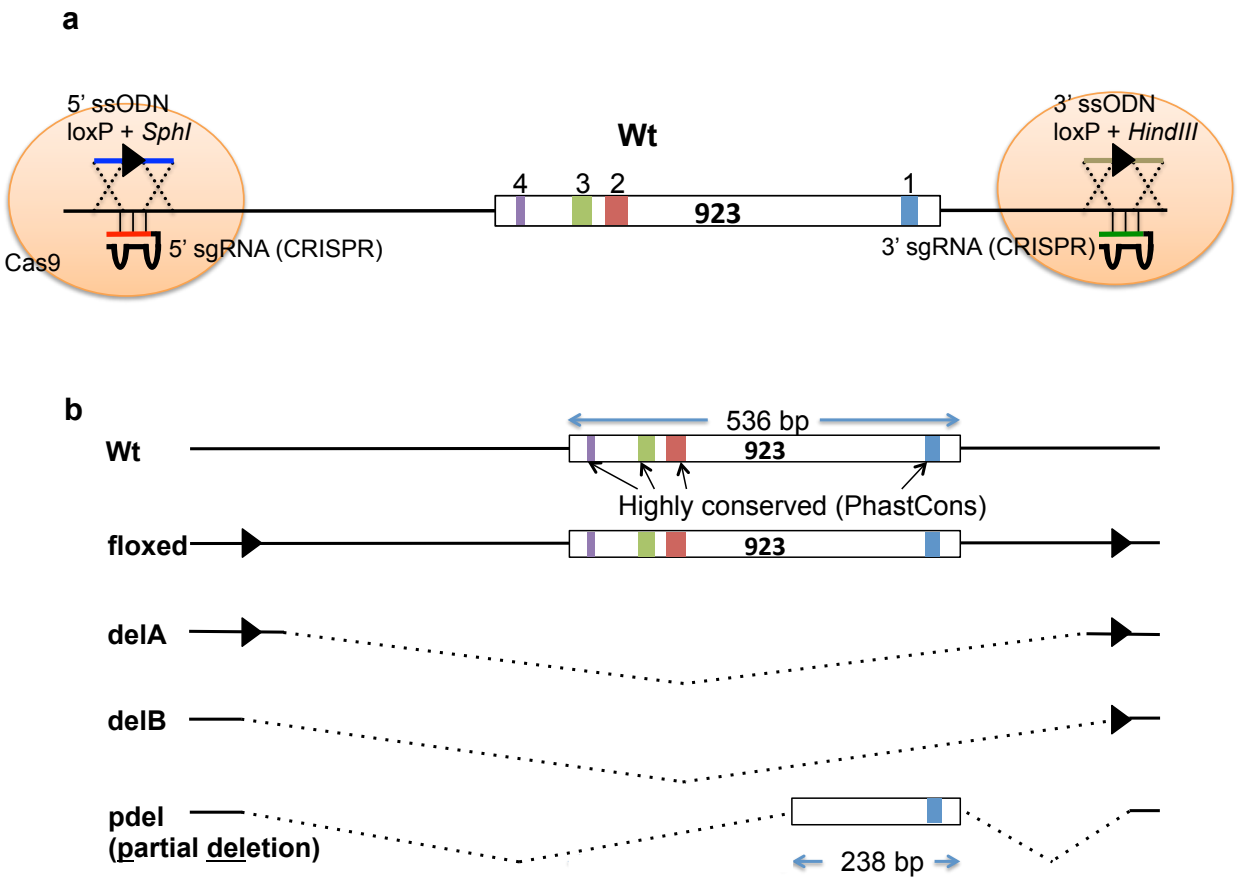


Figure 3.1. CRISPR/Cas9 targeting strategy. a) 2 sgRNAs were used to target Cas9 nuclease to specific sites flanking 923. 2 ssODNs were introduced to insert loxP sequences into the cleavage sites by homology-directed repair. b) Generation of multiple 923 enhancer alleles by CRISPR/Cas9 genome editing.

53-5 ($923^{fl/fl}$): Successfully introduced loxP sites flanking (positions) 923 in the mouse genome to create a floxed conditional deletion. Same strand 5' and 3' loxP sites confirmed by generation of F2s and cloning and sequencing.

923delA: Global deletion of whole 923 sequence. Insertion of flanking loxP sites.

923delB: Global deletion of whole 923 sequence. Larger region deleted in 59-10.

923pdel: Global deletion of partial 923 sequence. Loss of blocks 2, 3, and 4. Block 1 intact.

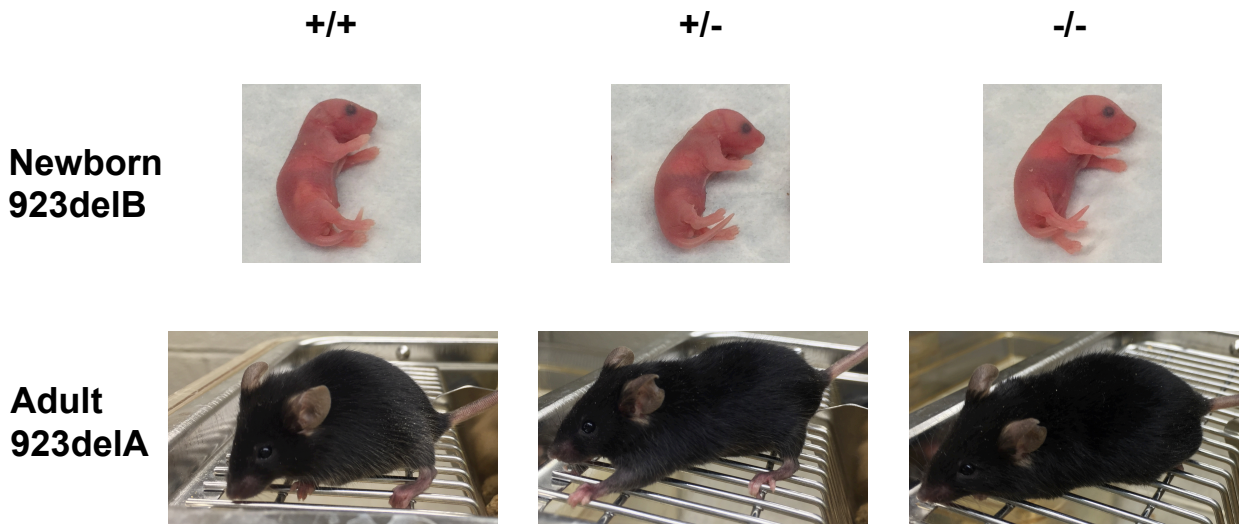


Figure 3.2. 923 deletion mice are viable and appear normal. Homozygous deletion mice do not appear different under normal housing conditions, when compared to heterozygous and homozygous wildtype littermates.

all 3 heterozygous parental crosses even after several generations of backcrossing to control for possible off-target effects (X^2 test, α=0.05) (Table S1). Normal fecundity was also observed for both 923^{delA} and 923^{delB} -/- mice (data not shown). Thus, the data demonstrate that the loss of the 923 enhancer does not affect overall viability under the conditions in which the mice were housed.

3.3.3 Skin barrier appears normal in 923 deletion mice

We next investigated the function of the skin barrier in the 923^{delA} and 923^{delB} -/- mice. No gross morphological differences in the skins of 923^{delA} and 923^{delB} +/- and -/- mice (both newborn and 8-week) compared to the respective wildtype littermates were observed (Figure 3.3).

We subsequently examined the morphology of the cornified envelopes from 923^{delA}, 923^{delB}, and 923^{pdel} -/- mice that contribute to the structural integrity of the skin barrier. Equal quantities of angular and balloon shaped cornified envelopes from newborn 923^{delA}, 923^{delB}, and 923^{pdel} -/- mice were observed compared to their wild type littermates (Figure 3.4). Similarly, the edges of the cornified envelopes appeared smooth for all 923 homozygous knockout mice, indicating no observable structural differences of the cornified envelopes as well.

As both 923^{delA} and 923^{delB} -/- mice appeared normal, we sought to determine if the development of skin barrier formation in the mouse embryo was delayed due to the loss of 923. Delayed skin barrier formation has been observed in several mouse models that revealed a compensatory mechanism to correct the skin barrier during development (Koch et al. 2000; Strong et al. 2006). No apparent differences in the patterning and temporal development of skin barrier formation were observed in the homozygous 923^{delA} and 923^{delB} -/- embryos compared to

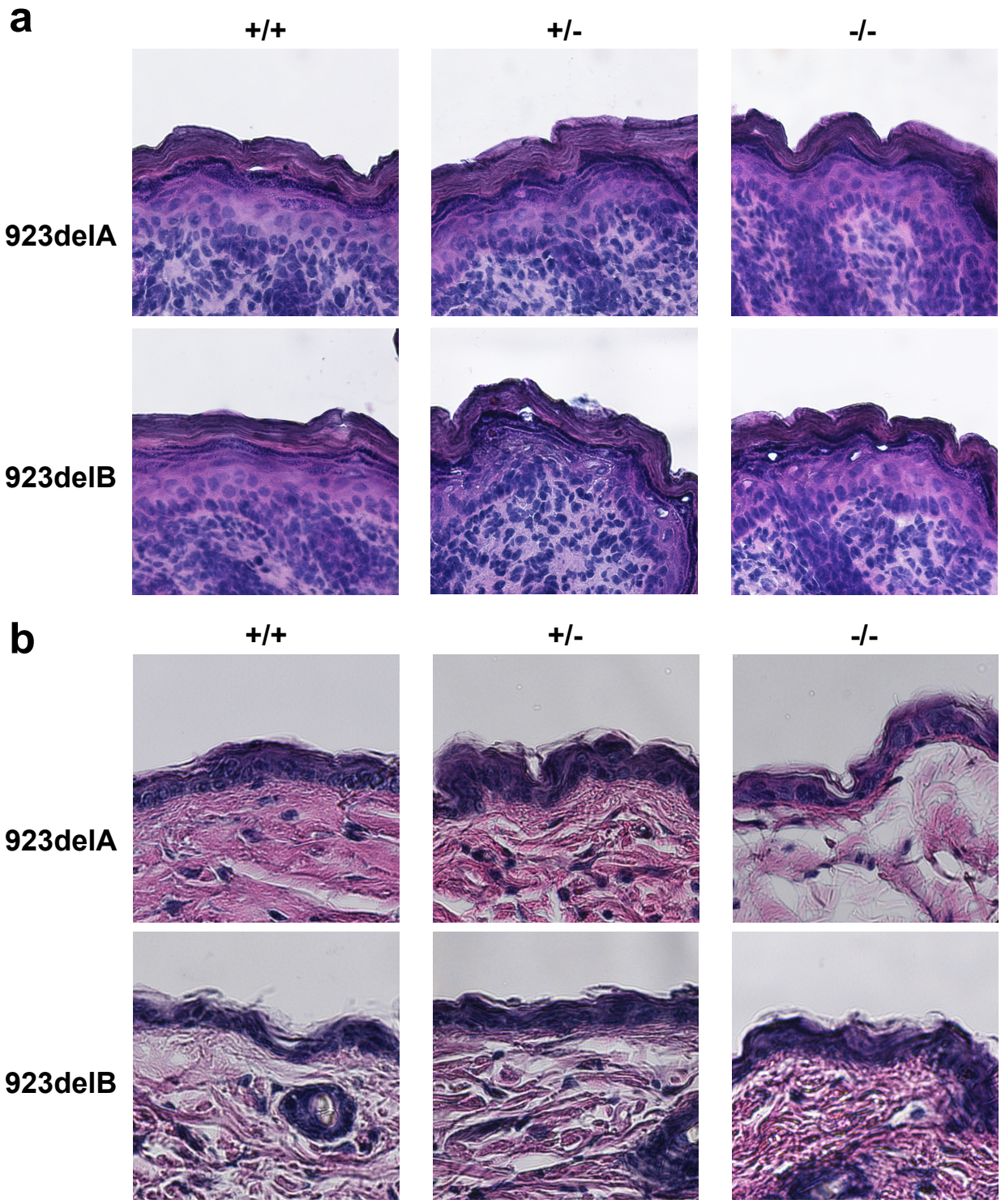


Figure 3.3. Normal histology of 923 deletion mouse epidermis. 923 deletion mice appear to have normal epidermal structure based on H&E staining of epidermal sections from a) newborn mice, and b) 8 week old mice.

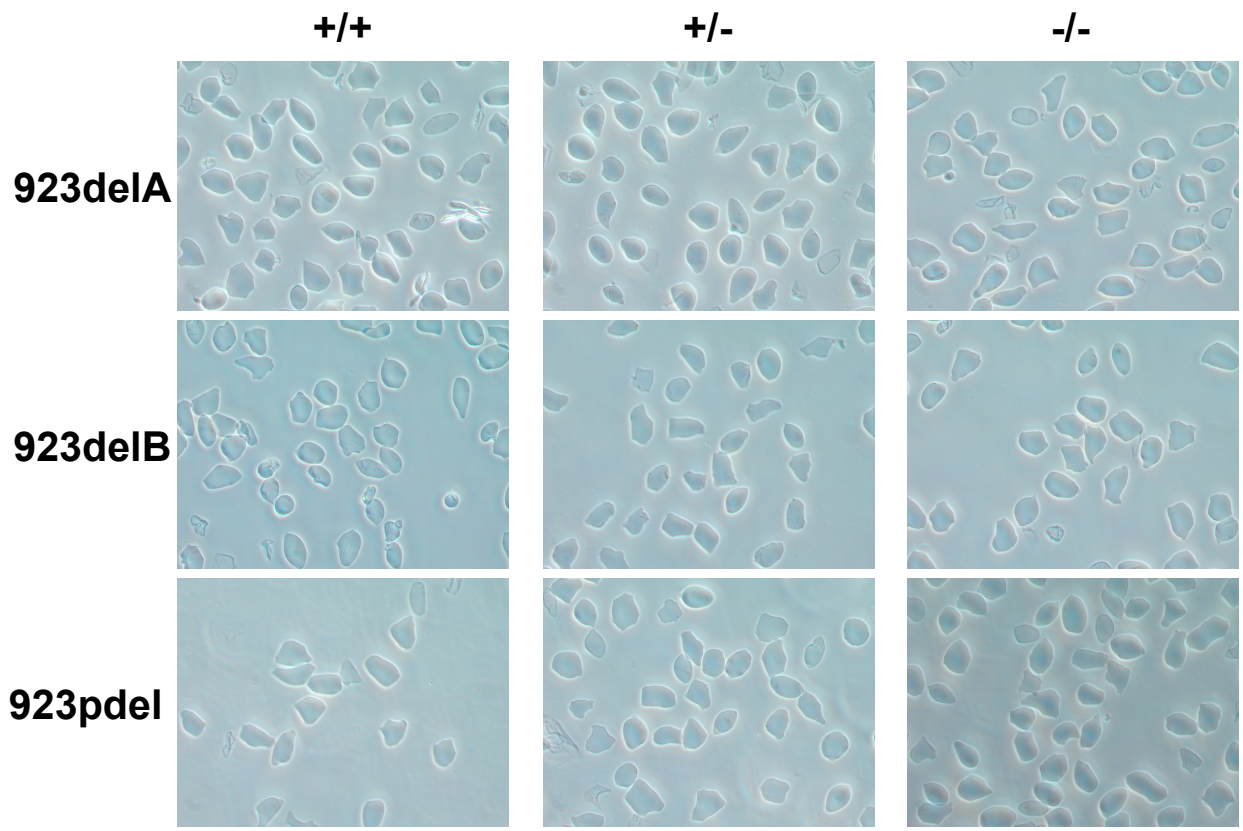


Figure 3.4. Normal cornified envelope morphology in 923 deletion and partial deletion mice. Similar quantities of angular and balloon shaped cornified envelopes with smooth edges were isolated from newborn skin of homozygous deletion, heterozygous and wildtype littermates of 923delA, 923delB, and 923pdel lines.

wild type and heterozygous littermates assessed by blue X-gal reactivity as a proxy for outside-in barrier formation (Fig. 3.5). Together, our results indicate normal skin barrier function with respect to skin morphology, cornified envelope formation, and skin barrier development with the loss of the 923 enhancer in mice.

3.3.4 The 923 enhancer is required for dose-dependent expression of proximal gene *Ivl*, *Smcp*, *Lce6a*, and distal *Crnn* and *Lce3* genes

To determine the effect of the loss of the 923 enhancer on the skin transcriptome, we performed RNA-seq on newborn skin. We treated the independent deletion lines, 923^{delA} and 923^{delB}, as biological replicates to account for stochastic changes in gene expression not due to the loss of 923. Our results identified a significant decrease (adjusted p-value <0.05) in the expression of the proximal gene *Ivl,* and the distal gene *Crnn* in 923^{del} -/- mice in comparison to heterozygous and wildtype newborn littermates (Figure 3.6a). Additionally, while statistical significance was not achieved, a decrease was noted in the expression of proximal genes *Smcp* and *Lce6a*, and distal genes *Lce3a*, *Lce3b*, *Lce3e* and *Lce3f*. Moreover, we determined that the decrease in *Ivl*, *Smcp* and *Crnn* expression was dose-dependent with 923^{del} +/- mice exhibiting approximately 50% less *Ivl* and *Crnn* expression compared to wild-type littermates. The magnitude of the difference in expression was greatest for *Ivl,* which displayed a 30-fold decrease in 923^{del} -/- relative to wild-type littermates. The dose-dependent decrease in *Ivl* was also observed by qPCR in cultured primary keratinocytes isolated from 923^{delB} -/- newborn mice in comparison to wild-type and heterozygous littermates, confirming a cell-intrinsic effect (Figure 3.6d). The observed decrease in *Ivl* gene expression appeared to thus lead to reduced involucrin protein expression in the epidermis (Figure 3.6a, c). By contrast, 923^{pdel}-/- newborn

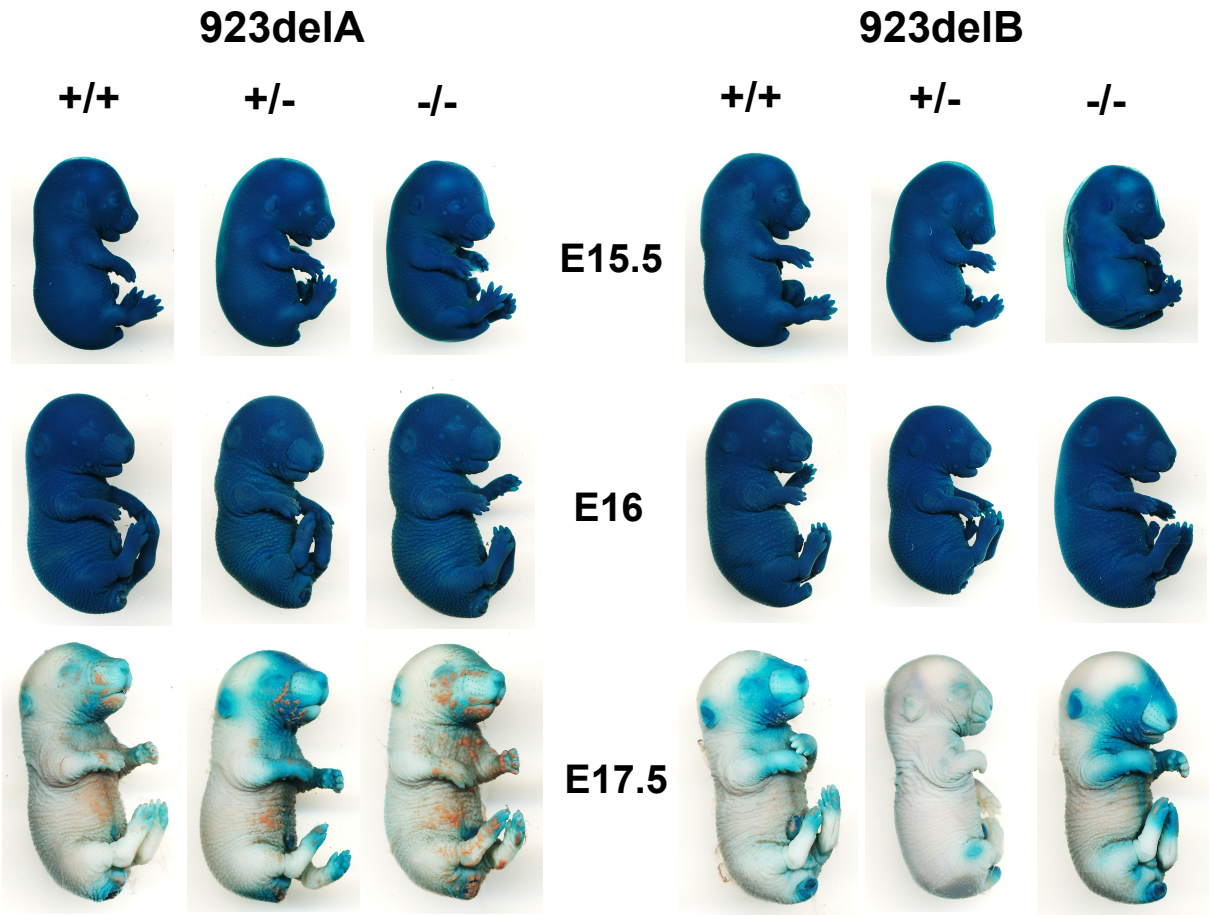


Figure 3.5. Normal patterning of skin barrier development in 923 deletion mice. The extent of skin barrier formation was assessed by an outside-in X-gal dye penetration assay. Blue stain indicates X-gal reactivity with endogenous β-galactosidase where the X-gal solution has penetrated the epidermis where the skin barrier has not formed.

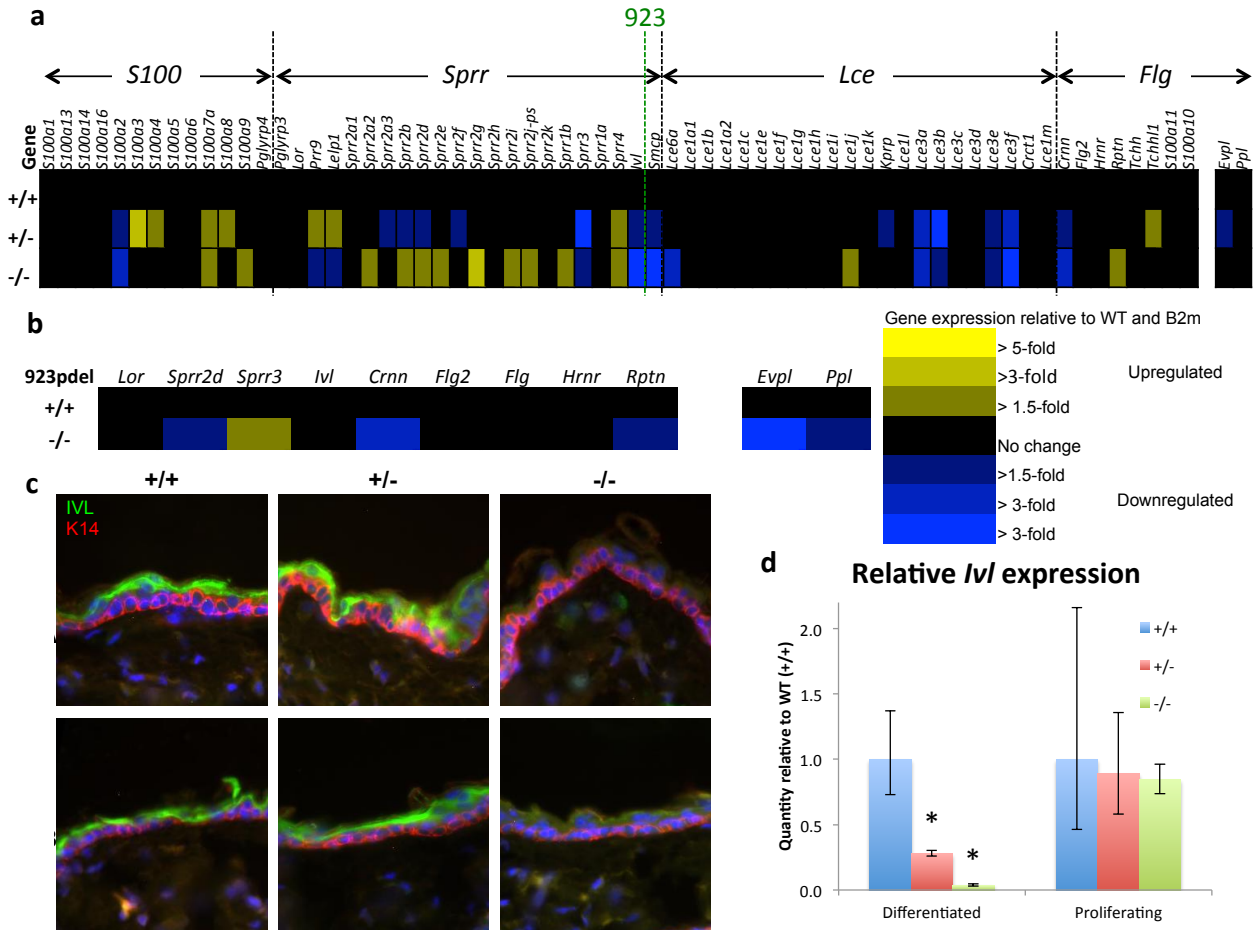


Figure 3.6. Deletion of 923 results in loss of proximal and distal gene expression. a) RNA-seq to assess gene expression in the skin of newborn mice revealed a dose-dependent decrease in neighbouring genes, *Ivl,* and *Smcp,* and *Lce6a,* and a compensatory increase in the expression of the *Sprr* gene family in both independent full deletion lines, 923^{delA} and 923^{delB}. b) In 923^{pdel} mice, no significant changes in gene expression were observed as assessed by qPCR, indicating sufficiency of the remaining 238bp of the 3' end of 923 for regulation of proximal gene expression. c) Immunofluorescent staining of 923 KO epidermal sections confirms decreased protein expression of IVL assessed by RNA-seq. d) Allele dose-dependent decrease in *Ivl* expression was observed in cultured primary keratinocytes isolated from 923delB wildtype, heterozygous, and homozygous deletion littermates, upon differentiation. Asterisk indicates statistical significance ($p < 0.05$).

mice displayed no significant changes in EDC gene expression, thus demonstrating that 238 bp of the 923 locus is sufficient for cis-regulatory gene expression (Figure 3.6b, 3.1b). In our previous study, we demonstrated a requirement for a PhastCons block (highly conserved sequence [mouse: 20 bp, human: 23 bp]) for enhancer activity and AP-1 transcription factor binding (Figure 3.1b) (Oh et al. 2014). As block 1 is intact in the 923pdel-/- mice, the lack of gene expression differences in these mice further validates our previous findings, indicating that 238 bp of the 923 locus that includes block 1 is sufficient for 923's enhancer activity both *in vitro* and *in vivo*.

3.3.5 Loss of 923-driven proximal gene expression results in a compensatory increase in the expression of *Sprr* gene family members

Gene expression analysis also revealed a compensatory upregulation of *Sprr* gene family members (*Sprr2a2, Sprr2e, Sprr2g, Sprr2i, Sprr2j, Sprr1b*) in 923^{delA} and 923^{delB} -/- mice in comparison to wild-type littermates (Figure 3.6a). The *Sprr* gene family which also contains *Ivl* (de Guzman Strong et al. 2010), encodes proteins that are precursors for cornified envelope formation (Hohl et al. 1995). Together, our data identifies a requirement for the 923 enhancer to regulate proximal and distal gene expression in the EDC, the loss of which induces coordinated upregulation of *Sprr* family members to reinforce the skin barrier.

3.4 Discussion

Despite large, publically available datasets by the ENCODE Consortium and others that have identified many putative enhancers on a genome-wide scale (Kellis et al. 2014), the functions of these enhancers in the context of relevant tissue types remain poorly understood. However, while techniques exist for high-throughput screens of enhancer activity and for the

validation of enhancer function, these methods have relied on exogenous introduction of the putative enhancer to *in vitro* or transgenic models, and might not truly reflect the function of the endogenous enhancer. The advent of CRISPR-Cas9 genome editing technology has greatly facilitated the functional analyses of enhancers *in vivo* and in a more efficient and cost-effective manner (Guo et al. 2015; Han et al. 2015; Lupiáñez et al. 2015; Yang et al. 2016) by way of the relative ease of designing gRNAs to target loci. Recent studies have successfully used CRISPR-Cas9 to introduce mutations of enhancers ranging from specific nucleotide substitutions to deletions over 10 Kb in a variety of human and mouse cell lines, as well as mouse models in order to study the roles of these enhancers (Li et al. 2014; Han et al. 2015; Zhang et al. 2016). Furthermore, a study employing CRISPR-Cas9 to invert CTCF binding sites within distal enhancers resulted in aberrant chromatin topology and gene expression (Guo et al. 2015). From investigating the effects of enhancers on nearby gene expression, to understanding their roles in maintaining the broader genomic architecture, these studies highlight the ability of the CRISPR-Cas9 system to enable investigators to dissect the functions of non-coding elements within their native loci.

Here, we apply the CRISPR/Cas9 strategy that enabled us to generate multiple alleles of the 923 enhancer. Mice harboring either of the two independent deletions of 923 (923^{delA}, 923^{delB}) demonstrated in nearby genes, *Smcp, Lce6a,* distally located *Lce3* family genes (*Lce3a*, *Lce3b*, *Lce3e* and *Lce3f*) and *Crnn*, and to the greatest extent, an allele dose-dependent decrease in *Ivl*, the gene most proximal to 923. Coincident with this, we observed an upregulation of *Sprr* gene family members, *Sprr2a2, Sprr2e, Sprr2g, Sprr2i, Sprr2j,* and *Sprr1b*. However, mice with only a partial deletion of 923 did not display any significant alterations in gene expression, indicating that the remaining 923 sequence is sufficient for 923 enhancer activity.

The deletion of 923 resulting in the loss of proximal (*Ivl, Smcp*, and *Lce6a*) and distal (*Lce3*-family and *Crnn*) gene expression, while intervening *Lce* genes remain unaffected, is supported by previous studies demonstrating that subgroups of the *Lce* genes have distinct expression patterns, indicating independent modes of regulation (Marshall et al. 2001; Jackson et al. 2005, 2011; Niehues et al. 2016). The positions of the affected genes relative to 923, also suggests that 923 mediates a non-linear mechanism such as chromatin looping, to bring the affected genes into spatial proximity in order to regulate their expression. Our previous study which identified 923, also identified additional conserved non-coding elements (CNEs) with enhancer activity located in the vicinity of *Crnn* and the distal *Lce3* genes (de Guzman Strong et al. 2010). CNE 531 located within the *Lce3* gene family most distal from 923, as well as CNEs 461 and 409 which cluster around *Crnn* in the *Flg*-like gene family, displayed significant enhancer activity in differentiating keratinocytes. CNEs 531, 461, and 409 all flank the boundary between the *Lce* and *Flg*-like gene families. A previous study observed an enrichment of chromatin interactions between conserved non-coding sequences, thought to be a result of their occupation of the same transcription factories to drive the expression of co-regulated genes (Robyr et al. 2011). The regulation of many EDC genes by a common core of transcription factors (p63, AP1, Sp1, CREB, Ezh2, C/EBP, Klf) suggest similar behavior within the EDC, where co-regulated genes are organized into transcription factories by regulatory elements for efficient gene activation. Another study observed spatial compression of the EDC locus during terminal differentiation of keratinocytes, proposed to represent the looping of chromatin into an active chromatin structure (Fessing et al. 2011). The loss of 923 resulting in loss of expression from proximal genes and distal genes flanking the *Lce*-*Flg* boundary could therefore indicate the

loss of a chromatin interaction between 923 and enhancers in the *Lce-Flg* boundary region that normally co-regulates the genes near the two regions.

The upregulation of the *Sprr* gene family in response to loss of 923/*Ivl* indicates a compensatory response to maintain the skin barrier. *Ivl* and the other *Sprr* genes form the initial scaffold of the cornified envelope (CE), which is the basic structural unit of the epidermal barrier (Kalinin et al. 2002). Previous studies have demonstrated that the *Sprr* genes are coordinately upregulated during stress, including a model of epidermal barrier deficiency resulting from the ablation of the Klf4 transcription factor (Patel et al. 2003). The observation that *Sprr* genes are upregulated upon the loss of 923/*Ivl* suggests that the normal appearance and function of the epidermal barrier in the 923 knockout mice is a result of a deficiency that has been compensated by increased quantities of other Sprr proteins. The redundancy of individual components of the cornified envelopes has been well demonstrated. For example, loricrin deficient mice display mild erythema at birth that resolves within 5 days, and fragile CEs that contain increased amounts of other CE proteins such as *Sprr2* and repetin that compensate for the loss of loricrin (Koch et al. 2000). Also, a previous study demonstrated a delayed barrier formation and abnormal CE ultrastructure in *Ivl-Evpl-Ppl* triple KO mice (Sevilla et al. 2007) compared to single *Ivl*, *Evpl*, and *Ppl* knockouts in mice that exhibited normal skin barrier formation (Djian et al. 2000; Määttä et al. 2001; Aho et al. 2004), suggesting that these three genes are able to compensate for each other's loss. Although the mechanism underlying the compensatory response is unclear, activation of *Sprr* genes in barrier-deficient epidermis was shown to correlate with an increased prevalence of DNaseI-hypersensitive sites within the locus, suggesting the activity of enhancers to coordinate this response (Martin et al. 2004). The upregulation of *Sprr* genes in 923 knockout mice could represent a shift from the use of a

preferred enhancer (923) to other nearby enhancer(s) that drive the expression of genes with related functions.

Together, these studies suggest a mechanism where resident enhancers within each gene family mediate looping of the EDC chromatin into an active chromatin hub for easy access to a common pool of transcription factors. Interactions between multiple enhancers across the EDC would exist as a "backup system" to allow rapid activation of genes of related function to compensate for each other's loss, in an organ such as the skin which is essential to survival. The existence of CTCF sites at the boundaries between gene families further supports the plausibility of this model, as CTCF has been observed to mediate the formation of chromatin loops in relation to gene expression (Tang et al. 2015).

The availability of CRISPR-Cas9 genome editing has never been timelier, as it will enable future functional analyses of additional enhancers that may be driving gene expression in the skin, and allow us to elucidate the mechanisms underlying the proposed network of enhancers that cooperate to maintain a functional skin barrier.

3.5 Materials and Methods

3.5.1 Mice

All mice were housed in pathogen-free, barrier facilities at NIH (Bethesda, MD) and Washington University School of Medicine (St Louis, MO). All animal procedures were approved by the Division of Comparative Medicine Animal Studies Committee at Washington University. All animal work was conducted in accordance with the Guide for the Care and Use of Laboratory

Animals of the National Institutes of Health. Morning observation of a vaginal plug was designated as embryonic day (E) 0.5.

3.5.2 Generation of 923 alleles in mice by CRISPR/Cas9 genome editing

To target the orthologous mouse 923 enhancer, two small guide RNAs and two ssODNs containing *loxP* sites with 80bp homology arms flanking *loxP* and restriction enzyme sites were designed to target insertion of *loxP* to Cas9 cut sites by homologous recombination (Integrated DNA Technologies, Coralville, IA). Three rounds of injection in 779 zygotes were performed. Founders were initially screened for large deletions via PCR using flanking X primers (m923Cas9delF and m923Cas9delR) and 5' and 3' *loxP*-specific primers that were resolved on 1% agarose gel electrophoresis. Positive candidates for *loxP* insertion and <1407 bp were subsequently cloned and Sanger sequenced. 923^{delA}, 923^{delB}, 923^{pdel} and 923^{flox} were backcrossed on to C57B6 mice at least 4 times and selected for the black coat phenotype to exclude potential off-target effects. Sequences of sgRNAs, ssODNs, and genotyping primers are listed in Supplementary Material.

$923^{delA}$and 923^{delB}: Primers flank entire region. Larger region deleted in 923^{delB}. Same primers identified 923^{pdel} due to partial deletion.

5' loxP: Left primer on upstream sequence, right primer on junction of 5' *loxP* and 923.

3' loxP; Left primer on junction of 923 and 3' loxP, right primer on downstream sequence.

3.5.3 Histology and Immunohistochemistry

Dorsal skin was excised from newborn mice and frozen in OCT compound prior to cryosectioning. Dorsal skin was excised from 8 week old mice and preserved in 4%

paraformaldehyde (Electron Microscopy Sciences, Hatfield, PA) prior to paraffin sectioning. Sections were stained with hematoxylin and eosin (H&E) by the Washington University Developmental Biology Histology Core. H&E stained sections were imaged on a Nikon Eclipse 80i brightfield microscope (Nikon, Tokyo, Japan).

Primary antibodies used for immunofluorescence are rabbit IVL (4b-KSCN, 1:200) and chicken K14 (5560, 1:500) (courtesy of J. Segre). Secondary antibodies used were goat anti-rabbit (Alexa Fluor 488, 1:500) and goat anti-chicken (Alexa Fluor 594, 1:500) IgG antibodies (Life Technologies, Frederick, MD). Sections were fixed in 4% paraformaldehyde (Electron Microscopy Sciences, Hatfield, PA) prior to permeabilization with 0.1% Triton X-100 and subsequent antibody incubation. Sections were counterstained with SlowFade Gold antifade reagent with DAPI (Life Technologies, Frederick, MD) prior to fluorescent imaging.

3.5.4 RNA analyses

Total RNA was isolated by TriZol extraction (Life Technologies, Frederick, MD) from epidermal isolates preserved in RNALater (Thermo Fisher Scientific, Waltham, MA).

Real-time qPCR on cDNA (generated using SuperScript II reverse transcriptase (Thermo Fisher Scientific)) using SYBR Green was performed in triplicate (Fast SYBR Green Master Mix ABI ViiA7, Thermo Fisher Scientific) and normalized to β2-microglobulin. Only CT values with single peaks on melt-curve analyses were included. Primers are listed in the supplementary material. RT-qPCR was performed on at least 2 individuals per genotype.

RNA-seq was performed on ribosome-depleted RNA libraries that were single-read sequenced on Illumina HiSeq3000 and analyzed by the Washington University Genome Technology Access Center (GTAC). Differential expression was determined based on normalization to the E14.5

developmental time point, with one individual from each independent deletion line (923^{delA}, 923^{delB}) representing biological replicates. Differential expression values corresponding to Figure 3.6a are listed in the Supplementary Material.

3.5.5 Dye penetration barrier assays

Mouse embryos were incubated in an X-gal solution for at least 4 hours at 37°C. The embryos were imaged on a CanoScan 5600F scanner (Canon, Melville, NY) to determine the extent and location of skin barrier formation.

3.5.6 Cornified envelope preparations

Skin was taken from newborn and postnatal day 1 mice and floated on 0.25% Trypsin to allow dissociation of the epidermis from the dermis. An approximately $1cm^2$ piece of epidermis was incubated at 95°C in a solution of 2% SDS to obtain cornified envelopes in a single-cell suspension. The suspension was placed on a slide and examined using phase contrast light microscopy on a Zeiss Axiovert 40 CFL (Carl Zeiss, Stockholm, Sweden). Images were captured with Zeiss AxioCam ERc 5s using Zeiss ZEN software (Carl Zeiss, Stockholm, Sweden).

3.6 Supplementary Material

Supplementary material can be viewed in Appendix B.

3.7 References

Aho, S, Li, K, Ryoo, Y, *et al.* (2004). Periplakin gene targeting reveals a constituent of the cornified cell envelope dispensable for normal mouse development. *Mol Cell Biol* 24: 6410–8.

D Hohl, P A de Viragh, F Amiguet-Barras, S Gibbs, C Backendorf, MH, Hohl, D, de Viragh, PA, *et al.* (1995). The small proline-rich proteins constitute a multigene family of differentially regulated cornified cell envelope precursor proteins. *J Invest Dermatol* 104: 902–9.

Djian, P, Easley, K, Green, H (2000). Targeted ablation of the murine involucrin gene. *J Cell Biol* 151: 381–8.

Fessing, MY, Mardaryev, AN, Gdula, MR, *et al.* (2011). p63 regulates Satb1 to control tissue-specific chromatin remodeling during development of the epidermis. *J Cell Biol* 194: 825–39.

Guo, Y, Xu, Q, Canzio, D, *et al.* (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* 162: 900–10.

de Guzman Strong, C, Conlan, S, Deming, CB, *et al.* (2010). A milieu of regulatory elements in the epidermal differentiation complex syntenic block: implications for atopic dermatitis and psoriasis. *Hum Mol Genet* 19: 1453–60.

Han, Y, Slivano, OJ, Christie, CK, *et al.* (2015). CRISPR-Cas9 genome editing of a single regulatory element nearly abolishes target gene expression in mice--brief report. *Arterioscler Thromb Vasc Biol* 35: 312–5.

Jackson, B, Brown, SJ, Avilion, AA, *et al.* (2011). TALE homeodomain proteins regulate site-specific terminal differentiation, LCE genes and epidermal barrier. *J Cell Sci* 124: 1681–90.

Jackson, B, Tilli, CMLJ, Hardman, MJ, *et al.* (2005). Late cornified envelope family in differentiating epithelia - Response to calcium and ultraviolet irradiation. *J Invest Dermatol* 124: 1062–70.

Kalinin, AE, Kajava, A V., Steinert, PM (2002). Epithelial barrier function: assembly and structural features of the cornified cell envelope. *Bioessays* 24: 789–800.

Kellis, M, Wold, B, Snyder, MP, *et al.* (2014). Defining functional DNA elements in the human genome. *Proc Natl Acad Sci U S A* 111: 6131–8.

Koch, PJ, de Viragh, PA, Scharer, E, *et al.* (2000). Lessons from Loricrin-Deficient Mice: Compensatory Mechanisms Maintaining Skin Barrier Function in the Absence of a Major Cornified Envelope Protein. *J Cell Biol* 151: 389–400.

Li, Y, Rivera, CM, Ishii, H, *et al.* (2014). CRISPR reveals a distal super-enhancer required for Sox2 expression in mouse embryonic stem cells. *PLoS One* 9: e114485.

Lupiáñez, DG, Kraft, K, Heinrich, V, *et al.* (2015). Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* 161: 1012–25.

Määttä, a, DiColandrea, T, Groot, K, *et al.* (2001). Gene targeting of envoplakin, a cytoskeletal linker protein and precursor of the epidermal cornified envelope. *Mol Cell Biol* 21: 7047–53.

Marshall, D, Hardman, MJ, Nield, KM, *et al.* (2001). Differentially expressed late constituents of the epidermal cornified envelope. *Proc Natl Acad Sci U S A* 98: 13031–6.

Martin, N, Patel, S, Segre, JA (2004). Long-range comparison of human and mouse Sprr loci to identify conserved noncoding sequences involved in coordinate regulation. *Genome Res* 14: 2430–8.

Niehues, H, van Vlijmen-Willems, IMJJ, Bergboer, JGM, *et al.* (2016). Late cornified envelope (LCE) proteins: distinct expression patterns of LCE2 and LCE3 members suggest nonredundant roles in human epidermis and other epithelia. *Br J Dermatol* 174: 795–802.

Oh, IY, Albea, DM, Goodwin, ZA, *et al.* (2014). Regulation of the Dynamic Chromatin Architecture of the Epidermal Differentiation Complex Is Mediated by a c-Jun/AP-1-Modulated Enhancer. *J Invest Dermatol* 134: 2371–80.

Patel, S, Kartasova, T, Segre, JA (2003). Mouse Sprr locus: A tandem array of coordinately regulated genes. *Mamm Genome* 14: 140–8.

Robyr, D, Friedli, M, Gehrig, C, *et al.* (2011). Chromosome conformation capture uncovers potential genome-wide interactions between human conserved non-coding sequences. *PLoS One* 6: e17634.

Sevilla, LM, Nachat, R, Groot, KR, *et al.* (2007). Mice deficient in involucrin, envoplakin, and periplakin have a defective epidermal barrier. *J Cell Biol* 179: 1599–612.

Strong, CDG, Wertz, PW, Wang, C, *et al.* (2006). Lipid defect underlies selective skin barrier impairment of an epidermal-specific deletion of Gata-3. *J Cell Biol* 175: 661–70.

Tang, Z, Luo, OJ, Li, X, *et al.* (2015). CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* 163: 1611–27.

Yang, R, Kerschner, JL, Gosalia, N, *et al.* (2016). Differential contribution of cis-regulatory elements to higher order chromatin structure and expression of the CFTR locus. *Nucleic Acids Res* 44: 3082–94.

Zhang, X, Choi, PS, Francis, JM, *et al.* (2016). Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. *Nat Genet* 48: 176–82.

Chapter 4

Genome-wide Chromatin Architecture of the EDC

4.1 Abstract

Enhancers are known to drive tissue-specific gene expression by forming physical interactions by chromatin looping with the gene promoters. Epidermal enhancer 923 was previously shown to interact with several EDC gene promoters selectively assayed in the context of epidermal development. However, we do not fully understand the extent to which 923 is involved in the chromatin architecture underlying expression of EDC and other genes essential for epidermal development. Here, I sought to determine the chromatin architecture within and surrounding the EDC, pertaining to keratinocyte differentiation. I approached this by using large-scale high-throughput chromosome conformation capture assays (4C-seq) to identify the keratinocyte-specific chromatin interactions with respect to 2 viewpoints in the EDC, the 923 enhancer and *Flg* promoter. A marked decrease in interaction frequencies beyond the EDC in all EDC-viewpoint libraries supports the notion of a topologically associated domain encompassing the EDC. Within the EDC, I identified an enrichment of interactions for the 923 enhancer with the *Sprr* and *Lce* gene families in the keratinocytes compared to the P5424 T-cell line that does not express EDC genes. Keratinocyte-specific enrichment of interactions was observed between each of the 923 enhancer and *Flg* promoter with the gene desert located between the *S100* and *Sprr*

families. Of note was a proliferating keratinocyte-specific interaction between 923 and a previously identified enhancer (CNE 531) near the *Crct1* gene. In addition, the *Flg* viewpoint displayed a differentiated keratinocyte-specific enrichment of interactions with the region between *Hrnr* and *Rptn* containing previously identified enhancer CNE 184. Trans-interactions enriched in keratinocytes were identified with non-EDC genes that are important for epidermal function including *Trp63,* an important regulator of keratinocyte differentiation. Together, my results identify a network of biologically relevant chromatin loops to the EDC to include enhancer-gene interactions in cis and in trans as well as enhancer-enhancer interactions. The evidence for chromatin architecture involving the EDC and other loci important for epidermal function further supports the notion of a "form follows function" principle for keratinocyte-specific gene expression.

4.2 Introduction

As described in Chapter 1, epidermal development involves the coordinated expression of many genes including the EDC genes, the keratins, desmogleins, and plakins. The regulation of these genes requires the binding of many of the same transcription factors (TFs) such as AP-1, KLF, and CEBP, to proximal and distal enhancers to drive their tissue-specific gene expression. While the importance of these individual components is known, the precise mechanisms by which they cooperate to drive such coordinate expression are not well understood. Investigations into the chromatin architecture surrounding an active gene or enhancer will provide insights about the molecular mechanisms that drive gene expression in a cell or tissue-specific manner. A recent analysis leveraged a focused high-resolution 3C variant known as chromosome conformation capture carbon-copy (5C) to investigate the relationship between the chromatin architecture and expression of the *CFTR* locus in a variety of cell types (Smith et al. 2016).

Based on the 5C interactions, they identified a region spanning the locus known as a topologically associated domain (TAD), where there was a high frequency of chromatin interactions within the TAD, and lower interaction frequencies with regions outside the TAD. The TADs and their boundaries were consistent across different cell-types. However, cell-specific gene expression coincided with cell-type-specific enhancer-promoter interactions within TADs, as well as cell-type-specific interactions between TADs that tended to involve loci clustered around TAD boundaries. This suggested that the key to elucidating the molecular mechanisms driving cell-specific expression lies in the analysis of cell-specific chromatin interactions. Additional studies have also definitively established a role for CTCF to maintain the boundaries of these TADs, thereby allowing genes that are close on the linear genome, but residents of neighboring TADs, to have different gene expression patterns (Guo et al. 2015; Lupiáñez et al. 2015; Yang et al. 2016).

Within the β-globin locus, temporal control of gene expression is a result of the organization of active chromatin hubs (ACH) by CTCF-mediated loops (TADs) and enhancer-gene interactions (Palstra et al. 2003). These studies in conjunction with ENCODE-annotated CTCF binding sites led me to hypothesize the existence of one or more TADs within the EDC. These TADs and the enhancer-promoter interactions within would thus form an ACH to mediate the concomitant activation of EDC genes brought into proximity with the enhancers during keratinocyte differentiation. Furthermore, intra-chromosomal (cis) and inter-chromosomal (trans) interactions revealed spatial association of the active globin genes with other co-regulated genes at “transcription factories”, and found that the transcription factor KLF1 was necessary for mediating the associations of KLF1 co-regulated genes (Schoenfelder et al. 2010). Given that the

same transcription factors regulate many of the co-expressed genes essential for epidermal differentiation, it seems likely that the EDC participates in such a transcription factory.

In Chapter 2, the characterization of 923's enhancer function revealed physical chromatin interactions between 923 and several of the EDC gene promoters. This was determined using low-throughput chromosome conformation capture assays (3C) that depend on PCR-based identification of individual ligation junctions. 3C is a very targeted approach that provides a limited view of all possible interactions in which a single genomic region might be a participant. It requires prior knowledge, or a strong suspicion, of the existence of an interaction in order for primers to be designed for the amplification of specific targets. The evidence of 923's enhancer activity described in Chapters 2 and 3 suggests a pivotal role for 923 in mediating the chromatin architecture of the EDC. Further interrogation of 923 will provide a more detailed view of its interactions within the EDC as well as a unique viewpoint for enhancer-centric EDC contacts with other loci. Given my previous work and the prior studies, I hypothesized that 923 formed both cis-interactions representing enhancer-promoter contacts within an EDC TAD or ACH and trans-interactions representing an epidermal differentiation transcription factory, in a keratinocyte-specific manner.

To capture genome-wide chromatin interactions with the EDC, I performed the circular chromosomal conformation capture assay coupled with next-generation sequencing (4C-seq) (Splinter et al. 2012), with respect to the 923 enhancer and the *Flg* promoter in proliferating keratinocytes, differentiating keratinocytes, and P5424 T-cells. The T-cells do not express EDC genes and thus provide background interaction frequencies for an inactive EDC locus. I aimed to discover chromatin interactions with the *Flg* promoter as coding mutations have been identified in ichthyosis vulgaris and atopic dermatitis, diseases with a component of skin barrier deficiency

component (Irvine and McLean 2006; Palmer et al. 2006; Smith et al. 2006; Nomura et al. 2009), but little is known about the enhancers that regulate *FLG*. 4C-seq is a variation of 3C that in theory detects "all" chromatin looping events for a specific "viewpoint" or "bait" sequence of interest. Thus, I sought to identify 923 enhancer- and *Flg* promoter-mediated chromatin interactions specific to keratinocytes that define both "within-TAD" cis-interactions and "between-TAD" trans-interactions.

The 4C strategy will enable me to 1) gain a genome-wide view of the native chromatin architecture mediated by 923 and the *Flg* promoter underlying normal keratinocyte differentiation, and 2) potentially identify cis- or trans-acting regulatory elements for *Flg* gene expression

4.3 Results

4.3.1 Generation and sequencing of 4C-seq libraries

4C-seq libraries were generated from proliferating and differentiated mouse primary keratinocytes with 3 biological replicates per condition/cell-type. Specifically, the 4C-seq libraries were derived from the same cells used in my previously cross-linked 3C experiments described in Chapter 2. To establish a background frequency of non-keratinocyte-specific interactions, I also generated libraries from the P5424 RAG1−/−, p53−/− pro-T-cell line. In addition to the 923 and *Flg* gene promoter viewpoints, I included the Eβ enhancer with known interactions within the *Tcrb* locus in the P5424 T-cells as a technical control for the 4C-seq technique (Oestreich et al. 2006).

The 4C-seq libraries were created as previously described (Splinter et al. 2012, see also Methods). However, I wanted to ensure enough unique reads to maximize mappability of the interacting sequences. Thus, I devised a unique strategy to increase the number of informative bases in the sequencing reads (interacting sequences) in my 4C-seq libraries. Instead of adding standard Illumina sequencing adaptor tails to the 4C amplicons, I designed custom nested index PCR primers to bind within the viewpoint or bait (923 or *Flg*) as close as possible to the restriction site junction with the interacting sequence (Fig. 4.1). This also necessitated the use of custom sequencing primers (Table S1). I successfully generated 4C-seq libraries from proliferating and differentiated mouse primary keratinocytes based on optimal digestion using *HindIII* (first digest) and *NlaIII* (second digest) restriction enzymes (>80% digestion assessed by qPCR across the restriction sites flanking the viewpoint fragment). I determined both efficient ligation (formation of high molecular weight products), and successful synthesis of nested PCR amplicons, by agarose gel electrophoresis (Fig. S1). The 21 4C-seq libraries were pooled and then sequenced on two lanes of an Illumina Hi-Seq 2500 to obtain 50 bp single-end reads. The use of two lanes represent technical replicates to account for batch effects.

4.3.2 Quality assessment of 4C sequencing data

The raw sequencing reads were initially demultiplexed by allowing for 1 mismatch for each of the index sequences and subsequently binned into the corresponding library, with a range of 47,000 reads – 6.2 million reads per library in one lane (lane 6), and 43,000 reads – 6.9 million reads per library in the second lane (lane 7) (Fig. S2). Many of the reads, 62 million (52%) from lane 6 and 58 million (46%) from lane 7, could not be assigned to any of the libraries. The demultiplexed reads were mapped to a "fragment end library" consisting of a reduced genome containing only genomic sequences flanking the recognition sites of the

restriction enzymes used in the experiment (*HindIII* and *NlaIII*), and limited only to fragments flanked by one *HindIII* and one *NlaIII* site (Splinter et al. 2012). Visual inspection of the demultiplexed RPM normalized reads in the UCSC genome browser revealed unexpectedly that

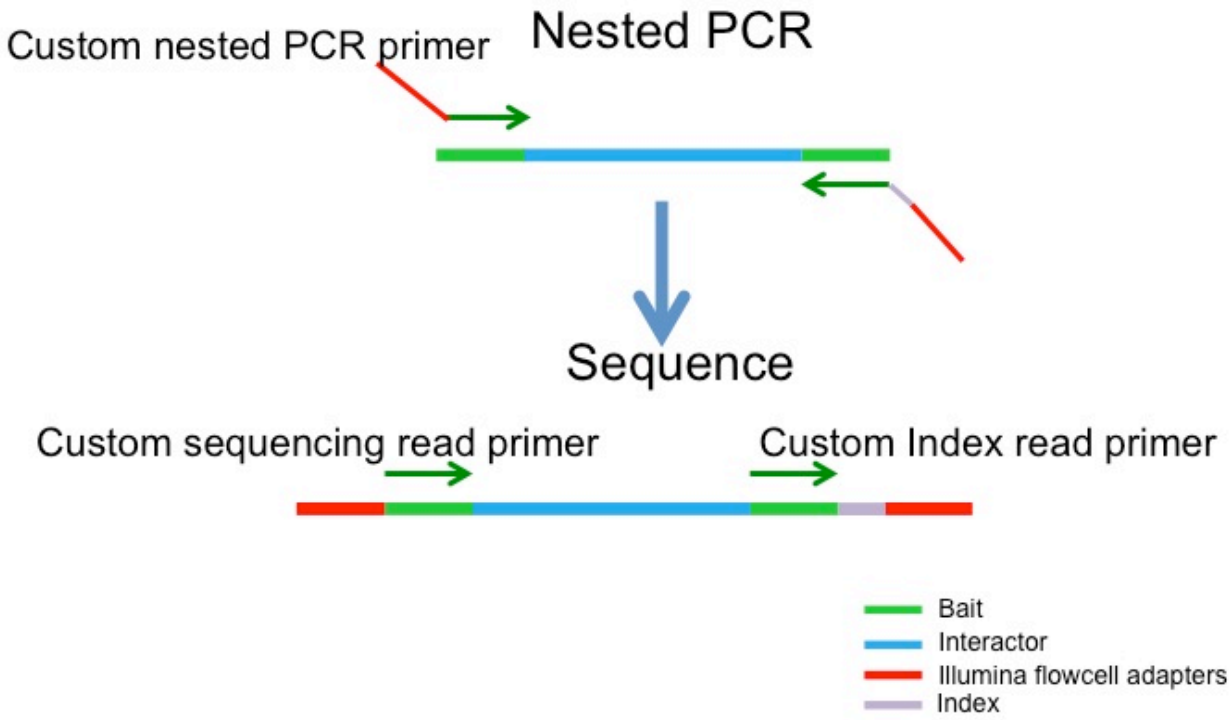


Figure 4.1. Custom primers for 4C-seq library preparation. Custom nested and indexed primers were designed to amplify off the inverse PCR product, as close as possible to the bait-interactor ligation junctions as possible. Primers for all libraries utilizing the same viewpoint are identical, except the index sequences on the reverse primer which are unique for each library. Libraries were submitted for sequencing with custom sequencing and index read primers.

in one of the *Flg* promoter viewpoint P5424 T-cell libraries, the region surrounding 923 contained the highest frequency of reads (Fig. S3). Since 4C experiments characteristically display the highest interaction frequencies close to the viewpoint fragment and decrease with increasing distance (Splinter et al. 2012), this unusual distribution of interactions suggested contamination of reads between the libraries. I hypothesized that this was due to the misassignment of reads during the demultiplexing process, and meant that any attempt at identification of significant interactions would most likely produce many false positives.

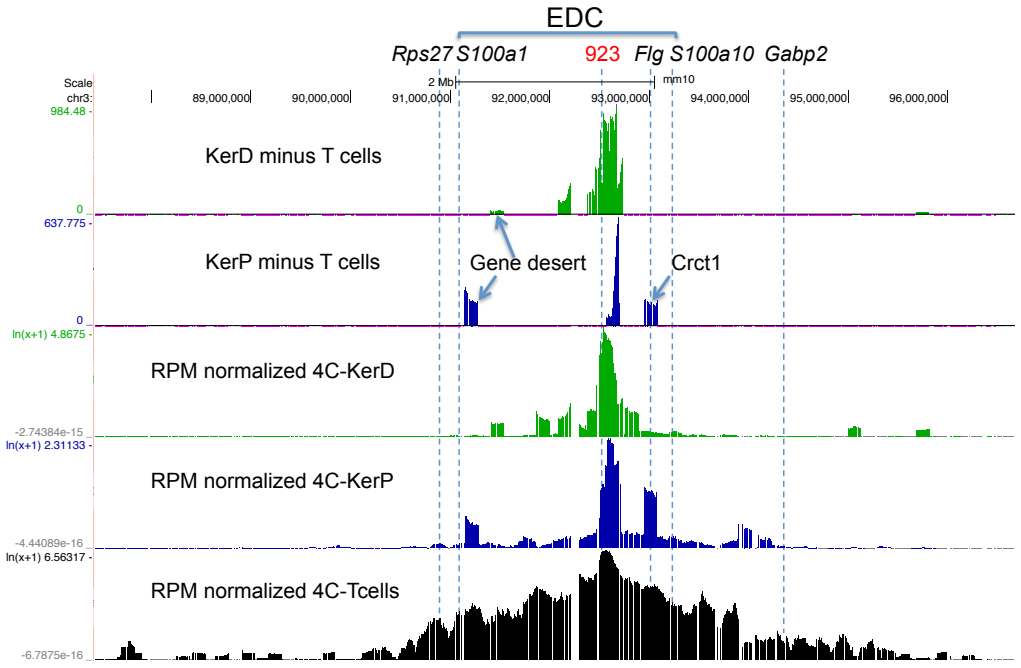
In order to increase stringency, I repeated the demultiplexing of the 4C-seq data to allow exact matches (0 mismatches) for each of the index sequences. As a result, the number of reads mapped to each library decreased to 5315 reads - 4.4 million reads/library (lane 6) and 6535 reads - 4.8 million reads/library (lane 7), with an increase in the number of unassignable reads to 81 million (68%) and 82 million (65%) respectively (Fig. S2). Again, the demultiplexed reads were mapped to the reduced genome as before. Visual inspection of the 0 mismatch demultiplexed RPM normalized reads in the UCSC genome browser revealed that the unusual peak previously observed near 923 in the *Flg* viewpoint P5424 T-cell libraries was no longer present (Fig. S3). Instead, each library now displayed the expected interaction pattern where the greatest number of reads mapped to the region close to its respective viewpoint (Fig. S3, S4, S5). This finding confirmed that the increased reads mapping to the 923 viewpoint in a library for the *Flg* viewpoint stemmed from reads obtained from an improper assignment of reads belonging to a 923 viewpoint library. The 1.4 – 10-fold loss in the number of reads resulting from the more stringent demultiplexing meant that the sparse coverage achieved in most of the libraries was prohibitive to the method of analyzing 4C that relies on identifying regions that are significantly over-represented relative to the neighboring regions (Splinter et al. 2012). Nevertheless, using

the 0 mismatch demultiplexed libraries, each of the two lanes produced 11 out of 21 libraries with sufficient 4C library quality (>40% cis-interactions, >1 million reads) for further analysis for the identification of keratinocyte specific cis- and trans-interactions (Fig. S6, Table S2) (van de Werken et al. 2012). From each lane, one out of three biological replicates of the 923 viewpoint differentiating keratinocyte libraries (923 KerD), *Flg* viewpoint proliferating keratinocyte libraries (Flg KerP), and *Flg* viewpoint differentiating keratinocyte libraries (Flg KerD), two out of three biological replicates of the 923 viewpoint P5424 T-cell libraries (923 Tcells) and *Flg* viewpoint P5424 T-cell libraries (Flg Tcells), and all three biological replicates of the Eβ enhancer viewpoint P5424 T-cell libraries (Eβ Tcells) had passed the quality threshold defined earlier. None of the 923 viewpoint proliferating keratinocyte libraries (923 KerP) surpassed the quality threshold, but one replicate that fell just short of the threshold was included in the downstream analyses for the purpose of completeness.

4.3.3 4C-seq identifies the EDC as a topologically associated domain

4C-seq libraries that passed the quality threshold as defined above were visualized on the UCSC Genome Browser using RPM-normalized reads, and represent chromatin interaction profiles in proliferating keratinocytes, differentiating keratinocytes and P5424 T-cells. In each of these libraries, the peak height at each mapped position corresponds to the number of reads per million (RPM) at that position, and represents the frequency of chromatin interactions between the viewpoint in question and the fragment to which the read belongs. The interaction profiles demonstrate the characteristic interaction frequencies that are highest in the regions closest to the respective viewpoints, and decline with increasing distance from the viewpoint (Fig. 4.2). The 923 and *Flg* promoter viewpoint in keratinocyte libraries showed a marked decline in the frequencies of interactions that occur within a 5Mb domain surrounding the EDC (*S100a1* –

a



b

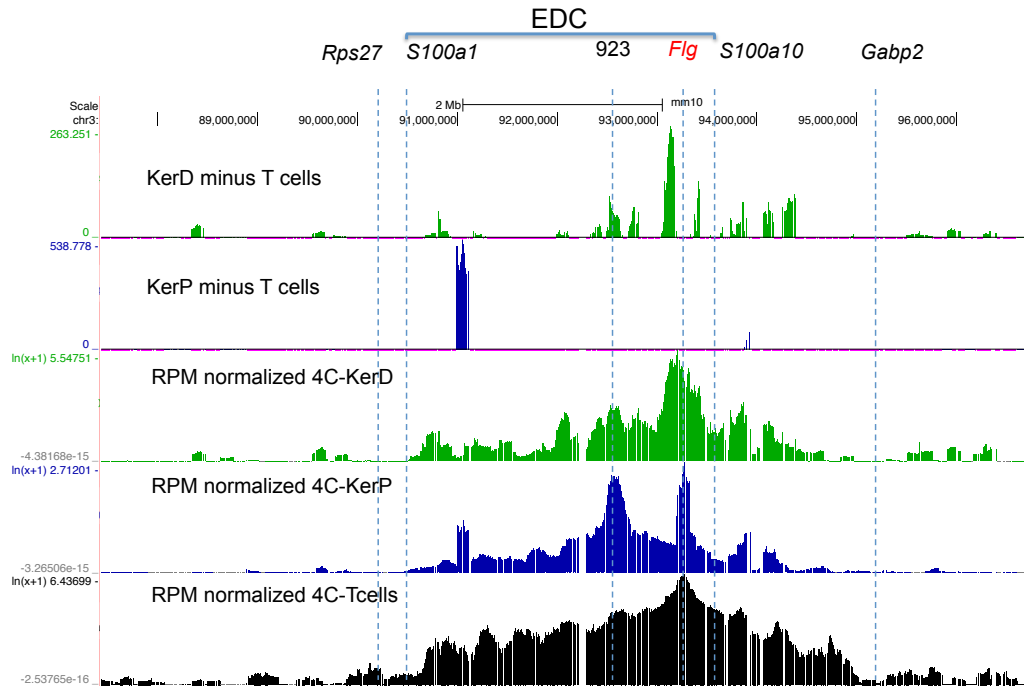


Figure 4.2. 4C-seq reveals an EDC TAD and an enrichment of EDC reads in keratinocyte libraries. After RPM normalization, peaks remain visible in the keratinocyte 4C libraries after subtracting T-cell reads, indicating an enrichment of interactions within the EDC in keratinocytes with both the a) 923 and b) *Flg* viewpoints.

S100a10) and demarcated by neighboring genes *Rps27* and *Gabp2* (Fig. 4.2). There also appeared to be localized peaks within the EDC indicating an overrepresentation of interactions in specific regions of the EDC. While interactions existed beyond the EDC, these were rare and sparse. These observed chromatin interactions in keratinocytes suggest a compact chromatin

structure within the bounds of *Rps27* and *Gabp2*, specific interactions within this structure, and obstacles to interactions with regions outside this 5Mb domain. On the other hand, the T-cell libraries representing an inactive EDC showed a more gradual decrease in interaction frequencies that appeared to decline as a function of distance. Beyond the boundaries of the locus marked by *Rps27* and *Gabp2*, **t**here was a low, albeit evenly distributed frequency of interactions (Fig. 4.2). This suggested a chromatin conformation of the EDC locus in T-cells that is absent of the specific interactions within the chromatin structure that were observed in the keratinocytes.

Together, these observations indicate an EDC TAD that appear to exhibit similar boundaries where interactions are more frequent within the TAD and less frequent outside the TAD, regardless of cell-type. Within the TAD in keratinocytes, however, there appear to be cell-specific interactions.

4.3.4 Keratinocyte-specific cis-interactions with the 923 enhancer and Flg promoter were observed within the EDC

The distribution of 4C-seq reads revealed distinct regions for which there were observable increased interactions frequencies that were specific to either keratinocytes compared to T cells (Fig 4.2). Thus, to identify chromatin interactions for the 923 enhancer and the *Flg* promoter within the EDC that were enriched specifically in keratinocytes, I subtracted 4C-seq reads obtained from the T-cells that represent the "background". This approach to identify

keratinocyte-specific chromatin interactions with either the 923 enhancer or the *Flg* promoter enables a more focused investigation on key chromatin interactions but does not completely exclude the importance of genomic regions associated with fewer reads for interactions. I also used the MACS peak caller to identify regions where 4C-seq reads for the 923 enhancer and *Flg* promoter were enriched in the keratinocytes compared to the P5424 T-cells (Feng et al. 2012).

The numbers of regions containing keratinocyte-enriched 4C-seq reads based on subtracting the background T-cell reads within the EDC are listed in Table 4.1. As shown in Table 4.2, which lists a subset of the reads accounted for in Table 4.1, only a small number of the keratinocyte-enriched 4C-seq reads overlap annotated regulatory elements. Interestingly, several of the keratinocyte-enriched 4C-seq reads correspond to annotated Cdx1 transcription factor binding sites (Lesurf et al. 2016) (Table 4.2). Cdx1 does not have a known role in keratinocyte differentiation, but is important for regulating the differentiation of gastrointestinal epithelia (Chan et al. 2009). The identification of an annotated binding site for Jarid2, Suz12, EZH2, Mtf2 within exon 2 of the *Tchh* gene (Table 4.2, UCSC mm10: chr3: 93443451-93443501) was particularly interesting since Jarid2, Suz12, and EZH2 are Polycomb group proteins that are important for epigenetic regulation of keratinocyte function (Eckert et al. 2011; Botchkarev et al. 2012), and Mtf2 has been known to complex with Polycomb Repressive Complex 2 (Li et al. 2010). Another region (Table 4.2, chr3:93,015,655-93,015,705), containing 4C-seq reads enriched specifically in the 923 viewpoint proliferating keratinocytes, overlaps the *Crct* gene. This region is orthologous to a previously identified CNE in the human genome, 531, that demonstrated enhancer activity in an *in vitro* reporter assay (de Guzman Strong et al. 2010).

A keratinocyte-specific enrichment of 4C-seq reads was also identified in both proliferating and differentiating keratinocytes within the 1.3 Mb gene desert that separates the

Table 4.1. Numbers of enriched 4C-seq reads in keratinocyte libraries based on subtraction of T-cell reads.

Gene family	Enriched in library			
	923 KerD	923 KerP	Flg KerD	Flg KerP
S100 5'	3	5	13	11
Gene desert	38	21	70	102
Pglyrp3	2	0	6	7
Sprr	15	5	36	42
Between Sprr and Lce	2	0	1	5
Lce	11	12	18	24
Between Lce and Flg	5	4	10	2
Flg	17	14	30	3
Between Flg and S100 3'	0	1	2	1
S100 3'	1	2	2	0

Table 4.2. Regulatory element annotations of enriched 4C-seq reads in keratinocyte libraries based on subtraction of T-cell reads.

Genomic position			Enriched in library	**Annotated Function** (Lesurf et al. 2016)
Chr	Start	End		
chr3	92303052	92303102	923 KerD	Cdx1 binding site upstream of *Sprr2b*
chr3	93443451	93443501	923 KerP, Flg KerD	Jarid2, Suz12, EZH2, Mtf2 binding site overlapping *Tchh* exon 2
chr3	91203067	91203117	Flg KerD, Flg KerP	Cdx1 binding site in gene desert
chr3	91262658	91262708	Flg KerD, Flg KerP	Cdx1 binding site in gene desert
chr3	93015655	93015705	923 KerP	Functional conserved regulatory element CNE 531 (de Guzman Strong et al. 2010)

S100 and *Sprr* gene families (Fig. 4.2a). In the 923 viewpoint proliferating keratinocyte library, one enriched region was located approximately 520 kb downstream from the *S100a9* gene (~800 kb upstream from *Pglyrp3*) (UCSC mm10: chr3:91,216,458-91,216,507), while one was observed in the differentiating keratinocytes approximately 780 kb downstream from *S100a9* (~540 kb upstream from *Pglyrp3*) (chr3: 91,471,110-91,471,159). With the *Flg* promoter viewpoint (Fig. 4.2b), the proliferating keratinocytes, showed an enriched region approximately 370 kb downstream from the *S100a9* gene (~950 kb upstream from *Pglyrp3*) (chr3: 91065728-91065777), while several were observed in the differentiating keratinocytes at approximately 185 kb, 540 kb, 760 kb, and 700 kb from *S100a9* (1135 kb, 780 kb, 560 kb, and 520 kb from *Pglyrp3*) (chr3:90880868-90880917, chr3: 91138413-91138462, chr3: 91230604-91230653, chr3: 91272119-91272168). The enrichment of reads within the gene desert might indicate as yet unannotated regulatory elements, or might be an artifact of the chromatin architecture of the rest of the EDC locus. The differentiated keratinocytes also showed specific enrichment of interactions between the *Flg* promoter and regions across the *Flg, Lce,* and *Sprr* gene families (Fig. 4.2b). One of these interaction peaks in particular lies between *Hrnr* and *Rptn*, in a region orthologous to a human EDC region harboring a cluster of previously identified CNEs (180, 184, and 189) with demonstrated regulatory activity *in vitro* (de Guzman Strong et al. 2010). Thus, I identified chromatin interactions between the 923 enhancer and the *Flg* promoter that correlate to functional regulatory elements in keratinocytes and suggests a potential role in these specific looping events for functional gene activation.

To determine the biological relevance of the chromatin interactions, I subsequently interrogated the MACS peaks using the Genomic Regions Enrichment of Annotations Tool (GREAT). Specifically, GREAT assigned potential biological functions to the regions where

4C-seq reads were enriched in keratinocytes relative to T-cells (McLean et al. 2010). Tables S3 – S6 list all Gene Ontology (GO) associations that achieve a significance threshold of binomial FDR <0.05, ranked in decreasing order of significance (smallest binomial FDR to largest binomial FDR). EDC genes are listed in bold.

This analysis identified an enrichment of 4C-seq reads in close proximity to genes associated with epidermal development-related GO terms (Table 4.3, Table 4.4). Keratinization, cornified envelope, and keratinocyte differentiation were the top three GO terms (based on FDR) for the 923 KerD (Table 4.3, Table S3). Cornified envelope was the top GO term associated with interactions enriched in *Flg* KerD, with keratinization, keratinocyte differentiation, epidermal cell differentiation, epidermis development, skin development, and structural constituent of epidermis achieving significance as well (Table 4.4, Table S4). This was consistent with the established role of *Flg* as an important keratinocyte differentiation gene, as well as 923's role in regulating key structural genes such as *Ivl*. In the *Flg* KerD library. Enriched reads were found in the *Flg*, *Sprr (*including *Ivl)*, *Lce,* and *S100* gene families (Table 4.4, Table S5), placing the *Flg* promoter within a network of chromatin interactions extending all across the EDC. In the 923 KerD library, most of the enriched reads were found within the *Sprr* gene family (Table 4.3, Table S3), indicating that within the EDC, the 923 enhancer mediates chromatin interactions mainly localized to the *Sprr* gene family to which it belongs. Considering the proximity of 923 to the *Ivl* gene, the absence of reciprocity between the 923 viewpoint and *Flg* was surprising, but could be due to the smaller number of reads obtained from the 923 KerD library.

The GREAT analysis also identified an enrichment of 4C-seq reads in both the 923 KerD (Table S3) and *Flg* KerD (Table S5) libraries near genes associated with detection and defense responses to bacteria, including GO terms such as peptidoglycan receptor activity and detection

Table 4.3. GREAT-predicted cis-regulatory targets of 4C-seq reads enriched in 923 enhancer viewpoint differentiating keratinocytes (923 KerD) relative to T-cells.

ID	Desc	BinomFdrQ	Genes
GO:0031424	keratinization	2.60E-14	**Ivl,Sprr1b,Sprr2a1,Sprr2a2,Sprr2a3,Sprr2i,Sprr3,Sprr4**
GO:0001533	cornified envelope	2.67E-14	**Ivl,Sprr1b,Sprr2a1,Sprr2a2,Sprr2a3,Sprr2i**
GO:0030216	keratinocyte differentiation	3.35E-05	**Ivl,Sprr1b,Sprr2a1,Sprr2a2,Sprr2a3,Sprr2i,Sprr3,Sprr4,**Yap1

Table 4.4. GREAT-predicted cis-regulatory targets of 4C-seq reads enriched in *Flg* promoter viewpoint differentiating keratinocytes (Flg KerD) relative to T-cells.

ID	Desc	BinomFdrQ	Genes
GO:0001533	cornified envelope	4.89E-18	Anxa1,**Flg,Hrnr,Ivl,Lor,Rptn,Sprr1b,Sprr2a1,Sprr2a2,Sprr2a3,Sprr2f,Sprr2g,Sprr2h,Sprr2i**
GO:0031424	keratinization	3.83E-13	Abca12,**Hrnr,Ivl,Lor,Sprr1b,Sprr2a1,Sprr2a2,Sprr2a3,Sprr2f,Sprr2g,Sprr2h,Sprr2i,Sprr3,Sprr4**
GO:0030216	keratinocyte differentiation	2.29E-05	Abca12,Anxa1,Ctnnd1,**Hrnr,Ivl,Lor,Sprr1b**,**Sprr2a1,Sprr2a2,Sprr2a3,Sprr2f,Sprr2g,Sprr2h,Sprr2i,Sprr3,Sprr4**,Trp63
GO:0009913	epidermal cell differentiation	2.45E-03	Abca12,Anxa1,Atoh1,Clic5,Clrn1,Ctnnd1,Ercc3,Gli2,Grxcr1,Hdac2,**Hrnr,Ivl,Lor**,Myo7a,Pcdh15,Slitrk6,**Sprr1b,Sprr2a1,Sprr2a2,Sprr2a3,Sprr2f,Sprr2g,Sprr2h,Sprr2i,Sprr3,Sprr4**,Trp63
GO:0008544	epidermis development	1.68E-02	Abca12,Anxa1,Atoh1,Barx2,Clic5,Clrn1,Ctnnd1,Ercc3,**Flg**,Foxq1,Gli2,Grxcr1,Hdac2,**Hrnr**,Igf1r,Inhba,**Ivl**,**Lce1g,Lce1i,Lce1l,Lce1m,Lce3a,Lce3c,Lce3f,Lor**,Myo7a,Pcdh15,Pdgfa,Pou3f2,Slitrk6,**Sprr1b,Sprr2a1,Sprr2a2,Sprr2a3,Sprr2f,Sprr2g,Sprr2h,Sprr2i,Sprr3,Sprr4**,Tgfb2,Trp63
GO:0043588	skin development	2.18E-02	Abca12,Anxa1,Arrdc3,Atoh1,Barx2,Clic5,Clrn1,Col1a1,Col1a2,Ctnnd1,Ercc3,**Flg**,Foxq1,Gli2,Grxcr1,Hdac2,**Hrnr**,Igf1r,Inhba,**Ivl,Lce1g,Lce1i,Lce1l,Lce1m,Lce3a,Lce3c,Lce3f,Lor**,Myo7a,Pcdh15,Pdgfa,Pou3f2,Slitrk5,Slitrk6,**Sprr1b,Sprr2a1,Sprr2a2,Sprr2a3,Sprr2f,Sprr2g,Sprr2h,Sprr2i,Sprr3,Sprr4,**Tcf7l2,Tgfb2,Trp63
GO:0030280	structural constituent of epidermis	3.82E-02	**Lor**

of bacterium. A closer look revealed that this was the result of an enrichment of 4C-seq reads within the gene desert which are associated with flanking genes *Pglyrp3* and *Pglyrp4*, and correspond to the enriched reads that had been identified in the gene desert by subtracting the background T-cell reads.

4.3.5 Keratinocyte-specific trans-interactions with the 923 enhancer and Flg promoter were observed near epidermal differentiation genes

The results of the MACS peak calling and subsequent GREAT analysis were also used to predict the functional significance of trans-interactions for the EDC in keratinocytes. Interestingly, within both 923 KerP and *Flg* KerP libraries, there were shared enrichments of 4C-seq reads relative to T-cells near *Vmn1r* family genes related to pheromone receptor activity (Tables S4 and S6). The *Flg* KerP libraries also displayed an enrichment of 4C-seq reads near genes belonging to the *Olfr* family (Table S6). Although the expression of the *Olfr* genes was originally thought to be limited to sensory neurons in the olfactory nasal epithelium, studies have indicated roles for the *Olfr* genes in other epithelial tissues (Gu et al. 2014). Therefore, the identification of 4C-seq reads near *Olfr* aside from the *Vmn1r* genes with respect to the EDC, could represent spatial proximity of these loci within a nuclear compartment for gene regulatory crosstalk.

I decided to focus on trans-interactions that would be most relevant to understanding the genome-wide interactions mediated by 923 and the *Flg* promoter in the context of keratinocyte differentiation by only considering MACS peaks that were associated with significantly enriched GO terms keratinocyte differentiation, keratinization, and cornified envelope, as determined by GREAT (Tables 4.3 and 4.4). These included an interaction between the 923 enhancer and

genomic fragments neighboring *Yap1* in differentiating keratinocytes, associated with the GO annotation of keratinocyte differentiation (Table 4.3, Table 4.5, Fig. 4.3). As well, MACS identified trans interactions between the *Flg* promoter with genomic regions near *Abca12, Anxa1, Ctnnd1,* and *Trp63* (Table 4.4, Table 4.6). Interestingly, *Trp63,* encodes p63, which is an important regulator of epidermal differentiation (Fig. 4.4). The identification of interactions with important epidermal differentiation genes on other chromosomes suggests possible co-regulation of *Flg* and 923-regulated genes by a wider network of chromatin interactions that extends beyond the EDC. However, additional experiments will need to be performed in the future to validate these findings.

4.4 Discussion

In summary, my results identify similar EDC TADs in proliferating versus differentiated keratinocytes and P5424 T-cells, consistent with the earlier studies that noted invariant TAD boundaries regardless of cell-type (Smith et al. 2016). However, within the boundaries of the EDC TAD, the keratinocyte libraries appeared to also contain distinctive peaks in interactions compared to the T-cells, likely representing interactions specific to an active EDC locus. Using 3D-FISH, a previous study demonstrated keratinocyte-specific compression of the EDC locus mediated by chromatin remodeler, Satb1, and predicted that Satb1 regulates EDC gene expression by forming a densely looped chromatin structure (Fessing et al. 2011). A related study further demonstrated that low EDC gene expression was correlated with distal positioning of the EDC from the neighboring genes *Rps27* and *Gabp2*, while high EDC expression correlated with proximity between the EDC and *Rps27* and *Gabp2* (Mardaryev et al. 2013). Using 923- and *Flg*-centric 4C-seq as independent functional assays for the chromatin architecture of the EDC, I identified a similar topology within the EDC that is unique to

Table 4.5. 4C-identified trans-interactions.

Bait	Target	Function
923	*Yap1*	Modulator of epidermal stem cell proliferation
Flg	*Abca12*	Keratinocyte lipid transporter, important for maintaining epidermal lipid barrier
	Anxa1	Interacts with S100A11 in epidermal growth factor-triggered growth signal pathway
	Ctnnd1	Important for cell adhesion
	Trp63	Master regulator of keratinocyte differentiation

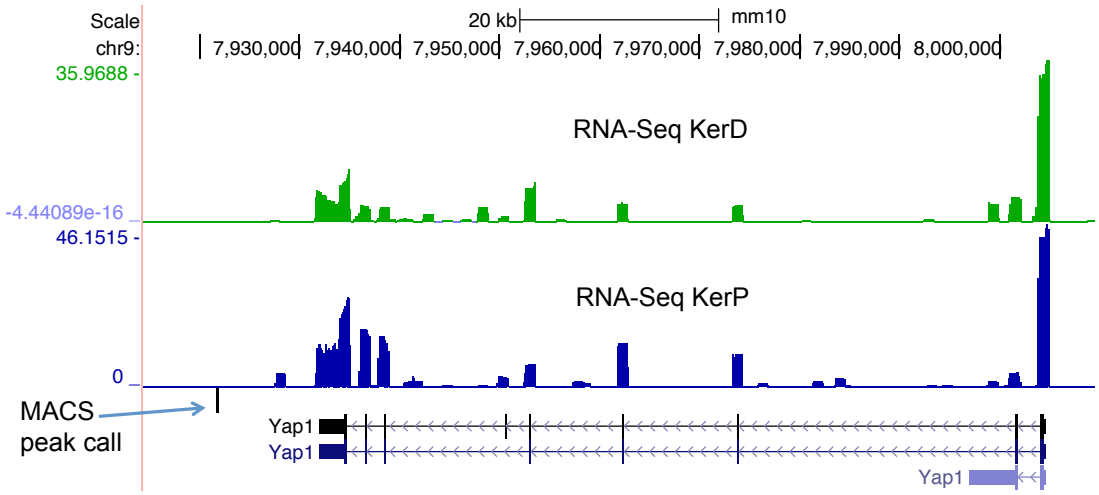


Figure 4.3. Keratinocyte-specific chromatin interactions identified between the 923 enhancer and *Yap1*. Example of a MACS-called significant interaction in 923 viewpoint differentiating keratinocyte library relative to T-cells. A 4C interaction with 923 mapped to a genomic region near *Yap1* and was identified by MACS as significantly enriched in a differentiating keratinocyte library relative to a T-cell library.

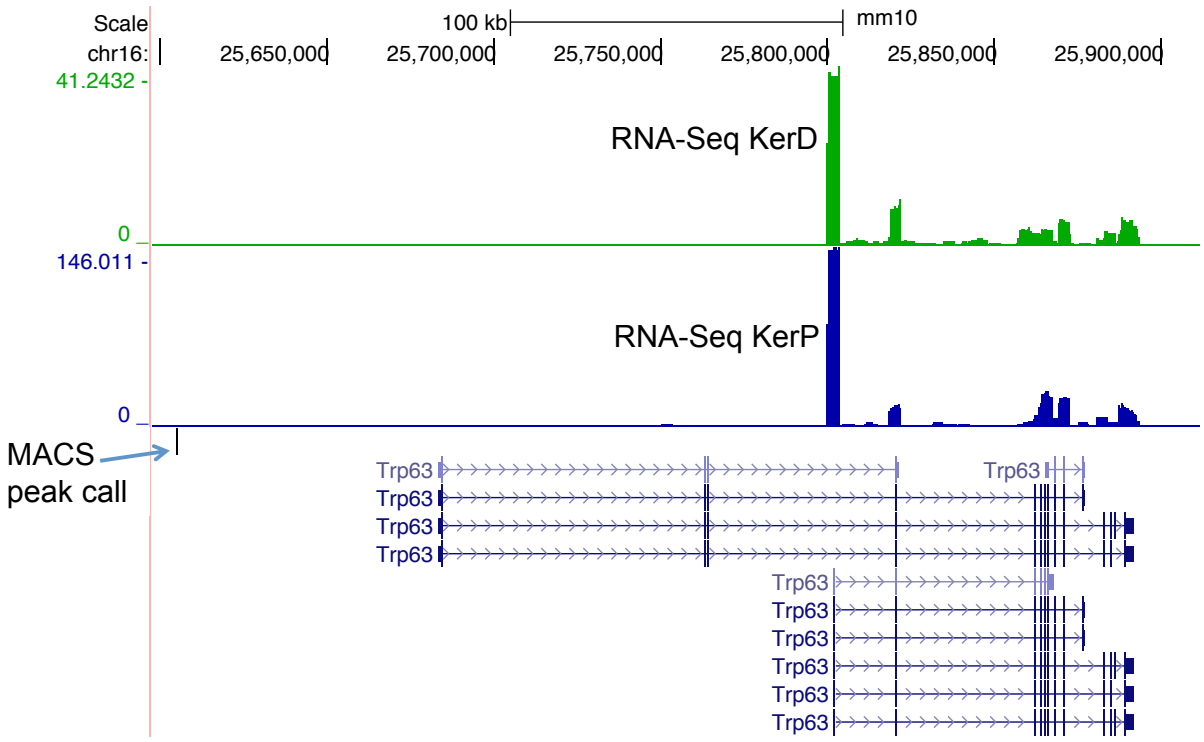


Figure 4.4. Keratinocyte-specific chromatin interactions identified between the *Flg* promoter and *Trp63*. Example of a MACS-called significant interaction in the *Flg* promoter viewpoint differentiating keratinocyte library relative to T-cells. A 4C interaction with the *Flg* promoter mapped to a genomic region near *Trp63* and was identified by MACS as significantly enriched in a differentiating keratinocyte library relative to a T-cell library.

Table 4.6. Numbers of enriched 4C-seq reads in keratinocyte libraries relative to T-cells based on MACS peak calls.

Gene family	Enriched in library			
	923 KerD	923 KerP	Flg KerD	Flg KerP
S100 5'	0	0	2	0
Gene desert	6	1	13	1
Pglyrp3	0	0	3	0
Sprr	21	1	10	2
Between Sprr and Lce	0	0	1	0
Lce	3	3	7	0
Between Lce and Flg	0	0	8	0
Flg	0	0	22	3
Between Flg and S100 3'	0	0	1	0
S100 3'	0	0	2	0

keratinocytes compared to T-cells that do not express the EDC genes. The EDC cis-interactions identified by subtraction of T-cell reads indicated that both 923 and the *Flg* promoter were capable of forming interactions throughout the EDC (Table 4.1). This method identified cis-interaction profiles that encompass similar regions with respect to both 923 and *Flg* baits, providing further evidence for the EDC chromatin domain identified in the previous studies (Fessing et al. 2011; Mardaryev et al. 2013). However, the more stringent MACS peak caller indicated that 923, formed more localized interactions mainly within the *Sprr* gene family and the nearby *Lce* family, while the *Flg* promoter displayed a more far-reaching interactions, extending throughout the EDC (Table 4.6).

The identification of chromatin looping events within the gene desert for both 923 and *Flg* loci suggests the co-opting of a regulatory element by both a gene and an enhancer. The chromatin looping events between 923 and CNE 531 enhancer support the tendency for regulatory elements to interact with each other, as previously noted (Robyr et al. 2011). The interactions between the *Flg* promoter and the gene desert, as well as the region harboring CNE enhancers 180, 184, and 189, also illustrate the use of 4C-seq data to discover putative functional regulatory elements for chromatin remodeling.

Recent studies have identified specific roles for CTCF to mediate chromatin architecture and gene expression. Depletion of the structural proteins CTCF and cohesin and the deletion of CTCF binding sites and a cis-regulatory element identified that while CTCF binding was necessary for maintaining a normal interaction profile, the regulatory element was required for normal gene expression (Yang et al. 2016). Additional studies definitively demonstrated that the deletion of a CTCF site, can alter the TAD structure caused aberrant enhancer-promoter interactions and thus aberrant gene expression (Guo et al. 2015; Lupiáñez et al. 2015). Consistent

with these findings, ENCODE-annotated clusters of CTCF binding sites are found between each of the 4 EDC gene families and suggest that up to 4 TADs may exist within the EDC (Fig. 4.5). The locations of the CTCF sites are largely conserved across a variety of tissues and cell-types, indicating a conservation of TADs.

Interestingly, the human orthologous region (UCSC hg38 chr1:152,162,963-152,173,456) for the approximate 12 kb region (UCSC mm10 chr3: 93,376,723-93,389,039) that interacts with the *Flg* promoter also encompasses a long noncoding RNA, FLG-AS1 (Flg anti-sense). FLG-AS1 has 9 isoforms. The overlap between the CNE 184 enhancer and exon 1 of FLG-AS1-005 (ENST00000593011.5) suggests the existence of an enhancer RNA for this particular splice variant of FLG-AS. Recent studies have demonstrated that enhancer RNAs (eRNAs) contribute to regulating gene expression, that include stabilization of chromatin loops and targeting mRNAs for degradation (Kim et al. 2010). The interaction between the *Flg* promoter and the proximal region of the CNE 184 enhancer could indicate co-regulation of *Flg* and FLG-AS transcription during keratinocyte differentiation. The resulting FLG-AS transcript could likely participate in a positive feedback loop by stabilizing the chromatin interaction.

As demonstrated in this chapter, chromatin conformation studies can help us elucidate mechanisms of gene activation by the discovery of putative regulatory elements. Often the regulatory elements that affect the expression of a gene expression might be distally located, making their target effects difficult to elucidate. Many common diseases cannot be completely explained by coding variants with an alternative hypothesis for regulatory variants for disease risks. Therefore, by identifying specific enhancer-promoter interactions and TADs within which genes and enhancers are likely to interact, chromatin conformation studies have the potential to pinpoint candidate regions for the identification of disease causing variants.

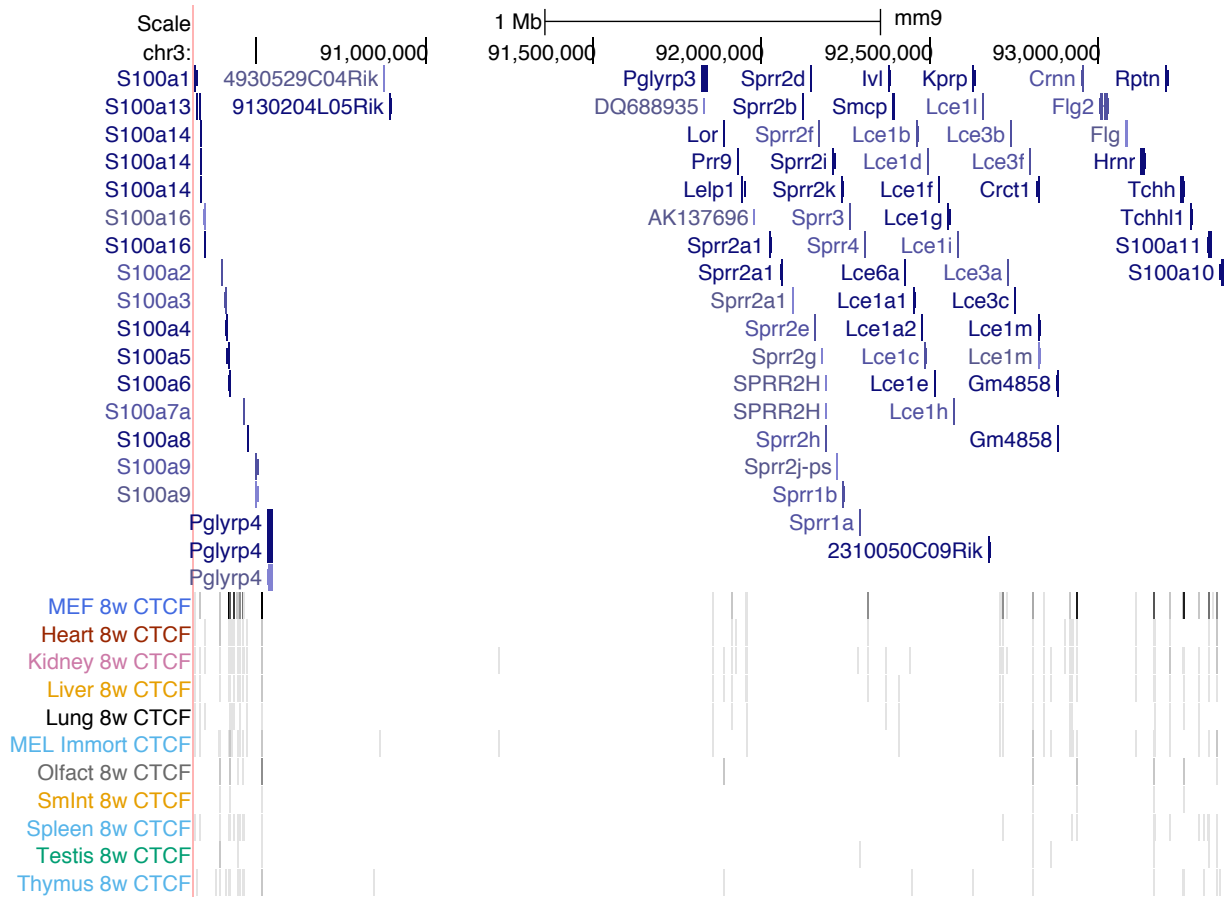


Figure 4.5. Distribution of CTCF binding sites across the EDC. ENCODE CTCF ChIP-seq reveals binding across a variety of cell lines that corresponds to the approximate boundaries of each of the EDC gene families. Binding intensities correspond to the darkness of the lines, and indicate varying intensities in different cell lines and tissues **(**UCSC Genome Browser build mm9).

4.5 Materials and Methods

4.5.1 Circular chromosome conformation capture library preparation

Primary keratinocytes were isolated from newborn mice as previously described (Lichti *et al.*, 2008) and plated under proliferating or differentiating (2.0 mM Ca^{2+}) conditions in custom keratinocyte media. The 4C assay was performed as previously described (Splinter et al., 2012). Approximately 10 million cells were harvested at 72 hours post-calcium treatment and cross-linked while in single-cell suspension with 2% formaldehyde. The cross-linking reaction was quenched with 0.125M glycine followed by cell lysis to recover the nuclei. The cross-linked nuclei were incubated at 37ºC for 4 hours with 200U *Hind*III (New England BioLabs, Ipswich, MA). An additional 200U of *Hind*III was supplemented before the nuclei were incubated overnight, and again before another 4h incubation the following day. Each 4C library was assessed for efficient digestion by qPCR, using primers to amplify across a *HindIII* site and comparing the Ct values to amplification within a restriction fragment (primers listed in Table S1). The libraries were then ligated with T4 DNA ligase (New England BioLabs) overnight at 16ºC. An aliquot from each library was electrophoresed on a 1.5% agarose gel and assessed for efficient ligation as determined by the presence of large molecular weight fragments. Libraries with satisfactory ligation efficiencies were then de-cross-linked using Proteinase K (IBI Scientific, Peosta, IA), and purified by phenol/chloroform (Life Technologies). A second digestion was performed at 37ºC overnight using 50U of *Nla*III (New England BioLabs, Ipswich, MA). Again, each library was assessed for efficient digestion by qPCR (primers listed in Table S1) before proceeding to a second round of ligation. After efficient ligation was confirmed, libraries were purified using the QIAquick PCR purification kit (Qiagen). The 4C libraries were

amplified by inverse PCR followed by a nested PCR using primers containing Illumina adaptor tails to generate 4C amplicon libraries (Table S1). Both rounds of PCR used the Expand Long Template Polymerase (Roche) and the following PCR conditions: 94ºC for 2 min followed by 29 cycles of 94ºC for 10 s, 55ºC for 1 min and 68ºC for 3min (Fig. S1). 4C amplicon libraries were sequenced using custom sequencing primers on an Illumina HiSeq 2500 (Illumina, Hayward, CA) (Fig. 4.1). See Table S1 for inverse primer sequences and nested sequences.

4.5.2 Data Analysis

Mapping of 4C-seq reads

The raw sequencing reads were first demultiplexed by allowing either 0, 1, or 2 mismatches in the index sequences. The successfully demultiplexed reads were then mapped to the mouse genome (UCSC mm10) using the Burrows-Wheeler Aligner (BWA). Using the Basic4Cseq software (Walter et al. 2014), the mapped reads were limited to a reduced genome containing only valid restriction fragments based on the restriction enzyme strategy used in the experiment and represent valid interactions.

Determination of 4C-seq library quality

Each 4C-seq library was assessed for quality using the Basic4Cseq software (Walter et al. 2014) prior to further analysis. The criteria for a good quality library was previously empirically determined as having at least 40% of cis-interactions and > 1 million reads (van de Werken *et al.,* 2012) (Figure S6, Table S2).

Identification of keratinocyte-specific 4C interactions

Libraries that passed the quality control criteria were used for the identification of keratinocyte-specific interactions. Enrichment of interactions in keratinocytes versus P5424 T-cells was identified using the MACS peak caller (v1.4.2) (Zhang, *et al.*, 2008). The default parameters were used for running MACS. For each bait, keratinocyte libraries (proliferating or differentiating) were used as "treatment" files, while T-cell libraries were used as "control" files.

Identification of trans-interactions

Peaks called by MACS were used as input for the Genomic Regulatory Enrichment of Annotations Tool (GREAT v3.0) (McLean, C., *et al.*, 2010). The UCSC mm10 build was used, and GREAT hits that were significant by the binomial test (false discovery rate <0.05) were further considered.

4.6 Supplementary Material

Supplementary material may be viewed in Appendix C.

4.7 References

Botchkarev, VA, Gdula, MR, Mardaryev, AN, *et al.* (2012). Epigenetic regulation of gene expression in keratinocytes. *J Invest Dermatol* 132: 2505–21.

Chan, CWM, Wong, N a, Liu, Y, *et al.* (2009). Gastrointestinal differentiation marker Cytokeratin 20 is regulated by homeobox gene CDX1. *Proc Natl Acad Sci U S A* 106: 1936–41.

Eckert, RL, Adhikary, G, Rorke, EA, *et al.* (2011). Polycomb group proteins are key regulators of keratinocyte function. *J Invest Dermatol* 131: 295–301.

Feng, J, Liu, T, Qin, B, *et al.* (2012). Identifying ChIP-seq enrichment using MACS. *Nat Protoc* 7: 1728–40.

Fessing, MY, Mardaryev, AN, Gdula, MR, *et al.* (2011). p63 regulates Satb1 to control tissue-specific chromatin remodeling during development of the epidermis. *J Cell Biol* 194: 825–39.

Gu, X, Karp, PH, Brody, SL, *et al.* (2014). Chemosensory functions for pulmonary neuroendocrine cells. *Am J Respir Cell Mol Biol* 50: 637–46.

Guo, Y, Xu, Q, Canzio, D, *et al.* (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* 162: 900–10.

de Guzman Strong, C, Conlan, S, Deming, CB, *et al.* (2010). A milieu of regulatory elements in the epidermal differentiation complex syntenic block: implications for atopic dermatitis and psoriasis. *Hum Mol Genet* 19: 1453–60.

Irvine, AD, McLean, WHI (2006). Breaking the (un)sound barrier: filaggrin is a major gene for atopic dermatitis. *J Invest Dermatol* 126: 1200–2.

Kim, T-K, Hemberg, M, Gray, JM, *et al.* (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465: 182–7.

Lesurf, R, Cotto, KC, Wang, G, *et al.* (2016). ORegAnno 3.0: a community-driven resource for curated regulatory annotation. *Nucleic Acids Res* 44: D126–32.

Li, G, Margueron, R, Ku, M, *et al.* (2010). Jarid2 and PRC2, partners in regulating gene expression. *Genes Dev* 24: 368–80.

Lupiáñez, DG, Kraft, K, Heinrich, V, *et al.* (2015). Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* 161: 1012–25.

Mardaryev, AN, Gdula, MR, Yarker, JL, *et al.* (2013). p63 and Brg1 control developmentally regulated higher-order chromatin remodelling at the epidermal differentiation complex locus in epidermal progenitor cells. *Development* 141: 101–11.

McLean, CY, Bristor, D, Hiller, M, *et al.* (2010). GREAT improves functional interpretation of

cis-regulatory regions. *Nat Biotechnol* 28: 495–501.

Nomura, T, Akiyama, M, Sandilands, A, *et al.* (2009). Prevalent and rare mutations in the gene encoding filaggrin in Japanese patients with ichthyosis vulgaris and atopic dermatitis. *J Invest Dermatol* 129: 1302–5.

Oestreich, KJ, Cobb, RM, Pierce, S, *et al.* (2006). Regulation of TCRbeta gene assembly by a promoter/enhancer holocomplex. *Immunity* 24: 381–91.

Palmer, CNA, Irvine, AD, Terron-Kwiatkowski, A, *et al.* (2006). Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat Genet* 38: 441–6.

Palstra, R-J, Tolhuis, B, Splinter, E, *et al.* (2003). The beta-globin nuclear compartment in development and erythroid differentiation. *Nat Genet* 35: 190–4.

Robyr, D, Friedli, M, Gehrig, C, *et al.* (2011). Chromosome conformation capture uncovers potential genome-wide interactions between human conserved non-coding sequences. *PLoS One* 6: e17634.

Schoenfelder, S, Sexton, T, Chakalova, L, *et al.* (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet* 42: 53–61.

Smith, EM, Lajoie, BR, Jain, G, *et al.* (2016). Invariant TAD Boundaries Constrain Cell-Type-Specific Looping Interactions between Promoters and Distal Elements around the CFTR Locus. *Am J Hum Genet* 98: 185–201.

Smith, FJD, Irvine, AD, Terron-Kwiatkowski, A, *et al.* (2006). Loss-of-function mutations in the gene encoding filaggrin cause ichthyosis vulgaris. *Nat Genet* 38: 337–42.

Splinter, E, de Wit, E, van de Werken, HJG, *et al.* (2012). Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. *Methods* 58: 221–30.

Walter, C, Schuetzmann, D, Rosenbauer, F, *et al.* (2014). Basic4Cseq: an R/Bioconductor package for analyzing 4C-seq data. *Bioinformatics* btu497 – .

van de Werken, HJG, de Vree, PJP, Splinter, E, *et al.* (2012). 4C technology: protocols and data analysis. *Methods Enzymol* 513: 89–112.

Yang, R, Kerschner, JL, Gosalia, N, *et al.* (2016). Differential contribution of cis-regulatory elements to higher order chromatin structure and expression of the CFTR locus. *Nucleic Acids Res* 44: 3082–94.

Chapter 5

Conclusions

5.1 Summary

Early studies towards understanding transcriptional regulation initially focused on the proximal sequences for a gene of interest. For epidermal development, genetic approaches to elucidate the mechanisms underlying the keratinocyte-specific gene expression identified a shared core of p63, AP1, SP1, and C/EBP transcription factors in the activation of *K5*, *K14*, *K1*, *K10*, *LOR*, and *IVL*. These early studies approached gene regulation from a proximal and gene-centric level. As seen in the regulation of EDC gene expression within the developing epidermis, proper control of gene expression is an intricate hierarchy of events. Due to the shared biology of the EDC genes, the EDC locus was eventually recognized as a gene cluster. The availability of whole genomes further identified sequence conservation of the EDC locus throughout mammalian evolution, and together with the conservation of expression patterns between mammalian species, indicated a selective pressure to maintain the locus as a regulatory module. Bioinformatics tools and high-throughput experimental techniques have subsequently allowed us to analyze the regulatory landscapes of the EDC locus as a whole.

My work sought to identify the link between concomitant EDC gene expression pattern with respect to the 923 epidermal-specific enhancer during keratinocyte differentiation. Based on the assumption that sequence conservation in noncoding regions likely implies function

(Lindblad-Toh et al. 2011), 923 was identified as one of the 48 conserved non-coding elements within the EDC locus that could serve as potential modulators (enhancers or repressors) of the gene cluster (de Guzman Strong et al. 2010). This study demonstrated enhancer function for the human 923 sequence as it was sufficient to increase reporter gene expression in keratinocytes as well as direct epidermal specificity in transgenic reporter mice.

In Chapter 2, I further demonstrated that the enhancer activity for 923 correlated with the dorsal to ventral patterning of skin barrier formation and the activation of marker genes of epidermal differentiation , *K1*, *Lor*, *Flg*, and *Ivl* (Oh et al. 2014). Using a tiling 3C assay to interrogate physical chromatin interactions between the 923 enhancer and EDC gene promoters, I discovered a 923-centric chromatin architecture of the EDC that was dynamic with respect to keratinocyte differentiation. I further identified a requirement for AP-1 transcription factor binding to the 923 enhancer for EDC gene activation and chromatin remodeling based on bioinformatics and genetics studies that mapped the functional relevance of the AP-1 binding site and 3C assays on differentiated keratinocytes treated with the pharmacological AP-1 binding inhibitor Tanshinone IIA. As distal enhancers have been known to regulate gene expression via chromatin looping interactions (Ong and Corces 2011), my findings demonstrated this paradigm for keratinocyte biology as mediated by 923. Furthermore, my identification of a role for AP-1 to affect EDC gene expression and chromatin looping provided a link between studies that highlighted AP-1 as an important regulator of epidermal differentiation (Mehic et al. 2005; Rorke et al. 2010), with cell biology studies that identified remodeling of the EDC and proper nuclear spatial organization as requirements for EDC gene expression (Fessing et al. 2011; Mardaryev et al. 2013).

In Chapter 3, my goal was to determine if the 923 enhancer is necessary for EDC gene expression. I addressed the function of 923 by deleting the orthologous 923 sequence in mice using CRISPR/Cas9 genome editing. Targeted deletion of the 923 enhancer *in vivo* as a functional assay in mice is more feasible given the ease of design, cost, and specificity of the CRISPR/Cas9 genome editing technology. Using two designed sgRNAs to target Cas9 nuclease activity to the flanking ends of 923 and two *loxP* ssODNs with specific homology arms to enable the insertion of flanking *loxP* sequences by homology-directed repair, I recovered at least four 923 mutation alleles. They included one floxed 923 (923^{flox}) with both *loxP* inserting in cis, two independent deletion alleles (923^{delA}, 923^{delB}), and one partial deletion with a remaining 238 bp of 923 (923^{pdel}). Although 923^{delA} and 923^{delB} mice demonstrated normal viability and normal epidermal barrier structure and function, gene expression analysis of both deletion lines indicated an allele-dose dependent decrease in the expression of 923's immediate neighbor, *Ivl*, and to a lesser extent, proximal genes *Smcp* and *Lce6a,* and distal genes *Crnn* and several of the *Lce* gene family members (*Lce3a*, *Lce3b*, *Lce3e* and *Lce3f)*. Correlating with this observation was a compensatory upregulation of several members of the *Sprr* gene family (*Sprr2a2, Sprr2e, Sprr2g, Sprr2i, Sprr2j,* and *Sprr1b*) that are important to the integrity of the epidermal skin barrier as well. Thus, based on these initial studies, I identified a requirement for the 923 enhancer for proximal and distal gene activation in the epidermis.

In Chapter 4, I sought to identify the chromatin architecture of the EDC on a genome-wide scale with respect to the 923 enhancer and *Flg* promoter in keratinocytes compared to non-EDC expressing T cells. My previous approach towards understanding the chromatin interactions of 923 using the 3C assays described in Chapter 2, provided a selective view on 923-mediated interactions that were limited to EDC promoter regions. With increasing evidence that chromatin

architecture is meditated by interactions between regulatory elements aside from enhance-promoter interactions (Robyr et al. 2011), I was compelled to expand my view of 923-mediated chromatin architecture by using an unbiased 4C-seq approach. This would not only allow me to interrogate interactions across the EDC at a higher resolution, but also allow me to identify genome-wide interactions that would help me place the role of 923 and the EDC within a broader context of the many other genes involved in epidermal differentiation. I also included a second viewpoint at the *Flg* gene promoter in order to elucidate the underlying chromatin architecture that affects its expression. I was interested in *Flg* as it encodes a major structural protein of the skin barrier. As mentioned previously, *FLG* mutations in humans are associated with atopic dermatitis (AD), but only account for about 40% of the disease incidence (Palmer et al. 2006). However, the persistent linkage of AD to the EDC even after excluding known *FLG* mutations suggests that additional causal variants reside within the EDC that could include regulatory elements. The use of the *Flg* gene promoter as a viewpoint for a 4C-seq experiment would therefore allow me to identify candidate enhancers that could be modulators of *Flg* expression, while also providing an alternate view of the EDC chromatin landscape. From these experiments, both the 923 and *Flg* promoter viewpoints revealed a topology domain in the EDC that was specific to keratinocytes compared to T cells. Moreover, there also existed distinct chromatin interactions in keratinocytes, and even specifically in proliferative versus differentiated states. Interaction between the 923 enhancer and non-coding regions within the gene desert between the *S100* and *Sprr* gene families were strongly enriched in keratinocytes and suggests the presence of as yet unannotated regulatory elements within the gene desert. Also, enhancer-enhancer chromatin interactions were observed in keratinocytes, suggesting a role for chromatin looping between enhancers to mediate keratinocyte-specific chromatin architecture of the EDC. These

studies emphasize the need to more closely examine these regions and their roles in shaping the EDC chromatin topology. Furthermore, the interaction identified between the *Flg* promoter and enhancer 184 in the EDC demonstrate the potential of 4C-seq and other chromosome conformation capture techniques for pinpointing target regions for elucidating causal variants for diseases such as AD.

Together, the results of my work establish a paradigm for enhancer-promoter chromatin interactions for gene activation specific to epidermal development. I characterized the AP-1-dependent enhancer 923, demonstrating its role in mediating spatial interactions with numerous EDC gene promoters to establish a poised conformation for coordinating EDC gene activation upon the induction of epidermal differentiation (Oh et al. 2014). In Chapter 3, I further discovered that the loss of the 923 enhancer results in a dose-dependent decrease in proximal genes (*Ivl*, *Smcp,* and *Lce6a*) and distally located genes (*Crnn* and distal *Lce* family genes), along with an increased expression of *Sprr* genes, to maintain the integrity of the skin barrier. The exact nature of these mechanisms have yet to be explored. However, interactions between 923 and other enhancer elements within the EDC (detailed in Chapter 4) suggest an intriguing model of chromatin looping mediated by enhancers in the *Sprr, Lce,* and *Flg* gene families to coordinate gene expression, and to rapidly activate functionally related genes as a "back-up" mechanism for overcoming the loss of functionally related proteins to maintain barrier integrity. These findings contributed to the field's understanding of the regulatory landscapes, chromosome territories and topologically associated domains (TADs) that bring regulatory enhancers and promoters into close proximity for the purpose of transcriptional regulation in the context of keratinocyte differentiation.

5.2 Future Directions

We have still much to learn about the regulatory principles that govern gene activation at the chromatin level and specifically in epidermal development. Elucidation of the molecular biology and biochemistry of enhancer-promoter interactions has opened up a new era to further investigate the paradigm of transcriptional regulation. Armed with new methodologies for genome sequencing and editing and protein engineering to both discover enhancers and to rapidly test their function, we approach a comprehensive understanding of the principles of genome architecture that modulate cellular transcriptomes at a faster rate than ever.

Large amounts of genomics data have been generated due to the availability of whole genome sequences from increasing numbers of species (Lindblad-Toh et al. 2011), high-throughput techniques such as ChIP-seq, RNA-seq, and DNase-seq. The accompanying development of bioinformatics tools to allow the integration of the genomic and epigenetic data (Hoffman et al. 2013), have made it possible to analyze gene regulation on a large-scale. These advances have enabled the field to identify enhancers across the genome based on correlative epigenetic marks. However, we are at a critical point in deciphering the function of these enhancers in skin biology and understanding the role that enhancers play in transcriptional regulation and genome organization.

Previous evidence has shown that cohesin and CTCF play a role in maintaining enhancer-promoter interactions in a multitude of cell-types and tissues (Hadjur et al. 2009; Mishiro et al. 2009; Hou et al. 2010; Seitan et al. 2011). Specifically, CTCF has been shown to bind at the boundaries of TADs in a directional manner, creating insulator boundary elements that allow neighboring genes to be regulated in different ways (Dixon et al. 2012; Phillips-Cremins et al.

2013; Guo et al. 2015; Tang et al. 2015). My own results from the 4C-seq experiments described in Chapter 4 hint at the involvement of both 923 and *Flg* in a network of chromatin interactions involving genes and regulatory elements throughout the genome to act cooperatively in order to drive proper gene expression for epidermal differentiation. Within the EDC, the boundaries of the gene families correspond approximately with clusters of CTCF binding sites (UCSC mm10). Further, the discrete temporal activation of each EDC gene family during epidermal differentiation (de Guzman Strong et al. 2010) indicates independent regulation of each gene family. Taken together this suggests that CTCF is likely to be a major player in mediating the formation of chromatin loops by defining TADs within the EDC. An intriguing hypothesis to test is a functional role for CTCF to maintain normal EDC gene expression during epidermal development.

Drawing from the expertise of the evolutionary biology field, we are now able to identify candidate regions with regulatory potential faster than ever (Siepel et al. 2005; Pennacchio et al. 2006; Hemberg et al. 2012). Techniques such as the massively parallel reporter assay (MPRA) that allow high-throughput assessment of enhancer activity have also made the task of *in vitro* testing of multiple putative enhancers far less laborious (Kwasnieski et al. 2014). In Chapter 1, the loss of AP-1 binding to 923 significantly diminished, but did not entirely ablate the ability of 923 to drive reporter gene expression or form chromatin interactions (Oh et al. 2014). This indicated a contribution of other unknown protein factors for 923's enhancer activity. Our lab performed targeted sequencing of the 923 enhancer in a cohort of atopic dermatitis patients and identified atopic dermatitis (AD) patient-specific variants that are predicted to disrupt the binding of transcription factors with key roles in epidermal development, including Klf4 and Smad2:3:4 (Quiggle, et al. unpublished data). These suggest targets for further dissection of the molecular

mechanisms governing 923's enhancer activity. Additionally, the observed 4C-seq interactions between enhancers 923 and 531 in proliferating keratinocytes, as well as between the *Flg* promoter and enhancer 184 in differentiating keratinocytes, suggest that enhancers in the EDC are important for physiologically sensitive chromatin looping. The MPRA technique would allow us to perform mutation saturation of 923 and other enhancers of interest in order to gain a high-resolution view of how variants within these elements could modulate EDC gene expression and even in complex diseases such as AD.

Advances in genome editing (CRISPR/Cas9, TALENs) available today have also made it easier than ever to functionally test the endogenous functions of enhancers (Fanucchi et al. 2013; Sander and Joung 2014; Lupiáñez et al. 2015). They are also an advancement over transgenic *in vivo* studies that lack the ability to address the function of enhancers in the appropriate genomic context. The precision of these genome editing methods enables us to more directly test our hypotheses in specific locus/loci within the regulatory landscape. Our CRISPR/Cas9 generated 923 knockout mice provide us with a unique opportunity to examine the importance and the necessity of 923 for mediating chromatin looping in the EDC by performing 3C or 4C assays on keratinocytes derived from the mice. By comparing the interactions observed in the homozygous mutants to wild-type littermates, we will be able to determine if the chromatin interactions observed previously using 923 as a bait are directly mediated by the 923 sequence, or if they were merely an artifact of a 923-independent chromatin architecture that brought 923 into proximity with the interacting regions. Furthermore, by overlaying the interaction data with RNA-seq data from these mice, we can draw conclusions about which interactions affect gene expression. We could also assess the allele-bias of 923 activity by using a heterozygous background to investigate the ability of the intact 923 to drive expression of the target genes on

the other allele harboring the deletion of 923. To test this, we would cross the 923 knockout mutants (C57BL/6xCBA) (backcrossed to select for the C57BL/6 background), to Balb/C mice and perform RNA-seq and 4C-seq on isolated epidermis from the heterozygous mice. The Balb/C line is one of the most phylogenetically distant strains from C57BL/6, for which over 3,300 SNPs across the EDC locus have been identified between the strains, including 39 within EDC genes (Eppig et al. 2015). The informative SNPs will allow us to phase the haplotypes, thus identifying allele-specific interactions and transcription from the 4C-seq and RNA-seq, respectively. The proof-of-principle for this approach has been demonstrated in human cell-lines (Tang et al. 2015), and holds great potential for elucidating the role of chromatin architecture in human disease, an area still largely unexplored.

The regulation of the *Flg*-like gene family is also as an interesting line of further investigation. As discussed earlier, human EDC enhancer 184 overlaps with an annotated splice variant of non-coding RNA, *FLG-AS1-005* (Ensembl release 84) (Yates et al. 2016), as well as ENCODE annotated H3K27Ac active enhancer marks (Hoffman et al. 2013). Further, there is evidence that 184 forms chromatin interactions with a number of EDC genes which are aberrantly expressed in AD-patients (Quiggle, et al. unpublished data). This suggests the possibility that 184/*FLG-AS1-005* is an eRNA that stabilizes the loop with *Flg* to drive expression. Future experiments such as RNA-ChIP can be employed to test the hypothesis that the *FLG-AS1-005* transcript interacts with the chromatin loop involving 184 and *Flg*.

5.3 References

Dixon, JR, Selvaraj, S, Yue, F, *et al.* (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485: 376–80.

Eppig, JT, Blake, JA, Bult, CJ, *et al.* (2015). The Mouse Genome Database (MGD): Facilitating mouse as a model for human biology and disease. *Nucleic Acids Res* 43: D726–36.

Fanucchi, S, Shibayama, Y, Burd, S, *et al.* (2013). Chromosomal Contact Permits Transcription between Coregulated Genes. *Cell* 155: 606–20.

Fessing, MY, Mardaryev, AN, Gdula, MR, *et al.* (2011). p63 regulates Satb1 to control tissue-specific chromatin remodeling during development of the epidermis. *J Cell Biol* 194: 825–39.

Guo, Y, Xu, Q, Canzio, D, *et al.* (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* 162: 900–10.

de Guzman Strong, C, Conlan, S, Deming, CB, *et al.* (2010). A milieu of regulatory elements in the epidermal differentiation complex syntenic block: implications for atopic dermatitis and psoriasis. *Hum Mol Genet* 19: 1453–60.

Hadjur, S, Williams, LM, Ryan, NK, *et al.* (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* 460: 410–3.

Hemberg, M, Gray, JM, Cloonan, N, *et al.* (2012). Integrated genome analysis suggests that most conserved non-coding sequences are regulatory factor binding sites. *Nucleic Acids Res*.

Hoffman, MM, Ernst, J, Wilder, SP, *et al.* (2013). Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Res* 41: 827–41.

Hou, C, Dale, R, Dean, A (2010). Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc Natl Acad Sci U S A* 107: 3651–6.

Kwasnieski, JC, Fiore, C, Chaudhari, HG, *et al.* (2014). High-throughput functional testing of ENCODE segmentation predictions. *Genome Res* 24: 1595–602.

Lindblad-Toh, K, Garber, M, Zuk, O, *et al.* (2011). A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478: 476–82.

Lupiáñez, DG, Kraft, K, Heinrich, V, *et al.* (2015). Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* 161: 1012–25.

Mardaryev, AN, Gdula, MR, Yarker, JL, *et al.* (2013). p63 and Brg1 control developmentally regulated higher-order chromatin remodelling at the epidermal differentiation complex locus in epidermal progenitor cells. *Development* 141: 101–11.

Mehic, D, Bakiri, L, Ghannadan, M, *et al.* (2005). Fos and jun proteins are specifically expressed during differentiation of human keratinocytes. *J Invest Dermatol* 124: 212–20.

Mishiro, T, Ishihara, K, Hino, S, *et al.* (2009). Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *EMBO J* 28: 1234–45.

Oh, IY, Albea, DM, Goodwin, ZA, *et al.* (2014). Regulation of the Dynamic Chromatin Architecture of the Epidermal Differentiation Complex Is Mediated by a c-Jun/AP-1-Modulated Enhancer. *J Invest Dermatol* 134: 2371–80.

Ong, C-T, Corces, VG (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet* 12: 283–93.

Palmer, CNA, Irvine, AD, Terron-Kwiatkowski, A, *et al.* (2006). Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat Genet* 38: 441–6.

Pennacchio, LA, Ahituv, N, Moses, AM, *et al.* (2006). In vivo enhancer analysis of human conserved non-coding sequences. *Nature* 444: 499–502.

Phillips-Cremins, JE, Sauria, MEG, Sanyal, A, *et al.* (2013). Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. *Cell* 153: 1281–95.

Robyr, D, Friedli, M, Gehrig, C, *et al.* (2011). Chromosome conformation capture uncovers potential genome-wide interactions between human conserved non-coding sequences. *PLoS One* 6: e17634.

Rorke, EA, Adhikary, G, Jans, R, *et al.* (2010). AP1 factor inactivation in the suprabasal epidermis causes increased epidermal hyperproliferation and hyperkeratosis but reduced carcinogen-dependent tumor formation. *Oncogene* 29: 5873–82.

Sander, JD, Joung, JK (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* advance on.

Seitan, VC, Hao, B, Tachibana-Konwalski, K, *et al.* (2011). A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. *Nature* 476: 467–71.

Siepel, A, Bejerano, G, Pedersen, JS, *et al.* (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 15: 1034–50.

Tang, Z, Luo, OJ, Li, X, *et al.* (2015). CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* 163: 1611–27.

Yates, A, Akanni, W, Amode, MR, *et al.* (2016). Ensembl 2016. *Nucleic Acids Res* 44: D710–6.

Appendix A

Supplementary Material for Chapter 2

Table S1. List of Mouse 3C Primers

Forward primers

923 (anchor)	AGAAGGCAAGGGGAGGATAA
2310050C09Rik	GACACGGGTCCAACTACCTC
Crct1	GTTGGTATAAGGGGCACAGG
Crnn	AGGACTCCTTCCTTCCCTCA
Flg2	ATTAAAATAAGACCTGCTGGTAAACT
Hrnr	CCAGATGATCCAGACACCGTA
Kprp	CACCCCTCCTGCACATAAAT
Lce1a1	TGATTTTTGTGTCCGTCTTCC
Lce1a2	ATGTACCCAAATGCCCAGAG
Lce1c	GGACCCTGTCCTGAAATTCG
Lce1d	GAGACTCTGCTATCTACAGGTGACA
Lce1e	TAAGCCCTCTGACGAGCATC
Lce1f	TCATGATGAATTCTTTAACCTTTCT
Lce3b	CATGTCAATAGCTCTGTCATGTGT
Lce3b_5’	CTGGGCATGGTCACTCCTAT
Lor	TGGAAGAGATTTAGGGAAGGAA
Lor_5’	CCTCCCAGGAGACCCATTTA
Pglyrp3	TTTTCTGAAGTACACCCTTTTGA
Pglyrp4	AAGCAATTGGTTGCAGCTCT
Rptn	GACCACAGGGACTCAACGAT
S100a1	GTGGGCTACGATCTCCTCTG
S100a1_5'	TGTGGTCTTTGCTGCTTTTG

S100a10	TTGTCTGGAGCTACCCCATC
S100a13	CTCTTCACCTCCCGTTCAAA
S100a14	GCAGGAAGATCTGGAAGTGC
S100a16	CTGTGCTCTGCTCCGTGTT
S100a2	CCAAGTTGTTAACATACCCAGCTA
S100a4	GAGGCAGAGGCTGTCAGTTT
S100a6	GGGCGTGTCAAGAAGGAGT
S100a7a	CAAAACAGTGCCTTGCTTCA
S100a8	TGTATCATTTGGTGCTGGGTA
S100a9	CACAGGTTGTTGGCTCTTTTT
Sprr1a	TCCTTTGACTCTGTCCCAAAA
Sprr1b	GGGGTGTGTACTCCGTGTTC
Sprr2a1	CGTTCCCTTGCTATACTCCATC
Sprr2b	CAGGGGTGAACTACCAGGAG
Sprr2e	CACATGAACTAACTCAAGGTAGAGG
Sprr2f	CCTCAGTGTTCAGATCTGGGTAA
Sprr3	AAGGGTGAAATCCAGGATGA
Sprr4	CACACCTTTATGAAGAACCAGGA
Tchh	TGTTTTAATCTTTTCATAGAGACCACA
Tchhl1_3’	TGGCCAATCTTGAACTCTGA

Reverse primers

923 (anchor)	AATGGAAAGCACACCCAGAA
Flg	TGAGTTAGAAGTTGTGATCACTAGGAA
Ivl	GTCTGAGGTTCCTGCAATCC
Lce1b	GGAGGCATAAGCAGTCAGGA
Lce6a	TTGGGGATTGAGGAAAACAG

S100a11	GCCCAGTGCTAGGTTCAAA
Sprr2d	GAACCTAATGGCTCAGGCTCT
Sprr2g	AAGAATGGAATGAGGAATAGGAAAC
Tchhl1	TTTTTCGGAAAGGTCACAACT

Control primers

cutF	CACACCTTTATGAAGAACCAGGA
cutR	GGCTTCAGAAAACCTCCAGA
uncutF	GGGTTGCCATTCAATACCAC
uncutR	TCTGGGCTCCTCCTCTGTAA
ERCC3-1	GCCCTCCCTGAAAATAAGGA
ERCC3-2	GACTTCTCACCTGGGCCTACA

Table S2. Mouse RT-PCR Primers

Gene	Forward	Reverse
2310050C09 Rik	CCTGCCAGGAAAGAAGATTT	ACAGTTTGAGGCTGGTAGGG
Crct1	TTCTGCCTAGCAGGTGTCAA	AGCAAGAGGAGGAGGAGGAC
Crnn	AACAGAAACTTGTCCCATCCTG	TGACGTCAGCAAACTCATGTT
Flg	CTCCTTCAGCTGCATTCGAT	TGCCTGTAGTTGTCCTTCCA
Flg2	TGCGTCAGGCCTTATCCTAC	TCCTTCTCCAGCAGTTCCTT
Hrnr	GCAAGCAACATCAGTCTCCA	CAGAATTTGGTGAAACTCTGTTACC
Ivl	GCCTTCTCCCTCCTGTGAGT	ATGTTTGGGAAAGCCCTTCT
Kprp	GCTCAGGTCCCAATCCAGTA	CCTTGGTCTCCACAACCACT
Lce1a1	CTCACCTTCCGAGGTATCCA	AAGACACAGGGGGACACTTG

Lce1a2	GCCCAAGGATCTTGTACTGC	CCAGGCTACAGCAGGAAGAC
Lce1b	TCCTCCTGAAGTGGCTACAGA	CCAGGCTACAGCAGGAAGAC
Lce1c	ACTGGCTGAGAAACCCACAG	CCAGGCTACAGCAGGAAGAC
Lce1d	ACTTCTCCTGAGGCGTCCAC	CCAGGCTACAGCAGGAAGAC
Lce1e	CCACTCACTGGGTGAGATACC	AACCCAAGCTACAGCAGGAA
Lce1f	ATCCACGCCATTAACACTGA	CCAGGCTACAGCAGGAAGAC
Lce3b	AGCATCCTCAGACACGGACT	TGTAGCACAGCAGGAAGAGG
Lce3c	TCTTCTCCTGCCTTTGCTGT	TGCTGATTCCTCCAGACTGT
Lce6a	ATTTCTGGCCCCATAAAACC	GCCTGCTGAACAAGAAGTCC
Lor	GGTCACCGGGTTGCAACGGA	GAGACACTAGAATTGGGAGG
Pglyrp3	TTGGCTTCTTGGCTCTCAGT	CAATGTCACACCAGCCTTTG
Pglyrp4	ATGCTGTCCTGGCTTCTTGT	TCAGCTTAGAGCTGCAACCA
Rptn	CCTGCCTCTTCTGCTCATTC	CCGAAGGATGTCTCCAAACT
S100a1	CAGTGGCCACATTTGCAG	TTCAGTTCTTTCTTGCTCAGCTT
S100a10	CTGAGAGTGCTCATGGAACG	TCCCCTTCTGCTTCATGTTT
S100a11	CCACCGTCAGCCACAGTC	ATCTAGCTGCCCGTCACAGT
S100a13	CCTTGCCTGGTGCTTATAAACTT	CCCTGCAAAGGTGAAGAAAG
S100a14	GGCAGGCTATAGGACAGACG	CCTCAGCTCCGAGTAACAGG
S100a16a	GAGGAGGTGGACTCACAGGA	TCCAGGTTCTGGATGAGCTT
S100a2	CGAGAGGCTCAAACACAACA	GGGACCCCTCAAGGAAGTTA
S100a4	CCCAAACCTCTCTATTCAGCA	CTTTTCCCCAGGAAGCTAGG
S100a6	ATCCCCTCGACCACTCCTT	AGATCATCCATCAGCCTTGC
S100a7a	GGATAGTGTGCCTCGCTTCA	AGACTGCCTGTCCTCCCTCT
S100a8	CTGAGTGTCCTCAGTTTGTGC	GACCTGAGATATGATGACTTTATTCTG

S100a9	GAAGGAAGGACACCCTGACA	TCAACTTTGCCATCAGCATC
Sprr1a	GAGAACCTGCTCTTCTCTGAGTAT	GCAGGGCTCAGGAACAGG
Sprr1b	GCGACCACACTACCTGTCCT	GGTGTCACAGGGTGTCTTGA
Sprr2a1	TTCCTGGTACTCAAGCATTGG	CAAGGCTCAAAGCACATGAC
Sprr2b	TTCCTGGTACTCAAGCATTGG	CTTGGGTGGACACTTCTGCT
Sprr2d	CAAGGCCGAGACTACTTTGG	ATGGCTCAGGACAAGGCTCT
Sprr2e	TCAGGTCCTAGGCTACTTTGG	TATGAGGGAGGAGGACATGG
Sprr2f	CTGGTACACACGTCCTGGAA	GCATTTCTGCTGGAATGAGG
Sprr2g	CAGGTCCTACACTACGTTGGAG	CTGGCATGGAGAAGGAAGAC
Sprr3	CCAAGAACCCAGTGATCTTCA	TGTTTCCTGGTTGTGGAACA
Sprr4	TCCCATCAGCATCAGAATCA	TGCTGTGCAGGACACTTCTC
Tchh	TGATGGAGCATCACTTAGCAA	GGCCTGATCGAGAGCATAAT
Tchhl1	CACATTGCCCCACATTCC	CCCTCAAGGAGCTGTCTCAG
Krt1	GACACCACAACCCGGACCCAAAACTTA	ATACTGGGCCTTGACTTCCGAGATGATG
Krt14	GGCCACCTACCTGGACAAG	GTCGATCTGCAGGAGGACAT
B2m	TGGTGCTTGTCTCACTGACC	CCGTTCTTCAGCATTTGGAT

Table S3. RNA-seq results

Gene Name	**log_2(Diff/Prolif)**	**P-Value**	**FDR (False Discovery Rate)**
2310050C09Rik	5.96	2.81E-27	1.59E-25
Crct1	4.77	1.15E-71	6.30E-69
Crnn	3.58	2.63E-05	1.48E-04
Flg	2.61	5.27E-05	2.77E-04

Flg2	8.65	8.08E-17	2.11E-15
Hrnr	5.64	9.03E-19	2.81E-17
Ivl	1.79	2.25E-10	2.95E-09
Kprp	3.21	7.22E-21	2.77E-19
Lce1a1	5.80	2.14E-27	1.23E-25
Lce1a2	6.77	6.86E-24	3.16E-22
Lce1b	5.24	2.34E-28	1.45E-26
Lce1c	7.24	9.99E-40	1.17E-37
Lce1d	6.09	1.64E-32	1.34E-30
Lce1e	4.76	3.90E-23	1.73E-21
Lce1f	6.89	1.42E-44	2.08E-42
Lce3b	0.74	3.05E-01	4.44E-01
Lce3c	2.58	2.49E-11	3.70E-10
Lce6a	6.48	5.34E-14	1.06E-12
Lor	6.25	1.70E-63	6.56E-61
Pglyrp3	0.63	5.00E-03	1.52E-02
Pglyrp4	3.93	5.18E-17	1.38E-15
Rptn	3.67	1.31E-12	2.23E-11
S100a1	1.05	6.87E-03	1.98E-02
S100a10	0.92	2.14E-05	1.22E-04
S100a11	1.40	7.85E-16	1.84E-14
S100a13	0.94	7.09E-02	1.41E-01
S100a14	1.82	2.15E-25	1.10E-23
S100a16	1.75	2.20E-19	7.32E-18
S100a2	-0.82	2.19E-01	3.47E-01
S100a4	2.77	5.09E-11	7.25E-10

S100a6	1.67	4.93E-15	1.08E-13
S100a7a	2.19	3.36E-12	5.48E-11
S100a8	3.54	3.04E-18	9.03E-17
S100a9	3.36	8.02E-36	7.89E-34
Sprr1a	3.54	1.41E-26	7.69E-25
Sprr1b	2.59	3.50E-28	2.15E-26
Sprr2a1c	4.16	1.70E-04	7.95E-04
Sprr2b	4.27	2.53E-25	1.28E-23
Sprr2d	0.70	1.40E-02	3.66E-02
Sprr2e	-1.50	4.50E-05	2.41E-04
Sprr2f	0.32	6.18E-01	7.46E-01
Sprr2g	1.72	1.04E-09	1.26E-08
Sprr3	1.12	3.27E-03	1.05E-02
Sprr4	6.54	4.88E-18	1.42E-16
Tchh	2.22	2.08E-18	6.27E-17
Tchhl1	4.34	4.04E-16	9.77E-15

Table S4. ChIP Primers

Name	**Coordinates (NCBI37/mm 9)**	**Forward**	**Reverse**
Block 1 (923 block 1)	Chr3: 92,380,676-92380831	CCTTGTGATGAATCCAAGAAA	TCTGGTCATATTCATCCCTTCA
Krt5	Chr15:	AGACGTGTGTCTGCATC	TTTGATGCGGTGAGCAA

(ENCODE annotated site within keratin 5 [Krt5])	101,540,819-101,540975	TGG	TTA
Neg (No ENCODE annotated site)	Chr3: 92,382,019-92,382,161	TGCTTGCTACTGGGTCTCAA	CCCTCCAAGTCCTATCATGC

Supplemental Materials and Methods

RNA-seq. Total RNA was isolated by TriZol extraction (Life Technologies, Frederick, MD) from a parallel set of cells corresponding to cells used for the 3C libraries. Poly-A mRNA was enriched (Dynabeads mRNA purification kit) and fragmented prior to SuperScript III 1st-strand synthesis system (Life Technologies, Frederick, MD) and subsequently the NEBNext mRNA Second Strand Synthesis Module (New England Biolabs, Ipswich, MA). RNA-seq libraries were generated according to the manufacturer's instructions (Illumina, Hayward, CA) and sequenced using the Illumina HiSeq 2000.

A transcriptome index of the UCSC mm10 (GRCm38, Dec. 2011) mouse genome assembly was constructed using TopHat v2.0.5 (Trapnell et al. 2009). Short, 50 bp single-end RNA-seq (Mortazavi et al. 2008) reads were aligned to the mouse genome with TopHat (parameter set: --bowtie1 -a 5 -m 1 -i 10 -I 500000 -x 20 -n 2), allowing only one splice mismatch, two mismatches within a read, an average intron length of 500,000 bp, a minimum anchor length of 5

bp and a minimum exon length of 10 bp. Reads that aligned to the transcriptome more than 20 times were discarded. Alignments were converted to SAM format using the samtools package (Li et al. 2009). Aligned transcripts in the SAM file were counted using the htseq-count program (parameter set: --stranded=no –quiet). Differential gene expression was calculated on the transcript counts using the Bioconductor edgeR package version_3.2.3 (Robinson et al. 2010). Significance cutoffs for differentially expressed genes were at FDR<0.001 and p<0.05.

References

Li, H, Handsaker, B, Wysoker, A, *et al.* (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–9.

Mortazavi, A, Williams, BA, McCue, K, *et al.* (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5: 621–8.

Robinson, MD, McCarthy, DJ, Smyth, GK (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–40.

Trapnell, C, Pachter, L, Salzberg, SL (2009). TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* 25: 1105–11.

Appendix B

Supplementary Material for Chapter 3

Table S1. X^2 test for deviation from expected Mendelian ratio reveals normal viability of 923 KO mice (α=0.05, df=2).

923^{delA} F_2

Genotype	**Observed**	**Ratio**	**Expected**	**X^2 test**	
Het	43	0.597222222	36	1.361111111	
WT (-)	14	0.194444444	18	0.888888889	
Mut (+)	15	0.208333333	18	0.5	
				2.75	**X^2 statistic**

923^{delA} F_5

Genotype	**Observed**	**Ratio**	**Expected**	**X^2 test**	
Het	71	0.489655172	72.5	0.031034483	
WT (-)	32	0.220689655	36.25	0.498275862	
Mut (+)	42	0.289655172	36.25	0.912068966	
				1.44137931	**X^2 statistic**

923^{delB} F_2

Genotype	**Observed**	**Ratio**	**Expected**	**X^2 test**	
Het	84	0.482758621	87	0.103448276	
WT (-)	42	0.24137931	43.5	0.051724138	
Mut (+)	48	0.275862069	43.5	0.465517241	
				0.620689655	**X^2 statistic**

923^{delB} F_4

Genotype	Observed	Ratio	Expected	X^2 test	
Het	79	0.512987013	77	0.051948052	
WT (-)	28	0.181818182	38.5	2.863636364	
Mut (+)	47	0.305194805	38.5	1.876623377	
				4.792207792	**X^2 statistic**

923^{pdel} F_6

Genotype	Observed	Ratio	Expected	X^2 test	
Het	8	0.216216216	18.5	5.959459459	
WT (-)	11	0.297297297	9.25	0.331081081	
Mut (+)	18	0.486486486	9.25	8.277027027	
				14.56756757	**X^2 statistic**

Small Guide RNA (sgRNA) targeting sequences

Upstream 5' sgRNA

5'- GAATACATCCCAGGAACAT -3'

Downstream 3' sgRNA

5'- CAGTAAGCTAGCGCTAGAC -3'

ssODN sequences

Upstream 5' ssODN

5'-
AGAAGTTTTTCAGTTCCCCATAGTTGTCCTGAGGAGCATATAATCTTTGTCTTAAGCAGATTTG
TTTACAATAATTCCCTATAACTTCGTATAGCATACATTATACGAAGTTATGCATGCTTTAAAGA
GATAGAGGACTGACATGACCCTCTGTCCTCTAAAACAAGTTTGCCAGGATTTCTCCATTCCCAG
AGCCATGA-3'

Downstream 3' ssODN

5’-
TCTCTGTTGTTAGAGTCCATCTCCTACACCGATAGAGACTGATTCTGAAAAAAAAGGAAGCTCC
CACTGTCCAAGTTCTAAAGCTTATAACTTCGTATAGCATACATTATACGAAGTTATTGGAAACC
AGACACCCTGGCTGCTGCTCTGAAGGCAACTCTTCCCTATCAGGCTCCTTAATAGGATTTGATC
AGTGTGAC-3’

SphI restriction site

HindIII restriction site

Homology arms

LoxP site

Genotyping primers

Mutant line		**F primer**	**R primer**	**Mutant allele**	**Wildtype allele**
923delA		CAGTTCCCCATAGTTGTCCTG	GGAAGAGTTGCCTTCAGAGC	317	1407
923delB				164	
923pdel				579	
923^{fl}	5' loxP	TCTTTAGTGCTCAGTTAACAGCT	AGTCCTCTATCTCTTTAAAGCATGCATAAC	259	-
	3' loxP	GTTCTAAAGCTTATAACTTCGTATAGCA	TGCCTCACCAAATTCTCACA	195	-

qPCR primers

Gene	Forward	Reverse
Crnn	AACAGAAACTTGTCCCATCCTG	TGACGTCAGCAAACTCATGTT
Flg	CTCCTTCAGCTGCATTCGAT	TGCCTGTAGTTGTCCTTCCA
Flg2	TGCGTCAGGCCTTATCCTAC	TCCTTCTCCAGCAGTTCCTT
Hrnr	GCAAGCAACATCAGTCTCCA	CAGAATTTGGTGAAACTCTGTTACC
Ivl	GCCTTCTCCCTCCTGTGAGT	ATGTTTGGGAAAGCCCTTCT
Lor	GGTCACCGGGTTGCAACGGA	GAGACACTAGAATTGGGAGG
Rptn	CCTGCCTCTTCTGCTCATTC	CCGAAGGATGTCTCCAAACT
Sprr2d	CAAGGCCGAGACTACTTTGG	ATGGCTCAGGACAAGGCTCT
Sprr3	CCAAGAACCCAGTGATCTTCA	TGTTTCCTGGTTGTGGAACA
Evpl	TCAAGGGGCTGAGCAAAG	AGCTTCTTCTGCGTCTCCAA
Ppl	ATACAGCCCAACGGTGCAG	CAGCACGTACAGCAGCTTTT
B2m	TGGTGCTTGTCTCACTGACC	CCGTTCTTCAGCATTTGGAT

RNA-seq results: Differential Expression relative to wildtype

Gene	logFC het	logFC mut	adj.P.Val het	adj.P.Val mut
S100a1	2.01E-01	5.78E-01	9.66E-01	9.84E-01
S100a13	5.79E-01	-6.02E-02	9.66E-01	9.84E-01
S100a14	1.26E-01	-8.39E-02	9.66E-01	9.84E-01
S100a16	-7.12E-02	4.08E-02	9.66E-01	9.84E-01
S100a2	-1.31E+00	-2.26E+00	9.66E-01	7.67E-01
S100a3	2.01E+00	-1.14E-02	1.60E-01	9.92E-01
S100a4	5.92E-01	1.76E-02	9.66E-01	9.89E-01
S100a5	1.09E-01	-6.38E-02	9.66E-01	9.84E-01
S100a6	5.35E-01	-5.49E-02	9.66E-01	9.84E-01
S100a7a	1.49E+00	7.49E-01	9.66E-01	9.84E-01
S100a8	1.52E+00	4.86E-01	9.66E-01	9.84E-01
S100a9	4.16E-01	1.05E+00	9.66E-01	9.84E-01

Pglyrp4	-2.24E-01	3.47E-01	9.66E-01	9.84E-01
Pglyrp3	7.35E-02	2.41E-01	9.66E-01	9.84E-01
Lor	-2.42E-01	-2.75E-02	9.66E-01	9.84E-01
Prr9	6.85E-01	-8.96E-01	9.66E-01	9.84E-01
Lelp1	6.74E-01	-7.17E-01	9.66E-01	9.84E-01
Sprr2a1	1.09E-01	-6.38E-02	9.66E-01	9.84E-01
Sprr2a1	1.09E-01	-6.38E-02	9.66E-01	9.84E-01
Sprr2a2	1.09E-01	8.27E-01	9.66E-01	8.44E-01
Sprr2a2	1.09E-01	8.27E-01	9.66E-01	8.44E-01
Sprr2a3	-6.55E-01	3.02E-01	9.66E-01	9.84E-01
Sprr2b	-7.32E-01	1.43E+00	9.66E-01	9.10E-01
Sprr2d	-7.78E-01	1.17E+00	9.66E-01	7.17E-01
Sprr2e	-2.66E-01	1.21E+00	9.66E-01	9.84E-01
Sprr2f	-1.39E+00	-1.55E-02	9.66E-01	9.95E-01
Sprr2g	1.24E-01	1.71E+00	9.68E-01	9.84E-01
Sprr2h	-5.80E-01	1.57E-01	9.66E-01	9.84E-01
Sprr2i	2.64E-01	1.20E+00	9.66E-01	9.84E-01
Sprr2j-ps	-1.09E-01	1.03E+00	9.66E-01	9.84E-01
Sprr2k	-5.84E-01	1.81E-01	9.66E-01	9.84E-01
Sprr1b	-5.04E-01	1.06E+00	9.66E-01	9.84E-01
Sprr3	-3.22E+00	-1.23E+00	9.04E-01	9.84E-01
Sprr1a	-1.53E-01	4.58E-01	9.66E-01	9.84E-01
Sprr4	9.98E-01	6.94E-01	9.66E-01	9.84E-01
Ivl	-8.40E-01	-4.93E+00	5.75E-01	4.58E-04
Smcp	-9.23E-01	-2.95E+00	9.66E-01	2.80E-01
Lce6a	-3.07E-01	-1.71E+00	9.66E-01	9.84E-01
Lce1a1	-2.33E-01	-8.65E-02	9.66E-01	9.84E-01
Lce1b	-6.27E-02	-9.21E-03	9.66E-01	9.90E-01
Lce1a2	-1.15E-01	4.03E-02	9.66E-01	9.84E-01
Lce1c	-1.70E-01	-4.23E-02	9.66E-01	9.84E-01
Lce1e	-2.34E-01	8.44E-02	9.66E-01	9.84E-01
Lce1f	-4.11E-01	3.62E-02	9.66E-01	9.84E-01
Lce1g	-1.88E-01	1.98E-01	9.66E-01	9.84E-01
Lce1h	-8.97E-02	4.04E-01	9.66E-01	9.84E-01
Lce1i	-8.13E-03	2.59E-01	9.94E-01	9.84E-01
Lce1j	-3.16E-02	6.84E-01	9.82E-01	9.84E-01
Lce1k	1.59E-01	3.93E-01	9.66E-01	9.84E-01
Kprp	-6.75E-01	-1.56E-01	6.63E-01	9.84E-01
Lce1l	-1.59E-01	4.21E-01	9.66E-01	9.84E-01
Lce3a	-2.08E+00	-1.59E+00	9.66E-01	9.84E-01
Lce3b	-2.33E+00	-5.95E-01	8.92E-01	9.84E-01

Lce3c	-4.84E-01	-9.96E-02	9.66E-01	9.84E-01
Lce3d	-5.77E-01	5.65E-01	9.66E-01	9.84E-01
Lce3e	-8.12E-01	-1.51E+00	9.66E-01	6.78E-01
Lce3f	-2.19E+00	-2.48E+00	9.66E-01	9.84E-01
Crct1	5.15E-02	-4.66E-02	9.66E-01	9.84E-01
Lce1m	3.96E-02	-1.16E-01	9.67E-01	9.84E-01
Crnn	-6.76E-01	-1.73E+00	8.45E-01	3.07E-02
Flg2	3.03E-01	4.76E-01	9.66E-01	9.11E-01
Hrnr	-8.69E-02	1.45E-01	9.66E-01	9.84E-01
Rptn	-3.00E-01	1.35E+00	9.66E-01	9.84E-01
Tchh	2.66E-01	-9.53E-02	9.66E-01	9.84E-01
Tchhl1	8.49E-01	-4.92E-01	9.66E-01	9.84E-01
S100a11	-2.21E-01	-3.22E-01	9.66E-01	9.84E-01
S100a10	-6.16E-02	-1.47E-01	9.66E-01	9.84E-01
Evpl	-1.48E+00	-1.33E-01	-4.90E-02	9.84E-01
Ppl	2.66E-01	-9.53E-02	1.16E-01	9.84E-01

Appendix C

Supplementary Material for Chapter 4

Table S1. Custom Primers for 4C-seq

Nested Inverse PCR Guide Primers	
Name	
923 Read HindIII	AATGATACGGCGACCACCGAGATCTGTCAGATCTCAACATTCCTG TCAAA
923 Index 1 NlaIII	CAAGCAGAAGACGGCATACGAGATAACCTCAACAACTTCCTAATA TTTTAATGTCAAACAA
923 Index 2 NlaIII	CAAGCAGAAGACGGCATACGAGATTCTAAGCACAACTTCCTAATA TTTTAATGTCAAACAA
923 Index 3 NlaIII	CAAGCAGAAGACGGCATACGAGATCTGTCATACAACTTCCTAATA TTTTAATGTCAAACAA
923 Index 4 NlaIII	CAAGCAGAAGACGGCATACGAGATGGAGGTGACAACTTCCTAATA TTTTAATGTCAAACAA
923 Index 5 NlaIII	CAAGCAGAAGACGGCATACGAGATGCTCGATACAACTTCCTAATA TTTTAATGTCAAACAA
923 Index 6 NlaIII	CAAGCAGAAGACGGCATACGAGATTAGAGTAACAACTTCCTAATA TTTTAATGTCAAACAA
923 Index 7 NlaIII	CAAGCAGAAGACGGCATACGAGATTCAGTCTACAACTTCCTAATA TTTTAATGTCAAACAA
923 Index 8 NlaIII	CAAGCAGAAGACGGCATACGAGATTTCCAAGACAACTTCCTAATA TTTTAATGTCAAACAA
923 Index 9 NlaIII	CAAGCAGAAGACGGCATACGAGATTAATCGGACAACTTCCTAATA TTTTAATGTCAAACAA
Flg Read HindIII	AATGATACGGCGACCACCGAGATCTTGAGTTAGAAGTTGTGATCA CTAGGAA
Flg Index 10 NlaIII	CAAGCAGAAGACGGCATACGAGATCGCTGCCAAATTGGAAGCTAC AAAAACATAGGTA
Flg Index 11 NlaIII	CAAGCAGAAGACGGCATACGAGATATGATGGAAATTGGAAGCTAC AAAAACATAGGTA
Flg Index 12 NlaIII	CAAGCAGAAGACGGCATACGAGATCTTGTTAAAATTGGAAGCTAC AAAAACATAGGTA
Flg Index 13 NlaIII	CAAGCAGAAGACGGCATACGAGATACGCCTCAAATTGGAAGCTAC AAAAACATAGGTA
Flg Index 14 NlaIII	CAAGCAGAAGACGGCATACGAGATAGTTAAAAAATTGGAAGCTAC AAAAACATAGGTA
Flg Index 15 NlaIII	CAAGCAGAAGACGGCATACGAGATGAGGACCAAATTGGAAGCTAC AAAAACATAGGTA
Flg Index 16 NlaIII	CAAGCAGAAGACGGCATACGAGATGCCACCGAAATTGGAAGCTAC AAAAACATAGGTA
Flg Index 17 NlaIII	CAAGCAGAAGACGGCATACGAGATCGACAGTAAATTGGAAGCTAC AAAAACATAGGTA

Flg Index 18 NlaIII	CAAGCAGAAGACGGCATACGAGATCAAATACAAATTGGAAGCTAC AAAAACATAGGTA
Eβ Read HindIII	AATGATACGGCGACCACCGAGATCTAGTGGTAGGAATTGTTAGGA AAAGAAG
Eβ Index 19 NlaIII	CAAGCAGAAGACGGCATACGAGATCGTACTCATTTCAGCTCTCAT CTATGAATGTAAGAGT
Eβ Index 20 NlaIII	CAAGCAGAAGACGGCATACGAGATTATCTGTATTTCAGCTCTCAT CTATGAATGTAAGAGT
Eβ Index 21 NlaIII	CAAGCAGAAGACGGCATACGAGATCATTGAGATTTCAGCTCTCAT CTATGAATGTAAGAGT
Sequencing Primers	
Name	
923 Read	GTCAGATCTCAACATTCCTGTCAAAGCTT
923 Index Read	TTGTTTGACATTAAAATATTAGGAAGTTGT
Flg Read	AGAAGTTGTGATCACTAGGAATACAAGCTT
Flg Index Read	TACCTATGTTTTTGTAGCTTCCAATTT
Eβ Read	AGTGGTAGGAATTGTTAGGAAAAGAAGCTT
Eβ Index Read	ACTCTTACATTCATAGATGAGAGCTGAAAT

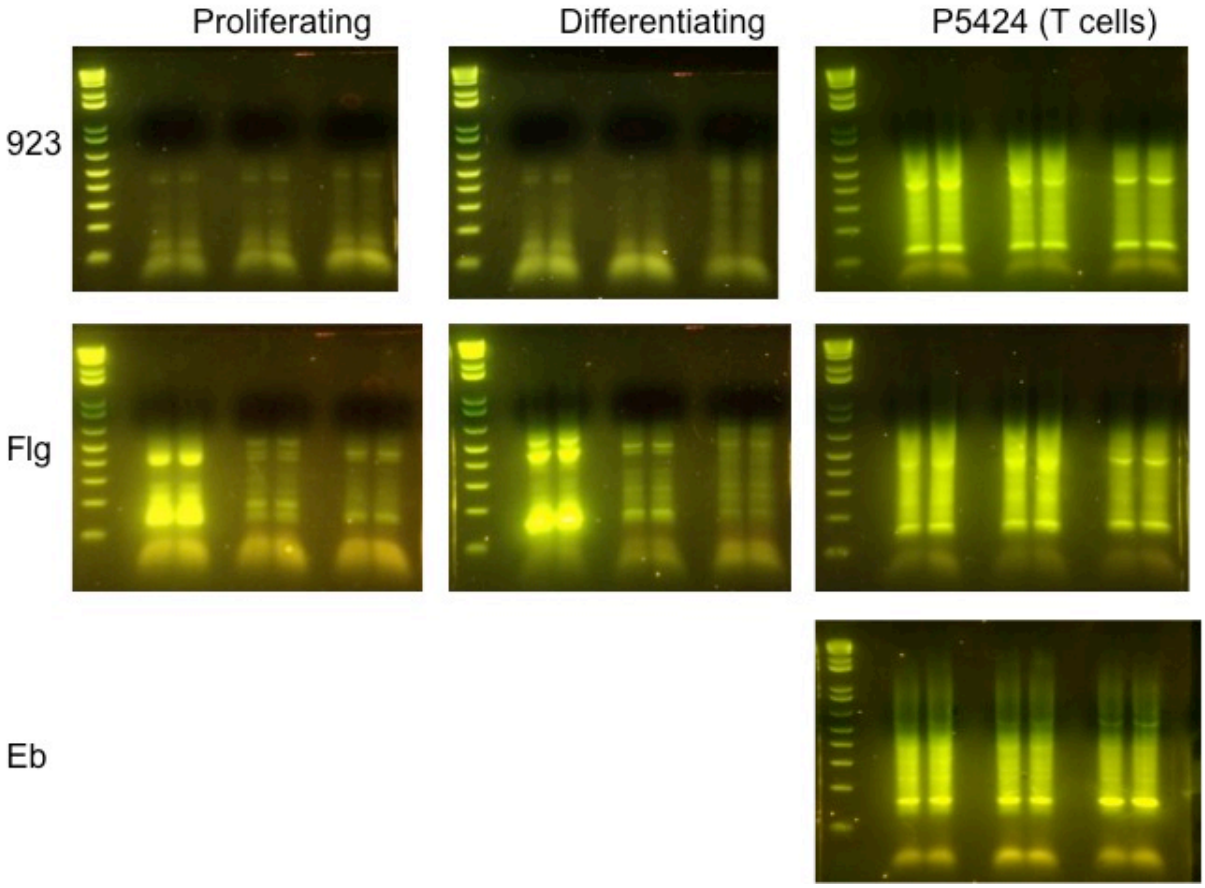


Figure S1. 4C-seq nested PCR products. The nested products were separated on a 1.5% agarose gel. The leftmost lane contains a 1 Kb Plus DNA Ladder (Invitrogen). The smears and discrete bands indicate that the 4C library contains amplicons of various sizes and quantities, representing a variety of interacting sequences and interaction frequencies. PCR products larger than ~120bp were cut out and submitted for sequencing to exclude primer dimers.

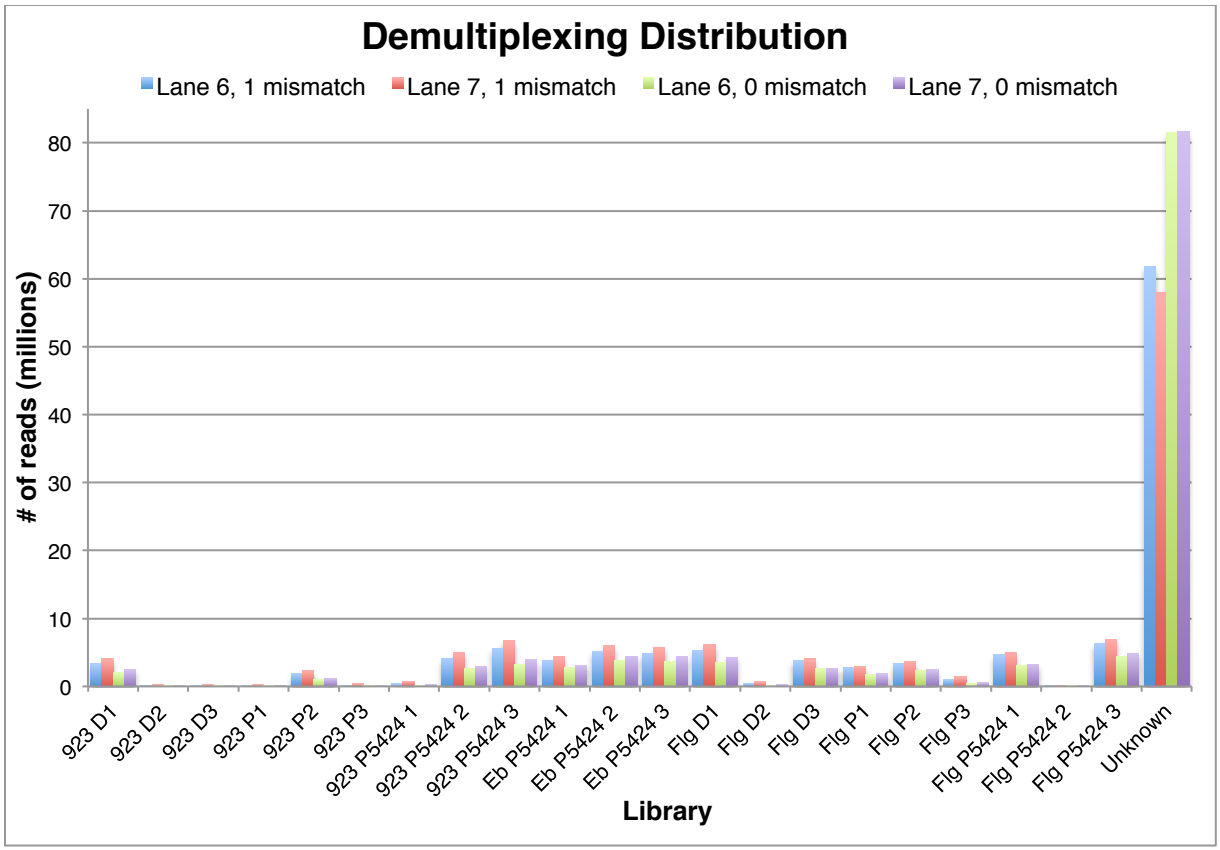


Figure S2. Demultiplexing distribution allowing for 1 or 0 mismatches in the index. Allowing for 1 mismatch in the index sequences resulted in a range of 47,000 reads – 6.2 million reads per library in lane 6, and 43,000 reads – 6.9 million reads per library in lane 7. When mismatches were not allowed, only 5315 reads - 4.4 million reads were obtained per library in lane 6, 6535 reads - 4.8 million reads per library in lane 7. Approximately 60 million reads were unassignable even allowing for 1 mismatch, while approximately 80 million reads were unassigned when no mismatches were allowed.

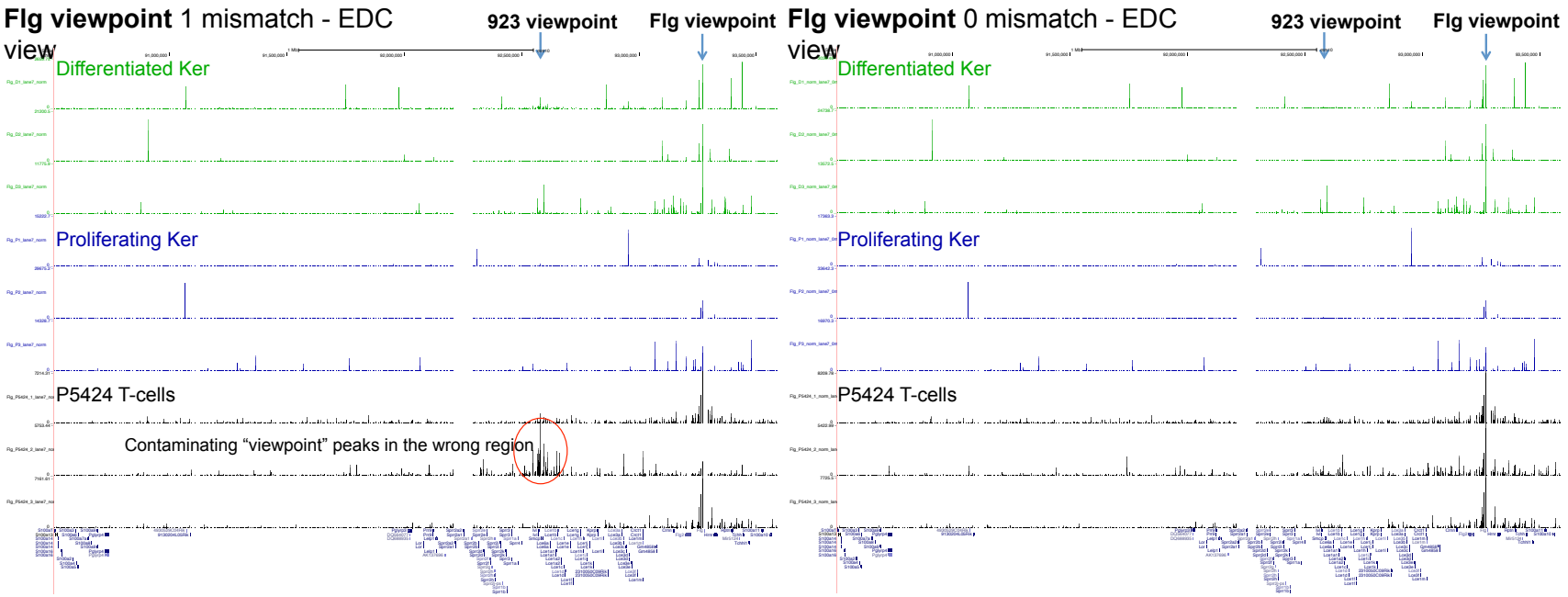


Figure S3. View of the EDC in Flg 4C-seq libraries demultiplexed by allowing 1 mismatch (left) and 0 mismatches (right) in index sequences. Different colored tracks indicate different cell types or conditions (green: differentiated keratinocytes, blue: proliferating keratinocytes, black: P5424 T-cells). A viewpoint-like peak is observed at 923 in one of the P5424 *Flg* libraries generated by allowing 1 mismatch in the index. This peak is not present in the same library generated by allowing 0 mismatch in the index, indicating that contamination between libraries was a result of mismatches allowed during demultiplexing.

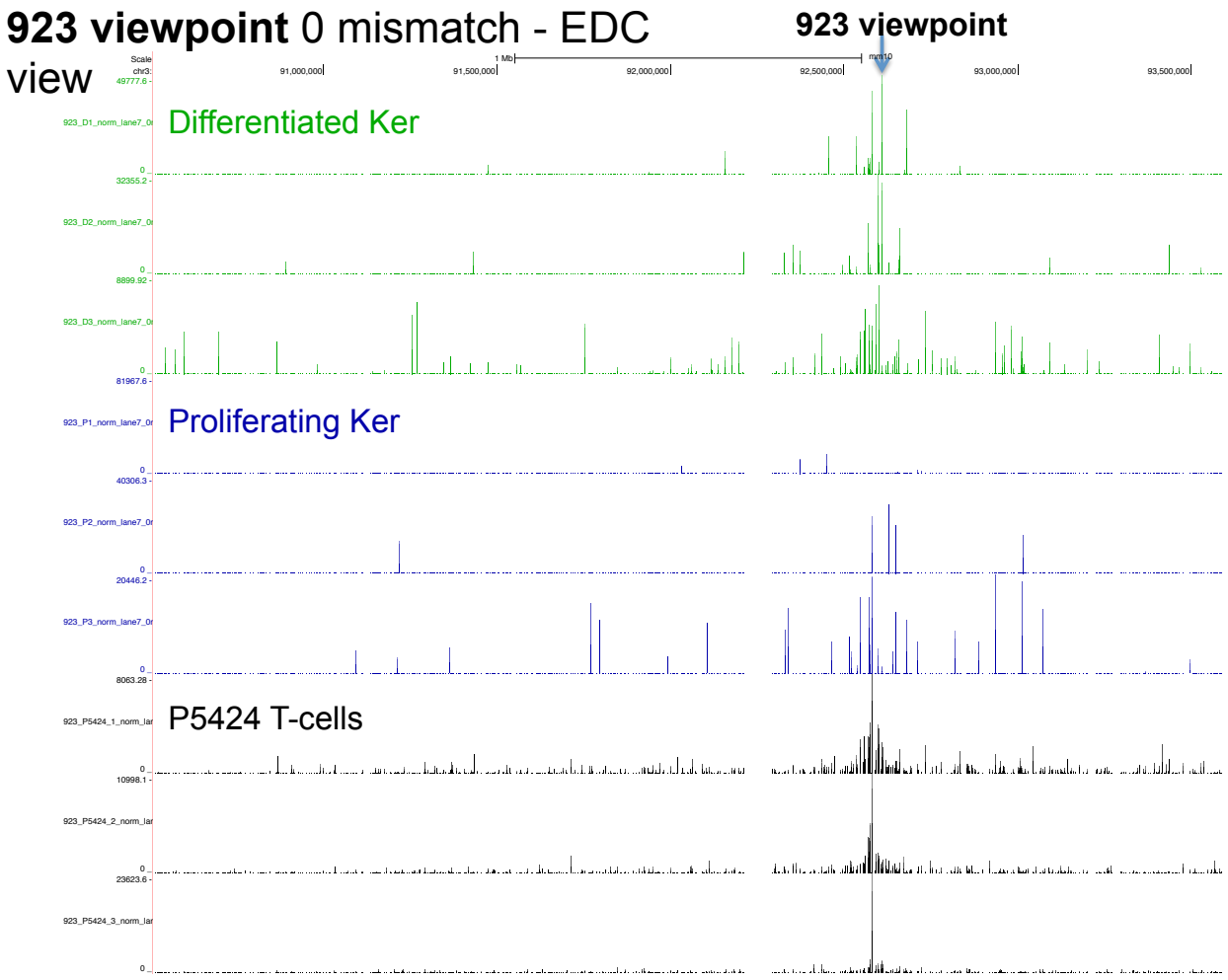


Figure S4. View of the EDC in 923 4C-seq libraries demultiplexed by allowing 0 mismatches in index sequences. Different colored tracks indicate different cell types or conditions (green: differentiated keratinocytes, blue: proliferating keratinocytes, black: P5424 T-cells).

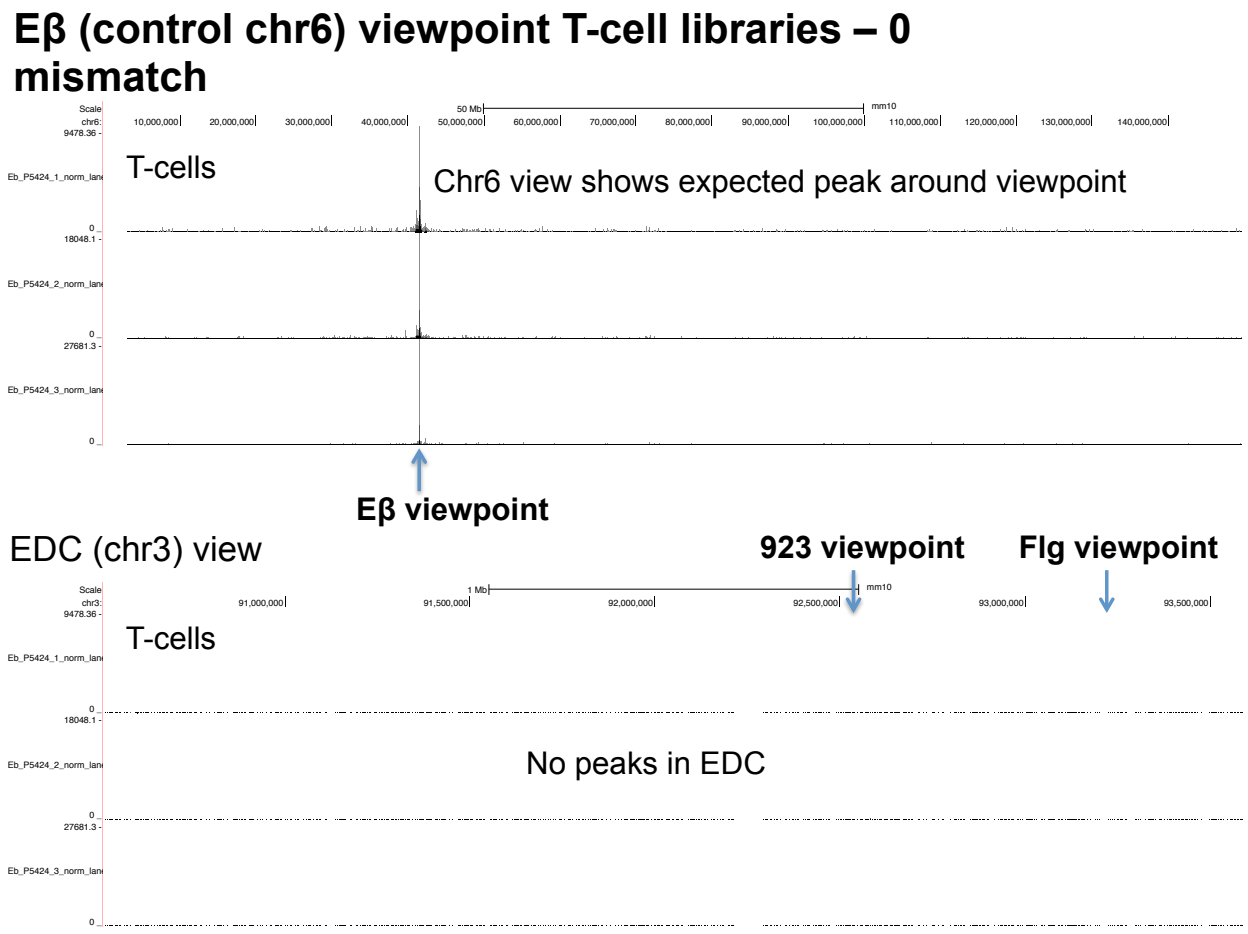


Figure S5. View of the EDC in Eβ 4C-seq libraries demultiplexed by allowing 0 mismatches (bottom) in index sequences. Different colored tracks indicate different cell types or conditions (green: differentiated keratinocytes, blue: proliferating keratinocytes, black: P5424 T-cells).

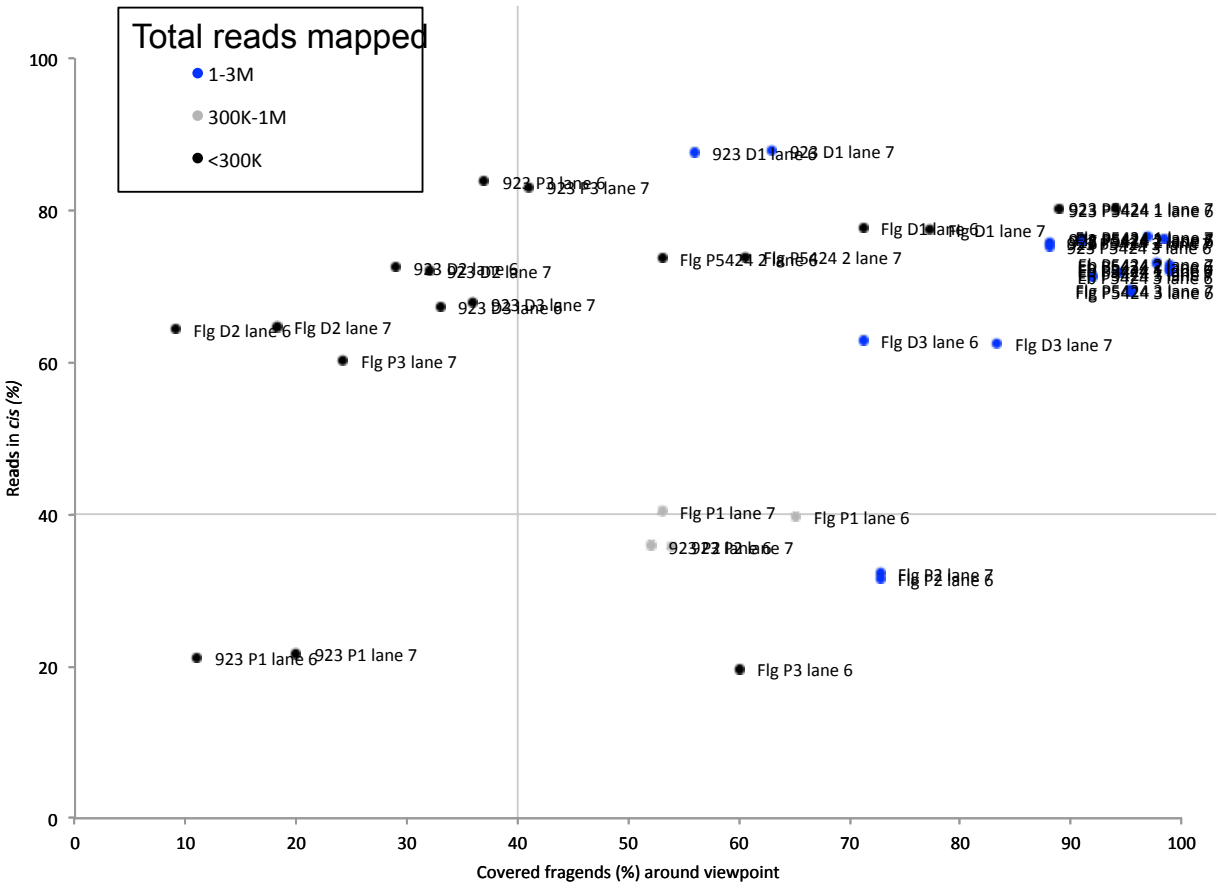


Figure S6. 4C-seq library quality. Libraries were assessed based on the criteria of having at least 1 million total reads mapping to fragments (blue), between 300,000 – 1 million reads (grey), or less that 300,000 reads (black), and the number of reads mapping in cis and near the viewpoint/bait. Good quality libraries have over 1 million reads (blue), over 40% of reads mapping in cis, and over 40% of the fragments near the viewpoint having reads mapping to them (upper right quadrant).

Table S2. 4C-seq library quality.

Lane 6 Library quality		Lane 7 Library quality	
Good	**Poor**	**Good**	**Poor**
923 D1	923 D2	923 D1	923 D2
923 P5424 2	923 D3	923 P5424 2	923 D3
923 P5424 3	923 P1	923 P5424 3	923 P1
Eβ P5424 1	923 P2	Eβ P5424 1	923 P2
Eβ P5424 2	923 P3	Eβ P5424 2	923 P3
Eβ P5424 3	923 P5424 1	Eβ P5424 3	923 P5424 1
Flg D3	Flg D1	Flg D3	Flg D1
Flg P5424 1	Flg D2	Flg P5424 1	Flg D2
Flg P5424 3	Flg P1	Flg P5424 3	Flg P1
	Flg P2		Flg P2
	Flg P3		Flg P3
	Flg P5424 2		Flg P5424 2

Table S3. GREAT-predicted cis-regulatory targets of 4C-seq reads enriched in 923 enhancer viewpoint differentiating keratinocytes (923 KerD) relative to T-cells.

ID	Desc	BinomFdrQ	Genes
GO:0031424	keratinization	2.60E-14	**Ivl,Sprr1b,Sprr2a1,Sprr2a2,Sprr2a3,Sprr2i,Sprr3,Sprr4**
GO:0001533	cornified envelope	2.67E-14	**Ivl,Sprr1b,Sprr2a1,Sprr2a2,Sprr2a3,Sprr2i**
GO:0030216	keratinocyte differentiation	3.35E-05	**Ivl,Sprr1b,Sprr2a1,Sprr2a2,Sprr2a3,Sprr2i,Sprr3,Sprr4,**Yap1
GO:0042834	peptidoglycan binding	2.14E-03	**Pglyrp3**,Tlr2
GO:0016019	peptidoglycan receptor activity	4.07E-03	**Pglyrp3**
GO:0008745	N-acetylmuramoyl-L-alanine amidase activity	4.07E-03	**Pglyrp3**
GO:0016045	detection of bacterium	4.16E-03	Naip6,**Pglyrp3**,Tlr2
GO:0006027	glycosaminoglycan catabolic process	4.93E-03	Hyal4,**Pglyrp3**
GO:0009253	peptidoglycan catabolic process	5.26E-03	**Pglyrp3**
GO:0044117	growth of symbiont in host	5.87E-03	**Pglyrp3**
GO:0032827	negative regulation of natural killer cell differentiation involved in immune response	5.87E-03	**Pglyrp3**
GO:0009595	detection of biotic stimulus	2.93E-02	Naip6,**Pglyrp3**,Tlr2,Tlr4
GO:0006026	aminoglycan catabolic process	3.25E-02	Hyal4,**Pglyrp3**
GO:0008329	signaling pattern recognition receptor activity	3.30E-02	**Pglyrp3**,Tlr2,Tlr4

Table S4. GREAT-predicted cis-regulatory targets of 4C-seq reads enriched in 923 enhancer viewpoint proliferating keratinocytes (923 KerP) relative to T-cells.

ID	Desc	BinomFdrQ	Genes
GO:0016503	pheromone receptor activity	1.35E-03	Gm4133,Gm4187,V1ra8,Vmn1r119,Vmn1r13,Vmn1r14,Vmn1r15,Vmn1r16,Vmn1r165,Vmn1r17,Vmn1r18,Vmn1r185,Vmn1r2,Vmn1r234,Vmn1r235,Vmn1r238,Vmn1r26,Vmn1r27,Vmn1r29,Vmn1r3,Vmn1r30,Vmn1r39,Vmn1r41,Vmn1r42,Vmn1r53,Vmn1r57,Vmn1r71,Vmn1r77,Vmn1r78,Vmn1r88,Vmn1r89

Table S5. GREAT-predicted cis-regulatory targets of 4C-seq reads enriched in *Flg* promoter viewpoint differentiating keratinocytes (Flg KerD) relative to T-cells.

ID	Desc	BinomFdrQ	Genes
GO:0001533	cornified envelope	4.89E-18	Anxa1,**Flg,Hrnr,Ivl,Lor,Rptn,Sprr1b,Sprr2a1,Sprr2a2,Sprr2a3,Sprr2f,Sprr2g,Sprr2h,Sprr2i**
GO:0016019	peptidoglycan receptor activity	2.87E-14	**Pglyrp3,Pglyrp4**
GO:0008745	N-acetylmuramoyl-L-alanine amidase activity	2.87E-14	**Pglyrp3,Pglyrp4**
GO:0044117	growth of symbiont in host	8.29E-14	**Pglyrp3,Pglyrp4**
GO:0032827	negative regulation of natural killer cell differentiation involved in immune response	8.29E-14	**Pglyrp3,Pglyrp4**
GO:0009253	peptidoglycan catabolic process	9.35E-14	**Pglyrp3,Pglyrp4**
GO:0031424	keratinization	3.83E-13	Abca12,**Hrnr,Ivl,Lor,Sprr1b,Sprr2a1,Sprr2a2,Sprr2a3,Sprr2f,Sprr2g,Sprr2h,Sprr2i,Sprr3,Sprr4**
GO:0042834	peptidoglycan binding	4.20E-13	**Pglyrp3,Pglyrp4**
GO:0006026	aminoglycan catabolic process	1.90E-12	BC051070,Chi3l3,Chi3l4,Chi3l7,Chia,**Pglyrp3,Pglyrp4**
GO:0006027	glycosaminoglycan catabolic process	2.28E-11	**Pglyrp3,Pglyrp4**
GO:0016045	detection of bacterium	6.80E-10	**Pglyrp3,Pglyrp4**
GO:0032823	regulation of natural killer cell differentiation	1.83E-08	**Pglyrp3,Pglyrp4**
GO:0008329	signaling pattern recognition receptor activity	4.90E-08	Colec12,**Pglyrp3,Pglyrp4**
GO:0032689	negative regulation of interferon-gamma production	4.72E-06	Gata3,**Pglyrp3,Pglyrp4**
GO:0032814	regulation of natural killer cell activation	5.76E-06	Il12b,Lyst,**Pglyrp3,Pglyrp4**
GO:0009595	detection of biotic stimulus	8.61E-06	**Pglyrp3,Pglyrp4**
GO:0016811	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amides	1.48E-05	Acer2,Acy1,Aga,Gls,Hdac2,Hdac9,Ngly1,**Pglyrp3,Pglyrp4**
GO:0030216	keratinocyte differentiation	2.29E-05	Abca12,Anxa1,Ctnnd1,**Hrnr,Ivl,Lor,Sprr1b**,**Sprr2a1,Sprr2a2,Sprr2a3,Sprr2f,Sprr2g,Sprr2h,Sprr2i,Sprr3,Sprr4**,Trp63

GO:0002698	negative regulation of immune effector process	1.64E-04	Bcl6,Irak3,Itch,**Pglyrp3,Pglyrp4**,Tgfb2
GO:0045620	negative regulation of lymphocyte differentiation	3.37E-04	Bcl6,Gli3,**Pglyrp3,Pglyrp4**
GO:0009913	epidermal cell differentiation	2.45E-03	Abca12,Anxa1,Atoh1,Clic5,Clrn1,Ctnnd1,Ercc3,Gli2,Grxcr1,Hdac2,**Hrnr,Ivl,Lor**,Myo7a,Pcdh15,Slitrk6,**Sprr1b,Sprr2a1,Sprr2a2,Sprr2a3,Sprr2f,Sprr2g,Sprr2h,Sprr2i,Sprr3,Sprr4**,Trp63
GO:0001818	negative regulation of cytokine production	3.41E-03	Bcl6,Chrna7,Gata3,Gata6,Ido1,Ifng,Il12b,Il6ra,Irak3,Irg1,**Pglyrp3,Pglyrp4**,Tgfb2,Trim27,Vsig4
GO:0050866	negative regulation of cell activation	3.91E-03	Bcl6,Cnr1,Gli3,Ido1,Itch,Pawr,Pde5a,Pdgfa,Pdgfb,**Pglyrp3,Pglyrp4**,Prdm1,Serpine2,Socs6,Sox11,Vsig4,Vtcn1
GO:0016810	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds	3.93E-03	Acer2,Acy1,Adar,Aga,Apobec4,Atic,Gls,Hdac2,Hdac9,Ngly1,**Pglyrp3,Pglyrp4**
GO:0050830	defense response to Gram-positive bacterium	4.46E-03	Il6ra,**Pglyrp3,Pglyrp4**
GO:0071229	cellular response to acid	5.37E-03	Col1a1,Col1a2,Col4a6,Ipo5,Pdgfc,Pdgfd,**S100a10**,Sh3bp4
GO:0005391	sodium:potassium-exchanging ATPase activity	5.39E-03	Atp1a1,Atp1a2,Atp1a4,Atp1b1
GO:0051250	negative regulation of lymphocyte activation	6.60E-03	Bcl6,Gli3,Ido1,Itch,Pawr,Pde5a,**Pglyrp3**,**Pglyrp4**,Prdm1,Socs6,Sox11,Vsig4,Vtcn1
GO:0008556	potassium-transporting ATPase activity	1.17E-02	Atp1a1,Atp1a2,Atp1a4,Atp1b1
GO:0008544	epidermis development	1.68E-02	Abca12,Anxa1,Atoh1,Barx2,Clic5,Clrn1,Ctnnd1,Ercc3,**Flg**,Foxq1,Gli2,Grxcr1,Hdac2,**Hrnr**,Igf1r,Inhba,**Ivl**,**Lce1g,Lce1i,Lce1l,Lce1m,Lce3a,Lce3c,Lce3f,Lor**,Myo7a,Pcdh15,Pdgfa,Pou3f2,Slitrk6,**Sprr1b,Sprr2a1,Sprr2a2,Sprr2a3,Sprr2f,Sprr2g,Sprr2h,Sprr2i,Sprr3,Sprr4**,Tgfb2,Trp63
GO:0002695	negative regulation of leukocyte activation	1.73E-02	Bcl6,Cnr1,Gli3,Ido1,Itch,Pawr,Pde5a,**Pglyrp3,Pglyrp4**,Prdm1,Socs6,Sox11,Vsig4,Vtcn1
GO:0009408	response to heat	2.12E-02	Casq1,Cetn1,Crnn,Psip1,Xylt1
GO:0043588	skin development	2.18E-02	Abca12,Anxa1,Arrdc3,Atoh1,Barx2,Clic5,Clrn1,Col1a1,Col1a2,Ctnnd1,Ercc3,**Flg**,Foxq1,Gli2,Grxcr1,Hdac2,**Hrnr**,Igf1r,Inhba,**Ivl,Lce1g,Lce1i,Lce1l,Lce1m,Lce3a,Lce3c,Lce3f,Lor**,Myo7a,Pcdh15,Pdgfa,Pou3f2,Slitrk5,Slitrk6,**Sprr1b,Sprr2a1,Sprr2a2,Sprr2a3,Sprr2f,Sprr2g,Sprr2h,Sprr2i,Sprr3,Sprr4,**Tcf7l2,Tgfb2,Trp63
GO:0004013	adenosylhomocysteinase activity	2.29E-02	Ahcy,Ahcyl1,Gm4737
GO:0005509	calcium ion binding	3.77E-02	**9130204L05Rik**,Adam8,Amy2a5,Anxa1,Bglap,Calb1,Casq1,Cdh10,Cdh12,Cdh18,Cdh19,Cdh2,Cdh20,Cdh6,Cdh7,Cdh9,Cetn1,Clstn3,**Crnn**,Dmd,Dnahc7b,Dsc3,Edem1,Edil3,Efemp1,Egfem1,Egfl6,Eltd1,Epdr1,Fat4,Fbln5,Fbn2,**Flg**,**Flg2**,Fscb,Fstl5,Gm5849,Gpd2,Grm7,**Hrnr**,Kcnip4,Lrp1b,Man1a,Man1a2,Mctp2,Notch2,Pcdh10,Pcdh15,Pcdh17,Pcdh18,Pcdh19,Pcdh20,Pcdhb22,Pclo,Plcl2,Pls3,**Rptn,S100a10,S100a11,S100a3**,Slc25a13,Slc25a24,Slit2,Sulf1,Syt1,**Tchhl1**,Wdr49
GO:0030280	structural constituent of epidermis	3.82E-02	**Lor**

GO:0016160	amylase activity	3.97E-02	Amy2a5,Mgam

Table S6. GREAT-predicted cis-regulatory targets of 4C-seq reads enriched in *Flg* promoter viewpoint proliferating keratinocytes (Flg KerP) relative to T-cells.

ID	Desc	BinomFdrQ	Genes
GO:0009593	detection of chemical stimulus	6.73E-04	Gm10081,Gnat3,Kcnmb2,Olfr1087,Olfr1089,Olfr109,Olfr1111,Olfr1112,Olfr1257,Olfr13,Olfr1326-ps1,Olfr135,Olfr138,Olfr1431,Olfr146,Olfr1505,Olfr17,Olfr191,Olfr193,Olfr195,Olfr196,Olfr199,Olfr201,Olfr229,Olfr292,Olfr356,Olfr357,Olfr366,Olfr368,Olfr38,Olfr389,Olfr435,Olfr453,Olfr48,Olfr714,Olfr715,Olfr724,Olfr725,Olfr736,Olfr738,Olfr743,Olfr744,Olfr76,Olfr794,Olfr796,Olfr855,Olfr890,Olfr891,Olfr936,Olfr945,Olfr948,Olfr972,Olfr974,Rtp4,Tlr2,Tlr4,Trpa1
GO:0050906	detection of stimulus involved in sensory perception	9.36E-04	Col11a1,Gm10081,Gnat3,Gpr98,Myc,Olfr1087,Olfr1089,Olfr109,Olfr1111,Olfr1112,Olfr1257,Olfr13,Olfr1326-ps1,Olfr135,Olfr138,Olfr1431,Olfr146,Olfr1505,Olfr17,Olfr191,Olfr193,Olfr195,Olfr196,Olfr199,Olfr201,Olfr229,Olfr292,Olfr356,Olfr357,Olfr366,Olfr368,Olfr38,Olfr389,Olfr435,Olfr453,Olfr48,Olfr714,Olfr715,Olfr724,Olfr725,Olfr736,Olfr738,Olfr743,Olfr744,Olfr76,Olfr794,Olfr796,Olfr855,Olfr890,Olfr891,Olfr936,Olfr945,Olfr948,Olfr972,Olfr974,Pcdh15,Rtp4,Sox2,Trpa1
GO:0051606	detection of stimulus	1.57E-03	Cadm1,Cngb3,Col11a1,Gm10081,Gnat3,Gpr98,Kcnmb2,Myc,Olfr1087,Olfr1089,Olfr109,Olfr1111,Olfr1112,Olfr1257,Olfr13,Olfr1326-ps1,Olfr135,Olfr138,Olfr1431,Olfr146,Olfr1505,Olfr17,Olfr191,Olfr193,Olfr195,Olfr196,Olfr199,Olfr201,Olfr229,Olfr292,Olfr356,Olfr357,Olfr366,Olfr368,Olfr38,Olfr389,Olfr435,Olfr453,Olfr48,Olfr714,Olfr715,Olfr724,Olfr725,Olfr736,Olfr738,Olfr743,Olfr744,Olfr76,Olfr794,Olfr796,Olfr855,Olfr890,Olfr891,Olfr936,Olfr945,Olfr948,Olfr972,Olfr974,Pcdh15,Rtp4,Sox2,Tlr2,Tlr4,Trpa1
GO:0050907	detection of chemical stimulus involved in sensory perception	2.09E-03	Gm10081,Gnat3,Olfr1087,Olfr1089,Olfr109,Olfr1111,Olfr1112,Olfr1257,Olfr13,Olfr1326-ps1,Olfr135,Olfr138,Olfr1431,Olfr146,Olfr1505,Olfr17,Olfr191,Olfr193,Olfr195,Olfr196,Olfr199,Olfr201,Olfr229,Olfr292,Olfr356,Olfr357,Olfr366,Olfr368,Olfr38,Olfr389,Olfr435,Olfr453,Olfr48,Olfr714,Olfr715,Olfr724,Olfr725,Olfr736,Olfr738,Olfr743,Olfr744,Olfr76,Olfr794,Olfr796,Olfr855,Olfr890,Olfr891,Olfr936,Olfr945,Olfr948,Olfr972,Olfr974,Rtp4,Trpa1
GO:0050911	detection of chemical stimulus involved in sensory perception of smell	2.32E-03	Gm10081,Olfr1087,Olfr1089,Olfr109,Olfr1111,Olfr1112,Olfr1257,Olfr13,Olfr1326-ps1,Olfr135,Olfr138,Olfr1431,Olfr146,Olfr1505,Olfr17,Olfr191,Olfr193,Olfr195,Olfr196,Olfr199,Olfr201,Olfr229,Olfr292,Olfr356,Olfr357,Olfr366,Olfr368,Olfr38,Olfr389,Olfr435,Olfr453,Olfr48,Olfr714,Olfr715,Olfr724,Olfr725,Olfr736,Olfr738,Olfr743,Olfr744,Olfr76,Olfr794,Olfr796,Olfr855,Olfr890,Olfr891,Olfr936,Olfr945,Olfr948,Olfr972,Olfr974
GO:0007606	sensory perception of chemical stimulus	2.39E-03	Gm10081,Gnat3,Grm7,Olfr1087,Olfr1089,Olfr109,Olfr1111,Olfr1112,Olfr1257,Olfr13,Olfr1326-ps1,Olfr135,Olfr138,Olfr1431,Olfr146,Olfr1505,Olfr17,Olfr191,Olfr193,Olfr195,Olfr196,Olfr199,Olfr201,Olfr229,Olfr292,Olfr356,Olfr357,Olfr366,Olfr368,Olfr38,Olfr389,Olfr435,Olfr453,Olfr48,Olfr714,Olfr715,Olfr724,Olfr725,Olfr736,Olfr738,Olfr743,Olfr744,Olfr76,Olfr794,Olfr796,Olfr855,Olfr890,Olfr891,Olfr936,Olfr945,Olfr948,Olfr972,Olfr974,Rtp4,Slc24a4,Tas2r114,Tas2r115,Tas2r122,Trpa1,V1ra8,Vmn1r53
GO:0004984	olfactory receptor activity	4.83E-03	Gm10081,Olfr1087,Olfr1089,Olfr109,Olfr1111,Olfr1112,Olfr1257,Olfr13,Olfr1326-ps1,Olfr135,Olfr138,Olfr1431,Olfr146,Olfr1505,Olfr17,Olfr191,Olfr193,Olfr195,Olfr196,Olfr199,Olfr201,Olfr229,Ol

			fr292,Olfr356,Olfr357,Olfr366,Olfr368,Olfr38,Olfr389,Olfr435,Olfr453,Olfr48,Olfr714,Olfr715,Olfr724,Olfr725,Olfr736,Olfr738,Olfr743,Olfr744,Olfr76,Olfr794,Olfr796,Olfr855,Olfr890,Olfr891,Olfr936,Olfr945,Olfr948,Olfr972,Olfr974
GO:0007608	sensory perception of smell	5.13E-03	Gm10081,Grm7,Olfr1087,Olfr1089,Olfr109,Olfr1111,Olfr1112,Olfr1257,Olfr13,Olfr1326-ps1,Olfr135,Olfr138,Olfr1431,Olfr146,Olfr1505,Olfr17,Olfr191,Olfr193,Olfr195,Olfr196,Olfr199,Olfr201,Olfr229,Olfr292,Olfr356,Olfr357,Olfr366,Olfr368,Olfr38,Olfr389,Olfr435,Olfr453,Olfr48,Olfr714,Olfr715,Olfr724,Olfr725,Olfr736,Olfr738,Olfr743,Olfr744,Olfr76,Olfr794,Olfr796,Olfr855,Olfr890,Olfr891,Olfr936,Olfr945,Olfr948,Olfr972,Olfr974,Slc24a4
GO:0004930	G-protein coupled receptor activity	4.86E-02	Adora2b,Agtr1b,Agtr2,Calcr,Celsr1,Chrm3,Cxcr7,Cysltr1,Ednrb,Emr4,F2rl3,Fzd8,Gm10081,Gm9268,Gpr87,Gpr98,Grm7,Hpgd,Htr1a,Htr2a,Lpar4,Nmur2,Npy2r,Npy6r,Olfr1087,Olfr1089,Olfr109,Olfr1111,Olfr1112,Olfr1257,Olfr13,Olfr1326-ps1,Olfr135,Olfr138,Olfr1431,Olfr146,Olfr1505,Olfr17,Olfr191,Olfr193,Olfr195,Olfr196,Olfr199,Olfr201,Olfr229,Olfr292,Olfr356,Olfr357,Olfr366,Olfr368,Olfr38,Olfr389,Olfr435,Olfr453,Olfr48,Olfr714,Olfr715,Olfr724,Olfr725,Olfr736,Olfr738,Olfr743,Olfr744,Olfr76,Olfr794,Olfr796,Olfr855,Olfr890,Olfr891,Olfr936,Olfr945,Olfr948,Olfr972,Olfr974,P2ry10,P2ry14,Pth2r,S1pr1,Sstr1,Tacr1,Tas2r114,Tas2r115,Tas2r122,V1ra8,Vmn1r169,Vmn1r170,Vmn1r230,Vmn1r231,Vmn1r3,Vmn1r5,Vmn1r53,Vmn1r6,Vmn1r64,Vmn1r65,Vmn1r68,Vmn1r69,Vmn1r71,Vmn1r72,Vmn2r18,Vmn2r19,Vmn2r20,Vmn2r23,Vmn2r24,Vmn2r53,Vmn2r54,Vmn2r55,Vmn2r6,Vmn2r63,Vmn2r7,Vmn2r77,Vmn2r78,Vmn2r79,Vmn2r92,Vmn2r94