2019

Establishment And Regulation Of Epithelial Enhancers By P63

Enrique Lin Shiao
University of Pennsylvania, linenrique@gmail.com

Recommended Citation

Establishment And Regulation Of Epithelial Enhancers By P63

Abstract

The transcription factor p63 is a key mediator of epidermal commitment, development, and differentiation. Point mutations in p63 in human patients lead to developmental defects, including orofacial clefting, one of the most common congenital defects in humans. To date, knowledge is limited about how pivotal is p63's role in human craniofacial development, due in part to a lack of tractable models to study de novo expression of p63 and its role at developmentally regulated genes. Using an inducible trans-differentiation model, combined with epigenomic sequencing and multi-cohort meta-analysis of GWAS data, we show that p63 establishes enhancers at craniofacial development genes to modulate their transcription. We further identify histone methyltransferase KMT2D as a key interacting partner of p63 at these enhancers and identify a novel role for this histone methyltransferase in maintaining epithelial homeostasis. Our results demonstrate that disease-specific substitution mutation in the DNA binding domain or SAM protein interaction domain of p63 respectively eliminate or reduce establishment of these enhancers. Finally, we show enhancers established by p63 are highly enriched for SNPs associated with nonsyndromic cleft lip +/- cleft palate (CL/P). These orthogonal approaches indicate a strong molecular link between p63 enhancer function and CL/P, illuminating molecular mechanisms underlying this developmental defect and revealing vital regulatory elements and new candidate causative genes.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Biochemistry & Molecular Biophysics

First Advisor
Shelley L. Berger

Keywords
Chromatin, Cleft lip, Cleft palate, Enhancer, Epigenetics, p63

Subject Categories
Biochemistry | Developmental Biology | Genetics

ESTABLISHMENT AND REGULATION OF EPITHELIAL ENHANCERS BY P63

Enrique Lin Shiao

A DISSERTATION

in

Biochemistry and Molecular Biophysics

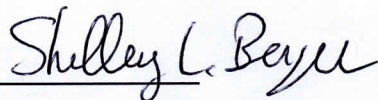Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2019
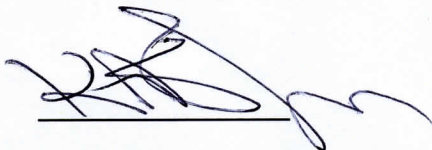
Supervisor of Dissertation

Shelley L. Berger

Shelley L. Berger, Ph.D.

Daniel S.Och University Professor

Graduate Group Chairperson

Kim A. Sharp, Ph.D., Associate Professor of Biochemistry and Biophysics

Dissertation Committee

Ronen Marmorstein, Ph.D., George W. Raiziss Professor

Gerd A. Blobel, M.D., Ph.D., Professor of Pediatrics

Roger A. Greenberg, M.D., Ph.D., J. Samuel Staub, M.D. Professor

Brian Capell, M.D., Ph.D., Assistant Professor of Dermatology and Genetics

To my parents Sofia Shiao Wu and Antonio Lin Huang who at my age were not pursuing a Ph.D. but instead were embarking on a far more challenging and courageous adventure, leaving their hometowns, families, titles, professions, language and culture to start a new life across the globe. Thanks for fueling my thirst for adventure and the unknown in both life and in science.

ACKNOWLEDGMENT

shared and for helping me maintain my German skills, as well as introducing me to a lot of Philly life through Tori. Special thanks to Mara, Kelly and Sam for all the times and experiences shared. I am also grateful to Meilin, Mariel, Matt and Laura for all the special moments.

To my colleagues at the Berger lab, my very sincere gratitude for all your help, conversations, feedback and shared moments in and outside of the lab in these past 5 years. Particular thanks goes to Lacey and Morgan who both mentored me directly for several months and taught me many of the techniques I have learned during my time here, as well as helping me improve as a scientist while maintaining a great environment to work in. Thanks to Kate who has also been a great mentor and colleague in the time we have shared, as well as a great friend. To all other members past and present, it has been a real privilege to work with so many people from all over the world and with such diverse expertise, I hope we stay in touch and find ways to collaborate in the future. I am also incredibly grateful to Sophia, your help and amazing organizational skills have been crucial in every step of my time here. Special shout out to Yemin, who was extremely helpful with bioinformatic questions and analyses.

My sincere gratitude to Hillary Nelson, Ph.D. and the Public Health Certificate Program for allowing me to complement my scientific training with broad training in public health, biostatistics and epidemiology. Through PHCP I was also connected to Puentes de Salud and PCCY where I had the great opportunity to volunteer and give back to Latin American communities in Philadelphia. I am also very thankful to ORDT for the work that you do and the experiences you have provided, particular shout out to Raquel Castellanos, Ph.D. and Arnaldo Diaz, Ph.D., whose incredible dedication and passion towards increasing diversity in science is incredibly contagious.

I am also extremely grateful to all the members of the Penn Science Diplomacy Group who joined me through the conception of the group and worked hard in all the different projects, working with embassies and institutions abroad (Cuba, Lithuania, North Korea, Costa Rica, among others) to create a more united world through science. My sincere gratitude also to Kevin Alicea, my cofounder in Caminos en Ciencia. It has been a great experience to work together in

this podcast and a real privilege to talk to researchers from across the globe and be able to share their experiences and pathways in science to the broader Latin American audience.

To my Carpenter roommates, Ari, Robert and Martin. It was a privilege sharing a home with you guys for so many years. I am grateful for all the time shared, from commiserating about grad school life, going to concerts, cooking, exploring the city, hosting piñata parties and so much more. I am grateful to my Ph.D. experience for allowing me to gain three more best friends. My sincere gratitude to my best friends abroad as well, particularly to Fabricio, Christopher, Andrea and Sofia who have stayed in touch and come to visit in multiple occasions.

Special thanks to my brothers Andres and Esteban, I know you are always there for me and I am glad we can always find some time to see each other every few months. I am also grateful to my nephews, Esteban, Luca and Maximo, its been really fun to see you all grow and develop and I wish the best for all of you always.

To my fiancé Caitlin, thank you so much for joining me in the final leg of my Ph.D. You have greatly enriched my life through the experiences we have shared. I really enjoy our conversations and all the different facts and knowledge I gain from our interactions together. I feel incredibly lucky to have met you and to be able to explore the world and different cuisines with you. Thank you as well for convincing me to adopt our dog Petri, these past few months living together have been awesome. I love you and look forward to the many more experiences we will share in the future.

Finally, my utmost gratitude goes to my parents and greatest role models Sofía and Antonio. Your support and dedication through every step of my life have been important to get to where I am. Your courage and fearlessness in leaving everything behind to start a new life and seek better opportunities in Costa Rica have been the greatest example for my own life. As you embarked in this adventure and without knowing, you taught me to not be afraid of uncertainty and also to always continue moving and seeking to push the boundaries of the known. I have taken this lesson with me in my life and in science. Words cannot convey how much I love you and how grateful I am for everything you have done.

ABSTRACT

ESTABLISHMENT AND REGULATION OF EPITHELIAL ENHANCERS BY P63

Enrique Lin Shiao

Shelley L. Berger

The transcription factor p63 is a key mediator of epidermal commitment, development, and differentiation. Point mutations in p63 in human patients lead to developmental defects, including orofacial clefting, one of the most common congenital defects in humans. To date, knowledge is limited about how pivotal is p63's role in human craniofacial development, due in part to a lack of tractable models to study de novo expression of p63 and its role at developmentally regulated genes. Using an inducible trans-differentiation model, combined with epigenomic sequencing and multi-cohort meta-analysis of GWAS data, we show that p63 establishes enhancers at craniofacial development genes to modulate their transcription. We further identify histone methyltransferase KMT2D as a key interacting partner of p63 at these enhancers and identify a novel role for this histone methyltransferase in maintaining epithelial homeostasis. Our results demonstrate that disease-specific substitution mutation in the DNA binding domain or SAM protein interaction domain of p63 respectively eliminate or reduce establishment of these enhancers. Finally, we show enhancers established by p63 are highly enriched for SNPs associated with nonsyndromic cleft lip +/- cleft palate (CL/P). These orthogonal approaches indicate a strong molecular link between p63 enhancer function and CL/P, illuminating molecular mechanisms underlying this developmental defect and revealing vital regulatory elements and new candidate causative genes.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

CHAPTER 1: Introduction

1.1: Epigenetics

One of the most fascinating hallmarks in biology is how multicellular organisms can form from a single cell and how cells with an identical genome can show so many different phenotypes. The main difference between distinct cells in an organism comes down to what genes are expressed in each cell and this is in part modulated by the organization of DNA into chromatin. In chromatin, DNA wraps around proteins known as histones that function as spools to allow DNA to be organized in a particular way within the nucleus. Both DNA and histones can be covalently modified to modulate the organization of DNA in the nucleus and ultimately regulate gene expression (Figure 1.1). The study of heritable changes in gene expression patterns that do not involve modifications in the DNA sequence is known as epigenetics[1,2]. Epi- comes from the Greek and means over, outer or above.
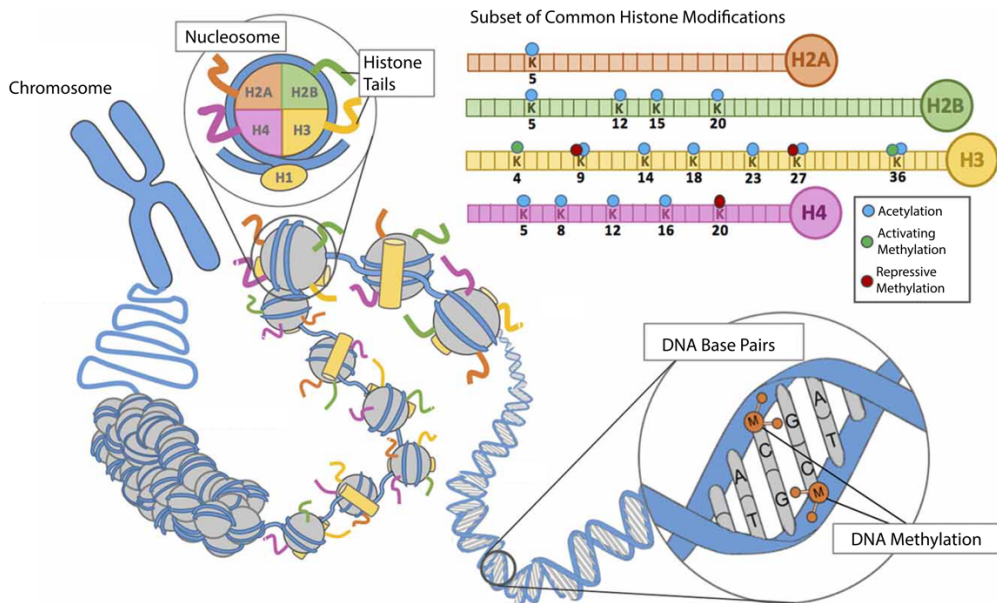
Figure 1.1: Chromatin can be covalently modified to modulate its organization in the nucleus. Figure adapted from Weaver et al.[3]

Many epigenetic mechanisms have been identified thus far, including the role of DNA methylation, histone modifications and long non-coding RNAs in modulating gene expression[1–4]. However, how all these different mechanisms act in concert to regulate complex processes, such as lineage specification and development, remains to be elucidated. Additionally, how disruption of these mechanisms can lead to disease is largely unclear. The completion of the human genome project and the advancement of next generation sequencing technologies have opened up a plethora of new avenues to study epigenetics and start to answer some of these questions.

1.2: From enhancers to lineage specification

In 1972 the geneticist Susumu Ohno coined the term "junk DNA" to refer to parts of the genome that did not encode for proteins and which many people believed had no function[5]. The completion of the human genome project in 2003 showed that only about 1% of the human genome are protein-coding regions, encompassing about 21.000 protein-coding genes[6,7]. However, a great amount of evidence has accumulated since then to demonstrate that many parts of the non-coding genome perform important regulatory functions to modulate gene expression[4,8–10].

Cis-regulatory elements, in particular, are regions of non-coding DNA that regulate transcription of neighboring genes[8]. The best characterized cis-regulatory elements include promoters and enhancers, both of which can be bound by transcription factors to modulate downstream gene expression. Promoters are primarily localized ~50 base pairs(bp) upstream or downstream of the transcription start site (TSS), while enhancers can be localized upstream, downstream, within intronic regions or even mega bases apart from the genes they regulate[11,12]. While promoters are sufficient to assemble RNA Pol II machinery to start transcription, transcription is often low in the absence of enhancers[11,13].

The first enhancer was identified about 40 years ago, as part of the sequence of the SV40 virus genome[14,15]. Isolation and introduction into a reporter in HeLa cells demonstrated over 100-fold enhancement of transcription of a reporter gene[14]. Since then, many enhancers have been identified in mammalian genomes and their biochemical and functional properties have been thoroughly investigated[10,12,13]. Comparative genomics facilitated the identification of enhancers by looking at highly conserved non-coding regions across vertebrates and mammals[16–18]. With the advent of next-generation sequencing, many more enhancers have been described in the past decade and key features have been found to *de novo* identify enhancers: open chromatin, decorated by histone modifications (H3K27ac, H3K4me1, H3K122ac and H4K16ac), bound by RNA Pol II and transcription factors; and expression of enhancer RNAs (eRNAs)[19–26]. (Figure 1.2)
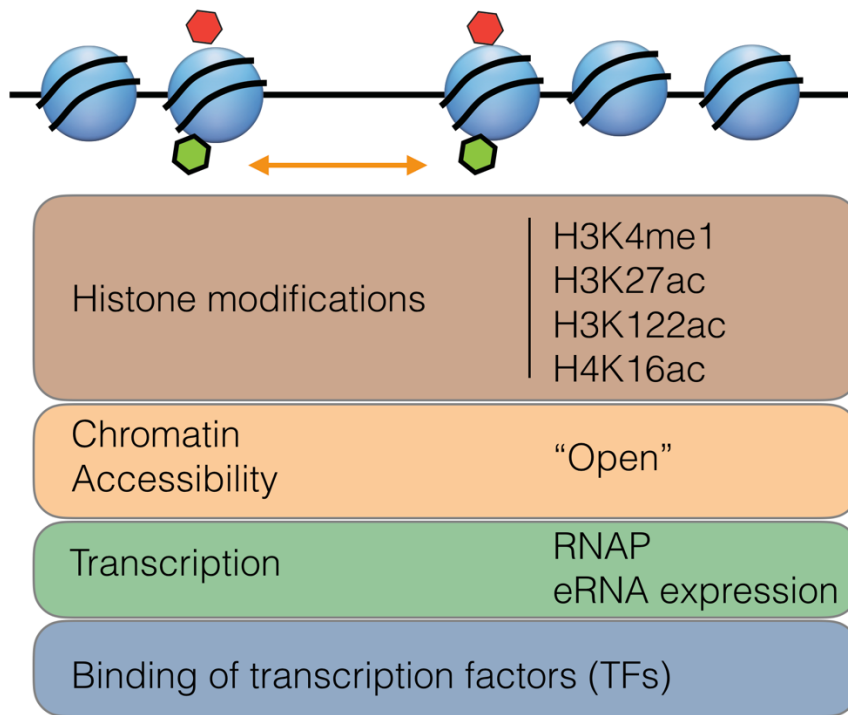
Figure 1.2: Anatomy of enhancers. Specific chromatin features have been found to be characteristic to enhancers, including activating histone modification marks, depletion of nucleosomes, active transcription and binding of transcription factors.

Enhancers and promoters are associated to cell fate determination, as they can be bound by lineage specific transcription factors and remodeled to regulate tissue specific transcription. The current model for how enhancers work posits that active enhancers bound by transcription factors can loop towards the promoter through protein-protein interactions at both sites to activate downstream transcription of specific genes[27–31] (Figure 1.3). Many questions remain around enhancers, particularly how enhancers are established, maintained and inherited through cell divisions is not fully understood[12].
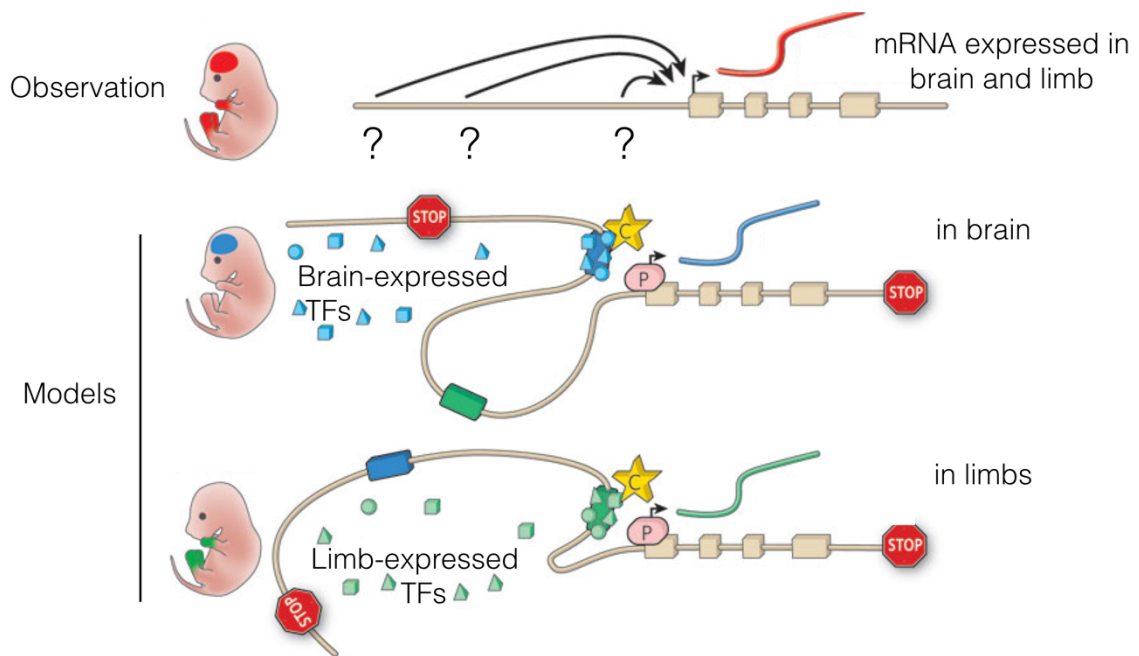
Figure 1.3: Enhancers are non-coding regions of DNA that regulate cell-type specific gene expression and maintain cell identity. Local expression of a gene during brain and limb development is achieved by binding of tissue specific transcription factors to distinct upstream enhancers and looping of the enhancer to the promoter region. Figure adapted from Visel et al.[27]

Hematopoiesis is perhaps one of the best studied models for lineage commitment and the β-globin locus involved in erythroid differentiation is among the first and most thoroughly characterized gene clusters at which long-range interactions between a distal enhancer and target β-globin promoters have been observed[29,32,33].During hematopoiesis, hematopoietic stem cells (HSC) undergo transcriptional and epigenetic changes associated with lineage restriction and cell specification to give rise to all the hematopoietic cell types[34–36]. Key transcription factors, such as GATA1 and GATA2 have been identified in this process that modify chromatin structure and gene expression to regulate cell fate decisions. Translocations in the the β-globin locus have been shown to result in thalassaemias even when substitution mutations have not been found in the the β-globin genes, showing that diseases can emerge as a consequence of disruption of enhancer function[37].

Genetic variation in distant enhancers has been linked to many more human Mendelian disorders, including pre-axial polydactyly due to mutations in an enhancer controlling sonic hedgehog (SHH)[38], Holt-Oram syndrome (exhibiting cardiac and limb malformations) due to mutations in an enhancer controlling TBX5[39] and many other disorders[40–42]. Importantly genome-wide association studies (GWAS) have shown that the majority of disease-associated variants are not localized within protein-coding regions but instead are localized at non-coding regions and many could possibly be localized at enhancers[43,44].

In order to understand how enhancers may influence different diseases, it is necessary to expand our knowledge about how enhancers are specified in different lineages and fates beyond the hematopoietic pathway and how mutations in enhancer regions can lead to dysregulation of gene expression in different cell types.

1.3: The epidermis

The surface epithelium of the skin or epidermis forms the outermost layer of cells in the body, providing a direct physical, chemical and immunological barrier between the body and the environment. The epidermis is self-renewing, with progenitor basal keratinocytes giving rise to all the differentiated layers of the stratified epidermis[45](Figure 1.4). This differentiation process occurs through an outward migration and terminal differentiation program that is tightly controlled by epigenetic changes and by spatiotemporal expression of epithelial specific genes.
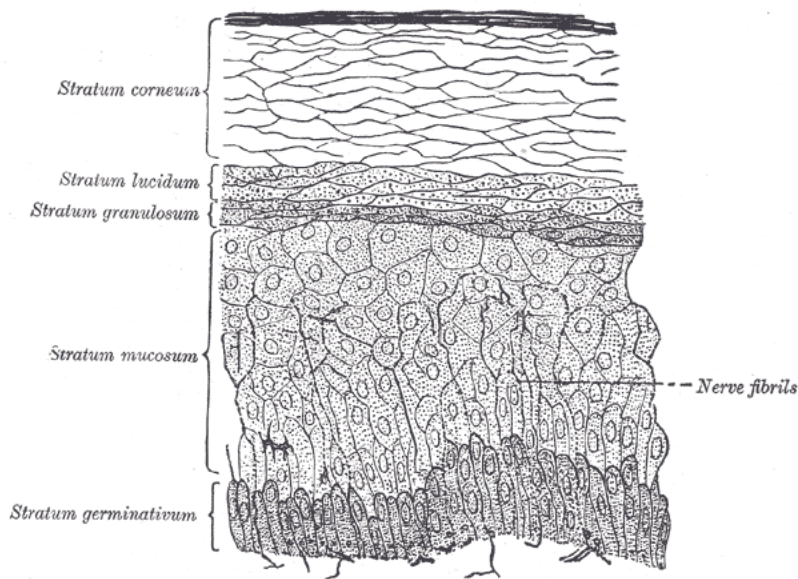
Figure 1.4: The epidermis is composed of stratified layers that stem from the progenitor basal keratinocytes. Schematic of an epidermis section showing the different layers. Figure taken from Anatomy of the Human Body[46].

Dysfunction of epidermal lineage specification and differentiation is associated with over a hundred human genetic diseases, impacting over 20% of the population[47]. Some of the most severe diseases include epidermolysis bullosa (manifested by skin blistering and erosions), squamous cell carcinomas, as well as ectodermal defects (showing orofacial clefting and limb malformations)[47]. Given the high incidence of epidermal disorders in humans it is imperative to

understand the molecular mechanisms that lead to epithelial specification and differentiation and study how these processes are affected in disease.

The mammalian epidermis derives from the ectoderm through a process of lineage and cell fate specification that results in the stratified squamous epithelium[48]. The consequent differentiation of basal epidermal cells into terminally differentiated cells of the cornified layer has been thoroughly studied and characterized[45,49]. However, the early events that lead to the commitment of epidermal stem cells and progenitor cell populations are still not completely understood.

Bone morphogenetic protein (BMP) and retinoic acid (RA)[50–52] initially specify the epidermal lineage from the lateral surface ectoderm. The surface ectoderm is a single-layered epithelium characterized by the expression of markers keratin 8(KRT8) and keratin 18(KRT18). In the presence of epidermal growth factor, fibroblast growth factor and insulin, the surface ectoderm commits to form the stratified epidermal progenitors known as basal membrane keratinocytes which express keratin 5(KRT5) and keratin 14(KRT14)[52,53]. The basal keratinocytes can then further differentiate to produce stratified multi-layered skin.

Mouse developmental models, human syndromes and in vitro epidermal differentiation of pluripotent stem cells (PSCs) have begun to uncover key transcription factors required for epidermal lineage specification and differentiation. A key transcription factor involved in regulating gene expression programs to modulate these processes is p63, which has been termed a master regulator of epidermal development[54,55].

1.4: Transcription factor p63

Transcription factor p63, a member of the p53 family of proteins, is almost exclusively expressed in epithelial cells[55]. Unlike p53, numerous N- and C-terminal variants of p63 can be

transcribed to yield polypeptides that exhibit different functions in transcriptional control, with ΔNp63α as the most abundant and best characterized isoform in epithelial cells (Figure 1.5)[55,56]. Importantly, the high sequence identity between the DNA binding domains (DBD) of p53 and p63 allow them to bind to the same DNA consensus motif.
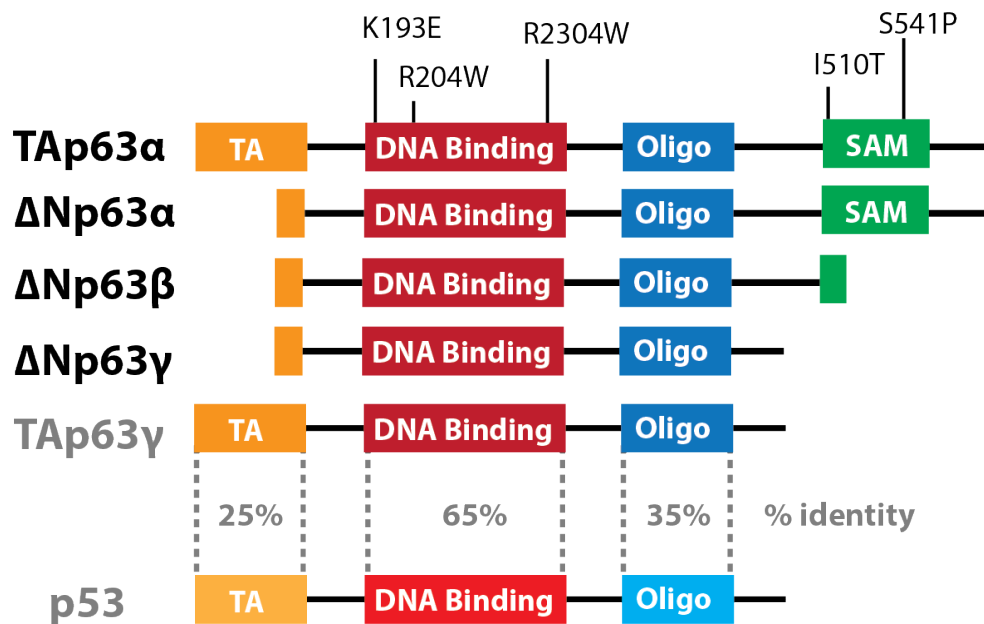
Figure 1.5: p63 can be transcribed as 6 isoforms. Including ΔN and TA (transcriptional activation domain) isoforms. Disease related mutations in the DBD and SAM domain of p63 are highlighted, as well as the high sequence identity with p53.

While deletion of *Trp53* (encoding the p53 protein) shows no phenotype at birth[57], deletion of *Trp63* (encoding the p63 protein) in mice leads to clear developmental and morphological defects in the squamous epithelia and epidermis, leading to abnormal craniofacial development, truncated limbs and loss of salivary glands, hair follicles and teeth[54,55] (Figure 1.6). In humans, germline missense mutations in p63 cause ectodermal dysplastic syndromes, leading to cleft lip/palate and limb malformations[47,58,59]. Interestingly, these mutations are largely

clustered in either the DBD or the sterile alpha motif (SAM) domain and show clear differences in the location and severity of phenotypes[60]. These genetic data strongly implicate p63 in critical epithelial cell processes, including establishment of epithelial identity.

Although p63 mutations have not directly been linked to tumorigenesis[61,62], loss of p63 has been observed in the epithelial to mesenchymal transition (EMT) in several types of cancers, including human breast and human prostate adenocarcinoma[63,64]. Our laboratory previously demonstrated that p53 family member proteins, including p63, can directly bind to nucleosome-occluded DNA sequences, displaying properties more commonly observed in pioneer factors[65]. Pioneer factors are transcription factors that can bind to condensed chromatin and recruit chromatin modifiers and other transcription factors to regulate the chromatin landscape and enhancer activity[66–69]. They are critical for specifying cell lineages, presumably through their ability to regulate cell type-specific enhancer networks[67]. While the importance of p63 in the development and differentiation of epithelial tissues has been demonstrated in mouse knockout models and human syndromes, the biochemical role of p63 in de novo epithelial lineage commitment and the molecular and epigenetic mechanisms by which it maintains epithelial cell identity though enhancer regulation and pioneer factor activity remain to be elucidated. A main reason for this is the difficulty of studying these processes during embryonic development and a lack of tractable models to do this[70].

1.4: p63 remodels chromatin during epidermal development

While the importance of p63 in regulating epidermal development has been known for almost 30 years now[54,55], the molecular mechanisms by which it achieves this have only recently begun to be elucidated. Early on, it was recognized that p63 could bind upstream of genes to either activate or repress their transcription[55,71]. This mechanism was shown to be crucial for proliferation and differentiation of human epidermal keratinocytes[72–74], regulation of signaling

pathways involved in ectodermal organogenesis[75] as well as maintenance of proliferative potential of epidermal progenitor stem cells[76] and more recently regulation of nuclear shape and expression of nuclear envelope-associated genes[77].

A more complete picture of how p63 engages with chromatin to regulate epidermal development emerged with the advent of genome-wide high throughput sequencing techniques to profile changes in chromatin, including chromatin-immunoprecipitation sequencing (ChIP-Seq)[78] to profile changes in histone modification profiles and binding profiles of transcription factors, as well as assay for transposase-accessible chromatin sequencing (ATAC-Seq)[79] to profile open and closed chromatin regions. Initial experiments demonstrated that p63 binds to over 7000 sites in the genome in human foreskin keratinocytes, primarily in intronic regions and upstream of promoters[80]. Furthermore, results demonstrated that p63 bound sites were enriched for H3K4 monomethylation, a histone modification characteristic of enhancers[80]. Integrative analysis suggested that genes closest to p63 bound sites were downregulated when p63 was knocked down[81], correlating genome-wide binding of p63 to transcriptional changes and identifying a plethora of new genes regulated by p63.

Other transcription factors and chromatin modifying proteins involved in co-regulation of epidermal development with p63 were also identified through genome-wide and biochemical experiments, including chromatin reorganizer SATB1[82], chromatin remodeler BRG1[83] and transcription factor TFAP2C[80]. Further experiments showed correlation between p63 binding and acetylation of histone tail H3 lysine 27 (H3K27ac) during keratinocyte proliferation and differentiation suggesting a key role for p63 in regulating epidermal specific enhancers[84]. A further study demonstrated cooperation between the BAF chromatin remodeling complex and p63 and showed that both are required in an invitro human epidermal differentiation assay to maintain open chromatin regions characteristic of keratinocytes[85].

More recently, experiments using an in vitro epidermal differentiation assay from PSCs suggested TFAP2C, but not p63, is sufficient to initiate surface ectoderm differentiation[52]. These

10

data did show, however, that p63 is necessary for maturation of this differentiation program to specify the epidermal lineage. Data from the same group showed p63 could specify different transcriptional programs depending on which morphogens it is exposed to[51]. Another recent single-cell study showed p63 is able to remodel open chromatin regions to specify the epidermal lineage independently of previously described co-regulatory transcription factors[86] and unpublished work from the Zhou lab suggests a novel interaction between p63 and CTCF to modulate chromatin architecture in skin keratinocytes[87]. Together the data thus far points towards a key role for p63 in regulating epidermal lineage specification, however, many questions remain around the mechanism and how p63 engages other transcription factors during this process.

An additional fundamental question is whether dysregulation of p63's interaction with chromatin plays a role in disease. As mentioned above, substitution mutations in p63 lead to ectodermal dysplastic syndromes, including orofacial clefting and limb malformation[54,58,88]. Furthermore, p63 is amplified in a subset of squamous cell carcinomas[89–91]. Nevertheless, the molecular mechanisms behind these disorders are only recently being elucidated. A study published last year demonstrated that enhancer reprogramming via p63 in the squamous subtype of pancreatic ductal adenocarcinoma promoted migration, invasion and tumor growth[92]. Another study showed that keratinocytes from patients with electrodactyl, ectodermal dysplasia, and cleft lip/palate syndrome (EEC), exhibit a clear downregulation of epidermal genes and loss of epidermal enhancers[93]. Finally, murine studies and human genome wide association studies (GWAS) have suggested a connection between p63 and clefting of the lip and the palate[80,94–96]; a molecular link however remains to be demonstrated.

1.5: Cleft lip/palate

Mutations in p63 in humans are rare, however, orofacial clefting [with its distinct subtypes cleft lip +/- cleft palate (CL/P) and cleft palate only (CPO)] is a key feature of mutant p63

syndromes, and is one of the most common congenital defects world-wide with an approximate prevalence of 1:700 births[97]. The etiology of orofacial clefting is complex, with involvement of genetic and environmental factors. To date, over 400 different human syndromes have been reported showing orofacial clefting as part of their phenotypic spectrum. However, more than 70% of CL/P cases and more than 50% of CPO cases are non-syndromic (ns), presenting as isolated phenotypes with absence of other organs and systems affected, and hence their etiology remains largely unclear[97–100]. The commonality and complexity of the constellation of orofacial clefting syndromes suggests involvement of p63-linked regulatory dysfunction.

Murine models have been crucial in understanding lip and palate development. Nonetheless, mutations and knockout of proteins involved in human CL/P in mice rarely show clefting of the lip, with clefting primarily observed in the secondary palate[101,102]. Genome-wide association studies (GWAS) and other genetic approaches in human nsCL/P patients have identified 40 risk loci that show robust associations of single nucleotide polymorphisms (SNPs) with nsCL/P[100,103–105]. A provocative finding is that the majority of SNPs associated to nsCL/P are not found within protein-coding exons but rather at intronic and intergenic regions, and, for the vast majority of these SNPs, the molecular mechanism underlying nsCL/P pathogenesis remains unknown. A single case in a family with CL/P has been reported in which a substitution mutation in p63 disrupts its binding to an enhancer and affects transcription of *IRF6*[95]. Whether CL/P associated SNPs function in a similar way is unknown, and most broadly, how general is enhancer disruption in developmental diseases, remains to be elucidated.

In the following chapters I will discuss two studies in which we investigated the role of p63 in human craniofacial development (Figure 1.6), and in the pathogenesis of orofacial clefting. In the first study, we developed an inducible trans-differentiation model that reveals p63 interaction with chromatin when expressed *de novo* and p63 regulation of gene expression. Our strategy encompassed a combination of epigenomic assays, including ATAC-seq, ChIP-seq of

p63 and the active enhancer histone modification H3K27ac, along with analysis of nsCL/P GWAS data. In parallel, for our second study, we combined mass spectrometry and genome-wide approaches to identify novel interactors of p63 at enhancers and characterized their function. Our results reveal p63 establishment of enhancers, displaying increased chromatin accessibility and de novo H3K27 acetylation, at critical craniofacial genes and demonstrate that substitution mutations in p63 abrogate this function. We identify KMT2D as a novel partner of p63 at epithelial enhancers and determine a novel role of this histone methyltransferase in coordinating epithelial homeostasis and differentiation. Further, we validate known genes and identify novel candidate genes and the enhancers that regulate them via p63, that might play a role in nsCL/P pathogenesis. Lastly, we show enhancers established by p63 are highly enriched for SNPs associated with nsCL/P, providing novel mechanistic insights into regulation of pathogenesis by CL/P-associated SNPs.

1. p63 is required for epithelial enhancer establishment at genes involved in CL/P. Mutations in p63 disrupt this process.
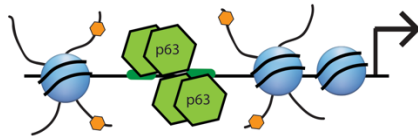
2. SNPs at p63 enhancer binding sites disrupt p63 recruitment and lead to dysregulation of genes involved in CL/P
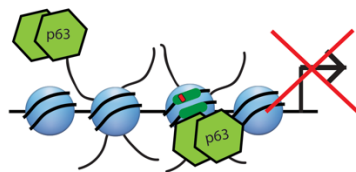
Figure 1.6: Central Hypotheses. While the importance of p63 in regulating epidermal development has been thoroughly studied, the molecular mechanisms by which it achieves this are not completely understood. Through investigation of p63's first encounter with chromatin I aim

to understand its role in epithelial enhancer establishment and how this process might be affected

in cleft lip/palate pathogenesis.

1.6: References

1.   Berger, S. L., Kouzarides, T., Shiekhattar, R. & Shilatifard, A. An operational definition of epigenetics. *Genes Dev.* **23**, 781–783 (2009).

2.   Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **128**, 693–705 (2007).

3.   Weaver, I. C. G. *et al.* Stress and the Emerging Roles of Chromatin Remodeling in Signal Integration and Stable Transmission of Reversible Phenotypes. *Front. Behav. Neurosci.* **11**, (2017).

4.   Nakagawa, S. & Kageyama, Y. Nuclear lncRNAs as epigenetic regulators—Beyond skepticism. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1839**, 215–222 (2014).

5.   Ohno, S. So much "junk" DNA in our genome. in *Evolution of Genetic Systems.* 366–370 (New York: Gordon and Breach.).

6.   Schmutz, J. *et al.* Quality assessment of the human genome sequence. *Nature* **429**, 365 (2004).

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

8.   Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Evolutionary changes in cis and trans gene regulation. *Nature* **430**, 85 (2004).

9.   Kleinjan, D. A. & Heyningen, V. van. Long-Range Control of Gene Expression: Emerging Mechanisms and Disruption in Disease. *Am. J. Hum. Genet.* **76**, 8–32 (2005).

10.   Ong, C.-T. & Corces, V. G. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.* **12**, 283–293 (2011).

11.   Maston, G. A., Evans, S. K. & Green, M. R. Transcriptional Regulatory Elements in the Human Genome. *Annu. Rev. Genomics Hum. Genet.* **7**, 29–59 (2006).

12.   Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A. & Bejerano, G. Enhancers: five essential questions. *Nat. Rev. Genet.* **14**, 288–295 (2013).

13.   Blackwood, E. M. & Kadonaga, J. T. Going the Distance: A Current View of Enhancer Action. *Science* **281**, 60–63 (1998).

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

15.   Moreau, P. *et al.* The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res.* **9**, 6047–6068 (1981).

16.   Bejerano, G. *et al.* Ultraconserved Elements in the Human Genome. *Science* **304**, 1321–1325 (2004).

17.   Pennacchio, L. A. *et al.* In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499 (2006).

18.   Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).

19.   Bowman, S. K. Discovering enhancers by mapping chromatin features in primary tissue. *Genomics* **106**, 140–144 (2015).

20.   Taylor, G. C. A., Eskeland, R., Hekimoglu-Balkan, B., Pradeepa, M. M. & Bickmore, W. A. H4K16 acetylation marks active genes and enhancers of embryonic stem cells, but does not alter chromatin compaction. *Genome Res.* **23**, 2053–2065 (2013).

21.   Zabidi, M. A. & Stark, A. Regulatory Enhancer–Core-Promoter Communication via Transcription Factors and Cofactors. *Trends Genet.* **32**, 801–814 (2016).

22.   Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318 (2007).

23.   Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci.* **107**, 21931–21936 (2010).

24.   Kim, T.-K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).

25.   Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).

26.   Melgar, M. F., Collins, F. S. & Sethupathy, P. Discovery of active enhancers through bidirectional expression of short transcripts. *Genome Biol.* **12**, R113 (2011).

27.   Visel, A., Rubin, E. M. & Pennacchio, L. A. Genomic views of distant-acting enhancers. *Nature* **461**, 199–205 (2009).

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

29.   Deng, W. *et al.* Controlling Long-Range Genomic Interactions at a Native Locus by Targeted Tethering of a Looping Factor. *Cell* **149**, 1233–1244 (2012).

30.   Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing Chromosome Conformation. *Science* **295**, 1306–1311 (2002).

31.   Cullen, K. E., Kladde, M. P. & Seyfred, M. A. Interaction between transcription regulatory regions of prolactin chromatin. *Science* **261**, 203–206 (1993).

32.   Tolhuis, B., Palstra, R.-J., Splinter, E., Grosveld, F. & Laat, W. de. Looping and Interaction between Hypersensitive Sites in the Active β-globin Locus. *Mol. Cell* **10**, 1453–1465 (2002).

33.   Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. & Fraser, P. Long-range chromatin regulatory interactions *in vivo*. *Nat. Genet.* **32**, 623–626 (2002).

34.   Krause, D. S. Regulation of hematopoietic stem cell fate. *Oncogene* **21**, 3262–3269 (2002).

35.   Orkin, S. H. & Zon, L. I. Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. *Cell* **132**, 631–644 (2008).

36.   Velten, L. *et al.* Human haematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol.* **19**, 271–281 (2017).

37.   Kleinjan, D. A. & Lettice, L. A. Chapter 13 Long-Range Gene Control and Genetic Disease. in *Advances in Genetics* **61**, 339–388 (Academic Press, 2008).

38.   Lettice, L. A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725–1735 (2003).

39.   Smemo, S. *et al.* Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Hum. Mol. Genet.* **21**, 3255–3263 (2012).

40.   Emison, E. S. *et al.* A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. *Nature* **434**, 857 (2005).

41.   Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci.* **106**, 9362–9367 (2009).

42.   Musunuru, K. *et al.* From noncoding variant to phenotype via *SORT1* at the 1p13 cholesterol locus. *Nature* **466**, 714–719 (2010).

43.   Ward, L. D. & Kellis, M. Evidence of Abundant Purifying Selection in Humans for Recently Acquired Regulatory Functions. *Science* **337**, 1675–1678 (2012).

44.   Way, G. P., Youngstrom, D. W., Hankenson, K. D., Greene, C. S. & Grant, S. F. Implicating candidate genes at GWAS signals by leveraging topologically associating domains. *Eur. J. Hum. Genet.* **25**, 1286–1289 (2017).

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

46.   Gray, Henry. *Anatomy of the Human Body*. (Lea and Febiger, 1918).

47.   Lopez-Pajares, V., Yan, K., Zarnegar, B. J., Jameson, K. L. & Khavari, P. A. Genetic pathways in disorders of epidermal differentiation. *Trends Genet.* **29**, 31–40 (2013).

48.   Troy, T.-C. & Turksen, K. Commitment of embryonic stem cells to an epidermal cell fate and differentiation in vitro. *Dev. Dyn.* **232**, 293–300 (2005).

49.   Fuchs, E. Chapter Nineteen - Epithelial Skin Biology: Three Decades of Developmental Biology, a Hundred Questions Answered and a Thousand New Ones to Address. in *Current Topics in Developmental Biology* (ed. Wassarman, P. M.) **116**, 357–374 (Academic Press, 2016).

50.   Metallo, C. M., Ji, L., de Pablo, J. J. & Palecek, S. P. Directed Differentiation of Human Embryonic Stem Cells to Epidermal Progenitors. in *Epidermal Cells: Methods and Protocols* (ed. Turksen, K.) 83–92 (Humana Press, 2010). doi:10.1007/978-1-60761-380-0_7

51.   Pattison, J. M. *et al.* Retinoic acid and BMP4 cooperate with p63 to alter chromatin dynamics during surface epithelial commitment. *Nat. Genet.* **50**, 1658 (2018).

52.   Li, L. *et al.* TFAP2C- and p63-Dependent Networks Sequentially Rearrange Chromatin Landscapes to Drive Human Epidermal Lineage Commitment. *Cell Stem Cell* **24**, 271-284.e8 (2019).

53.   Koster, M. I. & Roop, D. R. Mechanisms Regulating Epithelial Stratification. *Annu. Rev. Cell Dev. Biol.* **23**, 93–113 (2007).

54.   Mills, A. A. *et al. p63* is a *p53* homologue required for limb and epidermal morphogenesis. *Nature* **398**, 708–713 (1999).

55.   Yang, A. *et al.* p63, a p53 Homolog at 3q27–29, Encodes Multiple Products with Transactivating, Death-Inducing, and Dominant-Negative Activities. *Mol. Cell* **2**, 305–316 (1998).

56.   Levrero, M. *et al.* Structure, function and regulation of p63 and p73. *Cell Death Differ.* **6**, 1146–1153 (1999).

57.   Carbone, D. The P53 "knockout": Gone but not forgotten. *Hepatology* **16**, 1094–1096 (1992).

58.   Celli, J. *et al.* Heterozygous Germline Mutations in the p53 Homolog p63 Are the Cause of EEC Syndrome. *Cell* **99**, 143–153 (1999).

59.   Bokhoven, H. van & McKeon, F. Mutations in the p53 homolog p63: allele-specific developmental syndromes in humans. *Trends Mol. Med.* **8**, 133–139 (2002).

60.   Brunner, H. G., Hamel, B. C. J. & Bokhoven, H. van. The p63 gene in EEC and other syndromes. *J. Med. Genet.* **39**, 377–381 (2002).

61.   Keyes, W. M. *et al.* p63 heterozygous mutant mice are not prone to spontaneous or chemically induced tumors. *Proc. Natl. Acad. Sci.* **103**, 8435–8440 (2006).

62.   Ikawa, S., Nakagawara, A. & Ikawa, Y. p53 family genes: structural comparison, expression and mutation. *Cell Death Differ.* **6**, 1154–1161 (1999).

63.   Lindsay, J., McDade, S. S., Pickard, A., McCloskey, K. D. & McCance, D. J. Role of ΔNp63γ in Epithelial to Mesenchymal Transition. *J. Biol. Chem.* **286**, 3915–3924 (2011).

64.   Olsen, J. R. *et al.* p63 Attenuates Epithelial to Mesenchymal Potential in an Experimental Prostate Cell Model. *PLOS ONE* **8**, e62547 (2013).

65.   Zhu, J. *et al.* Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth. *Nature* **525**, 206–211 (2015).

66.	Iwafuchi-Doi, M. & Zaret, K. S. Pioneer transcription factors in cell reprogramming. *Genes Dev.* **28**, 2679–2692 (2014).

67.	Iwafuchi-Doi, M. *et al.* The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. *Mol. Cell* **62**, 79–91 (2016).

68.	Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **25**, 2227–2241 (2011).

69.	Soufi, A. *et al.* Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate Reprogramming. *Cell* **161**, 555–568 (2015).

70.	Chen, Y., Mistry, D. S. & Sen, G. L. Highly Rapid and Efficient Conversion of Human Fibroblasts to Keratinocyte-Like Cells. *J. Invest. Dermatol.* **134**, 335–344 (2014).

71.	Yang, A. *et al.* Relationships between p63 Binding, DNA Sequence, Transcription Activity, and Biological Function in Human Cells. *Mol. Cell* **24**, 593–602 (2006).

72.	Truong, A. B., Kretz, M., Ridky, T. W., Kimmel, R. & Khavari, P. A. p63 regulates proliferation and differentiation of developmentally mature keratinocytes. *Genes Dev.* **20**, 3185–3197 (2006).

73.	Romano, R.-A. *et al.* ΔNp63 knockout mice reveal its indispensable role as a master regulator of epithelial development and differentiation. *Development* **139**, 772–782 (2012).

74.	Shalom-Feuerstein, R. *et al.* ΔNp63 is an ectodermal gatekeeper of epidermal morphogenesis. *Cell Death Differ.* **18**, 887–896 (2011).

75.	Laurikkala, J. *et al.* p63 regulates multiple signalling pathways required for ectodermal organogenesis and differentiation. *Development* **133**, 1553–1563 (2006).

76.	Senoo, M., Pinto, F., Crum, C. P. & McKeon, F. p63 Is Essential for the Proliferative Potential of Stem Cells in Stratified Epithelia. *Cell* **129**, 523–536 (2007).

77.	Rapisarda, V. *et al.* p63 Transcription Factor Regulates Nuclear Shape and Expression of Nuclear Envelope-Associated Genes in Epidermal Keratinocytes. *J. Invest. Dermatol.* **137**, 2157–2167 (2017).

78.	Robertson, G. *et al.* Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* **4**, 651–657 (2007).

79.	Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).

80.	McDade, S. S. *et al.* Genome-wide analysis of p63 binding sites identifies AP-2 factors as co-regulators of epidermal differentiation. *Nucleic Acids Res.* **40**, 7190–7206 (2012).

81.	Carroll, D. K. *et al.* p63 regulates an adhesion programme and cell survival in epithelial cells. *Nat. Cell Biol.* **8**, 551–561 (2006).

82.	Fessing, M. Y. *et al.* p63 regulates Satb1 to control tissue-specific chromatin remodeling during development of the epidermis. *J Cell Biol* **194**, 825–839 (2011).

83.   Mardaryev, A. N. *et al.* p63 and Brg1 control developmentally regulated higher-order chromatin remodelling at the epidermal differentiation complex locus in epidermal progenitor cells. *Development* **141**, 101–111 (2014).

84.   Kouwenhoven, E. N. *et al.* Transcription factor p63 bookmarks and regulates dynamic enhancers during epidermal differentiation. *EMBO Rep.* **16**, 863–878 (2015).

85.   Bao, X. *et al.* A novel ATAC-seq approach reveals lineage-specific reinforcement of the open chromatin landscape via cooperation between BAF and p63. *Genome Biol.* **16**, 284 (2015).

86.   Fan, X. *et al.* Single Cell and Open Chromatin Analysis Reveals Molecular Origin of Epidermal Cells of the Skin. *Dev. Cell* **47**, 133 (2018).

87.   Qu, J., Yi, G. & Zhou, J. H. p63 cooperates with CTCF to modulate chromatin architecture in skin keratinocytes. *bioRxiv* 525667 (2019). doi:10.1101/525667

88.   Yang, A. *et al.* p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. *Nature* **398**, 714–718 (1999).

89.   Lo Muzio, L. *et al.* p63 overexpression associates with poor prognosis in head and neck squamous cell carcinoma. *Hum. Pathol.* **36**, 187–194 (2005).

90.   Watanabe, H. *et al.* SOX2 and p63 colocalize at genetic loci in squamous cell carcinomas. *J. Clin. Invest.* **124**, 1636–1645 (2014).

91.   Tanière, P. *et al.* TP53 mutations, amplification of P63 and expression of cell cycle proteins in squamous cell carcinoma of the oesophagus from a low incidence area in Western Europe. *Br. J. Cancer* **85**, 721 (2001).

92.   Somerville, T. D. D. *et al.* TP63-Mediated Enhancer Reprogramming Drives the Squamous Subtype of Pancreatic Ductal Adenocarcinoma. *Cell Rep.* **25**, 1741-1755.e7 (2018).

93.   Qu, J. *et al.* Mutant p63 Affects Epidermal Cell Identity through Rewiring the Enhancer Landscape. *Cell Rep.* **25**, 3490-3503.e4 (2018).

94.   Khandelwal, K. D. *et al.* Deletions and loss-of-function variants in TP63 associated with orofacial clefting. *Eur. J. Hum. Genet.* 1 (2019). doi:10.1038/s41431-019-0370-0

95.   Fakhouri, W. D. *et al.* An etiologic regulatory mutation in IRF6 with loss- and gain-of-function effects. *Hum. Mol. Genet.* **23**, 2711–2720 (2014).

96.   Richardson, R. *et al.* p63 exerts spatio-temporal control of palatal epithelial cell fate to prevent cleft palate. *PLOS Genet.* **13**, e1006828 (2017).

97.   Mangold, E. *et al.* Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. *Nat. Genet.* **42**, 24–26 (2010).

98.   Beaty, T. H. *et al.* A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near *MAFB* and *ABCA4*. *Nat. Genet.* **42**, 525–529 (2010).

99.   The genetics of isolated orofacial clefts: from genotypes to subphenotypes - Jugessur - 2009 - Oral Diseases - Wiley Online Library. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1601-0825.2009.01577.x. (Accessed: 3rd October 2018)

100.   Ludwig, K. U. *et al.* Imputation of orofacial clefting data identifies novel risk loci and sheds light on the genetic background of cleft lip ± cleft palate and cleft palate only. *Hum. Mol. Genet.* **26**, 829–842 (2017).

101.   Lough, K. J., Byrd, K. M., Spitzer, D. C. & Williams, S. . Closing the Gap: Mouse Models to Study Adhesion in Secondary Palatogenesis. *J. Dent. Res.* **96**, 1210–1220 (2017).

102.   Yu, K., Deng, M., Naluai-Cecchini, T., Glass, I. A. & Cox, T. C. Differences in Oral Structure and Tissue Interactions during Mouse vs. Human Palatogenesis: Implications for the Translation of Findings from Mice. *Front. Physiol.* **8**, (2017).

103.   Beaty, T. H., Marazita, M. L. & Leslie, E. J. Genetic factors influencing risk to orofacial clefts: today's challenges and tomorrow's opportunities. *F1000Research* **5**, 2800 (2016).

104.   Mostowska, A. *et al.* Common variants in DLG1 locus are associated with non-syndromic cleft lip with or without cleft palate. *Clin. Genet.* **93**, 784–793 (2018).

105.   Yu, Y. *et al.* Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. *Nat. Commun.* **8**, 14364 (2017).

CHAPTER 2: p63 establishes epithelial enhancers at critical craniofacial development genes

2.1: Introduction

The transcription factor p63, a member of the p53 family of proteins, is a key mediator of epithelial cell processes, including establishment and maintenance of epithelial identity[1]. Deletion of *Trp63* (encoding the p63 protein) in mice leads to developmental and morphological defects in the squamous epithelia and epidermis, leading to abnormal craniofacial development, truncated limbs, and loss of salivary glands, hair follicles and teeth[2,3]. In humans, germline missense mutations in p63 cause ectodermal dysplastic syndromes, which are characterized by orofacial clefting and limb malformations[4–6]. Interestingly, these mutations are largely clustered in either the DNA Binding Domain (DBD) or the protein-protein interaction Sterile Alpha Motif (SAM) domain; amino acid substitutions show clear differences in the physical location and severity of phenotypes[7]. p63 can bind directly to chromatin to change the chromatin landscape during keratinocyte differentiation[8–10]. However, the role of p63 in *de novo* epithelial lineage commitment and how p63 amino acid mutations lead to defects in this process, remain unknown.

Mutations in p63 in humans are rare, however, orofacial clefting [with its distinct subtypes cleft lip +/- cleft palate (CL/P) and cleft palate only (CPO)] is a key feature of mutant p63 syndromes, and is one of the most common congenital defects world-wide with an approximate prevalence of 1:700 births[11]. The etiology of orofacial clefting is complex, with involvement of genetic and environmental factors. To date, over 400 different human syndromes have been reported showing orofacial clefting as part of their phenotypic spectrum. However, more than 70% of CL/P cases and more than 50% of CPO cases are non-syndromic (ns), presenting as isolated phenotypes with absence of other organs and systems affected, and hence their etiology remains largely unclear[11–14]. The commonality and complexity of the constellation of orofacial clefting syndromes suggests involvement of p63-linked regulatory dysfunction.

Murine models have been crucial in understanding lip and palate development. Nonetheless, mutations and knockout of proteins involved in human CL/P in mice rarely show clefting of the lip, with clefting primarily observed in the secondary palate[15,16]. Genome-wide association studies (GWAS) and other genetic approaches in human nsCL/P patients have identified 40 risk loci that show robust associations of single nucleotide polymorphisms (SNPs) with nsCL/P[14,17–20]. A provocative finding is that the majority of SNPs associated to nsCL/P are not found within protein-coding exons but rather at intronic and intergenic regions, and, for the vast majority of these SNPs, the molecular mechanism underlying nsCL/P pathogenesis remains unknown. A single case in a family with CL/P has been reported in which a substitution mutation in p63 disrupts its binding to an enhancer and affects transcription of *IRF6*[21]. Whether CL/P associated SNPs function in a similar way is unknown, and most broadly, how general is enhancer disruption in developmental diseases, remains to be elucidated.

In this study we investigated the role of p63 in human craniofacial development, and in the pathogenesis of orofacial clefting. We used an inducible trans-differentiation model that reveals p63 interaction with chromatin when expressed *de novo* and p63 regulation of gene expression. Our strategy encompassed a combination of epigenomic assays, including ATAC-seq, ChIP-seq of p63 and the active enhancer histone modification H3K27ac, along with analysis of nsCL/P GWAS data. Our results reveal p63 establishment of enhancers, displaying increased chromatin accessibility and de novo H3K27 acetylation, at critical craniofacial genes and demonstrate that substitution mutations in p63 abrogate this function. We validate known genes and identify novel candidate genes and the enhancers that regulate them via p63, that might play a role in nsCL/P pathogenesis. Lastly, we show enhancers established by p63 are highly enriched for SNPs associated with nsCL/P, providing novel mechanistic insights into regulation of pathogenesis by CL/P-associated SNPs.

2.2: Results

2.2.1: p63 remodels chromatin to establish enhancers, upregulating epithelial and inflammation genes

To determine whether p63 has the ability to remodel the chromatin landscape to establish epithelial enhancers, we set up a doxycycline-inducible system to express the ΔNP63α isoform (hereafter referred to as p63) in human BJ dermal fibroblasts (ATCC CRL-2522). After 48 hours, p63 reaches an expression level comparable to endogenous levels observed in undifferentiated normal human epidermal keratinocytes, and expression levels are maintained without increase at 72 hours (Supplemental Figure 2.1 a). This p63 isoform was selected because it is the most abundant isoform in epithelial cells, is not expressed in fibroblasts, and has been demonstrated to play a fundamental role as a regulator of epithelial fate and keratinocyte differentiation[22–24].

Expression of p63 was induced for 72 hours, followed by p63 chromatin precipitation (ChIP-seq to assess p63 binding genome-wide), ATAC-seq (to assess open chromatin), and H3K27 acetylation (H3K27ac) ChIP-seq (Fig 2.1a). We observed an increase in ATAC-seq chromatin accessibility and H3K27 acetylation (H3K27ac) at p63-bound sites in fibroblasts expressing p63 compared to control fibroblasts with empty vector (Fig 2.1b, Supplemental Fig. 2.1b). 40K sites were bound by p63, of which 40% were previously closed chromatin sites becoming open only after ectopic expression of p63 (Fig 2.1c, Supplemental Fig. 2.1b; 17156/27441 peaks), while a smaller proportion were unchanged or closed after p63 expression (14991 open before and after; 10302 closed). Further, 40% of these p63 bound sites were flanked by H3K27ac with approximately 60% of these sites showing _de novo_ H3K27ac (Fig 2.1d, Supplemental Fig. 2.1b; 12071/20190). Partitioning of p63 bound sites by distance to transcription start sites (TSS) revealed that the majority of p63 sites are located at intergenic and intronic regions at a median distance of 27kb (Figs 2.1e,f), a distribution that is characteristic of enhancers. Overlaying published micrococcal nuclease (MNase) datasets in human fibroblasts[25] with our p63 ChIP sites revealed that 55% of sites bound by p63 when ectopically expressed are

nucleosome-occupied in fibroblasts, and 75% of these become open and devoid of nucleosomes upon ectopic expression of p63 (Supplemental Fig. 2.1c), expanding on previous work that showed p63 has the ability to bind to closed chromatin and remodel the enhancer landscape during keratinocyte differentiation[8]. Overall, these data reveal that *de novo* binding of p63 to chromatin leads to global changes in the enhancer landscape, specifically, to the establishment of enhancers characterized by *de novo* chromatin accessibility and by H3K27ac flanking p63-bound sites.
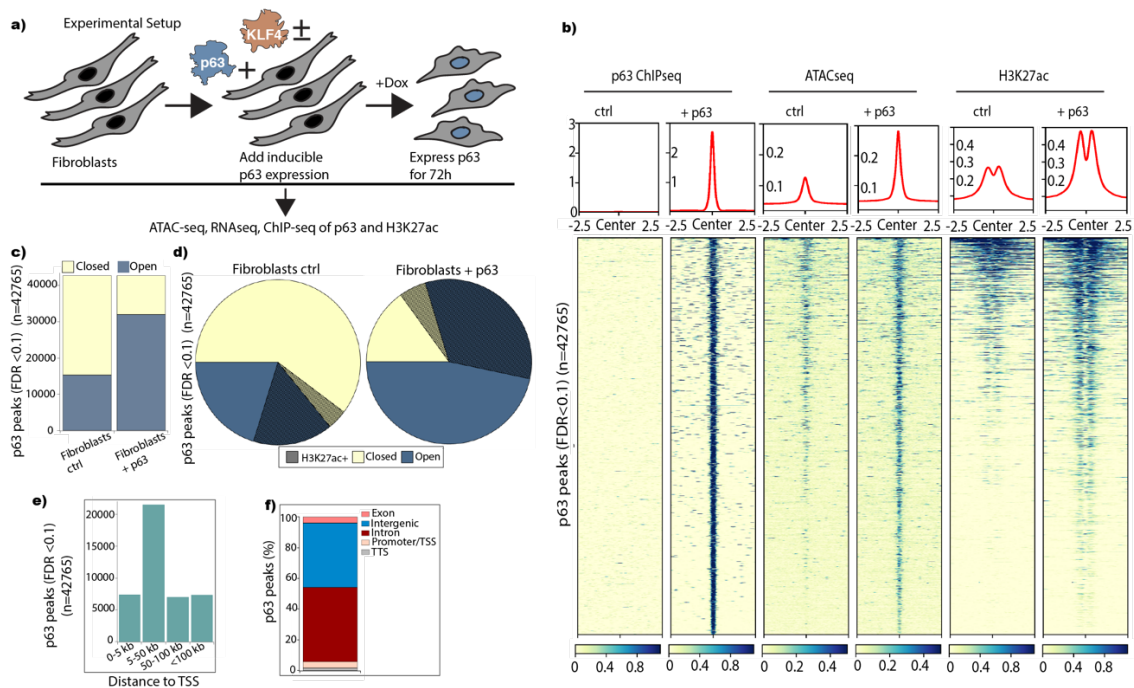
Figure 2.1. p63 remodels chromatin to establish enhancers A) Experimental setup showing inducible expression of p63 and downstream epigenomic analyses. **B)** Heatmap of p63 ChIPseq, ATACseq and H3K27ac ChIPseq (+/- 2.5kb from peak center) in control fibroblasts expressing an empty vector(ctrl) and in fibroblasts expressing p63 for 72 hours. **C)** Graph of called p63 peaks partitioned into open and closed chromatin according to ATACseq results in fibroblasts ctrl and fibroblasts + p63. **D)** Pie chart depicting differences in chromatin landscape at p63 called peaks in ctrl and after ectopic expression of p63 showing clear increase in chromatin accessibility (blue:

open) and H3K27ac (patterned lines). **E)** Distance to nearest TSS for all p63 peaks. **F)** Partitioning of p63 peaks into different genomic features.

We examined the transcriptional consequence of these changes in the chromatin landscape, via RNA-seq after ectopic expression of p63 for 72 hours in fibroblasts. 1960 genes were upregulated upon ectopic expression of p63 (Fig 2.2a; fold change > 1.5 and FDR < 0.05), validating known p63 targets *(IRF6, LCE1C, IVL, ITGB4 and SHH)*[26,27] and identifying over 1000 genes previously not known to be regulated by p63 (full list in Supplementary Table 2.1). Gene ontology (GO) analysis of all upregulated genes showed enrichment of cell-cell adhesion categories characteristic of epithelial cells and important for epithelial lineage specification (Fig 2.2a). Other categories were interferon, inflammation, and immune pathways (Fig 2.2a), consistent with previous reports showing p63 modulates epidermal development and inflammation programs, and plays a crucial role in T-cell development through establishment of the thymic epithelium[28–31]. Additionally, 1479 genes were downregulated upon expression of p63 (Supplemental Fig 2.1d, fold change > 1.5 and FDR < 0.05), consistent with previous reports of p63 playing a repressive role at specific genes (*COL6A2, CITED2, CDH13*)[27,28] and uncovering over 1000 new genes downregulated by p63.

We compared changes in the chromatin landscape to transcription, finding chromatin accessibility and H3K27ac increased at enhancers and promoters of genes that became upregulated upon expression of p63 in fibroblasts (Fig 2.2b,c). The largest fold change in chromatin accessibility was found at genes that became upregulated (Fig 2.2b, Supplemental Fig 2.1e). H3K27ac decreased at genes downregulated and increased dramatically at genes that were upregulated (Fig 2.2c, Supplemental Fig 2.1f). To validate these findings, we examined the previously known p63-regulated gene *IRF6*[21,32,33]: the p63-dependent *IRF6* enhancer showed *de novo* chromatin accessibility and H3K27ac, and establishment of this enhancer was accompanied by *de novo* activation of *IRF6* expression in fibroblasts (Fig 2.1d). We also observed enhancer

establishment and transcriptional activation of genes encoding the F11 receptor (*F11R)* and integrin subunit alpha 7 (*ITGA7*) (Fig 2.2e, Supplemental Fig 2.1g). Thus, our data provide the first evidence that p63 can establish enhancers, revealing known and new gene targets of p63 linked to epithelial cells and inflammation, and uncovering enhancers that might regulate their expression.
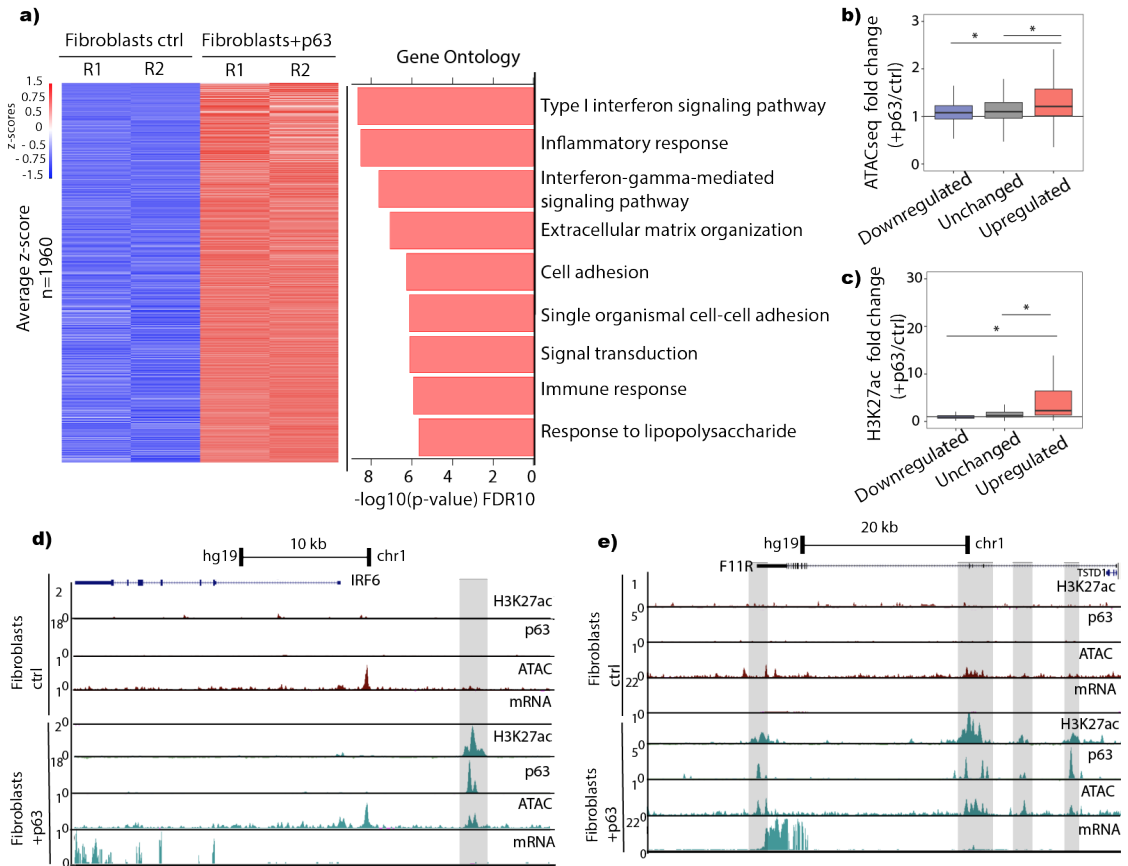
Figure 2.2: p63 upregulates epithelial and inflammation genes downstream of enhancer establishment. A) Heatmap of z-scored RNAseq results for 1960 genes upregulated (fold change > 1.5 and FDR < 0.05) upon ectopic expression of p63, and top gene ontology categories showing enrichment of epithelial and inflammations categories among upregulated genes. **B)C)** Boxplots showing increased chromatin accessibility and H3K27ac at p63/KLF4 sites +/- 250bp near upregulated genes (* p-Value < 10E-10). **D)** UCSC genome browser tracks showing transcriptional activation of *IRF6* and de novo H3K27ac and ATACseq signal at an enhancer

26

upstream of *IRF6* in fibroblasts with ectopic expression of p63 (grey highlighted box). **E)** UCSC genome browser tracks showing transcriptional activation of *F11R* and de novo H3K27ac and ATACseq signal within introns of *F11R in fibroblasts with ectopic expression of p63 (grey highlighted boxes).*

2.2.2: p63 and KLF4 co-establish keratinocyte specific enhancers to convert fibroblasts into keratinocyte-like cells.

These results demonstrate that p63 can establish chromatin structures required for expression of key epithelial specification genes. However, previous findings show conversion of fibroblasts into keratinocyte-like epithelial cells requires co-expression of p63 and Kruppel-Like Factor 4 (KLF4), and these keratinocyte-like epithelial cells share significant morphology, functional characteristics, and gene expression patterns with basal epithelial keratinocytes[34]. To understand the molecular mechanisms and chromatin changes that underlie this crucial conversion, we extended our system to temporal co-expression of p63 and KLF4 in fibroblasts driven by the same doxycycline-induced promoter (see Fig 2.3a, Supplemental Fig 2.2a). Indeed, keratin 14 (KRT14), a key biomarker of basal keratinocytes was upregulated only when p63 and KLF4 were co-expressed, and was not upregulated with expression of either p63 or KLF4 (Fig 2.3a,2.3c, Supplemental Fig 2.2b). Our results thus confirm the requirement for both factors to achieve conversion of fibroblasts into keratinocyte-like cells.

Broad GO analysis of the 2213 genes upregulated upon conversion showed high enrichment of gene classes encompassing epidermal development, cell adhesion, and other epithelial specific programs (Fig 2.3b, Supplementary Table 2.2). Many genes critical for keratinocytes, including *FOXN1*, *DSC2* and *DSG1,* were newly expressed (Fig 2.3c and Supplementary Table 2.2), and others became upregulated over 1000-fold, including *KRT14*, *KRT6a*, *GJB4* and *LAD1* (Fig 2.3c). As with ectopic expression of p63 alone, approximately 1000

genes were downregulated with ectopic expression of p63+KLF4 (Supplemental Fig 2.2c, fold change > 1.5 and FDR < 0.05), including previously known repressed targets (*CITED2*, *SATB2*, *TGFB1I1*, *SHOX2*)[27]. Overall, the fold decreased enrichment and GO category specificity of downregulated genes was much lower compared to the upregulated genes (Supplemental Fig 2.2c). Importantly, many key genes required for epidermal development and epithelial lineage specification were only upregulated in the combined system, including *JAG2, GRHL3* and *SFN* (Supplemental Fig 2.2d, Supplementary Table 2), and, importantly, their dysregulation has been implicated in developmental malformations including skeletal dysplasias and orofacial clefting[29,35,36]. Additionally, the top 15 upregulated genes (via fold change p63+KLF4/ctrl) by the combined system includes keratinocyte specific genes (e.g. *KRT6a, KRT14*) that are key mediators of cell-cell adhesion programs (e.g. *CDSN, LAD1, GJB4*), as well as keratinocyte differentiation (e.g. *TGM5, FA2H*), all of which were not upregulated when p63 was expressed without KLF4 (Fig 2.3c).
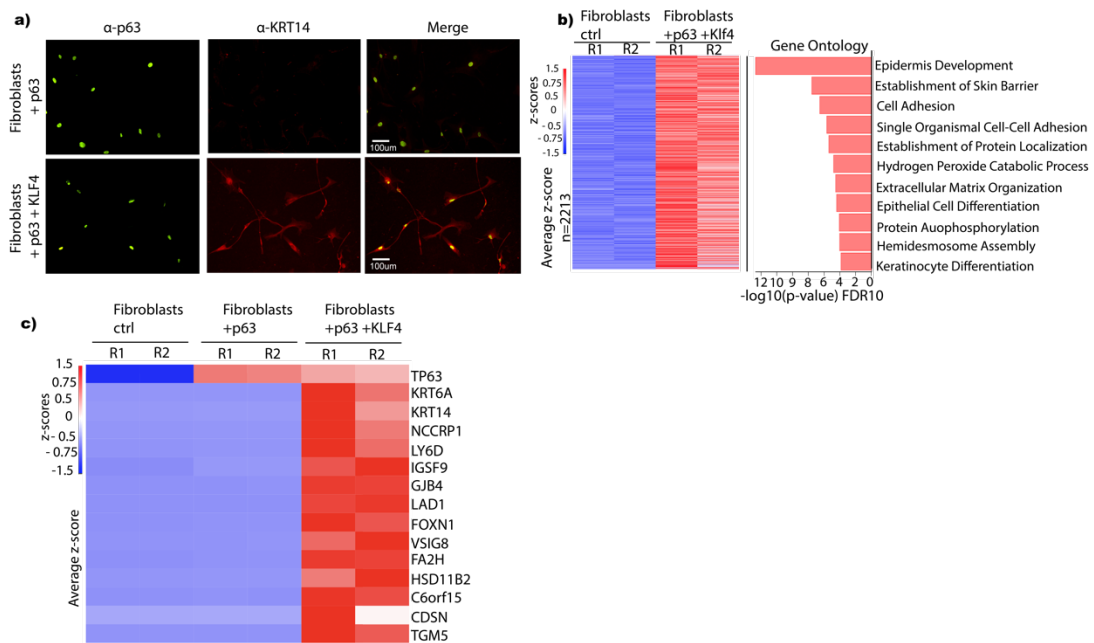
Figure 2.3: Ectopic co-expression of p63 and KLF4 convert fibroblasts into keratinocyte-like cells. **A)** Immunofluorescence of p63 and KRT14 in fibroblasts + p63 and fibroblasts

+p63+KLF4 showing KRT14 is only upregulated when both p63 and KLF4 are expressed. **B)** Heatmap of z-scored RNAseq results for 2213 genes upregulated (fold change > 1.5 and FDR < 0.05) upon ectopic expression of p63+KLF4 showing high enrichment of epidermal and skin related GO categories. **C)** Comparison of transcriptional regulation of the top 15 genes upregulated in fibroblasts + p63+KLF4 showing both factors are required for upregulation of keratinocyte specific genes.

To determine whether changes in transcription were a result of upstream enhancer establishment by p63 and KLF4, we induced expression of both proteins for 72 hours, followed by ATAC-seq and ChIP-seq of p63, KLF4 and H3K27ac. Our data revealed p63 bound to over 40K sites in the genome and KLF4 to approximately 25K sites (Fig 2.4a, Supplemental Fig 2.2e,2.2f). Because both proteins are needed for trans-differentiation, we hypothesized that co-bound sites might play a crucial role in establishing key keratinocyte-specific enhancers. Our ChIP-seq experiments revealed 13488 sites were co-bound by p63 and KLF4 (Fig 2.4a), primarily at intronic, intergenic and promoter regions (Supplemental Fig 2.2g,2.2h). Sites co-bound by p63 and KLF4 showed a clear increase in chromatin accessibility at 80% of the sites, with 30% of the sites became newly open after ectopic expression of p63 and KLF4 (Fig 2.4b, Supplemental Fig 2.2i), and 20% of shared sites became newly decorated by H3K27ac (Fig 2.4c, Supplemental Fig 2.2i). Direct comparison of specific shared p63/KLF4 bound peaks in basal keratinocytes[8] to fibroblasts, to fibroblasts + p63, and to fibroblasts + p63+KLF4 revealed highest similarity between H3K27ac marked enhancers in keratinocytes and those in fibroblasts +p63+KLF4 (Fig 2.4d, Supplemental Fig 2.2j). These results show p63 and KLF4 function together to establish chromatin structures concordant with those in basal keratinocytes.

Comparison of chromatin changes to gene expression revealed decreased chromatin accessibility at downregulated genes, as well as at genes that remained unchanged (Supplemental Fig 2.2j). Chromatin accessibility was increased only at promoters and enhancers

of genes becoming upregulated upon conversion (Fig 2.4e, Supplemental Fig 2.2k). Furthermore, H3K27ac decreased at genes that became downregulated, and profound increase in H3K27ac in upregulated genes, reaching more than 20-fold increase in converting cells compared to controls (Fig 2.4f, Supplemental Fig 2.2l). At the local level, a genome browser view of the *KRT14* locus exemplifies these dramatic changes in chromatin landscape and transcription (Fig. 2.5a). In fibroblasts +p63+KLF4, establishment of enhancers occurred at intronic and intergenic regions around *KRT14*. The establishment of these enhancers was accompanied by a stark increase in *KRT14* expression (Fig 2.5a). As mentioned above, *IRF6* is a key gene for CL/P, and KLF4 was also co-bound with p63 to the *IRF6* enhancer with accompanying chromatin changes (Fig 2.5b).
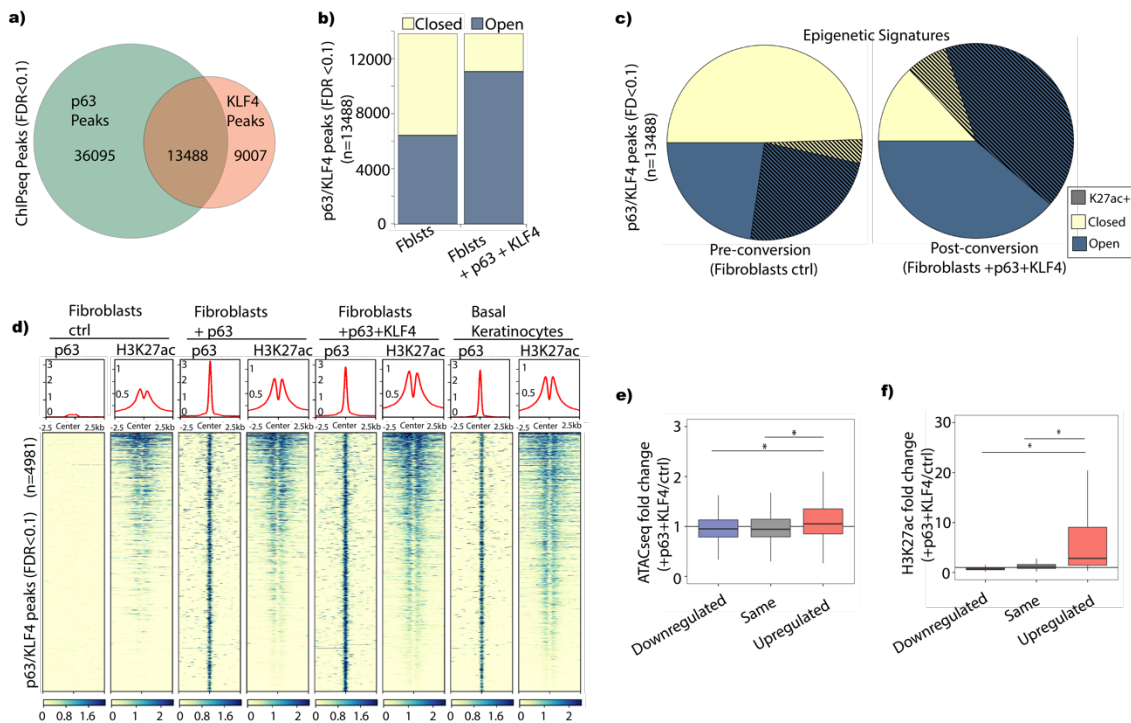
Figure 2.4: p63 and KLF4 co-establish keratinocyte specific enhancers. A) Venn diagram showing p63 and KLF4 are cobound in 13488 loci in the genome. **B)** Graph of p63/KLF4 peaks partitioned into open and closed chromatin according to peaks called from ATACseq showing increased chromatin accessibility after 72 hours at co-bound sites. **C)** Pie charts showing stark

increase in chromatin accessibility and H3K27ac at p63/KLF4 peaks. **D)** Heat map showing

H3K27ac enrichment flanking p63/KLF4 peaks (+/- 2.5kb from peak center) shared in p63-

ChIPseq of basal keratinocytes, comparing fibroblasts, fibroblasts +p63, fibroblasts +p63+KLF4

and reanalyzed basal keratinocyte data from[8] **E)F)** Boxplots showing increased chromatin

accessibility and H3K27ac at p63/KLF4 sites near upregulated genes (* p-Value < 10E-10).
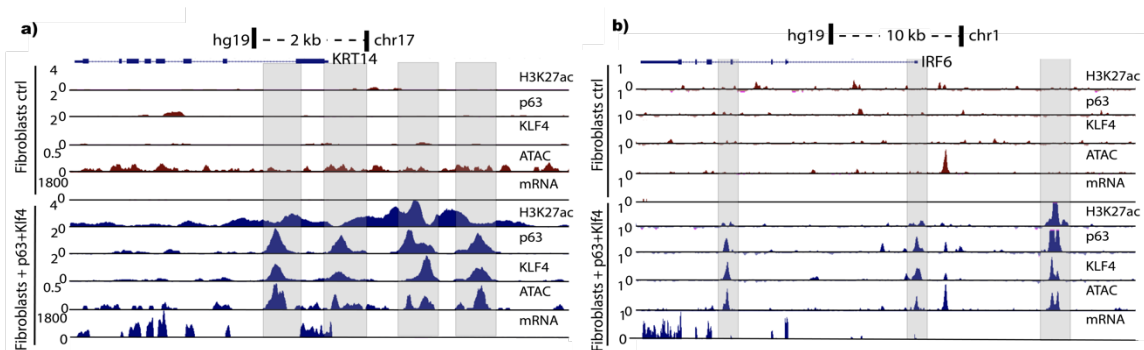
Figure 2.5: Local chromatin and transcriptional changes catalyzed by p63 and KLF4. A)B)
UCSC genome browser tracks showing *de novo* H3K27ac and ATACseq signal at enhancer and
promoter regions close to keratinocyte genes *KRT14* and *IRF6*, as well as transcriptional
activation (grey highlighted box).

Overall, our results show p63 and KLF4 convert fibroblasts into keratinocyte-like cells via

co-establishment of keratinocyte-specific enhancers. Specifically, p63 alone was able to establish

epithelial enhancers to upregulate 32% of genes upregulated by p63 + KLF4 (710/2214

upregulated genes) (Fig 2.1f, Supplementary Table 2.1, e.g. *IRF6, F11R, ITGA7*). However,

importantly, both p63 and KLF4 were required to establish keratinocyte-specific enhancers and

upregulate their respective genes, such as *KRT14*, *KRT6A* and *JAG2*. Because fibroblasts do not

express p63, this trans-differentiation model constitutes a tractable system to investigate

mechanisms underlying wild type and mutant p63 function in keratinocyte enhancer specification

and to uncover key disease-specific processes. Because p63 represents a key transcription

factor driving epithelial specification and because p63 mutations are implicated in cleft lip palate and other orofacial developmental anomalies, we focused on further characterization of p63.

2.2.3: Disease-specific p63 mutations found in patients with craniofacial malformations show defects in enhancer establishment.

Heterozygous point mutations in the DBD and SAM domains of p63 have been shown to lead to limb and craniofacial malformations in human patients[4–6]. While mutations within the DBD are likely to be deleterious due to loss of DNA binding[37–39], mutations within the SAM domain are likely to lead to dysregulation of protein-protein interactions needed for epithelial lineage specification[40–42].

We selected two point mutations for this study: p63R304W (Fig 2.6a, orange), a point mutation in the DBD occurring in patients with ectrodactyly ectodermal dysplasia cleft lip/palate syndrome (EEC)[7] and p63I537T(Fig 2.6a, blue), a point mutation in the SAM domain found in patients with ankyloblepharon-ectodermal dysplasia and cleft lip/palate syndrome (AEC)[39]. Since characterized patients carry heterozygous p63 mutations, our homozygous model captures a severe scenario, with mutation of all p63 molecules within a tetramer; this approach allows investigation of substitution mutations in a simplified context.

We introduced p63R304W (mtDBD) or p63I537T (mtSAM) into our inducible conversion system with co-expressed KLF4, for 72 hours (Fig 2.1a). Cellular immunofluorescence for KRT14 revealed defects in conversion for both mutants (Fig 2.6b). As expected, the deleterious effects of mtDBD were most severe, showing no induction of KRT14 expression after 72 hours, whereas mtSAM showed reduced KRT14 expression compared to WT p63 (Fig 2.6b). RNAseq revealed that either substitution mutation poorly induced the top 15 genes upregulated by WTp63+KLF4, including epithelial-specific genes *KRT6A, GJB4, LAD1* and *CDSN* (Supplemental Fig 2.3a; Supplementary Table 2.3). Further, among the 3367 differentially expressed genes in

WTp63+KLF4, nearly all failed to be induced in the mtDBD+KLF4 compared to control fibroblasts (Fig 2.6c). Hence, loss of p63 binding to DNA leads to complete loss of epithelial enhancer establishment in fibroblasts and nearly abrogates differential gene expression [16/3367 genes became upregulated (Supplementary Table 2.3)]. In contrast, mtSAM showed an intermediate expression pattern between WT+KLF4 and control fibroblasts, indicating specific defects in enhancer establishment and downstream epithelial gene induction (Fig 2.6c). Given this latter interesting specificity in defects, we focused on the SAM mutant p63I537T to dissect which deficiencies underlie the disease-specific phenotypes observed in AEC.
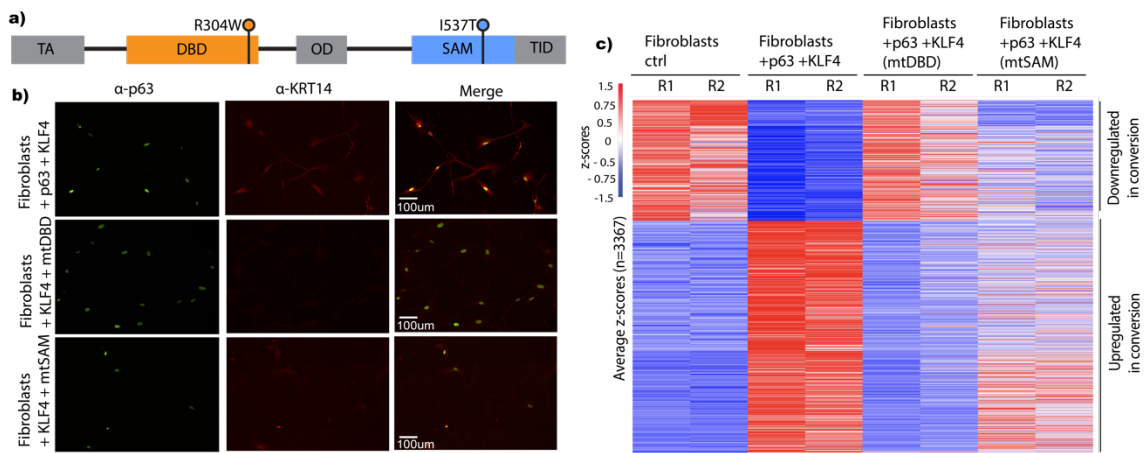
Figure 2.6: Disease specific p63 mutations found in patients with craniofacial malformations show defects in converting fibroblasts to keratinocyte-like cells. A) Linear schematic of p63, showing domain structure and selected patient derived mutations **B)** Immunofluorescence of p63 and KRT14 in fibroblasts + fibroblasts +KLF4 and +WTp63, mtDBD or mtSAM showing KRT14 is not upregulated when the DBD is mutated and induction of KRT14 is lower in mtSAM **C)** Heatmap of z-scored 3367 differentially expressed genes (fold change > 1.5 and FDR < 0.05) between fibroblasts control and fibroblasts +p63+KLF4 showing mtDBD has almost no change in transcriptional profile and mtSAM shows a transcriptional profile in between fibroblasts ctrl and fibroblast +WTp63+KLF4.

We investigated the effect of the mtSAM on p63 binding genome-wide. Among the 13488 p63/KLF4 DNA binding sites in WTp63+KLF4 conversion (Fig 2.4a), most (71%; 9513) of the p63 binding peaks were retained in mtSAM+KLF4, and only 29% (3975) of p63 bound peaks were lost (Supplemental Fig 2.3b). The partial loss of mtSAM binding to DNA may be due to increased protein aggregation caused by SAM domain mutations within p63[43]. However, despite largely maintaining binding to DNA (Fig 2.7a, left panel), there was no increase in chromatin accessibility (Fig 2.7a, ATAC-seq) at retained mtSAM/KLF4 sites (Fig. 2.7a, 2.7b left). Furthermore, while most H3K27ac peaks were retained (Fig 2.7a), overall the amplitude of H3K27ac peaks increased only modestly compared to WT+KLF4 (Fig 2.7b right). Taken together with the observation that KLF4 binding is retained in the majority of these peaks (Fig 2.7a), these results suggest that the SAM domain of p63 is crucial for chromatin accessibility and establishment of H3K27ac in epithelial enhancer specification.

We examined how these chromatin changes affect downstream gene regulation of the closest genes to p63/KLF4 bound sites. RNA-seq analysis revealed 50% of genes upregulated by p63/KLF4 were no longer upregulated in mtSAM (~1200 genes; Fig 2.7c), including key genes involved in craniofacial development, such as *GRHL3*, *JAG2*, *LAMB3* and *IRF6* (Supplemental Figure 2.3c, Supplementary Table 2.3). We assessed whether genes that were able to maintain activation in mtSAM/KLF4 had pre-existing enhancers, which might explain their continued expression. Indeed, genes that maintained expression with mtSAM/KLF4 were near pre-established and H3K27ac pre-decorated enhancers (Fig 2.7d left). In contrast, genes near newly established enhancers failed to be upregulated in mtSAM/KLF4 (Fig 2.7d right). Additionally, ~200 genes were aberrantly upregulated in mtSAM/KLF4, including genes involved in craniofacial development in non-epithelial cell types, e.g. *BMP2*, *VAX2* and *GAD1* (Fig. 2.7b, Supplementary Table 2.3).

At the local level, *FOXN1* (Fig 2.7e), *IRF6* (Fig 2.7f) and *LAMB3* (Supplemental Fig 2.3d) showed defects in enhancer establishment despite retained binding of mtSAM and KLF4. In

particular, chromatin remained largely closed with mtSAM (in red highlighted box) and reduced H3K27ac flanking these sites, and these defects in establishing normal enhancer chromatin landscape lead to profound failure in upregulating associated genes *FOXN1, IRF6* and *LAMB3* (Fig 2.7e,2.7f and Supplemental Fig 2.3a,2.3c).
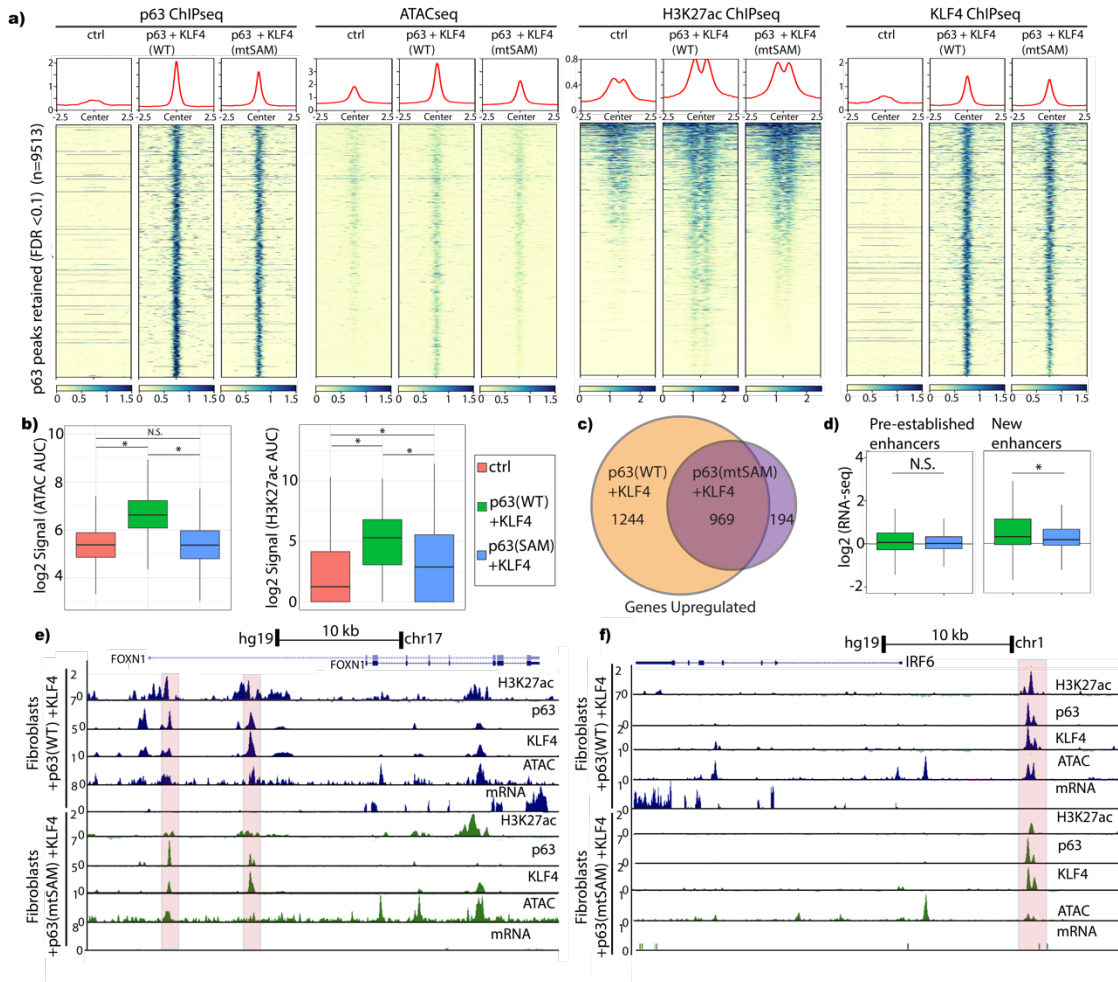
Figure 2.7: Mutations in the SAM domain of p63 found in AEC syndrome show defects in establishing enhancers and downstream activation of epithelial genes. A) Heatmap showing a stark reduction in chromatin accessibility (ATACseq signal) in mtSAM compared to WTp63, decreased H3K27ac and retained p63 and KLF4 binding. **B)** Box plots showing no increase in chromatin accessibility for mtSAM, as well as reduced enrichment of H3K27ac flanking p63/KLF4 peaks. **C)** Venn diagram showing over 1000 genes are no longer upregulated in mtSAM+KLF4

and about 200 genes are upregulated *de novo* by this mutant. **D)** Box plots showing RNA-seq signal of genes close to p63/KLF4 retained peaks separated into peaks at pre-established enhancers in fibroblasts and newly established enhancers after ectopic expression of p63/KLF4. **E)F)** UCSC genome browser tracks showing defects (red boxes) in establishing open chromatin and inducing gene expression at *FOXN1* and *IRF6*.

Together our chromatin and transcription data show a dramatic difference in the roles of the DBD and the SAM domains of p63. The DBD is essential for enhancer establishment due to loss of function in DNA binding. Importantly, while the SAM mutant is able to bind to DNA, there are profound defects in both chromatin opening and establishment of H3K27ac, correlating with defective upregulation of proximal epithelial genes, Thus, because the SAM domain mutant I537T is found prominently in AEC syndrome, our results identify a potential explicit molecular mechanism, in which mtSAM drives pathogenesis through aberrant enhancer establishment (despite maintaining DNA binding) leading to low levels of specific critical craniofacial genes, such as *IRF6, JAG2* and *GRHL3*.

2.2.4: p63 regulates expression of critical genes associated with cleft lip/palate.

While EEC and AEC are uncommon syndromes[5,6], orofacial clefting is a key feature of these pathologies and clefting is a common developmental defect throughout the world[11]. Mouse genetic models (including knockouts of *Trp63* and *Irf6*) have been crucial in understanding lip and palate development, however they do not fully recapitulate defects observed in human patients[15,16]. To date, 16 genes have been strongly implicated in nonsyndromic cleft lip+/- cleft palate (nsCL/P) via related phenotypes in mouse KO and knock down (KD) studies[44] (Supplementary Table 2.4a). Of the 16 reported genes, 5 were upregulated in our trans-differentiation model (*ESRP1*, *IRF6*, *FOLR1*, *FGFR8*, *WNT9B*) (see Fig 2.5b, Supplemental Fig

2.2d, Supplementary Table 2.2 and Supplementary Table 2.4a), and *ESRP1, IRF6* and *FOLR1* also showed robust dysregulation in mtSAM (Supplemental Fig 2.4a). Importantly, of these genes, only *IRF6* was previously reported to be regulated by p63[21,27,28], providing novel mechanistic insights into clefting pathogenesis via p63 mtSAM from our genome-wide study.

Genetic Association Database studies[45] of genes upregulated in our trans-differentiation model identified cleft lip and/or cleft palate as the most highly enriched disease associated category (Fig 2.8), with 76 associated genes identified, including genes known to cause nsCL/P (e.g.*WNT9B, JAG2, and CDH1*; full list in Supplementary Table 2.4b)[29,35,46,47]. Many of these 76 genes have been identified as causal for orofacial clefting through human syndromes and family studies, while others are in proximity of SNPs involved in CL/P but are not known to be causally or functionally involved with orofacial clefting.
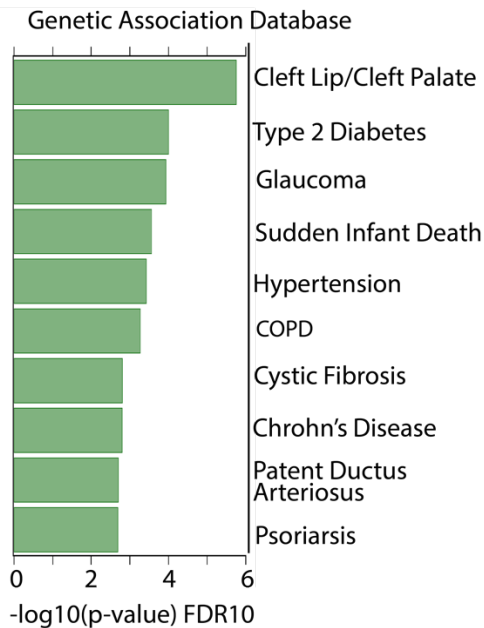
Figure 2.8: Transdifferentiation model shows genes upregulated by p63 and KLF4 have been associated to cleft lip/palate. Genetic association database results showing cleft lip/cleft palate is the most enriched disease category among the 2213 genes upregulated (fold change > 1.5 and FDR < 0.05) by p63 and KLF4.

To date, genome-wide association studies (GWAS) and other approaches have identified 40 human risk loci for nsCL/P with conclusive evidence (Supplementary Table 2.4c)[14,17–20]. These 40 loci are estimated to explain about 25% of the heritability reported for nsCL/P. However, the causative genes and mechanisms at the majority of these loci are unknown. While previous studies have shown the importance of neural crest cells in CL/P development[14], different tissue systems are expected to contribute to this complex disease. In particular, keratinocytes from patients with CL/P have recently been reported to display differentiation defects[48–51]. Based on the key observation that p63+KLF4 establish enhancers to regulate keratinocyte specific genes, including critical craniofacial genes (see Supplemental Fig 2.2d, Supplementary Table 2.2), we investigated the genes located within the topologically associated domain (TAD)[52,53] overlapping each of the 40 risk loci. The reasoning to examine gene association within TADs is because gene regulatory enhancer-promoter contacts occur primarily within TADs[52–54]. We found that within the associated TAD of 17 of these risk loci, at least 1 gene, and a total of 37 genes were upregulated in our trans-differentiation model (Fig 2.9, Supplementary Table 2.4c). We conclude that crucial regulatory interactions occur within TADs encompassing genes induced by p63/KLF4.
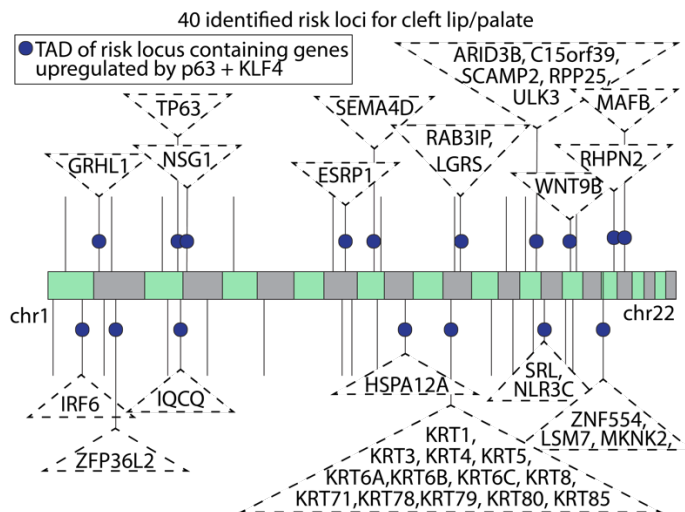
Figure 2.9: p63 and KLF4 upregulate genes within TADs containing cleft lip/palate risk loci. 40 identified risk loci for cleft lip/palate indicated by black lines at the respective chromosomes

with dark blue circles at TADs containing at least 1 gene upregulated by both p63 & KLF4(respective genes in dotted triangles).

To analyze whether any of the 37 genes (Fig 2.9b) located at the risk loci and/or their enhancer regions might contribute to nsCL/P, we used SNPs and their p-values from a recent GWAS on nsCL/P[14] to compute gene-based association p-values using MAGMA[55](Supplementary Table 2.4d). We first validated our approach by examining genes upregulated by p63+KLF4 within and outside of the 40 risk loci that have been shown to cause nsCL/P[21,47,56]. We identified significant p-values for known causal genes within the 40 risk loci, such as in the 1q32 locus (*IRF6*), 17q21 (*WNT9B*) and 8q22.1 (*ESRP1*) loci (Fig 2.10, red). Our approach also identified significant p-values for known causal genes outside of the risk loci, including *JAG2, CDH1, and CTNND1*[29,46,57] (Fig 2.10, red with *).
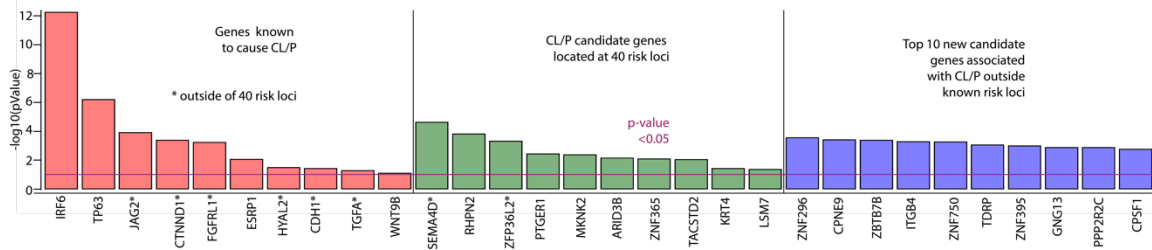
Figure 2.10: p63 regulates expression of critical genes associated with orofacial clefting. List showing association values of top 10 genes calculated by gene-based analysis with MAGMA as implemented in FUMA[73] in the following categories: (red) genes known to cause CL/P with * marking genes outside of the known 40 risk loci; (green) CL/P candidate genes located at 40 risk loci with significant (p<0.05) association to CL/P and (blue) new candidate genes associated with CL/P outside of known risk loci.

Within the 40 risk loci we identified 10 candidate genes upregulated by p63 with significant gene based association p-values to nsCL/P (p-value<0.05), including *SEMA4D*, *BEST3*, *ZFP36L2,* and *KRT4* (Fig 2.10, green). Interestingly, in three of the risk loci, our data point towards a risk gene for nsCL/P that is different from the gene previously suggested: At 2p21, *THADA* was previously suggested to be the risk gene[58], our data suggests it is *ZFP36L2* (p<0.05) (Fig 2.10, green with *), a gene shown to play a role in basal keratinocytes[59]. At 2p25.1 *KLF11* had been suspected to be the nsCL/P risk gene, our data suggests it is *GRHL1* (p<0.5) a close family member of *GRHL3*, a well-established risk gene for cleft palate[60,61]. Further, at risk loci 9q22.2 our data prioritize *SEMA4D* (p<10E-05) (Fig 2.10, green with *) instead of *GADD45G*[61], as the probable nsCL/P risk candidate.

Finally, to identify novel risk loci and novel causal candidate genes for nsCL/P, we extended our gene based association p-value study to genes upregulated in our trans-differentation model outside of the known 40 risk loci. We identified 111 novel candidate genes showing high association to nsCL/P (p-values <0.05 computed as above, Supplementary Table 2.4d). Several identified genes, including *ZNF296, CPNE9, ITGB4,* and *ZNF750* (Fig 2.10, blue) play key roles in epithelial cells but have no known role in CL/P. Regional association plots for *ZNF296* and *CPNE9* demonstrating the local association structures at these genes in the GWAS data reveal SNPs with nominal statistical significance at each of these genes and also in the surrounding non-coding areas (Fig 2.11a,b). The location of these SNPs suggest that common genetic variants within or near these genes regulated by p63 might alter gene function or expression and, thus, may contribute to nsCL/P, but the genes have escaped detection in genetic studies due to lack of statistical power (such as sample size).

Overall, our trans-differentiation model in human cells provides an orthogonal approach to identify candidate genes in nsCL/P directly regulated by p63. Indeed, even at established risk loci, we have used our approach to find key p63-regulated genes, as the most likely CL/P-associated genes (e.g. *ESRP1, GRHL1*, *RHPN2,* and *ARID3B*). Further we use this approach to

identify new nsCL/P specific candidate genes and the enhancers that regulate them (e.g. *ZNF296, CPNE9, ITGB4, and TDRP*).
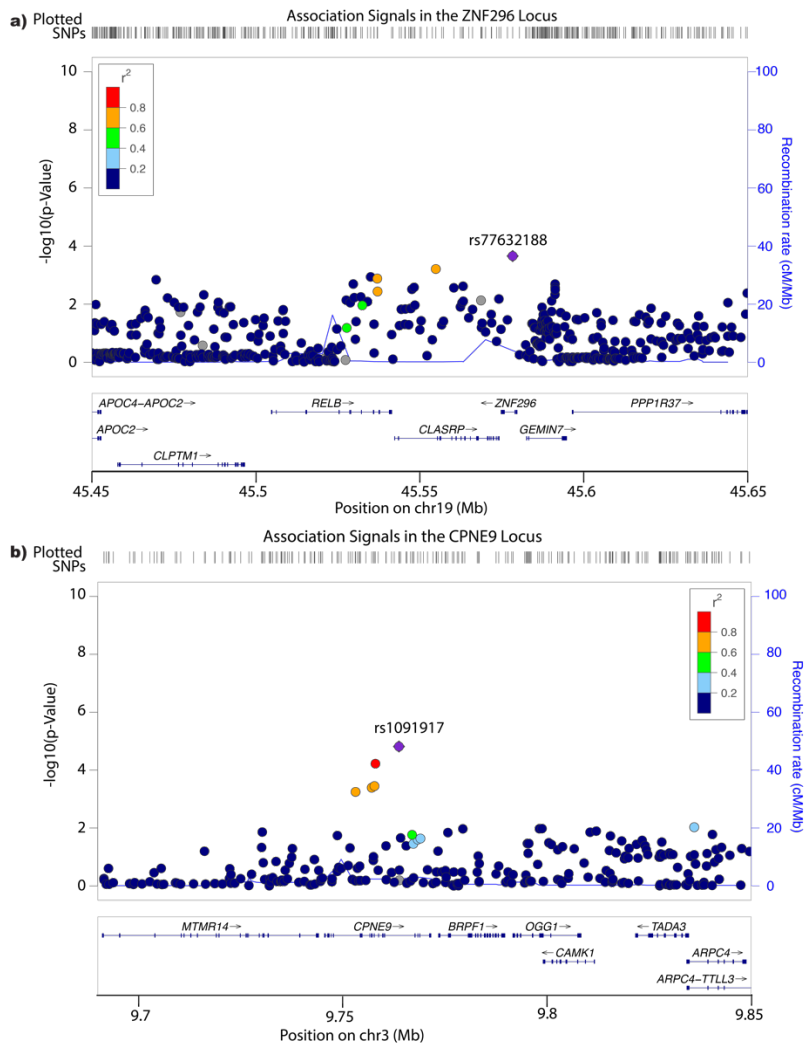
Figure 2.11: Transdifferentiation model reveals novel nsCL/P candidate genes and the enhancers that regulate them. A)B) Regional association plots of *ZNF296* and *CPNE9*, showing nsCL/P associated SNPs with nominal significant gene-based p-value (<0.05) within the genes and up to 200kb away from transcription start sites.

2.2.5: SNPs associated with cleft lip/palate are highly enriched in enhancers established by p63 and KLF4

GWA studies of SNPs have identified several genetic variants strongly associated to nsCL/P[14,17–20]. Many of these disease-associated SNPs are located outside of exons, thus suggesting a gene expression regulatory role, such as enhancers and promoters. Since orofacial clefting is a major phenotype in patients with point mutations in p63, and, as we show here, since p63 plays a key role in establishing epithelial enhancers, we explored potential overlap between p63 binding sites and nsCL/P-associated SNPs identified in GWA studies. We curated three lists of p63 binding sites: (A) all p63 peaks in fibroblasts +p63; (B) all p63 peaks in fibroblasts +p63+KLF4; (C) p63/KLF4 shared peaks in fibroblasts +p63+KLF4 (i.e. a subset of (B)). Using the same GWAS dataset as above[14] and the enrichment-tool GREGOR[62], we determined that list C - p63 peaks co-bound by KLF4 - showed strongest enrichment of nominal significant (p<0.05) nsCL/P SNPs (Fig 2.12a). We focused on H3K27ac flanking these sites as a mark of active enhancers, partitioning H3K27ac peaks into enhancers that are preestablished (already marked by H3K27ac in fibroblasts) or into new enhancers (only become marked by H3K27ac after fibroblast to keratinocyte-like conversion) (see Fig 2.4c, 2.4d). Enrichment analysis showed significant enrichment (p<0.05) of nominal significant nsCL/P SNPs only for newly established enhancers (Fig 2.12a).

Based on the results from GREGOR we performed colocalization analysis of nsCL/P SNPs, focusing on the p63/KLF4 shared peaks[14]. This analysis further accounts for the strength of association to nsCL/P of each SNP, with highly associated SNPs contributing to a lower overall co-localization p-value (see Methods). We observed significant colocalization (p-value <0.05) for p63/KLF4 peaks in which active enhancers are established (newly marked by H3K27ac), whereas shared p63/KLF4 peaks at preestablished enhancers and peaks without H3K27ac did not achieve statistical significance (Fig 2.12b).

We next focused on H3K27ac peaks - which are broader than p63/KLF4 peaks - because SNPs close to, but outside of p63/KLF4 peaks, might be in linkage disequilibrium and contribute to risk of CL/P. We performed colocalization analysis of nsCL/P SNPs in H3K27ac peaks in control fibroblasts, in preestablished enhancers, and in new enhancers flanking p63/KLF4 peaks (Fig. 2.12c). Here only enhancers newly established by p63/KLF4 showed strong colocalization (p-value <10E-4) (Fig 2.12c), confirming the results obtained in the GREGOR analysis.
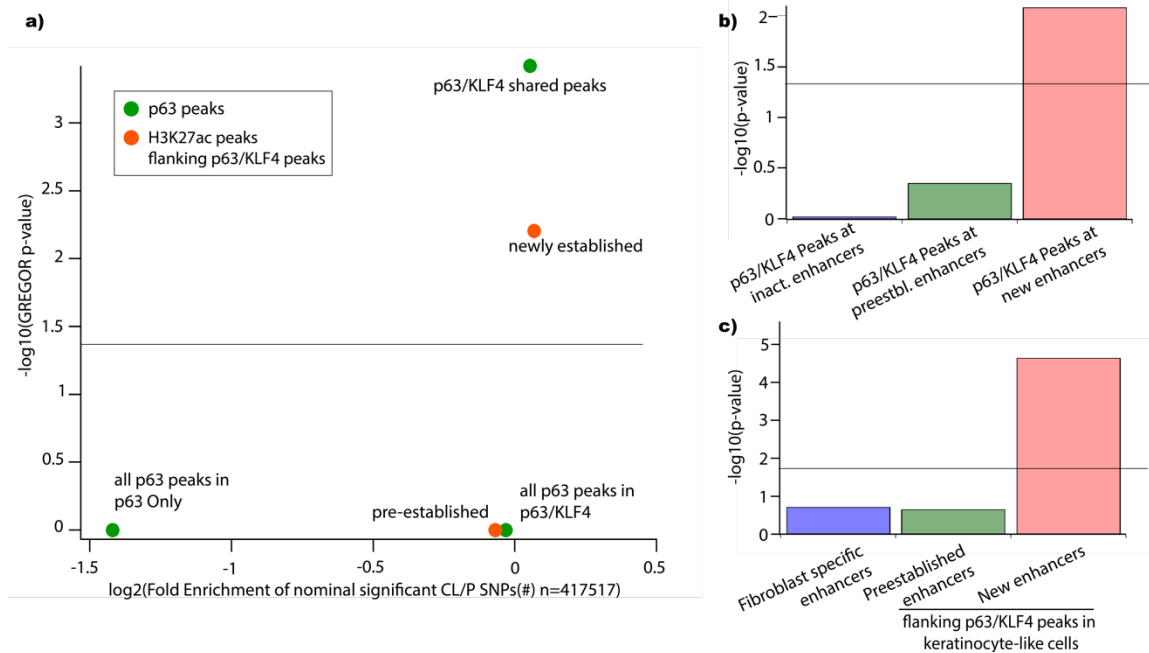
Figure 2.12: SNPs associated with non-syndromic cleft lip with or without cleft palate are highly enriched in enhancers established by p63 and KLF4. A) GREGOR analysis showing p63/KLF4 shared peaks (green) and newly established H3K27ac (orange) flanking these peaks are significantly (p<0.05) enriched for SNPs associated to nsCL/P. B) Co-localization of shared p63/KLF4 peaks at inactive/preestablished/new enhancers, with nominal significant nsCLP associated GWAS SNPs. Graph shows only p63/KLF4 peaks flanked by *de novo* H3K27ac peaks are enriched for nsCL/P associated SNPs. C) Co-localization of H3K27ac peaks with nominal significant nsCLP associated GWAS SNPs. Graph shows only new enhancers in converted cells flanking p63/KLF4 peaks are significantly enriched for nsCL/P associated SNPs.

Genome browser views overlaid with nsCL/P SNPs allowed visualization of the high colocalization of SNPs at p63/KLF4 peaks and newly established enhancers (ATAC-seq and H3K27ac peaks) close to *MAFB* (Fig 2.13a), at the known 8q24 risk locus[63] (Fig 2.13b), at the nsCL/P causing gene *IRF6* (Supplemental Fig 2.5a) and near *RHPN2* (Supplemental Fig 2.5b). Importantly, the genes downstream of these regulatory elements are activated in our trans-differentiation model (see RNA-seq in Fig 2.13a,b and Supplementary Fig 2.5a,b), indicating they are upregulated by establishment of new enhancers by p63/KLF4.
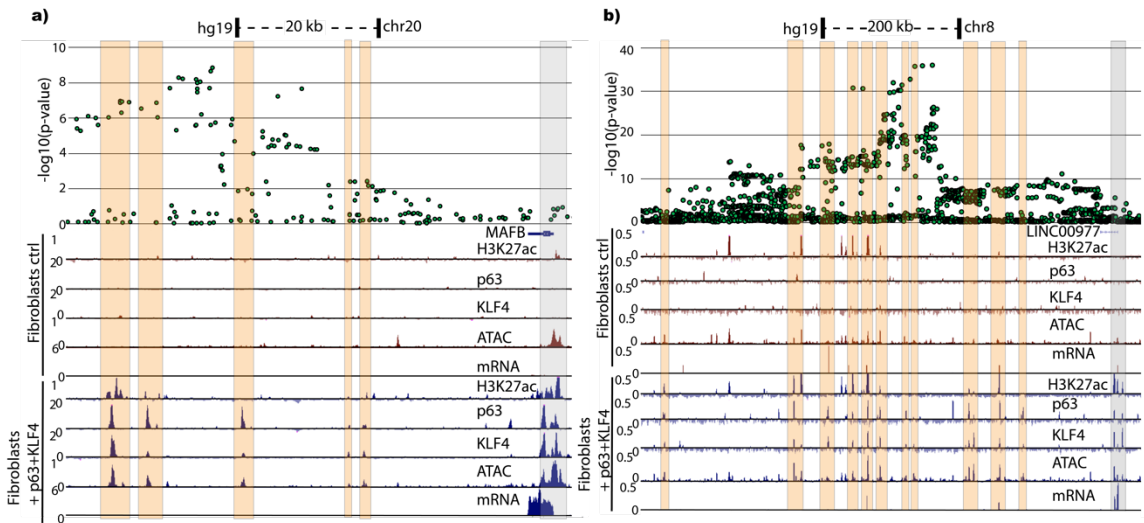
Figure 2.13: Enhancers established by p63 and KLF4 at MAFB and the 8q24 locus are enriched for SNPs associated with nsCL/P. A) and B) Overlay of SNPs and UCSC genome browser tracks highlighting in orange that p63/KLF4 peaks strongly colocalize with highly associated nsCL/P SNPs near *MAFB* and the known 8q24 locus, in grey binding of both proteins at promoters.

Overall, GREGOR analysis and colocalization results show enhancers established by p63/KLF4 are highly enriched for SNPs associated with nsCL/P. These SNPs likely affect downstream expression of epithelial specific genes and lncRNAs required for proper lip and palate development, as observed with *IRF6* and mutations in an upstream regulatory element of

this gene established by p63[21]. Thus our unbiased trans-differentiation model in conjunction with GWAS data provides an independent method to identify driver genes. The method also provides novel mechanistic insights into how nsCL/P-related SNPs might be regulating pathogenesis via disruption of p63/KLF4 binding sites crucial for enhancer establishment and downstream gene regulation.

2.3: Discussion

Here we report a trans-differentiation model as a novel approach to determine the role of p63 in craniofacial development. We show that p63 can establish epithelial enhancers to regulate downstream genes and that this process is disrupted when p63 carries mutations derived from human disease. In particular, our results show a novel mechanism underlying pathogenic p63 mutations, in which failure to open and establish enhancers leads to transcriptional dysregulation. We integrate our epigenomic data with GWAS datasets derived from nsCL/P patients to identify novel CL/P candidate genes regulated by p63. We also uncover strong enrichment of nsCL/P-associated SNPs at enhancers established by p63, providing mechanistic insight into how these SNPs lead to human pathogenesis.

A key role of transcription factor p63 in mediating epidermal commitment, development, and differentiation has been extensively demonstrated through mouse models and causative point mutations in human disease. In mice, knockout of p63 leads to developmental and morphological defects in the squamous epithelia and epidermis, causing abnormal craniofacial development, truncated limbs, and loss of salivary glands, hair follicles, and teeth[2,3]. To date, how p63 engages and modifies chromatin to promote normal development during these early stages is still being elucidated. Here, we report key observations that p63 can establish enhancers and that defects in this process may underlie human craniofacial malformations.

45

Previous studies of p63's engagement with chromatin have focused on its role in keratinocyte differentiation[8,10,51,64,65], while *de novo* engagement with chromatin has not been extensively addressed due in part to lack of tractable human models. Thus, to examine the first encounter of p63 with chromatin, we expressed p63 ectopically in a temporally controlled manner in human BJ fibroblasts, a cell line that does not express p63. We show p63 establishes enhancers to upregulate epithelial and inflammation genes. We next developed an inducible trans-differentiation model based on published reports showing p63 co-expressed with KLF4 can convert fibroblasts into keratinocyte-like cells[34]. We determined that co-binding of p63 and KLF4 leads to robust establishment of enhancers that correspond remarkably well to *bone fide* enhancers in basal keratinocytes. Thus, this model system recapitulates the epigenetic conditions of early epithelial development.

Mutations in the p63 DBD in EEC patients have been previously shown to cause loss of downstream transcriptional activation of known p63 targets[40,51,66]. Further studies showed that p63 mutations in AEC also disrupted transcription and caused p63 aggregation[43,40–42]. In particular, it has been proposed that aggregation of p63 in AEC leads to decreased binding of p63 to its DNA targets and that this is a key underlying disease mechanism [43]. However, it remains unclear how enhancer regulation by p63 is altered by substitution mutations. Our approach shows that one prevalent substitution mutation in the DBD leads to complete loss of p63 binding to chromatin and loss of enhancer establishment. On the other hand, one frequent substitution mutation in the SAM domain largely retains DNA binding but results in defects in chromatin opening, causing dysfunctional enhancer establishment and defective upregulation of key craniofacial genes. These results suggest the SAM domain is involved in recruiting chromatin remodelers to displace nucleosomes and establish open chromatin, in line with previous reports suggesting p63 and the BAF complex cooperate to open chromatin during epidermal differentiation[64]. Further experiments are required to identify proteins interacting with the SAM domain of p63 to perform these functions. However, our results point towards a novel function of the p63 SAM domain, as well as a novel molecular mechanism underlying AEC.

46

While EEC and AEC are relatively rare, one phenotype of both syndromes is cleft lip palate, which also occurs in the absence of p63 mutations in 1:700 births[11]. A connection between p63 and CL/P has previously been proposed[32,67,68], however, these studies were largely carried out in mice, a model in which KO and KD of p63 does not accurately recapitulate clefting of the lip or palate[2,3]. This could be in part due to extensive rewiring of the p63 response element network between mouse and human, as well as human specific craniofacial enhancers that were recently reported[69,70]. Further, mouse models of known disease-causing genes such as *IRF6, PVRL1*[71]*,* and *CDH1*[46] in human also fail to recapitulate clefting of the lip[15,16,35,44], suggesting stark differences in these processes between species and highlighting the need for orthogonal human approaches. In humans, GWAS and other approaches have identified 40 human risk loci for nsCL/P with conclusive evidence. These 40 loci represent about 25% of heritability in nsCL/P cases reported, however at the majority of these loci, the causative variants and mechanisms are unknown, due in part to a lack of mechanistic data within non-coding regions. Thus, by integrating epigenomic datasets from our trans-differentiation keratinocyte model with GWAS datasets, we gained novel insights into molecular mechanisms underlying nsCL/P.

Overlapping of our results with GWAS datasets from nsCL/P identified key known candidate genes regulated by p63 and newly established enhancers at known risk loci, including *IRF6* in the 1q32 locus and *WNT9B* at 17q21.32. Importantly, we identified *ZFP36L2* and *GRHL1* at 2p21 and 2p25.1 respectively as the lead risk nsCL/P genes despite previous reports focusing on other genes in those loci[58,61]. Our data also prioritized *ESRP1* and *SEMA4D,* known genes in epithelial regulation, as nsCL/P risk candidates at 8q22.1 and 9q22.2 loci in which the risk genes had not been identified. Knockout studies of *ESRP1* in particular showed full penetrance of CL/P in mice[56]. Further, by combining GWAS datasets with genes upregulated in our trans-differentiation model, we identified over 100 new candidate genes and risk loci for nsCL/P validation studies, including *ZNF296, CPNE9,* and *ITGB4.* These genes play important roles in epithelial cells but their role in nsCL/P had, prior to our study, been unknown.

47

In addition to identifying genes downstream of p63, our analysis also identified the enhancers that regulate them. For instance, *FOLR1* and *FOLR3* were both upregulated in our trans-differentiation model following establishment of upstream enhancers by p63 and KLF4. Knockout of the mouse homolog of *FOLR1* gene was shown to lead to orofacial clefting[72]; however, mutations in these genes have not been found in humans with CL/P. It is possible that previously uncharacterized mutations at the enhancers of these genes may increase risk of CL/P, which could be mitigated by folic acid supplementation, as is the case in mouse models with KO of *folbp1*[72].

Finally, using two different co-localization analyses, we found that enhancers established by p63+KLF4 are enriched for SNPs associated with nsCL/P, providing a novel molecular mechanism for how these SNPs may contribute to disease at a molecular level. For instance, at the nsCL/P risk loci 20q12, *MAFB* has been proposed to be the main risk gene for nsCL/P[12,73]. However, no functional or mechanistic data has supported this claim. Our integrated data shows p63 and KLF4 co-bind and establish enhancers upstream of *MAFB*, leading to robust *de novo* expression of *MAFB*. These enhancers are enriched for nsCL/P associated SNPs, providing a novel mechanism underlying how these SNPs may affect proper expression of *MAFB* during lip and palate development. Another example is the known 8q24 risk locus, which shows the strongest effect size across populations of Asian and European origin[63,74]. Previous reports conducted in murine models suggested that this region contains enhancers important for *MYC* regulation[75]; however, deletion of this region resulted in clefting of the lip or palate in less than 7% of murine embryos, despite robust downregulation of *MYC*. Our data shows this region contains a number of p63+KLF4 bound and established enhancers that are enriched for SNPs associated (p<5.0E-8) with nsCL/P. Further, we show establishment of enhancers in this region lead to upregulation of a long noncoding RNA *LINC00977* which is not present in mice. In human keratinocytes, it was previously observed that knockdown of p63 leads to downregulation of *MYC*[76]. Thus, it is possible that p63 regulates enhancer establishment at 8p24 to upregulate *LINC00977* and modulate *MYC* expression during lip and palate development. While mechanistic

experiments are necessary to confirm this, our data provide novel candidate disease mechanisms underlying nsCL/P.

Taken together, our findings address how p63 binds and remodels chromatin with its first encounter, identifies and prioritizes nsCL/P candidate genes and the enhancers that regulate them, and provides new explanations for how p63 mutations and SNPs associated with nsCL/P underlie disease.

2.4: Materials and methods

2.4.1: Cell culture

BJ foreskin fibroblasts (CRL-2522) were obtained from American Type Culture Collection (ATCC), and were cultured in a 37˚C incubator at 3% oxygen, in tissue culture medium containing DMEM with 10% FBS, 100µg/ml streptomycin and 100 units per ml penicillin.

2.4.2: Lentivirus and retrovirus infection

Stable cell lines were made by lentivirus and retrovirus infection, as previously described[74]. Tetracycline On System was introduced in two steps. First the reverse tetracycline-controlled transactivator (rtTA) was introduced via lentivirus Lenti CMV rtTA3 Blast (w756-1) (Addgene plasmid #26429). Lentivirus was transfected with packaging plasmids to HEKT293T cells, viral supernatant was filtered through a 0.45µm filter, supplemented with 8µg/ml polybrene, and mixed with trypsinized BJ cells for 24 hours. Cells were selected for 1 week using blasticidin at 4µg/ml.

We then cloned ΔNP63alpha-Flag (Addgene plasmid #26979) and ΔNP63α-Flag cloned with a T2A self-cleaving peptide linker to human KLF4 (plasmid #19777) into a retroviral construct pRetroX-TRE3G-puro construct. We introduced point mutations (I537T and R304W) into p63 via Quick Change II site directed mutagenesis from Agilent (Catalog # 200523). Retrovirus constructs

were transfected to Phoenix packaging cell line, purified and used as described above. Infected cells were selected with puromycin for approximately a week.

2.4.3: Trans-differentiation

BJ stable cell lines infected with rtTA and either an empty pRetrox-TRE3G vector, pRetrox-TRE3G-ΔNP63α-Flag or pRetrox-TRE3G-ΔNP63α-Flag-T2A-hKLF4 were induced for 72 hours using 1ug/ml doxcycycline replenished every 24 hours. A low dose of hydroxyurea (HU) 1mM was added after 24 and 48 hours, as we observed this improves conversion without affecting control cells. Cells were then collected for western blot, ATAC, RNA and ChIPseq.

2.4.4: Western Blot

Cells were lysed in RIPA buffer containing 1% NP-40, 150mM NaCl, 50mM Tris-Cl, pH 8.0 and 1% SDS supplemented with cOmplete, EDTA-free protease inhibitor (Roche 11873580001). Protein concentration was determined by BCA protein assay (Life technologies, #23227), after which equal amount of proteins were loaded and separated by polyacrylamide gels. Proteins were then transferred to nitrocellulose membranes. Antibodies used in this study were as follow: Flag (Sigma, M2, F1804), hKLF4 (R&D Systems, AF3640), p63 (Cell Signaling Technology, D2K8X), GAPDH (Fitzgerald, 10R-G109a) and KRT14 (Abcam, ab7800).

2.4.5: Immunofluorescence

Trans-differentiated and control fibroblasts grown on cover glasses were fixed in 4% paraformaldehyde in PBS for 30min. Cells were then washed twice with PBS, and permeabilized with 2.5% Triton X-100 in PBS for 10min. After two washes, cells were blocked in 10% BSA in PBS for 1 hour at room temperature. Cells were then incubated at 4°C overnight with primary antibodies in 5% BSA in PBS with 0.1% Tween 20 (PBST). Primary antibodies used were: Flag (Sigma, M2, F1804), hKLF4 (R&D Systems, AF3640), p63 (Cell Signaling Technology, D2K8X)

and KRT14 (Abcam, ab7800). The next day, cells were washed 4x10min with PBST, followed by incubation with Alexa Fluor-conjugated secondary antibodies (Life Technologies) in 5% BSA in TBST for 45 minutes at 37°C. Cells were then washed 3x10min in PBS, incubated with 0.5µg/ml DAPI in PBS for 5 min, and washed twice with PBS. The slides were then mounted using VECTASHIELD antifade mounting medium (H-1000, vector Laboratories) and incubated overnight in the dark at room temperature. The slides were then imaged with a Nikon Eclipse 80i fluorescent microscope.

2.4.6: Chromatin immunoprecipitation and RNA preparation

Trans-differentiated and control BJ fibroblasts (population doubling 20-30) were crosslinked with formaldehyde (1% final) for 5 min at room temperature or lysed in TRIzol (ThermoFischer Scientific, #15596018) and snap frozen for RNA isolation. Crosslinked cells were quenched with glycine (125mM final) for 5 min, followed by 2 x washes in cold PBS. Nuclei were then isolated from 20 million cells as previously described[75], and chromatin was sheared to 250bp average size using a Covaris S220. Immunoprecipitations were performed using 500µg sheared chromatin lysate and 5 µg antibodies preconjugated to Protein G beads (Invitrogen): H3K27ac (Active Motif, 39133), Flag (Sigma, M2, F1804) and hKLF4 (R&D Systems, AF3640). ChIP reactions were incubated for 16 hours at 4°C with rotation and then washed 4x in wash buffer (50mM HEPES-HCl pH 8, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% sodium deoxycholate, 0.5% N-laurylsarcosine), followed by 1 wash in ChIP final wash buffer (1xTE, 50mM NaCl). Immunoprecipitated DNA was eluted from washed beads, reverse crosslinked overnight, purified and used to construct libraries. Snapfrozen RNA samples were further extracted using chloroform and QIAGEN RNeasy Mini Kit (#74106) including DNase digestion. Purified RNA was then reverse transcribed to cDNA using High Capacity cDNA Reverse Transcription Kit from Applied Biosystems (4368814), cDNA was then quantified and used to construct sequencing libraries.

2.4.7: ChIP/RNA sequencing and data analysis

Sequencing libraries for ChIP experiments were prepared using NEBNext Ultra reagents (New England Biolabs). All ChIP samples and input were double-end sequenced on an Illumina NextSeq 500. Reads were mapped to the reference human genome assembly GRCh37/hg19 using Bowtie2 (maximum fragment size set to be 2000 bp). Properly aligned, non-duplicated pairs were retained for peak calling using MACS2 (q-value cutoff set to be 0.01). Fragment pileup per million reads for each ChIP library and corresponding input were also generated and subtracted by MACS2 for visualization. Only overlapping peaks shared by biological replicates were used for analysis.

Poly(A) + RNA was isolated using double selection with poly-dT beads, followed by first and second-strand synthesis. Reads were mapped to the reference human genome assembly GRCh37/hg19 using splice-aware tool STAR (maximum fragment size set to be 2000 bp). Alignments with a MAPQ score over ten were counted towards RefSeq genes using featureCounts (default parameters). Fragment pileup per ten million reads for each RNA library were generated by BedTools ("genomecov -split -bg -scale") and UCSC tool package ("bedGraphToBigWig"). Pairwise differential expression analysis was implemented using R package DESeq2 with FDR cutoff set to be 0.05 and fold-change cutoff set to be 1.5.

2.4.8: Assay for transposase-accesible chromatin (ATAC-seq)

ATACseq experiments were performed as previously described[76] using 100,000 cells and 5µL Tn5 transposase (Nextera XT Kit, Illumina) to tagment intact nuclei. Reads were mapped to the reference human genome assembly GRCh37/hg19 using Bowtie2 (maximum fragment size set to be 2000 bp). Properly aligned, non-duplicated pairs were retained and reads mapped to mitochondria or ENCODE blacklist regions were excluded. For peak calling, we adjusted the read start sites to represent the center of the transposon binding event (all reads aligning to the + strand were offset by +4 bp, and all reads aligning to the - strand were offset -5 bp). Broad peaks were called using MACS2 (q-value cutoff set to be 0.01). Fragment pileup per million reads for each ATAC library were also generated by MACS2 for visualization. Similar to the ChIP-seq data,

only overlapping peaks shared by biological replicates were used for analysis. To acquire sub-nucleosomal peaks, only fragments shorter than 120 nt were used.

2.4.9: Gene ontology analysis

GO terms of differentially regulated genes were determined using DAVID. We selected the top 10 GO categories based on the lowest FDR.

2.4.10: Intersection

All p63 peak sets were combined by taking the union using BedTools ("bedtools merge") and categorized into different subsets based on whether they overlap with ATAC-seq or ChIP-seq peaks (overlap by at least 1bp). Peaks were annotated to genes by the nearest RefSeq TSS using HOMER (default parameters for "annotatePeaks" command). Motif analysis was implemented using HOMER ("findMotifsGenome -size given").

2.4.11: Statistical analysis

Metaplots and heatmaps centered around any peak set were generated using DeepTools ("computeMatrix" and "plotHeatmap"). ChIP-seq or ATAC-seq signal over each peak region was computed using Bwtool (unit: per million reads per kb) in order to generate boxplots and bar graphs, and Mann-Whitney test was applied unless otherwise specified.

2.4.12: Analysis of GWAS-data for nsCL/P

GWAS-data for nsCL/P were retrieved from a previously published study[14] which had included 399 cases, 1.318 controls and 1.461 case-parent trios. The SNP data comprised about 8.01 million common variants at a minor allele frequency (MAF) > 1%. All analyses for the present manuscript were performed using the overall data, i.e., without stratification for ethnicity.

2.4.13: Gene-based p-values.

While single variants alone might be statistically underpowered to show significant effects, their contribution in sum can be analyzed by aggregating individual variants within a predefined region (for instance, genes). Based on the p-values obtained from the nsCL/P GWAS we used MAGMA (v1.06)[51] a principal-component-based regression framework that allows for gene-based analysis. We run the gene-based analysis for the p63 upregulated genes (n=250) as implemented in the FUMA online tool (Functional Mapping and Annotation of Genome-Wide Association Studies) platform (v1.3.2)[73], with default parameters (SNP-wide mean model) and with the 1000 Genome Phase3 as reference panel. In the gene set analyses, curated gene sets and GO terms from MsigDB are tested. For FUMA (≥ v1.3.1), 10655 gene sets (curated gene sets: 4738, GO terms: 5917) from MsigDB v6.1 are used. Bonferroni correction was performed for the all tested gene sets.

2.4.14: Enrichment analyses

To evaluate the enrichment of all nominal significant nsCL/P GWAS SNPs[70] in *à priori* selected regulatory elements, the GREGOR (v1.4.0)[58] software tool was used. Selection of the regions was performed based on the presence of p63 and/or KLF4 and/or H3K27ac peaks. For analysis of overlay of specific SNPs to specific elements, a colocalization approach was performed as previously suggested[14]. Briefly, SNPs were mapped to either "positive regions" (i.e., in p63+KLF4 +/- H3K27ac binding regions) or "negative regions" (all others) and then grouped into nine categories $10^{-k+1} \geq P > 10^{-k}$ for k = 1, . . ., 8 and $10^{-8} \geq P$. The frequencies of SNPs being located in the positive regions and those outside the annotated regions were counted within each category. The one-sided Cochran-Armitage trend test was applied in R environment with function prop.trend.test of stats package to test for an enrichment of smaller P-values within annotated regions on the basis of the resulting 9×2 contingency table.

2.5: Supplemental Figures
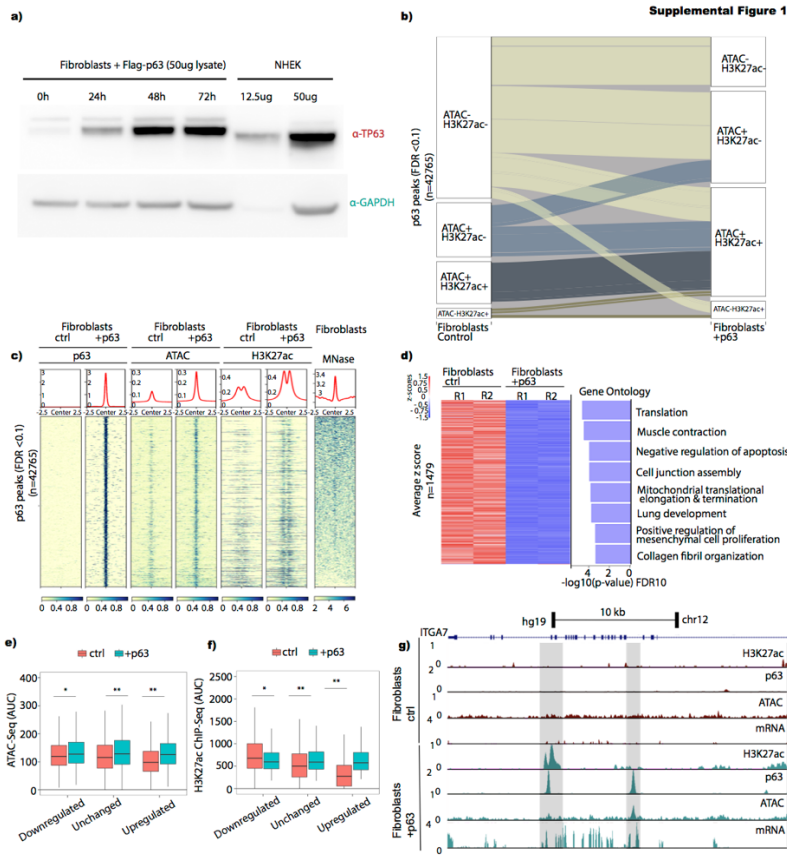
Supplemental Figure 2.1: **Ectopic expression of p63 in fibroblasts leads to establishment of enhancers and downstream upregulation of inflammation and epithelial genes. A)** Western blot showing p63 ectopically expressed in fibroblasts reaches an expression level comparable to endogenous levels observed in undifferentiated normal human epidermal keratinocytes after 48h **B)** Alluvial plot depicting changes in chromatin state after ectopic expression of p63 at p63 bound peaks for chromatin transitions with >200 peaks. **C)** Heatmap showing high overlap between newly accessible chromatin (ATACseq) in fibroblasts + p63 and

published MNaseSeq in human fibroblasts[25]. **D)** Heatmap showing 1479 downregulated genes in fibroblasts + p63 compared to control and GO categories enriched among downregulated genes. **E)F)** Boxplots showing increased chromatin accessibility and H3K27ac across downregulated, unchanged and upregulated genes in fibroblasts + p63, with the highest fold change for upregulated genes (*p-Value<10E-5, **p-Value<10E-16). **G)** UCSC genome browser tracks showing transcriptional activation of ITGA7 and *de novo* H3K27ac and ATACseq signal within introns of ITGA7 in fibroblasts with ectopic expression of p63 (grey highlighted boxes)
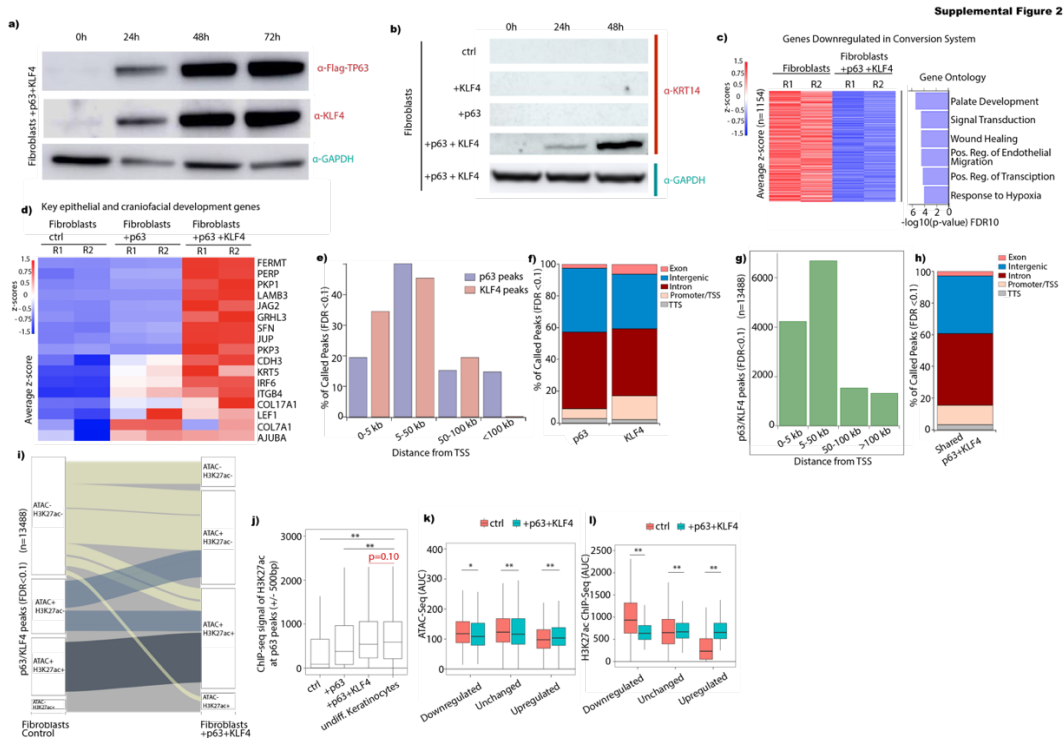
Supplemental Figure 2.2: Ectopic co-expression of p63 and KLF4 leads to establishment of keratinocyte specific enhancers and upregulation of epithelial genes. A) Western Blot showing robust co-expression of p63 and KLF4 in the inducible trans-differentiation model (fibroblasts+p63+KLF4) **B)** Western Blot showing KRT14 is only upregulated when both p63 and KLF4 are expressed and not in control fibroblasts, fibroblasts + KLF4 or +p63 alone. **C)** Heatmap showing 1154 downregulated genes in fibroblasts + p63+KLF4 compared to control and GO categories enriched among downregulated genes. **D)** Heatmap showing co-expression of p63 and KLF4 is required for upregulation of a subset of key epithelial and craniofacial genes. **E)** Distance to nearest TSS for all p63 and all KLF4 peaks in fibroblasts +p63+KLF4. **F)** Partitioning of p63 and KLF4 peaks into different genomic features showing intergenic and intronic regions are the most abundant for both **G)** Distance to closest TSS for p63/KLF4 shared peaks. **H)**

Partitioning of 13488 p63/KLF4 shared peaks showing enrichment of intronic and intergenic regions, consistent with enhancers. **I)** Alluvial plot depicting changes in chromatin state after ectopic expression of p63 and KLF4 at p63/KLF4 bound peaks for chromatin transitions with >200 peaks. **J)** Boxplot showing H3K27ac is not statistically different (p=0.01) at shared p63 peaks between p63+KLF4 in fibroblasts and p63 in undifferentiated keratinocytes (ChIP-seq signal at p63 peaks (n=4981) +/- 500bp) (**p-Value<10E-16) **K)L)** Boxplots showing increased chromatin accessibility and H3K27ac only at genes upregulated in fibroblasts + p63+KLF4 (*p-Value<0.05, **p-Value<10E-16). Very high fold change in H3K27ac at p63/KLF4 peaks close to upregulated genes.
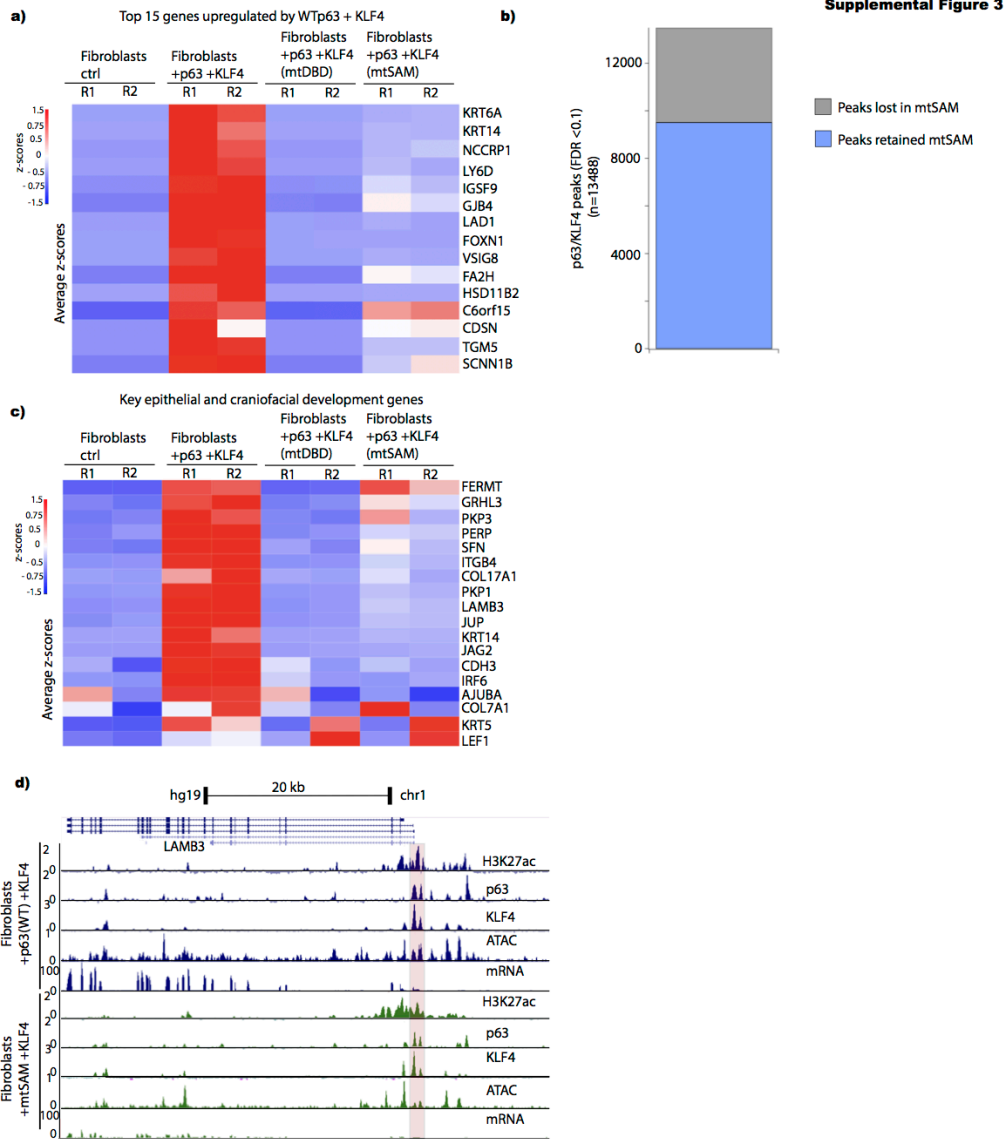
Supplemental Figure 2.3: Mutations in p63 disrupt enhancer establishment and lead to loss of upregulation of key keratinocyte specific genes. A) Heatmap showing mtDBD and mtSAM have disrupted or loss of upregulation of the top 15 genes upregulated by WTp63+KLF4. **B)** Category plot showing 71% (9513/13488) of p63/KLF4 sites are retained when p63 carries mutations in the SAM domain. **C)** Heatmap showing mtDBD and mtSAM have disrupted or loss of upregulation of a subset of epithelial and craniofacial genes. **D)** UCSC genome browser tracks showing defects (red box) in establishing open chromatin and inducing gene expression at LAMB3, a gene involved in craniofacial development.

a) DEGs directly linked to CL/P in mouse KO/KD models
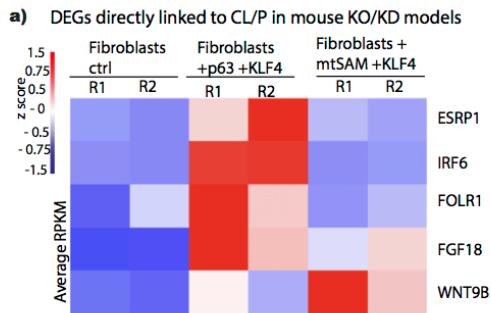
Supplemental Figure 2.4: Heatmap of genes regulated by p63+KLF4 that have been strongly linked to CL/P in murine KD and KO models. A) Heatmap of genes regulated by p63+KLF4 that have been strongly linked to CL/P in murine KD and KO models.
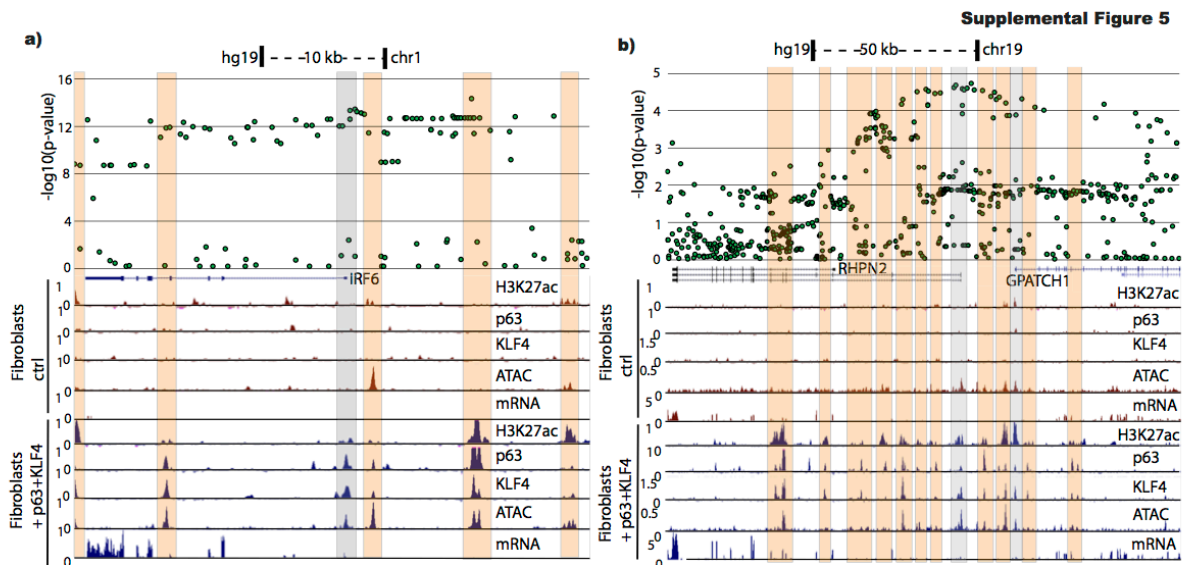
Supplemental Figure 2.5: p63/KLF4 peaks colocalize strongly with highly associated CL/P SNPs near *IRF6, RHPN2* and *GPATCH1*. A)B) Overlay of SNPs and UCSC genome browser tracks highlighting in orange that p63/KLF4 peaks colocalize strongly with highly associated CL/P SNPs near *IRF6*, *RHPN2* and *GPATCH1*, in grey binding of both proteins at promoters.

2.6: Supplemental Table Legends

Supplemental Table 2.1: Genes upregulated in fibroblasts +p63 compared to control fibroblasts, fold change > 1.5 and FDR < 0.05

Supplemental Table 2.2: Genes upregulated in fibroblasts +p63+KLF4 compared to control fibroblasts, fold change > 1.5 and FDR < 0.05

Supplemental Table 2.3: Column 1) Genes upregulated in fibroblasts +mtDBD+KLF4 compared to control fibroblasts, column 2) Genes upregulated in fibroblasts +mtSAM+KLF4 compared to control fibroblasts, column 3) Genes no longer upregulated in fibroblasts +mtSAM+KLF4 compared to WTp63+KLF4, column 4) Newly upregulated genes in fibroblasts +mtSAM+KLF4 compared to WTp63+KLF4; fold change > 1.5 and FDR < 0.05

Supplemental Table 2.4: A) Representative genes involved in isolated CL/P or CPO in murine KO/KD models and transcriptional regulation in control fibroblasts, fibroblasts + WTp63+KLF4 and fibroblasts +mtSAM+KLF4 and replicates. **B)** List of 76 genes upregulated by p63+KLF4 identified by genetic association database. **C)** List of 40 known risk loci for nsCL/P and genes within the TAD domains of each loci, in orange, genes upregulated by p63+KLF4. **D)** List of p63/KLF4 regulated genes and respective gene based CL/P association p-values for known causal genes, genes within the 40 risk loci and genes outside of the known risk loci.

2.7: References

1.	Yang, A. *et al.* p63, a p53 Homolog at 3q27–29, Encodes Multiple Products with Transactivating, Death-Inducing, and Dominant-Negative Activities. *Mol. Cell* **2**, 305–316 (1998).

2.	Yang, A. *et al.* p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. *Nature* **398**, 714–718 (1999).

3.	Mills, A. A. *et al. p63* is a *p53* homologue required for limb and epidermal morphogenesis. *Nature* **398**, 708–713 (1999).

4.	Lopez-Pajares, V., Yan, K., Zarnegar, B. J., Jameson, K. L. & Khavari, P. A. Genetic pathways in disorders of epidermal differentiation. *Trends Genet.* **29**, 31–40 (2013).

5.	Celli, J. *et al.* Heterozygous Germline Mutations in the p53 Homolog p63 Are the Cause of EEC Syndrome. *Cell* **99**, 143–153 (1999).

6.	Bokhoven, H. van & McKeon, F. Mutations in the p53 homolog p63: allele-specific developmental syndromes in humans. *Trends Mol. Med.* **8**, 133–139 (2002).

7.	Brunner, H. G., Hamel, B. C. J. & Bokhoven, H. van. The p63 gene in EEC and other syndromes. *J. Med. Genet.* **39**, 377–381 (2002).

8.	Kouwenhoven, E. N. *et al.* Transcription factor p63 bookmarks and regulates dynamic enhancers during epidermal differentiation. *EMBO Rep.* **16**, 863–878 (2015).

9.	Rapisarda, V. *et al.* p63 Transcription Factor Regulates Nuclear Shape and Expression of Nuclear Envelope-Associated Genes in Epidermal Keratinocytes. *J. Invest. Dermatol.* **137**, 2157–2167 (2017).

10.	Lin-Shiao, E. *et al.* KMT2D regulates p63 target enhancers to coordinate epithelial homeostasis. *Genes Dev.* **32**, 181–193 (2018).

11.	Mangold, E. *et al.* Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. *Nat. Genet.* **42**, 24–26 (2010).

12.	Beaty, T. H. *et al.* A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near *MAFB* and *ABCA4*. *Nat. Genet.* **42**, 525–529 (2010).

13.	The genetics of isolated orofacial clefts: from genotypes to subphenotypes - Jugessur - 2009 - Oral Diseases - Wiley Online Library. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1601-0825.2009.01577.x. (Accessed: 3rd October 2018)

14.	Ludwig, K. U. *et al.* Imputation of orofacial clefting data identifies novel risk loci and sheds light on the genetic background of cleft lip ± cleft palate and cleft palate only. *Hum. Mol. Genet.* **26**, 829–842 (2017).

15.	Lough, K. J., Byrd, K. M., Spitzer, D. C. & Williams, S. . Closing the Gap: Mouse Models to Study Adhesion in Secondary Palatogenesis. *J. Dent. Res.* **96**, 1210–1220 (2017).

16.   Yu, K., Deng, M., Naluai-Cecchini, T., Glass, I. A. & Cox, T. C. Differences in Oral Structure and Tissue Interactions during Mouse vs. Human Palatogenesis: Implications for the Translation of Findings from Mice. *Front. Physiol.* **8**, (2017).

17.   Beaty, T. H., Marazita, M. L. & Leslie, E. J. Genetic factors influencing risk to orofacial clefts: today's challenges and tomorrow's opportunities. *F1000Research* **5**, 2800 (2016).

18.   Leslie, E. J. *et al.* Genome-wide meta-analyses of nonsyndromic orofacial clefts identify novel associations between FOXE1 and all orofacial clefts, and TP63 and cleft lip with or without cleft palate. *Hum. Genet.* **136**, 275–286 (2017).

19.   Mostowska, A. *et al.* Common variants in DLG1 locus are associated with non-syndromic cleft lip with or without cleft palate. *Clin. Genet.* **93**, 784–793 (2018).

20.   Yu, Y. *et al.* Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. *Nat. Commun.* **8**, 14364 (2017).

21.   Fakhouri, W. D. *et al.* An etiologic regulatory mutation in IRF6 with loss- and gain-of-function effects. *Hum. Mol. Genet.* **23**, 2711–2720 (2014).

22.   Shalom-Feuerstein, R. *et al.* ΔNp63 is an ectodermal gatekeeper of epidermal morphogenesis. *Cell Death Differ.* **18**, 887–896 (2011).

23.   Romano, R.-A. *et al.* ΔNp63 knockout mice reveal its indispensable role as a master regulator of epithelial development and differentiation. *Development* **139**, 772–782 (2012).

24.   Soares, E. & Zhou, H. Master regulatory role of p63 in epidermal development and disease. *Cell. Mol. Life Sci.* **75**, 1179–1190 (2018).

25.   Soufi, A. *et al.* Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate Reprogramming. *Cell* **161**, 555–568 (2015).
26.   Viganò, M. A. *et al.* New p63 targets in keratinocytes identified by a genome-wide approach. *EMBO J.* **25**, 5105–5116 (2006).

27.   Pozzi, S. *et al.* Transcriptional Network of p63 in Human Keratinocytes. *PLOS ONE* **4**, e5008 (2009).

28.   Carroll, D. K. *et al.* p63 regulates an adhesion programme and cell survival in epithelial cells. *Nat. Cell Biol.* **8**, 551–561 (2006).

29.   Jiang, R. *et al.* Defects in limb, craniofacial, and thymic development in Jagged2 mutant mice. *Genes Dev.* **12**, 1046–1057 (1998).

30.   Yang, X. *et al.* ΔNp63 Versatilely Regulates a Broad NF-κB Gene Program and Promotes Squamous Epithelial Proliferation, Migration, and Inflammation. *Cancer Res.* **71**, 3688–3700 (2011).

31.   ΔNp63 regulates thymic development through enhanced expression of FgfR2 and Jag2 | PNAS. Available at: http://www.pnas.org/content/104/29/11999.long. (Accessed: 16th October 2018)

32.   Thomason, H. A. *et al.* Cooperation between the transcription factors p63 and IRF6 is essential to prevent cleft palate in mice. *J. Clin. Invest.* **120**, 1561–1569 (2010).

33.	Ferretti, E. *et al.* A Conserved Pbx-Wnt-p63-Irf6 Regulatory Module Controls Face Morphogenesis by Promoting Epithelial Apoptosis. *Dev. Cell* **21**, 627–641 (2011).

34.	Chen, Y., Mistry, D. S. & Sen, G. L. Highly Rapid and Efficient Conversion of Human Fibroblasts to Keratinocyte-Like Cells. *J. Invest. Dermatol.* **134**, 335–344 (2014).

35.	Suzuki, A., Jun, G., Abdallah, N., Gajera, M. & Iwata, J. Gene datasets associated with mouse cleft palate. *Data Brief* **18**, 655–673 (2018).

36.	Peyrard-Janvid, M. *et al.* Dominant Mutations in GRHL3 Cause Van der Woude Syndrome and Disrupt Oral Periderm Development. *Am. J. Hum. Genet.* **94**, 23–32 (2014).

37.	McGrath, J. A. *et al.* Hay–Wells syndrome is caused by heterozygous missense mutations in the SAM domain of p63. *Hum. Mol. Genet.* **10**, 221–230 (2001).

38.	Chung, J., Grant, R. I., Kaplan, D. R. & Irwin, M. S. Special AT-rich Binding Protein-2 (SATB2) Differentially Affects Disease-causing p63 Mutant Proteins. *J. Biol. Chem.* **286**, 40671–40680 (2011).

39.	Sathyamurthy, A., Freund, S. M. V., Johnson, C. M., Allen, M. D. & Bycroft, M. Structural basis of p63α SAM domain mutants involved in AEC syndrome. *FEBS J.* **278**, 2680–2688 (2011).

40.	Rostagno, P. *et al.* Embryonic stem cells as an ectodermal cellular model of human p63-related dysplasia syndromes. *Biochem. Biophys. Res. Commun.* **395**, 131–135 (2010).

41.	Huang, Y.-P. *et al.* AEC-Associated p63 Mutations Lead to Alternative Splicing/Protein Stabilization of p63 and Modulation of Notch Signaling. *Cell Cycle* **4**, 1440–1447 (2005).

42.	Fomenkov, A. *et al.* p63α Mutations Lead to Aberrant Splicing of Keratinocyte Growth Factor Receptor in the Hay-Wells Syndrome. *J. Biol. Chem.* **278**, 23906–23914 (2003).

43.	Russo, C. *et al.* Protein aggregation of the p63 transcription factor underlies severe skin fragility in AEC syndrome. *Proc. Natl. Acad. Sci.* **115**, E906–E915 (2018).

44.	Juriloff, D. M. & Harris, M. J. Mouse genetic models of cleft lip with or without cleft palate. *Birt. Defects Res. A. Clin. Mol. Teratol.* **82**, 63–77 (2008).

45.	Becker, K. G., Barnes, K. C., Bright, T. J. & Wang, S. A. The Genetic Association Database. *Nat. Genet.* **36**, 431–432 (2004).

46.	Frebourg, T. *et al.* Cleft lip/palate and CDH1/E-cadherin mutations in families with hereditary diffuse gastric cancer. *J. Med. Genet.* **43**, 138–142 (2006).

47.	Juriloff, D. M., Harris, M. J., Mager, D. L. & Gagnier, L. Epigenetic mechanism causes Wnt9b deficiency and nonsyndromic cleft lip and palate in the A/WySn mouse strain. *Birt. Defects Res. A. Clin. Mol. Teratol.* **100**, 772–788 (2014).

48.	Way, G. P., Youngstrom, D. W., Hankenson, K. D., Greene, C. S. & Grant, S. F. Implicating candidate genes at GWAS signals by leveraging topologically associating domains. *Eur. J. Hum. Genet.* **25**, 1286–1289 (2017).

49.	Pervjakova, N. & Prokopenko, I. The TAD-pathway for GWAS signals. *European Journal of Human Genetics* (2017). doi:10.1038/ejhg.2017.148

50.   Phillips-Cremins, J. E. *et al.* Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. *Cell* **153**, 1281–1295 (2013).

51.   Leeuw, C. A. de, Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Comput. Biol.* **11**, e1004219 (2015).

52.   Bebee, T. W. *et al.* The splicing regulators Esrp1 and Esrp2 direct an epithelial splicing program essential for mammalian development. *eLife* **4**, e08954 (2015).

53.   Cox, L. L. *et al.* Mutations in the Epithelial Cadherin-p120-Catenin Complex Cause Mendelian Non-Syndromic Cleft Lip with or without Cleft Palate. *Am. J. Hum. Genet.* **102**, 1143–1157 (2018).

54.   Beaty, T. H. *et al.* Confirming genes influencing risk to cleft lip with/without cleft palate in a case–parent trio study. *Hum. Genet.* **132**, 771–781 (2013).

55.   Prenzler, F., Fragasso, A., Schmitt, A. & Munz, B. Functional analysis of ZFP36 proteins in keratinocytes. *Eur. J. Cell Biol.* **95**, 277–284 (2016).

56.   Cura, F. *et al.* Replication analysis of 15 susceptibility loci for nonsyndromic cleft lip with or without cleft palate in an italian population. *Birt. Defects Res. A. Clin. Mol. Teratol.* **106**, 81–87 (2016).

57.   Leslie, E. J. *et al.* Identification of Functional Variants for Cleft Lip with or without Cleft Palate in or near PAX7, FGFR2, and NOG by Targeted Sequencing of GWAS Loci. *Am. J. Hum. Genet.* **96**, 397–411 (2015).

58.   Schmidt, E. M. *et al.* GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics* **31**, 2601–2606 (2015).

59.   Grant, S. F. A. *et al.* A Genome-Wide Association Study Identifies a Locus for Nonsyndromic Cleft Lip with or without Cleft Palate on 8q24. *J. Pediatr.* **155**, 909–913 (2009).

60.   Bao, X. *et al.* A novel ATAC-seq approach reveals lineage-specific reinforcement of the open chromatin landscape via cooperation between BAF and p63. *Genome Biol.* **16**, 284 (2015).

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

62.   Kirschner, R. D., Rother, K., Müller, G. A. & Engeland, K. The retinal dehydrogenase/reductase retSDR1/DHRS3 gene is activated by p53 and p63 but not by mutants derived from tumors or EEC/ADULT malformation syndromes. *Cell Cycle* **9**, 2177–2188 (2010).

63.   McDade, S. S. *et al.* Genome-wide analysis of p63 binding sites identifies AP-2 factors as co-regulators of epidermal differentiation. *Nucleic Acids Res.* **40**, 7190–7206 (2012).

64.   Richardson, R. *et al.* p63 exerts spatio-temporal control of palatal epithelial cell fate to prevent cleft palate. *PLOS Genet.* **13**, e1006828 (2017).

65.   Wilderman, A., VanOudenhove, J., Kron, J., Noonan, J. P. & Cotney, J. High-Resolution Epigenomic Atlas of Human Embryonic Craniofacial Development. *Cell Rep.* **23**, 1581–1597 (2018).

66.   Sethi, I., Gluck, C., Zhou, H., Buck, M. J. & Sinha, S. Evolutionary re-wiring of p63 and the epigenomic regulatory landscape in keratinocytes and its potential implications on species-specific gene expression and phenotypes. *Nucleic Acids Res.* **45**, 8208–8224 (2017).

67.   Suzuki, K. *et al.* Mutations of *PVRL1*, encoding a cell-cell adhesion molecule/herpesvirus receptor, in cleft lip/palate-ectodermal dysplasia. *Nat. Genet.* **25**, 427–430 (2000).

68.   Piedrahita, J. A. *et al.* Mice lacking the folic acid-binding protein Folbp1 are defective in early embryonic development. *Nat. Genet.* **23**, 228–232 (1999).

69.   Xiao, Y. *et al.* Evidence for SNP-SNP interaction identified through targeted sequencing of cleft case-parent trios. *Genet. Epidemiol.* **41**, 244–250 (2017).

70.   Ludwig, K. U. *et al.* Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. *Nat. Genet.* **44**, 968–971 (2012).

71.   Uslu, V. V. *et al.* Long-range enhancers regulating *Myc* expression are required for normal facial morphogenesis. *Nat. Genet.* **46**, 753–758 (2014).

72.   Wu, N., Rollin, J., Masse, I., Lamartine, J. & Gidrol, X. p63 Regulates Human Keratinocyte Proliferation via MYC-regulated Gene Network and Differentiation Commitment through Cell Adhesion-related Gene Network. *J. Biol. Chem.* **287**, 5627–5638 (2012).

73.   Watanabe, K., Taskesen, E., Bochoven, A. van & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).

74.   Zhu, J. *et al.* Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth. *Nature* **525**, 206–211 (2015).

75.   Shah, P. P. *et al.* Lamin B1 depletion in senescent cells triggers large-scale changes in gene expression and the chromatin landscape. *Genes Dev.* **27**, 1787–1799 (2013).

76.   Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).

77.   Watanabe, K., Taskesen, E., Bochoven, A. van & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).

78.   Zhu, J. *et al.* Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth. *Nature* **525**, 206–211 (2015).

79.   Shah, P. P. *et al.* Lamin B1 depletion in senescent cells triggers large-scale changes in gene expression and the chromatin landscape. *Genes Dev.* **27**, 1787–1799 (2013).

80.   Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).

Chapter 3: KMT2D regulates p63 target enhancers to coordinate epithelial homeostasis

3.1: Introduction

Proper regulation of epigenetic information through the precise control of gene enhancer and transcriptional networks is essential for establishing and maintaining cell fate and identity[1]. Disruption of these pathways critical for differentiation can lead to a loss of proliferative control, ultimately driving carcinogenesis[2]. In line with this, the gene encoding the chromatin modifier and master enhancer regulator *KMT2D* (*MLL4*) is amongst the most frequently mutated genes in human cancer[3,4,5,6,7]. There is an exceptionally high prevalence of *KMT2D* mutations in self-renewing epithelial tissues ranging from the lung and esophagus to the skin and bladder, with the highest frequencies seen in keratinocyte cancers (cutaneous squamous cell carcinoma and basal cell carcinoma), the two most common forms of cancer world-wide[7,8].

Functionally, KMT2D, along with the related protein, KMT2C (MLL3), are the major histone methyltransferases that deposit histone H3 lysine 4 (H3K4) monomethylation (H3K4me1), which poises enhancers for activation and transcription factor binding[9,10,11]. Furthermore, in addition to its catalytic role in H3K4me1 deposition, KMT2D is also required for the binding of the histone acetyltransferase complex, CBP/p300, which acetylates H3K27 (H3K27ac) at enhancers, resulting in enhancer activation during cell identity gene induction and changes in cellular fate[12,13,14]. In mice, complete deletion of *Kmt2d* is embryonic lethal[11]. Tissue-specific deletion studies have demonstrated that KMT2D is essential for the normal development and function of enhancers during adipogenesis and myogenesis[11], cardiac development[15], and B-cell lymphocyte development[16,17].

Despite the high incidence of *KMT2D* mutations in the epidermis, and established roles in enhancer function and developmental programs in other tissues, little is understood about KMT2D function in normal epithelial homeostasis. Given the poor understanding yet potential

66

importance of KMT2D in development and cancer, here we explore the role of KMT2D in self-renewing epithelia and provide the first description of KMT2D's role at epithelial enhancers, including its genome-wide interactions with the transcription factor p63, master regulator of epithelial transcription, in coordinating key gene expression programs required for both epithelial cell identity and the maintenance of epithelial progenitors.

3.2: Results

3.2.1: KMT2D depletion leads to reduced proliferation and reduced expression of epithelial development and adhesion genes

To study KMT2D in the epithelium, we reduced KMT2D levels in primary neonatal human epidermal keratinocytes (NHEKs) and in spontaneously immortalized human epidermal keratinocytes (HaCaTs), using short hairpin RNAs directed against *KMT2D* (shKMT2D) compared to scrambled control shRNA (shSC). RNA-seq carried out in both the NHEK and the HaCaT shRNA-treated keratinocytes demonstrated a significant reduction in KMT2D expression (Fig. 3.1A). This was confirmed by employing three distinct shKMT2D hairpins which displayed a consistent 2-3 fold reduction in KMT2D expression using RT-qPCR, western blotting, and immunofluorescence (IF) (Supplemental Figs. S3.1A-C, Fig. 3.1B). In addition, proliferative growth of shKMT2D keratinocytes was reduced by KMT2D loss compared to shSC keratinocytes (Fig. 3.1C), consistent with previous reports in other cell types[18,19]. To confirm that the reduced proliferative capacity did not reflect a simple global stress response, we examined levels of the senescence-associated tumor suppressor, p16/CDKN2A, and the stress and senescence-associated transcription factor, phospho-NF-κB p65 (Ser536), and did not observe any differences between either shSC or shKMT2D keratinocytes (Supplemental Fig. S3.1C).
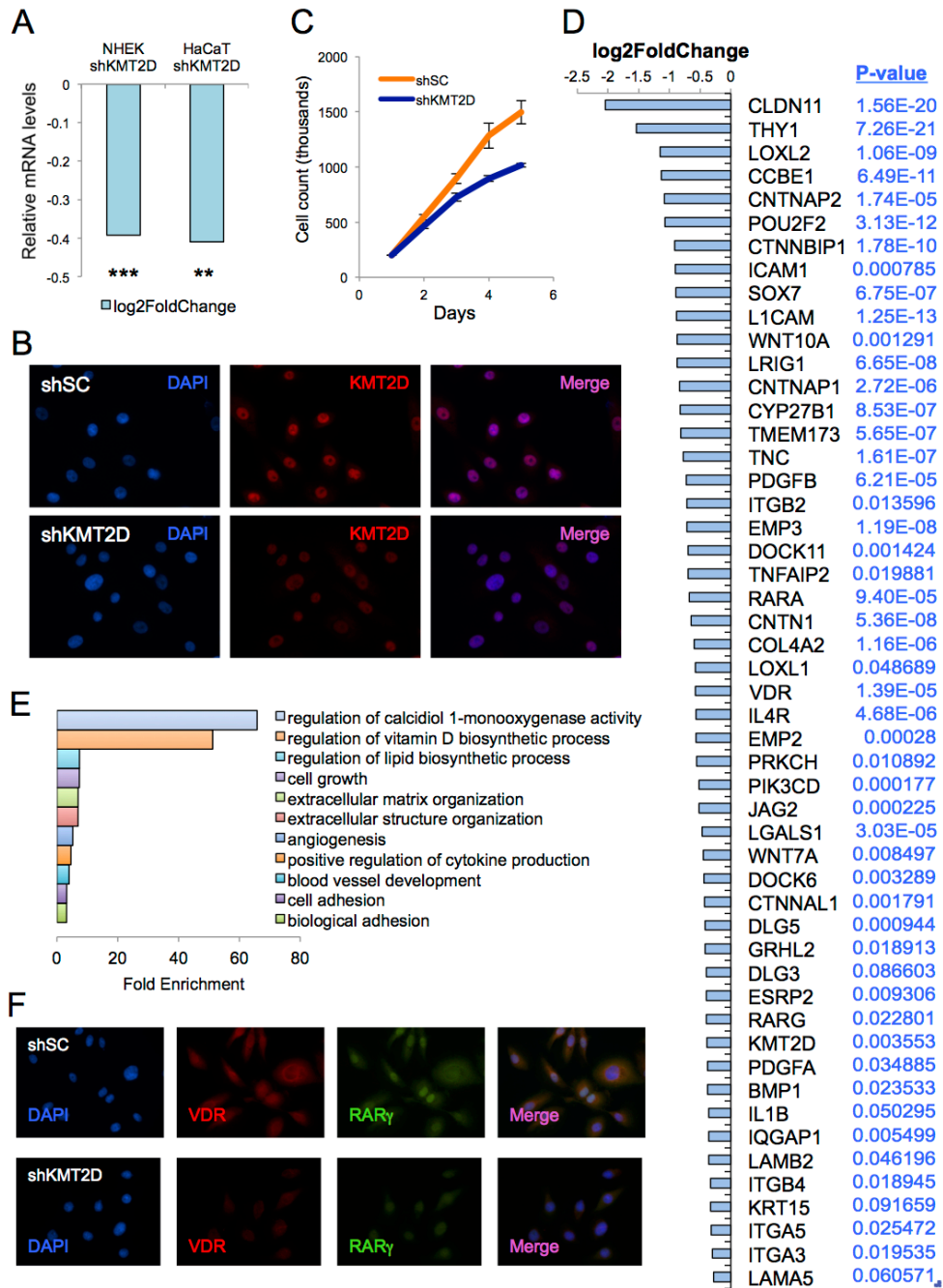
Figure 3.1: KMT2D loss leads to reduced keratinocyte proliferation and a broad loss of epithelial development and adhesion genes. *(A)* KMT2D mRNA expression is significantly reduced in both NHEK (p = 0.0036) and HaCaT (p = 0.0312) keratinocytes treated with shKMT2D as determined by RNA-seq. *(B)* shKMT2D keratinocytes display reduced levels of KMT2D, but

normal-appearing nuclear morphology. *(C)* shKMT2D keratinocytes display reduced proliferation in comparison to shSC keratinocytes. *(D)* Representative genes from shKMTD NHEKs that are significantly reduced in expression along with corresponding adjusted p-values. *(E)* Gene Ontology (GO) analysis of the 134 common genes lost with shKMT2D-treatment between both NHEK and HaCaT shKMT2D keratinocytes demonstrates that genes involved in key epithelial signaling pathways (*RARG*, *VDR*), epithelial cell growth and morphogenesis, polarity, and adhesion are enriched in amongst those genes with reduced expression. *(E)* shKMT2D keratinocytes display reduced expression of RARγ and VDR by IF (40x).

Given KMT2D's importance in transcriptional enhancer function, we examined the transcriptional consequences of KMT2D depletion in keratinocytes. We performed 3 biological replicates of RNA-seq from NHEKs treated with shKMT2D. Strikingly, we found a broad reduction in the expression of genes critical for epithelial development and differentiation, as well as adherence to the epidermal basement membrane (Fig. 3.1D). These encode proteins involved in epithelial development, migration, and differentiation (*CYP27B1*, *GRHL2*, *PDFGB*, *PRKCH*, *RARG*, *SOX7*, *VDR, WNT10A*)[20,21,22,23,24,25], the establishment and maintenance of cell polarity (*DLG3*, *DLG5*, *WNT7A*)[26] , cellular adhesion and the extracellular matrix organization (*CLDN11*, *CNTN1*, *EMP2*, *IL4R*, *LOXL1*, *LOXL2*, *SERPINE1*)[27,28] , and included disease-associated markers of interfollicular basal epidermal stem cells essential for the function of the epidermal basement membrane (*COL4A2*, *ITGA3*, *ITGA5*, *ITGB2*, *ITGB4*, *LAMA5*, *LAMB2*)[29,30] (representative genes in Fig. 3.1D, and full list in Supplementary Table S3.1A). Similarly, we also performed 3 RNA-seq replicates in the HaCaT keratinocytes treated with shSC and shKMT2D. Overlap of all significantly reduced genes (adjusted p-value < 0.1) from both sets of keratinocytes demonstrated a significant overlap in the transcriptional alterations observed with shKMT2D treatment (134 overlapping genes, p = 3.57e-76) (Supplementary Fig. S3.1D). Gene Ontology (GO) analysis was carried out on the common 134 genes and showed enrichment of genes

69

related to cellular growth, extracellular matrix organization, and adhesion, including key genes involved in epithelial development, differentiation and morphogenesis. Notably, the vitamin D signaling pathway, a key cell signaling pathway involved in epithelial differentiation was also among the most enriched pathways (Fig. 3.1E and Supplemental Table S3.2A)[23]. Visualization of the keratinocytes by IF confirmed the loss of expression of representative genes such as *RARG* and *VDR* (Fig. 3.1F, G). Collectively, these results show that KMT2D reduction drives a broad loss of key genes typically expressed by epidermal stem and progenitor cells of the basal layer, as well as genes involved in proper epidermal differentiation, together suggesting a role for KMT2D in maintaining the hallmark gene expression program and proliferative capacity characteristic of epithelial progenitors.

3.2.2: KMT2D is enriched genomewide at p63 target enhancers

To gain further insight into the function of KMT2D in the epidermis and the mechanism underlying keratinocyte gene expression, we mapped KMT2D binding genome-wide using chromatin immunoprecipitation (ChIP)-seq (full mapping information in Supplementary Table S3.3A). Consistent with its role in gene regulation at enhancers, a total of 1,887 KMT2D ChIP-seq binding sites were analyzed by DiffBind[31] and HOMER suite, among which half of KMT2D peaks (52%) bound to intergenic regions, while introns (and likely intronic enhancers) made up the majority of the remaining binding sites (40%) (Fig. 3.2A). Analysis of the nearest gene to each peak showed that 76% of peaks were associated with protein-coding genes, while the majority of the others occurred near non-coding RNAs (22%) (Fig. 3.2B).

We evaluated transcription factors binding motifs that were enriched at KMT2D binding regions. HOMER motif analysis of all KMT2D peaks identified p63, considered the master regulator of epithelial development and differentiation[32], as the most highly enriched transcription factor motif at KMT2D binding sites (Fig. 3.2C). To examine the overlap of KMT2D peaks with

p63 binding sites, we intersected the nearest gene to our KMT2D peaks with consensus previously published p63 binding site-associated genes[33,34]. We verified that KMT2D peaks indeed overlapped directly with p63 binding sites in shSC keratinocytes using ChIP-seq. Although our p63 antibody identified fewer total binding sites than observed in previous reports[33,34], our data suggested that p63 and KMT2D bound to overlapping sites across the genome, as 49% of p63 peaks contained a directly overlapping KMT2D peak, while another 13% of p63 peaks were with 100kb of a KMT2D peak. Overlapping peaks included known p63 target genes such as *MINK1*, *SERINC2*, and *TP63* itself (Figs. 3.2E-G).
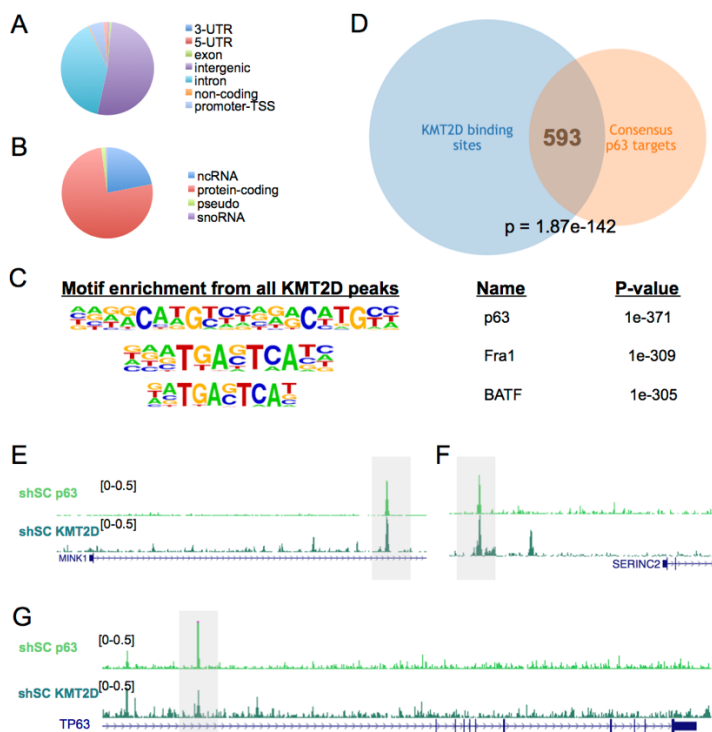
Figure 3.2: KMT2D co-localizes with p63 genome-wide at p63 target genes. *(A)* Consistent with its role in enhancer function, KMT2D binding is enriched in control (shSC) keratinocytes at intergenic and intronic regions genome-wide. *(B)* Over 75% of KMT2D binding sites in control (shSC) keratinocytes are associated with protein-coding genes, with the majority of the remaining binding sites associated with non-coding RNAs (ncRNAs). *(C)* HOMER motif analysis of KMT2D binding sites in shSC keratinocytes reveals a strong enrichment at p63 binding motifs. *(D)*

Intersection of consensus established p63 binding sites in keratinocytes with KMT2D binding sites in shSC keratinocytes demonstrates a significant overlap (p = 1.87e-142). *(E)* ChIP-seq for both KMT2D and p63 in shSC keratinocytes demonstrates that KMT2D binds at the same sites as p63 at known p63 target genes such as *MINK1*, *SERINC2 (F)*, and *TP63* itself *(G)*.

3.2.3: KMT2D and p63 interact on chromatin

We next wanted to test for a potential direct interaction between KMT2D and p63. Importantly, chromatin co-immunoprecipitation experiments in NHEKs were consistent with the ChIP-seq data in showing that KMT2D interacts with p63 on chromatin (Supplemental Fig. S3.2A). In line with data showing that the ΔNp63 isoform of p63 is the predominant form of p63 active in the epidermis, using a ΔNp63 specific antibody we determined that KMT2D interacts with the α, β and γ isoforms of ΔNp63, and suggesting that this interaction is not mediated by the C-terminal sterile alpha motif (SAM) or the transactivation (TA) domains absent in the ΔN forms of p63 (Supplemental Fig. S3.2B). Formaldehyde crosslinking followed by immunoprecipitation and mass spectrometry (IP-Massspec) further confirmed that p63 and KMT2D interact on chromatin (Figs. 3.3A,B). Intriguingly, these data also suggest that this interaction may in fact be mediated by a shared interacting partner of the more than 1000 that we identify here for the first time (Figs. 3.3A, B, Supplemental Fig. S3.3A-C, and Supplemental Table S3.4).
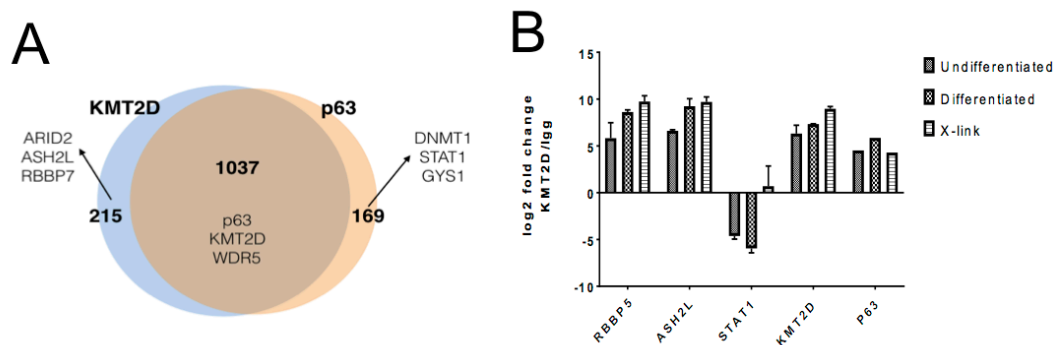
Figure 3.3: KMT2D interacts with p63 on chromatin. *(A)* Formaldehyde crosslinking followed

by IP-Mass Spec shows p63 and KMT2D interact on chromatin, as immunoprecipitated KMT2D is able to pull down p63, and immunoprecipitated p63 is able to pull down KMT2D. We further observe over 1000 shared interacting partners. *(B)* IP-Mass Spec identifies new and known (i.e. RBBP5 and ASH2L) interactors of KMT2D. p63 is identified as an interactor of KMT2D in both crosslinked NHEKs and uncrosslinked and DNAse treated undifferentiated and differentiated NHEKs.

3.2.4: KMT2D enrichment at p63 targets maintains expression of many key genes involved in epithelial homeostasis

To test the effects of KMT2D loss, we also performed KMT2D ChIP-seq on the shKMT2D keratinocytes. Hierarchal clustering showed that both shSC and the shKMT2D samples clustered with their respective replicate dataset (Fig. 3.4A). According to DiffBind[31], which employs a stringent peak-based approach, of the 1,887 total KMT2D peaks, 198 were identified as having at least a 2-fold loss of KMT2D peak size (Fig. 3.4B heatmap and black line in Fig. 3.4C metaplot). In addition, in order to capture more of the KMT2D changes in enrichment that we observed in our primary data with shKMTD-treatment, we also employed a non-peak-based approach by comparing the tag density (AUC) alterations in ChIP-seq enrichment with shKMT2D treatment. This approach identified 2,271 regions of at least a 2-fold decrease in KMT2D enrichment in shKMT2D keratinocytes in comparison to shSC keratinocytes (Supplementary Table S3.3A). In contrast, just 43 regions across the genome demonstrated a 2-fold or greater increase in KMT2D enrichment (Supplementary Table S3.3B). Consistent with KMT2D's interaction with p63 and binding at the same sites on chromatin, motif analysis of KMT2D reduced enrichment regions in shKMT2D keratinocytes once again showed that p63 was the most highly enriched transcription factor motif (Fig. 3.4D). In contrast, regions that gained KMT2D were enriched for transcription factors associated with epidermal differentiation including CEBPβ and AP1-associated factors, BATF and Fra1 (Fig. 3.4E)[35,36,37].
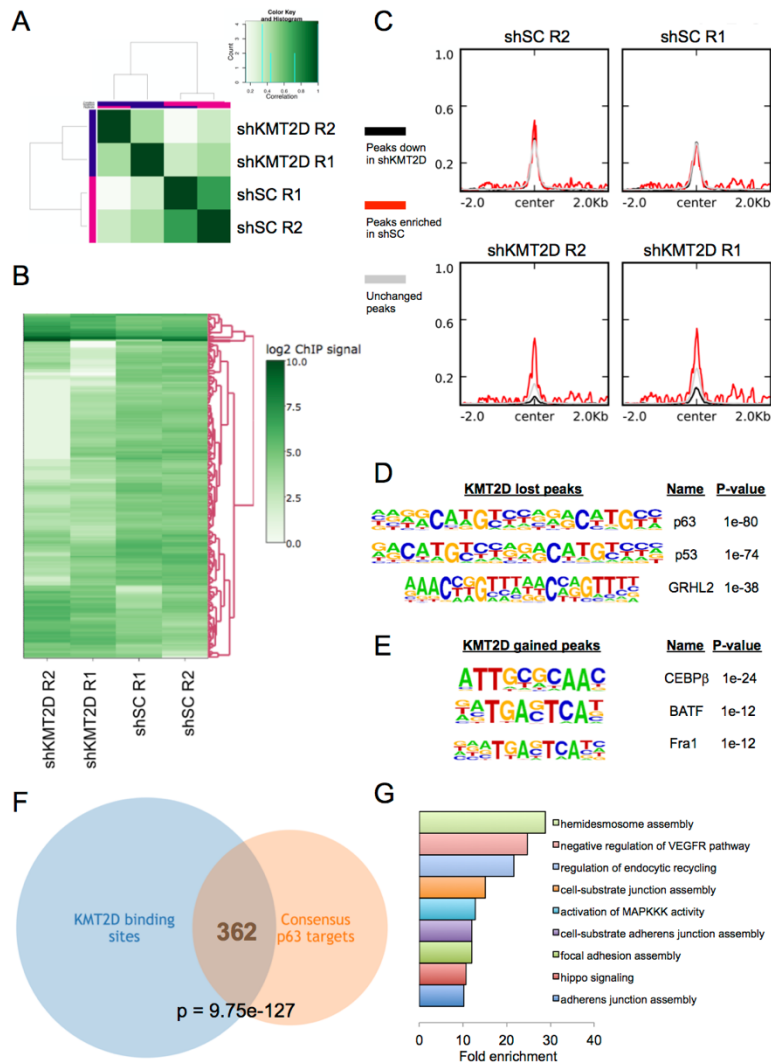
Figure 3.4: KMT2D depletion leads to a preferential loss of KMT2D at p63-target gene enhancers. *(A)* Two replicates of ChIP-seq were performed on shSC and shKMT2D keratinocytes, with shSC and shKMT2D clustering separately with their respective replicate dataset. *(B)* Heatmap of DiffBind-identified regions of at least 2-fold changes in enrichment demonstrates extensive reductions of KMT2D enrichment in shKMT2D keratinocytes. *(C)* Metaplots showing average ChIP signal near differential KMT2D binding sites across both replicates for shSC and shKMT2D keratinocytes. Black line indicates KMT2D-depleted regions, while red indicates enrichment, and gray indicates unchanged enrichment. *(D)* HOMER motif analysis of significantly lost KMT2D regions identifies p63-binding motifs as the most enriched

regions of KMT2D loss. *(E)* The few regions that gain KMT2D enrichment are enriched for CEBPβ motifs. *(F)* Intersection of all KMT2D bound region-associated genes (that lose at least 25% enrichment averaged across both ChIP-seq replicates) and established consensus p63-binding site-associated genes demonstrates a highly significant overlap (9.75e-127). *(G)* GO analysis of the 362 genes that are known p63-binding sites and identified here with reduced KMT2D enrichment demonstrates that genes associated with hemidesmosomes, cell and adherens junctions, focal adhesion, are highly enriched.

p63 has a critical role in regulating dynamic enhancers, including super-enhancers critical for cell identity and cell fate specification, during epidermal development and differentiation[34,38]. Hence, we explored overlaps between established keratinocyte p63 binding sites and target genes[33,34], and genes near KMT2D bound regions, focusing on genes that lost more than 25% enrichment with shKMT2D across both replicates. Here, the intersection of p63 binding sites and KMT2D-depleted genes was highly significant (p = 9.75e-127)(Fig. 3.4F and Supplemental Table S3.3C). GO analysis of these 362 "KMT2D-depleted/p63-target" genes included pathways similar to above: hemidesmosome assembly, cell and adherens junctions, and focal adhesion (Fig. 3.4G and Supplemental Table S3.2B); again including genes involved in epidermal development, stemness, polarity and adhesion (*CNTN1*, *CTNNBIP1*, *DLG5*, *ITGA3*, *ITGB4*, *LAMA5*, *LAMC1*, *PRKCH*, *RARG*, *SERPINE1*, *SOX7*, *VDR*, *WNT7A*, *WNT10A*), many of which are known to be p63 regulated[27,33,34,39]. Together, these data demonstrate that in keratinocytes KMT2D is enriched at p63 targets and maintains the expression of many key genes involved in epithelial homeostasis, including pathways involved in epidermal commitment and stratification such as the vitamin D and retinoid pathways, as well as numerous genes involved in regulating cell junctions and adhesion to the epidermal basement membrane.

75

3.2.5: KMT2D can drive the expression of p63-target genes through both catalytic histone
methylation activity as well as catalytically independent mechanisms at enhancers.

KMT2D is a histone methyltransferase, thus we assayed relevant histone modifications at
key genes, using ChIP-seq for key enhancer chromatin modifications, H3K4me1 (catalyzed by
KMT2D), and H3K27ac (KMT2D recruits the acetyltransferase CBP/p300)[12,13,14]. We performed
two replicates for each histone modification, and as expected H3K4me1 and H3K27ac clustered
separately (Supplemental Fig. S3.4A). We identified the top 5% depleted regions (and the
nearest associated gene) in each H3K27ac and H3K4me1 ChIP-seq replicate (Supplemental
Table S3.5). Consistent with both the RNA-seq and KMT2D binding data, the top 5% regions of
loss of both H3K4me1 (which KMT2D catalyzes) and H3K27ac (which KMT2D helps coordinate
through recruitment of p300/CBP), were enriched for GO categories involved in focal adhesion,
adherens junctions, and cell migration and proliferation (Supplemental Table S3.5). Overlap of
the KMT2D-depleted/p63-target genes that lost gene expression by RNA-seq with the genes that
lost enhancer histone modification enrichment identified 82 genes (Fig. 3.5A and Supplemental
Table S3.6A). The most highly significant GO categories enriched within these 82 genes involved
development and differentiation, consistent with both our current understanding of the roles of
both p63 and KMT2D (Fig. 3.5B, Supplemental Table S3.2C) as well as our other results
demonstrating the effects of KMT2D depletion on the expression of key genes involved in the
epidermal basement membrane (*COL4A2*, *ITGA3*, *ITGB4*, *LAMA5*) and epithelial development
and differentiation such as *CTNNBIP1*, *DLG5*, *EHD1*, *EHD2*, *EMP2*, *PRKCH*, *RARG*, *VDR*, and
SERPINE1 (Figs. 3.5C, D and Supplemental Fig. S3.4B-G, and Supplemental Table S3.6A). In
addition, there were 20 KMT2D-depleted/p63-target genes losing expression that did not lose
enhancer histone modifications (at least within the top 5% loss for either H3K4me1 and H3K27ac)
(Supplemental Table S3.6B). These genes were also involved in development and adhesion and
included *ITGA3*, *KRT15*, *IL1B*, and *SPARC*, and *TJAP1* (Supplemental Table S3.6B). These
latter observations are consistent with recent evidence suggesting that KMT2D depletion can lead
to extensive gene expression changes that are independent of its histone methyltransferase

catalytic activity[40]. Collectively, these results indicate that KMT2D can drive the expression of p63-target genes broadly involved in multiple aspects of epithelial development and differentiation, both through catalytic histone methylation activity as well as through other catalytically independent mechanisms at enhancers.
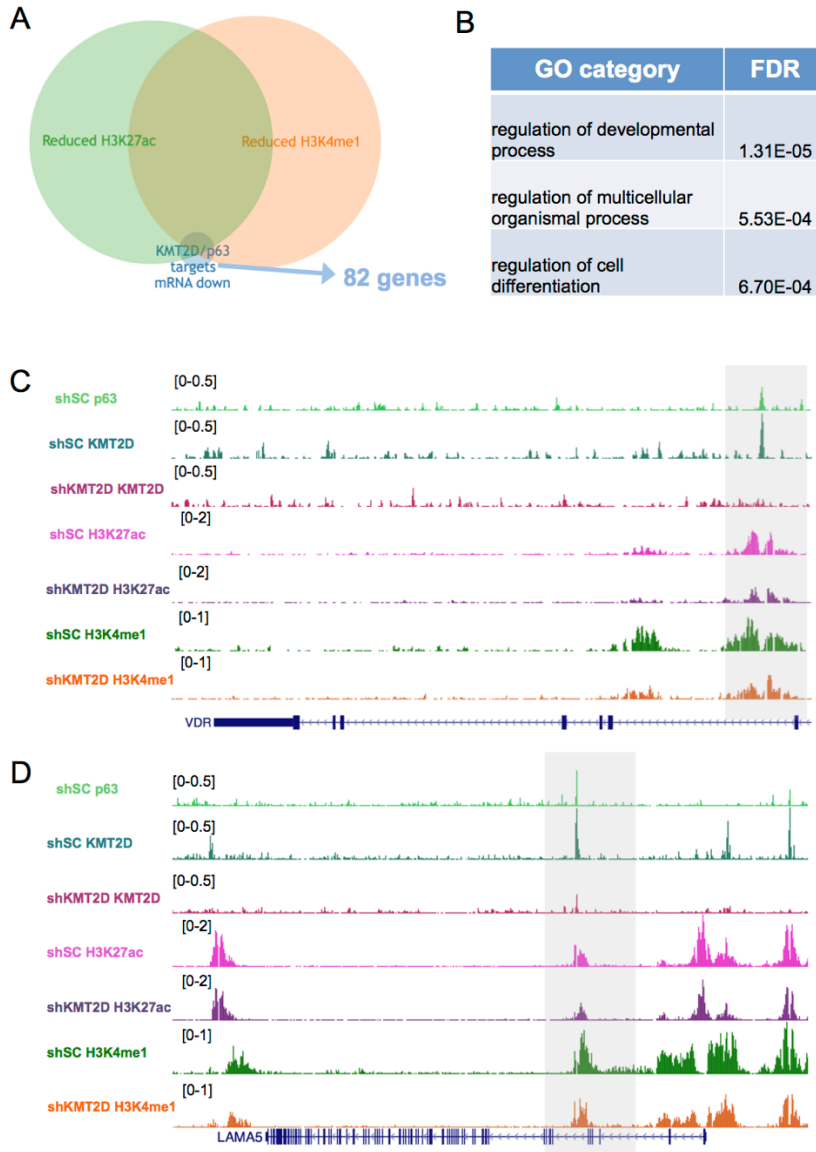
Figure 3.5: KMT2D depletion provokes a loss of enhancer histone modifications and p63-target genes. *(A)* Intersection of the KMT2D-depleted/p63-target genes that lose gene expression with those that lose enhancer histone modifications (H3K4me1 or H3K27ac) in

shKMT2D keratinocytes identifies 82 genes. *(B)* GO analysis of these 82 KMT2D-depleted/p63-target genes that lose enhancer histone modifications shows an enrichment for genes broadly involved in development and differentiation. ChIP-seq tracks from representative genes that lose KMT2D enrichment at p63 bound regions, as well as enhancer histone modifications. Examples include *VDR (C)* and *LAMA5 (D)*.

3.2.6: KMT2D coordinates with p63 to maintain epidermal progenitor gene expression and proper epidermal differentiation and stratification.

We then examined the genes that were upregulated upon shKMT2D-treatment of keratinocytes. GO analysis of the common genes that exhibited a significant increase ($p<0.1$) in the shKMT2D keratinocytes identified a striking enrichment of genes involved in keratinocyte differentiation, with the GO category "cornification" displaying an over 40-fold enrichment (Fig. 3.6A). These genes included markers of early epidermal stratification such as *KRT10* as well as canonical transcription factors involved in epidermal differentiation and stratification (*KLF4*, *GRHL1*, *GRHL3*, *ZNF750*) (Fig. 3.6B). Other genes significantly increased in shKMT2D keratinocytes included genes such as *KRT16* and *KRT17*, which are typically associated with hyperproliferative conditions such as squamous cell carcinoma and psoriasis (Fig. 3.6B and Supplemental Fig. S3.5A). The combination of data showing a loss of expression of genes involved in the basal epidermal stem cell layer and the progenitor state, along with the upregulation of genes associated with keratinocyte differentiation, together suggest that in undifferentiated keratinocytes, KMT2D expression actively coordinates with p63 to maintain epidermal progenitor gene expression and thus prevent terminal differentiation.
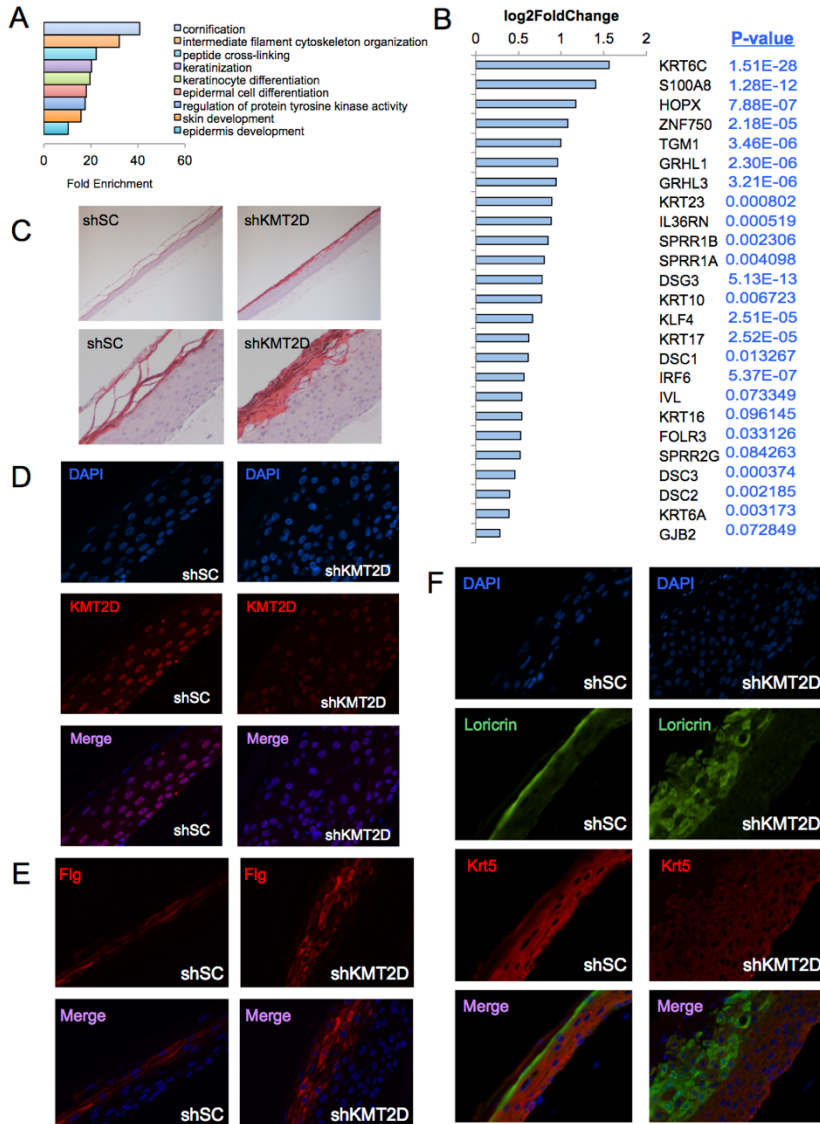
Figure 3.6: KMT2D depletion provokes premature and disorganized epidermal stratification. *(A)* GO analysis of the common genes that significantly (p<0.1) gain expression in shKMT2D keratinocytes identifies genes involved in epithelial cornification, keratinization, differentiation, and development as being the most enriched. *(B)* Representative genes that are upregulated in shKMT2D keratinocytes includes genes encoding proteins that are characteristic of stratifying epidermis such as *KRT10* and *IRF6*, as well as numerous members of the small proline rich (SPR) family of genes. Similarly upregulated are transcription factors such as *KLF4*, *GRHL1*, and *GRHL3* associated with keratinocyte differentiation. *(C)* Three-dimensional (3D)

human organotypic skin cultures demonstrate that shKMT2D keratinocytes display altered differentiation dynamics, including random premature cornification, a loss of polarity, and disorganized stratification in comparison to shSC keratinocytes (Hematoxylin and eosin stain, 10x). *(D)* IF of 3D human skin cultures demonstrates reduced levels of KMT2D throughout the epidermis in shKMT2D keratinocytes in comparison to shSC keratinocytes (40x). *(E)* Filaggrin and loricrin staining of 3D human skin cultures suggests both increased levels as well as disorganized epidermal stratification (20x).

To address how these chromatin and gene expression alterations may affect human skin, we utilized three-dimensional human organotypic skin cultures[41]. These cultures allow for the analysis of epidermal genes and proteins during differentiation in a more human and architecturally relevant context than standard submerged cultures[41]. While shSC-treated keratinocytes displayed a relatively normal appearing epidermis, shKMT2D-treated keratinocytes appeared highly disorganized, with cells showing a loss of polarity, irregular morphologies, and premature cornification (Fig. 3.6D), consistent with many of the observed gene expression changes. IF of these sections confirmed both decreased levels of KMT2D expression (Fig. 3.6C). Staining for other markers associated with differentiation such as filaggrin and loricrin demonstrated disorganized stratification dynamics with premature and elevated expression of these proteins, as well as notable reduction in the expression of keratin 5, a marker of the epidermal progenitor cells of the basal epidermis (Figs. 3.6D-F). In order to determine whether a loss of p63-target gene expression was solely responsible for the premature activation of differentiation-associated genes, we overexpressed p63 in the shKMT2D keratinocytes. While significant overexpression of p63 was achieved, it was unable to reverse the gene expression changes observed in the KMT2D-depleted keratinocytes (Supplemental Fig. S3.5B). These results were not entirely surprising given KMT2D's critical role in enhancer, transcription, and

80

transcription factor dynamics, and suggest that the effects of KMT2D on epithelial homeostasis and stratification dynamics are more broad than acting simply through p63.

3.3: Discussion

An abundance of evidence has shown that p63 is critical for limb development and epidermal specification[42,43,44], and mutations in TP63 are associated with several developmental disorders that display epithelial abnormalities[45]. Over time, p63 has come to be known as the master regulatory transcription factor of epithelial tissues, playing key roles at multiple levels of development and differentiation[46,47]. Excitingly, here we present evidence that KMT2D, a major enhancer histone methylation enzyme, and one of the most frequently mutated genes in epithelial cancer, is a critical regulator of p63-dependent gene expression programs and epithelial homeostasis. Our data indicate that KMT2D interacts with p63 on chromatin and at a broad array of enhancers involved in epithelial development, adhesion, and differentiation. Further, we also identify numerous potential new interacting proteins that may play a role in mediating the KMT2D-p63 interaction. In turn, depletion of KMT2D in human keratinocytes leads to extensive alterations in normal epidermal transcriptional programs, characterized by a broad loss of expression of genes typically expressed in the self-renewing basal epidermal stem cell compartment such as basal integrin and laminin genes, as well as key regulators of proper epidermal stratification, including the retinoid and vitamin D signaling pathways[21,23]. In support of our findings, p63 has been shown to regulate many of the same genes we observe to be regulated by KMT2D here, including those involved in adhesion and vitamin D signaling[39,48,49]. In response to the loss of the expression of these genes, genes involved in keratinocyte proliferation (*KRT1*, *KRT10*, *KRT6*, *KRT16*, *KRT17*) and transcription factors that promote terminal differentiation (*KLF4*, *GRHL1*, *GRHL3*, *ZNF750*) are prematurely turned on.

As one of the largest organs in the body, the skin serves as a major barrier to the external environment. Similar to other epithelial surfaces, in order to maintain homeostasis, the epidermis undergoes regular, highly coordinated self-renewal that involves dramatic alterations in gene transcription and cellular architecture[50]. This self-renewal ability is maintained in the interfollicular epidermis by stem cells in the basal layer that are typically slow cycling, but with a high proliferative potential and self-renewing capacity, serving to ultimately supply the more differentiated suprabasal epidermis[37,51]. Our data indicates that KMT2D plays a critical role in epithelial homeostasis by maintaining the adhesion and proliferative capacity of the epidermal basement membrane through its interaction with p63 to regulate critical enhancers for genes involved in both maintaining the basal epidermis as well as in the proper coordination of epithelial stratification.

Similar to both p63, while KMT2D may be critical for epithelial stem cell maintenance and gene expression, we cannot rule out that it may also play key roles later in coordinating later steps in terminal differentiation, as does p63[46]. Indeed, given its effects on genes also involved in the promotion of epidermal differentiation, such as the vitamin D and retinoid signaling pathways, it is likely KMT2D also participates in the coordination of all stages of epidermal commitment and differentiation. Consistent with this model, IF staining of normal skin sections for KMT2D (from both our lab as well as publicly available datasets) shows the presence of KMT2D in both basal epidermal stem cells as well as in the nuclei of cells nearing terminal differentiation (Supplemental Figs. S3.6A, B). This observation may offer insight into why KMT2D has been shown to serve both as an oncogene as well as a tumor suppressor in cancer, depending upon the cellular context[4,16,17,53,54]. Here we observe that KMT2D depletion reduces keratinocyte proliferation, which is consistent with reports that KMT2D deficiency slows the proliferation of both breast and colon cancer cell lines[19,55,56] , and correlates with improved outcomes in pancreatic cancer[53]. In contrast, our data suggesting a role for KMT2D in regulating genes involved in vitamin D and retinoid signaling would suggest a potential tumor suppressive role[23,57]. Indeed,

deciphering these potential oncogenic versus tumor suppressive functions of KMT2D in epithelial cancers remains a critical question and important direction for future studies.

3.4: Materials and methods

3.4.1:Keratinocyte cultures

Normal primary human epidermal keratinocytes (NHEKs) were isolated from de-identified discarded neonatal human foreskin obtained by Core B of the Penn Skin Biology and Diseases and Resource-based Center. Foreskin was incubated for 12 hours at 4C in 2.4 U/ml Dispase II. Sterile forceps were used to separate the underlying dermis. The epidermal sheet was transferred to a 60 mm tissue culture plate and incubated in 0.25% Trypsin for 10 minutes at 37C, then neutralized with 1 ml fetal bovine serum. Sterile forceps were used to scrape the epidermal sheet against the dish to dissociate cells. The suspension was passed through a 40 um strainer, and then centrifuged at 200g for 5 minutes. The cell pellet was re-suspended in 5 ml medium (see below). HaCaT keratinocytes were obtained from AddexBio. Both were cultured in a 50:50 mix of Keratinocyte-SFM (1x) supplemented with human recombinant Epidermal Growth Factor 1-53 and Bovine Pituitary Extract combined with Medium 154 supplemented with Human Keratinocyte Growth Supplement and 1% Penicillin-Streptomycin (10,000 U/mL) (all from Thermo Fisher Scientific). For differentiation experiments, NHEKs were cultured in media containing 1.22 mM calcium chloride for 48 hours and then harvested.

3.4.2: Viral transduction

Lentiviral particles expressing scrambled control (SC) or KMT2D (MLL4)-specific shRNAs (D4, D5, D9) were produced using the pLKO.1-Puro viral backbone plasmid generously provided by the Xianxin Hua Lab at the University of Pennsylvania. Three KMT2D shRNAs were employed: KMT2D shRNA D4:
 CCGGCCCACCTGAATCATCACCTTTCTCGAGAAAGGTGATGATTCAGGTG; KMT2D shRNA D5: CCGGCCTGAATTGAACAACAGTCTTCTCGAGAAGACTGTTGTTCAATTCA; KMT2D

shRNA D9: CCGGCGAATGGAAGAACAACGTGTACTCGAGTACACGTTGTTCTTCCATT. For the viral transductions, keratinocytes were infected with virus at 60% confluency on 10cm^2 plates for 16 hours in the presence of polybrene. 48 hours following infections, cells underwent selection with puromycin to obtain completely puromycin-resistant cell populations. For the p63-overexpression studies, NHEKs at ~60% confluency were transfected with 10µg of pcDNA3.1 deltaNp63alpha-FLAG (Addgene #26979) and 10µl of Lipofectamine2000 diluted in 1ml of OptiMEM overnight (~16 hours). Cells were placed in fresh media and RNA was harvested 48 hours following the initial transfection.

3.4.3: Growth curve measurement

200,000 cells were seeded on 950 mm^2 surface area (one well of 6-well plate) on day 0. Cell number was measured every two days with Countess automated cell counter (Life Technologies) following standard procedures and default parameter settings, after which 200,000 cells were plated back for the next count.

3.4.4: Reverse-transcriptase quantitative PCR (RT-qPCR)

RNA was extracted using RNeasy kit (Qiagen) following the manufacturer's instructions. cDNA was obtained by and then was analyzed by standard qPCR methods on a 7900HT Fast-Real-Time PCR (ABI). Primer sequences are available upon request. RT-qPCR data analysis was performed by first obtaining the normalized QT values (normalized to 18S ribosomal RNA). The average and standard deviations of biological replicates were obtained and assessed for significance using an unpaired t test (graphpad.com). For all experiments, 3 stars indicates $P<0.01$, 2 stars indicates $P<0.05$, 1 star indicates $P<0.1$, and NS indicates non-significance, or $P>0.1$.

3.4.5: RNA-seq

RNA was extracted using RNeasy kit (Qiagen catalog no. 74014) and following the manufacturer's instructions. All RNA-seq libraries were prepared using the NEBNext Poly(A) mRNA Magnetic Isolation Module followed by NEBNext Ultra Directional RNA Library Prep Kit for Illumina (both from NEB). Library quality was checked by Agilent BioAnalyzer 2100 (Agilent) and

libraries were quantified using KAPA Library Quantification Kits (KAPA Biosystems). Libraries were then sequenced using a NextSeq500 platform (75bp, single-end reads) (Illumina). All RNA-seq was aligned using RNA STAR under default settings to Homo sapiens UCSC hg19 (RefSeq & Gencode gene annotations). FPKM generation and differential expression analysis was performed using DESeq2. NHEK and HaCaT RNA-seq each consisted of three independent biological replicates.

3.4.6: Immunofluorescence

Cultured keratinocytes were fixed in 4% PFA in PBS for 30 minutes at room temperature. Cells were washed twice with PBS, and permeabilized with 0.5% Triton X-100 in PBS for 10 minutes. After washing two times, cells were blocked in 10% BSA in PBS for 1 hour at room temperature. Cells were incubated with primary antibodies in 5% BSA in PBS supplemented with 0.1% Tween 20 (PBST) overnight at 4°C. Antibodies are listed below. Then cells were washed four times with PBST, each for 10 minutes, followed by incubation with fluorophore-conjugated secondary antibody in 5% BSA in PBST for 1 hour at room temperature. Cells were then washed three times in PBST, once with PBS, and incubated with 1 µg/ml of DAPI for 5 minutes. The cells were then washed twice with PBS, and mounted with ProLong Gold (Invitrogen). For the 3D-organotypic skin cultures and normal human skin sections, following exposure xylene and ethanol, sections were treated with target unmasking fluid to deparaffinize the tissues. Sections were incubated in primary antibody overnight following blocking for two hours in BlockAid (Thermo Fisher Scientific). Following secondary antibody incubation and washes, the sections were incubated with 1 µg/ml of DAPI for 5 minutes, then washed twice with PBS, and mounted with ProLong Gold with DAPI (Invitrogen). The slides were observed and imaged using a Nikon Eclipse microscope.

3.4.7: Antibodies

Antibodies included Anti-Histone H3 (Abcam ab1791), Anti-Histone H3 (mono methyl K4) (Abcam ab8895), Anti-Histone H3 (acetyl K27) (Abcam ab4729), anti-KMT2D (HPA035977) (Sigma-Aldrich), anti-p63-α (D2K8X) (#13109 Cell Signaling), anti-p63 (4A4) (Santa Cruz sc-

8431), Anti-ΔNp63 (Biolegend #619001), Anti-Cytokeratin 16 antibody (EP1615Y) (Abcam ab76416), Anti-Vitamin D Receptor antibody - ChIP Grade (Abcam ab3508), RARγ Antibody (G-1) (Santa Cruz sc-7387).

3.4.8: Mass spectrometry (MS) analysis and protein identification

FLAG-tagged KMT2D protein was transfected (Life Technologies, #11668019) and expressed in HEK293T cells and FLAG-tagged ΔNp63α was generated by *in vitro* translation (Thermo, #88881). Subsequently, both were separately subjected to immunoprecipitation with FLAG antibody conjugated protein G Dynabeads (Life Technologies, #10004D). Following stringent washes, equal amount of lysates of undifferentiated or differentiated NHEKs were lysed for 1 hour at 4°C in modified RIPA buffer containing 150 mM NaCl, 1% NP-40, 50 mM Tris-Cl, pH 8.0, and 1% SDS, supplemented with protease inhibitors or lysate of undifferentiated crosslinked NHEKs (crosslinked/lysed/sonicated as above) were added. After overnight co-immunoprecipitation, samples were thoroughly washed using a buffer containing: 20 mM Tris, pH 8.0, 137 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 1% NP-40, and protease inhibitors. Beads were resuspended in 50μl of 50mM Ammonium bicarbonate (NH4HCO3), samples were reduced with 0.5μl of 1M DTT at 55°C for 30min and then cooled down to room temperature. Samples were then alkylated using 2.5μl of 500mM iodacetamide for 30 min at room temperature and protected from light. Following this, 1.5μg of trypsin (Promega REF:V511C) was added to each sample and allowed to digest overnight. The next day samples were subjected to 80°C for 10min to stop trypsin digestion. Tubes were placed on a magnet and peptides were transferred to a new tube. Beads were washed twice with 50% acetonitrile and transferred to the same tube. Finally samples were dried on a SpeedVac and desalted by stage tipping. After in-beads digestion, samples were analyzed using nano Liquid Chromatography coupled online to tandem Mass Spectrometry (nLC-MS/MS). Peptides were loaded onto a 75um I.D x 17cm in-house packed column (ReproSil-Pur C_{18}-AQ, 3 μm) using an Easy-nLC-1000 (Thermo Scientific, San Jose, CA).

Peptides were eluted over 75 minutes gradient, 3-35% solvent B (A- 0.1% Formic Acid, B- 80% Acetonitrile, 0.1% Formic Acid) with a flow rate of 300 nl/min. The nLC was coupled to an Orbitrap Fusion (Thermo Scientific, San Jose, CA) operating in data-dependent mode (DDA). The spray voltage was set at 2.3 kV and the capillary temperature 275°C. Full MS scans were performed within the range 350-1200 m/z at a resolution of 120,000 (at 200 m/z). Data-dependent MS/MS was performed in the ion trap using top speed mode set to 3 seconds; signals with charge state 2^+ and higher with an intensity above 10000 counts were automatically selected for MS/MS fragmentation performed using high-energy collision dissociation (HCD) with normalized collision energy of 27. MS raw files were analyzed using MaxQuant software (1.6.0.16)[58] . Parameters for MS/MS database searching included the following: precursor mass tolerance 4.5 ppm; product mass tolerance 0.5 Da; enzyme trypsin; missed cleavages allowed 2; static modifications carbamidomethyl (C); variable modifications acetyl (protein N-term) and oxidation (M); label-free quantification method iBAQ; database Homo sapiens (Uniprot, November 2017). PSMs and protein false discovery rate was filtered for < 0.01.

3.4.9: Chromatin immunoprecipitation followed by sequencing (ChIP-seq)

ChIP-seq was performed as previously described[59]. Briefly, keratinocytes cultured in $10cm^2$ dishes were fixed in 1% formaldehyde for 5 minutes and fixation was quenched with addition of glycine to 125mM for an additional 5 minutes. Cells were harvested by scraping from plates, and washed twice in 1x PBS before storage at -80°C. ChIP extracts that were sonicated 9 times for 5 minutes each round (30 seconds sonication with intermediate incubation of 30 seconds per round) using a Bioruptor (Diagenode). All ChIPs were performed using 500 µg of extract and 2µg of antibody per sample. 30 µl of Protein G Dynabeads (Invitrogen) were used per ChIP. ChIP DNA was also used to make sequencing libraries using NEBNext Ultra DNA Library Prep Kit for Illumina (NEB). Library quality was checked by Agilent BioAnalyzer 2100 (Agilent) and libraries were quantified using KAPA Library Quantification Kits (KAPA Biosystems). Libraries were then sequenced using a NextSeq500 platform (75bp, single-end reads) (Illumina). Post sequencing, all data was demultiplexed from the raw reads using Illumina's BCL2FASTQ from

BaseSpace. Fastq files were aligned to Homo sapiens UCSC hg19. Tracks were visualized using the UCSC genome browser as described further below.

3.4.10: ChIP-seq bioinformatics analyses

Raw reads were mapped to the reference human genome assembly GRCh37 (hg19) using Bowtie2 local alignment[60]. The alignment result was further processed by SAMtools to remove low quality alignments (a minimum mapping quality of 10), duplicate reads, reads mapped to the mitochondria, unplaced (chrUn) and random (chr_random) contigs, and reads mapped to the ENCODE blacklisted genomic regions. MACS2 was used to identify ChIP-Seq enriched regions (or peaks)(Zhang et al. 2008), requiring peak calling FDR lower than 0.01. Distribution of resulting peaks in different genomic regions and motif analysis were conducted using HOMER suite ("annotatePeaks -size given" and "findMotifsGenome")[61] . Gene target for each peak was identified as gene with nearest TSS to the peak. Genome-wide tag density was computed using BEDTools and UCSC command line utilities to generate tracks for visualization. Each individual track was normalized to 1 million reads. For each ChIP sample, the corresponding H3 (for H3K27ac and H3K4me1) or input control (for KMT2D and p63) was subtracted from track using deepTools. Differential binding sites in shSC and shKMT2D conditions were identified by the DiffBind Bioconductor package (default parameters). Heatmaps were also generated by DiffBind. For KMT2D differential binding, we retrieved all sites with 2-fold differences for downstream analysis. Metaplot showing average ChIP signal near differential binding sites were generated using deepTools. Further, to include more differential binding sites that were not identified due to the stringency of DiffBind, a statistical test based differential analysis was performed. Tag density (AUC) was computed from each normalized, input-subtracted KMT2D sample using bwtool. For each merged KMT2D peak region, we used Student's T-test to examine whether its tag density is significantly different between shSC and scKMT2D samples. To identify the top 5% regions of histone modification loss, sequence tags for two replicates of H3K27ac and H3K4me1 ChIP-seq were aligned to the NCBI v37 assembly of

the human genome (hg19) using bowtie v1.1.1 (parameters -m 1 --best). Peaks were called for the two replicates using MACS2 callpeak, with matching whole histone subunit H3 ChIP-seq as a sonication efficiency control (parameters -s 75 -q 0.01). The peak sets for both histone PTMs were further filtered to use a 0.1% FDR. Peaks were then assessed for losses in the shKMT2D keratinocytes, and those peaks with a quantitative loss in the 95th percentile were associated to their nearest gene by distance to the TSS.

3.4.11: Statistics

For Venn diagrams, hypergeometric testing was performed. For RT-qPCR and the growth curve measurements, the standard error of the mean of three replicates is presented. Additional statistical methods listed in the "bioinformatics analyses" section above.

3.4.12: Chromatin co-immunoprecipitation and western blot

NHEKs were crosslinked in 1% formaldehyde (FisherScientific) in PBS for 5 minutes at room temperature. Glycine was used to quench the reaction followed by pelleting, collecting and lysing the cells as previously described[59] . Lysed cells were sonicated with a Covaris sonicator (S220). The supernatant was then diluted in the final sonication buffer, split equally, and subjected to immunoprecipitation at 4°C overnight with the corresponding antibodies: anti-KMT2D (HPA035977) (Sigma-Aldrich) and rabbit anti-IgG (Abcam ab46540). 10% of the supernatant was utilized for an Input control. The next day beads were washed thoroughly with RIPA buffer consisting of 50mM HEPES-KOH (pH 7.5), 500mM LiCl, 1mM EDTA, 1% NP-40 and 0.7% Na-Deoxycholate. Proteins were eluted from the beads in 50mM Tris-HCL (pH 8.0), 10mM EDTA and 1% SDS. Samples were then loaded and separated in polyacrylamide gels followed by transfer to a nitrocellulose membrane. Membranes were incubated in anti-p63-α antibody (Cell Signaling D2K8X) and imaged on a GE Amersham Imager 600.

3.4.13: Three-dimensional (3D) organotypic human skin cultures

The 3D organotypic human skin cultures were performed as previously described(Simpson et al. 2010). Briefly, J2 3T3 fibroblasts were grown in DMEM+10% FBS. Cells were released from culture plates using 0.25% trypsin for 5 min at 37C, then re-suspended in DMEM+10% FBS and counted using a hemacytometer to determine the volume needed to obtain 0.75-1 million fibroblasts per organotypic culture. The required volume was centrifuged in a 50 ml sterile conical tube at 200g for 5 minutes and the supernatant was removed. The fibroblast cell pellet was re-suspended in 1/10 the final required volume (2 ml per culture) of 10x Collagen Re-suspension Buffer (1.1 g NaHCO3 plus 2.39 g HEPES in 50 ml 0.05 N NaOH) and held on ice. 1/10th the final volume of 10x DMEM (Sigma) was then added and the cells were mixed by vigorous pipetting. Purified high-concentration rat tail collagen I (Corning) was added and diluted with sterile dH2O to a final concentration of 4 mg/ml of the final volume. 0.05 N NaOH was added to a pH of approximately 7. The collagen-fibroblast slurry was mixed by inverting, then 2 ml was pipetted into the upper chamber of a 6-well transwell insert (Corning) placed within a deep-well 6-well tissue culture plate (Corning). The fibroblast-collagen matrices were allowed to polymerize at 37C for 60 min. Then, the matrices were submerged in DMEM+10% FBS and placed at 37C overnight. The next day, NHEKs were trypsinized, re-suspended in DMEM+10% FBS and counted to collect 1 million cells per culture, centrifuged at 200 g for 5 min, and the supernatant was discarded. The NHEK pellet was re-suspended in E-medium supplemented with 5 ng/ml EGF (Sigma) to a volume of 2 ml per culture. The DMEM was removed from both the upper and lower chambers of the transwell plates containing the collagen-fibroblast matrices. 2 ml of NHEKs (1 million cells) were seeded atop each matrix in the upper transwell chamber and 14 ml of E-medium with 5 ng/ml EGF was added to the bottom chamber. The cultures were placed at 37C overnight. The next day, the medium was aspirated from both the top and bottom chambers of the transwell. To place the NHEK monolayers at an air-liquid interface and induce stratification, 10 ml E-medium (without EGF supplementation) was added only to the bottom chamber of the transwell and the cultures were grown at 37C for up to 12 days, feeding 10 ml E-medium every

other day. To generate protein lysates, the transwell apparatus was removed from the plate and the organotypic culture was separated from the underlying matrix using sterile forceps. The culture was transferred into urea sample buffer (8 M urea, 1% SDS, 10% glycerol, 60 mM Tris, 5% beta-mercaptoethanol, pH 6.8) and dissolved by vigorous pipetting using a 25 gauge needle and a 1 ml syringe. Organotypic cultures were prepared for routine histology by submerging the culture in 10% neutral buffered formalin for 24-48 hrs.

3.4.14: Data availability

The relevant data sets generated during this study have beenuploaded to NCBI Gene Expression Omnibus and are availableunder accession numbers GSE110475, GSE110453, GSE110467, GSM2991421–GSM2991435, and GSM2991609–GSM2991620.

3.4.15: Gene Ontology analyses

All Gene Ontology (GO) analyses were performed using PANTHER at http://geneontology.org/page/go-enrichment-analysis.

3.5: Supplemental Figures
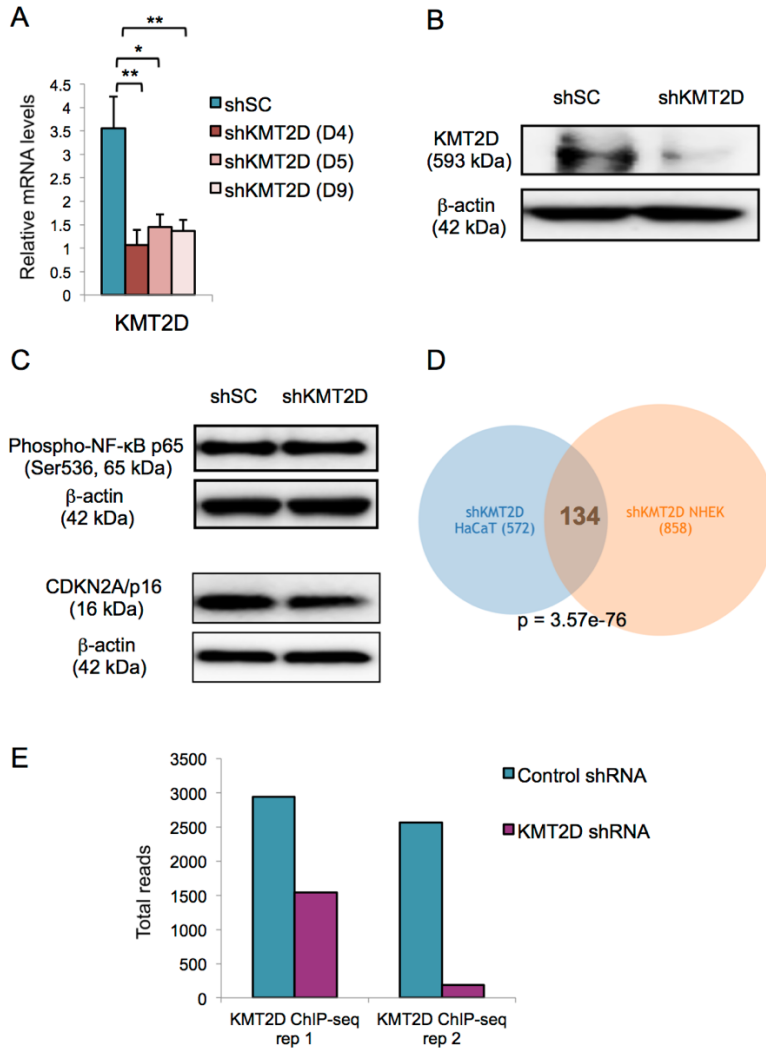
Supplemental Figure S3.1: *(A)* RT-qPCR confirms a significant loss of KMT2D mRNA expression with shKMT2D treatment using 3 unique shRNAs in comparison to shSC keratinocytes. *(B)* Western blotting demonstrates reduced levels of KMT2D in shKMT2D keratinocytes. *(C)* Western blotting also demonstrates no differences in the levels of the tumor suppressor p16/CDKN2A nor the senescence-associated transcription factor phospho-NF-κB p65 (Ser536). *(D)* Intersection of the significant (p<0.1) genes with reduced expression in both shKMT2D NHEK and HaCaT keratinocytes in comparison to shSC keratinocytes demonstrates a significant overlap of 134 genes (p = 3.57e-76).
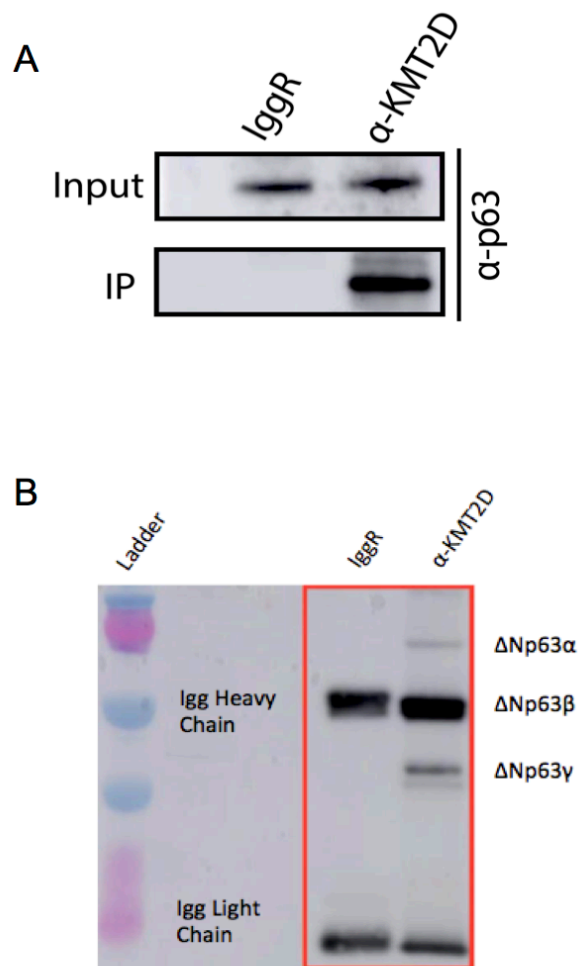
Supplemental Figure S3.2: *(A)* Formaldehyde crosslinking followed by co-immunoprecipitation (Co-IP) experiments show that p63 and KMT2D interact on chromatin, as immunoprecipitated KMT2D is able to pull down p63, while control IgG is not. *(B)* Using a ΔNp63 specific antibody we determined that KMT2D interacts with the α, β and γ isoforms of ΔNp63, suggesting the interaction is not mediated by the C-terminal sterile alpha motif (SAM) domain or the transactivation (TA) domain absent in the ΔN forms.
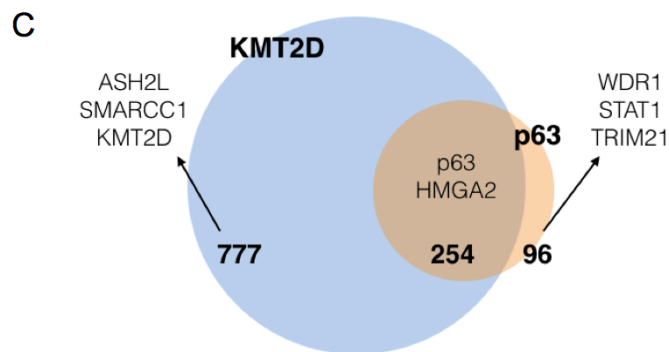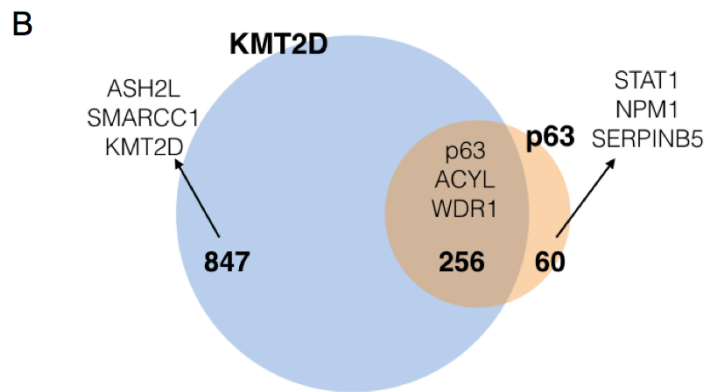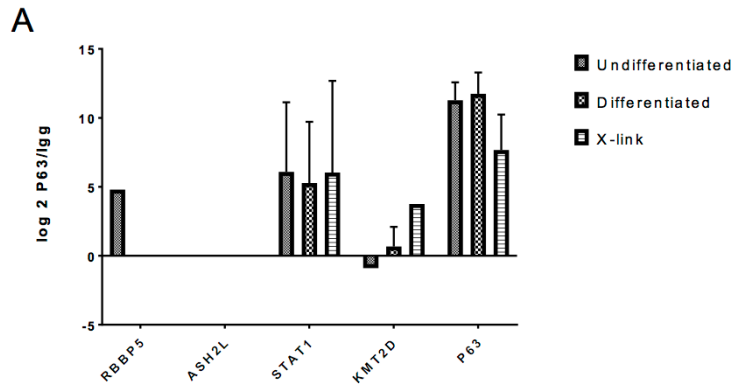
Supplemental Figure S3.3: *(A)* IP-Mass Spec identifies interactors of p63, e.g. STAT1. KMT2D

is identified as an interactor of p63 in both crosslinked NHEKs and uncrosslinked and DNAse

treated differentiated NHEKs. *(B)* IP-Massspec of p63 and KMT2D in undifferentiated,

uncrosslinked and DNAse treated NHEKs reveals 256 shared interacting genes, including p63.

(C) IP-Massspec of p63 and KMT2D in differentiated, uncrosslinked and DNAse treated NHEKs

reveals 254 shared interacting genes, including p63.
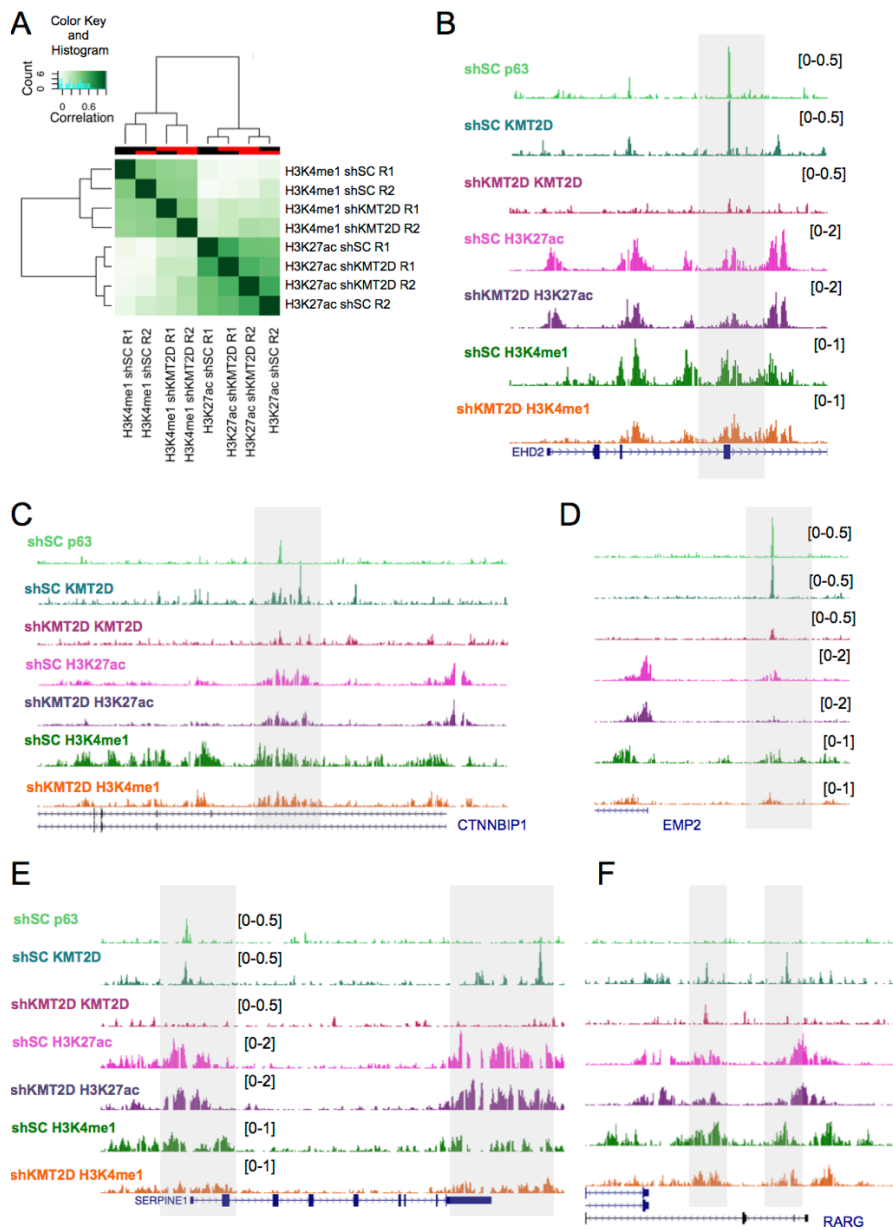
Supplemental Figure S3.4: *(A)* Hierarchal clustering of two replicates of both H3K4me1 and H3K27ac demonstrates separate clustering of the two unique histone modifications. ChIP-seq tracks from representative genes that lose KMT2D enrichment at p63 bound regions, as well as enhancer histone modifications. Examples include *EHD2 (B)*, *CTNNBIP1 (C)*, *EMP2 (D)*, *SERPINE1 (E)*, and *RARG (F)*.

A

B
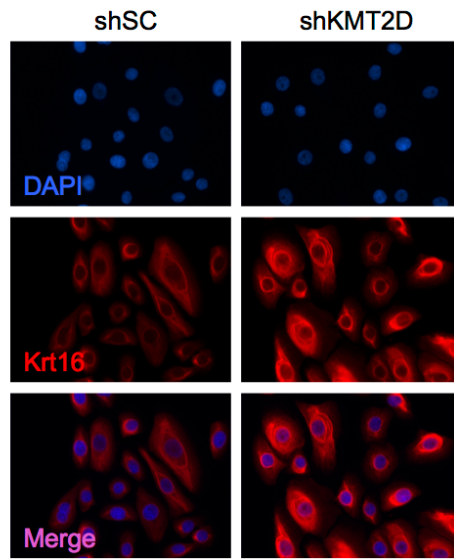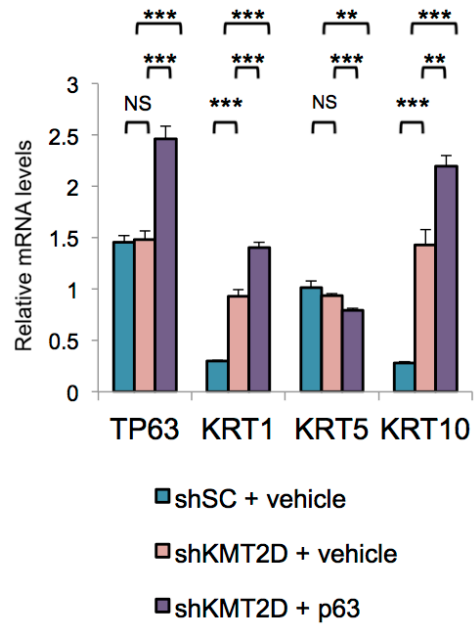
Supplemental Figure S3.5: shKMT2D keratinocytes display increased levels of Krt16 staining by IF in comparison to shSC keratinocytes (40x).
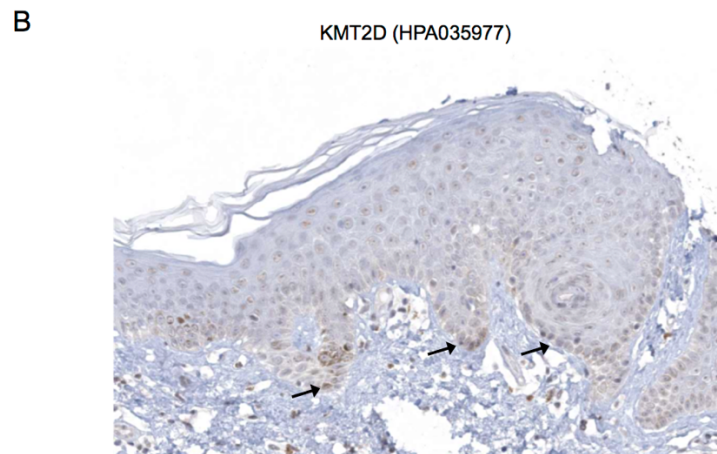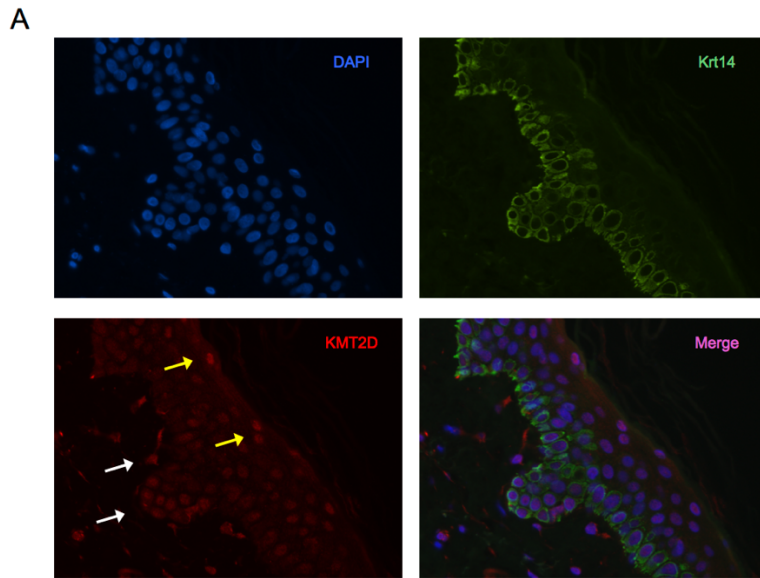
Supplemental Figure S3.6: *(A)* IF staining of normal human epidermis with KMT2D antibodies demonstrates that KMT2D (red) is present both in the basal epidermis (white arrows) that is also keratin 14 positive, but also in the outer more differentiated layers (yellow arrows), suggesting that it may play a role at all levels of epithelial stratification and differentiation. *(B)* Publicly available immunohistochemistry (IHC) data (antibody HPA035977) from the Human Protein Atlas displays similar findings of KMT2D's presence at all levels of the epidermis, including the basal epidermal layer. *(C)* Overexpression of p63 in shKMT2D keratinocytes does not rescue the premature activation of differentiation-associated genes such as *KRT1* and *KRT10*.

3.6: References

1. Sur I, Taipale J. 2016. The role of enhancers in cancer. *Nat Rev Cancer* **16**: 483-493.

2. Gonda TJ, Ramsay RG. 2015. Directly targeting transcriptional dysregulation in cancer. *Nat Rev Cancer* **15**: 686-694.

3. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA et al. 2013. Mutational landscape and significance across 12 major cancer types. *Nature* **502**: 333-339.

4. Ford DJ, Dingwall AK. 2015. The cancer COMPASS: navigating the functions of MLL complexes in cancer. *Cancer Genet* **208**: 178-191.

5. Rao RC, Dou Y. 2015. Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. *Nature reviews Cancer* **15**: 334-346.

6. Sze CC, Shilatifard A. 2016. MLL3/MLL4/COMPASS Family on Epigenetic Regulation of Enhancer Function and Cancer. *Cold Spring Harb Perspect Med* **6**.

7. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E et al. 2012. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**: 401-404.

8. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E et al. 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**: pl1.

9. Herz HM, Mohan M, Garruss AS, Liang K, Takahashi YH, Mickey K, Voets O, Verrijzer CP, Shilatifard A. 2012. Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. *Genes & development* **26**: 2604-2620.

10. Hu D, Gao X, Morgan MA, Herz HM, Smith ER, Shilatifard A. 2013. The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Mol Cell Biol* **33**: 4745-4754.

11. Lee JE, Wang C, Xu S, Cho YW, Wang L, Feng X, Baldridge A, Sartorelli V, Zhuang L, Peng W et al. 2013. H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. *Elife* **2**: e01503.

12. Wang C, Lee JE, Lai B, Macfarlan TS, Xu S, Zhuang L, Liu C, Peng W, Ge K. 2016. Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. *Proc Natl Acad Sci U S A* **113**: 11871-11876.

13. Lai B, Lee JE, Jang Y, Wang L, Peng W, Ge K. 2017. MLL3/MLL4 are required for CBP/p300 binding on enhancers and super-enhancer formation in brown adipogenesis. *Nucleic Acids Res*.

14. Wang SP, Tang Z, Chen CW, Shimada M, Koche RP, Wang LH, Nakadai T, Chramiec A, Krivtsov AV, Armstrong SA et al. 2017. A UTX-MLL4-p300 Transcriptional Regulatory Network Coordinately Shapes Active Enhancer Landscapes for Eliciting Transcription. *Mol Cell* **67**: 308-321 e306.

15. Ang SY, Uebersohn A, Spencer CI, Huang Y, Lee JE, Ge K, Bruneau BG. 2016. KMT2D regulates specific programs in heart development via histone H3 lysine 4 di-methylation. *Development* **143**: 810-821.

16. Ortega-Molina A, Boss IW, Canela A, Pan H, Jiang Y, Zhao C, Jiang M, Hu D, Agirre X, Niesvizky I et al. 2015. The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development. *Nat Med* **21**: 1199-1208.

17. Zhang J, Dominguez-Sola D, Hussein S, Lee JE, Holmes AB, Bansal M, Vlasevska S, Mo T, Tang H, Basso K et al. 2015. Disruption of KMT2D perturbs germinal center B cell development and promotes lymphomagenesis. *Nat Med* **21**: 1190-1198.

18. Issaeva I, Zonis Y, Rozovskaia T, Orlovsky K, Croce CM, Nakamura T, Mazo A, Eisenbach L, Canaani E. 2007. Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. *Mol Cell Biol* **27**: 1889-1903.

19. Guo C, Chen LH, Huang Y, Chang CC, Wang P, Pirozzi CJ, Qin X, Bao X, Greer PK, McLendon RE et al. 2013. KMT2D maintains neoplastic cell proliferation and global histone H3 lysine 4 monomethylation. *Oncotarget* **4**: 2144-2153.

20. Rollman O, Jensen UB, Ostman A, Bolund L, Gustafsdottir SM, Jensen TG. 2003. Platelet derived growth factor (PDGF) responsive epidermis formed from human keratinocytes transduced with the PDGF beta receptor gene. *J Invest Dermatol* **120**: 742-749.

21. Lee DD, Stojadinovic O, Krzyzanowska A, Vouthounis C, Blumenberg M, Tomic-Canic M. 2009. Retinoid-responsive transcriptional changes in epidermal keratinocytes. *J Cell Physiol* **220**: 427-439.

22. Hara T, Miyazaki M, Hakuno F, Takahashi S, Chida K. 2011. PKCeta promotes a proliferation to differentiation switch in keratinocytes via upregulation of p27Kip1 mRNA through suppression of JNK/c-Jun signaling under stress conditions. *Cell Death Dis* **2**: e157.

23. Bikle DD. 2012. Vitamin D and the skin: Physiology and pathophysiology. *Rev Endocr Metab Disord* **13**: 3-19. -. 2014. The vitamin D receptor: a tumor suppressor in skin. *Adv Exp Med Biol* **810**: 282-302.

24. Chen W, Xiao Liu Z, Oh JE, Shin KH, Kim RH, Jiang M, Park NH, Kang MK. 2012. Grainyhead-like 2 (GRHL2) inhibits keratinocyte differentiation through epigenetic mechanism. *Cell Death Dis* **3**: e450.

25. Xu M, Horrell J, Snitow M, Cui J, Gochnauer H, Syrett CM, Kallish S, Seykora JT, Liu F, Gaillard D et al. 2017. WNT10A mutation causes ectodermal dysplasia by impairing progenitor cell proliferation and KLF4-mediated differentiation. *Nat Commun* **8**: 15397.

26. Rodriguez-Boulan E, Macara IG. 2014. Organization and execution of the epithelial polarity programme. *Nat Rev Mol Cell Biol* **15**: 225-242.

27. Ferone G, Mollo MR, Missero C. 2015. Epidermal cell junctions and their regulation by p63 in health and disease. *Cell Tissue Res* **360**: 513-528.

28. Petridou NI, Spiro Z, Heisenberg CP. 2017. Multiscale force sensing in development. *Nat Cell Biol* **19**: 581-588.

29. Ko MS, Marinkovich MP. 2010. Role of dermal-epidermal basement membrane zone in skin, cancer, and developmental disorders. *Dermatol Clin* **28**: 1-16.

30. Simpson CL, Patel DM, Green KJ. 2011. Deconstructing the skin: cytoarchitectural determinants of epidermal morphogenesis. *Nat Rev Mol Cell Biol* **12**: 565-580.

31. Stark R, Brown, G. 2011. DiffBind: differential binding analysis of ChIP-Seq peak data.

32. Romano RA, Smalley K, Magraw C, Serna VA, Kurita T, Raghavan S, Sinha S. 2012. DeltaNp63 knockout mice reveal its indispensable role as a master regulator of epithelial development and differentiation. *Development* **139**: 772-782.

33. McDade SS, Henry AE, Pivato GP, Kozarewa I, Mitsopoulos C, Fenwick K, Assiotis I, Hakas J, Zvelebil M, Orr N et al. 2012. Genome-wide analysis of p63 binding sites identifies AP-2 factors as co-regulators of epidermal differentiation. *Nucleic Acids Res* **40**: 7190-7206.

34. Kouwenhoven EN, Oti M, Niehues H, van Heeringen SJ, Schalkwijk J, Stunnenberg HG, van Bokhoven H, Zhou H. 2015. Transcription factor p63 bookmarks and regulates dynamic enhancers during epidermal differentiation. *EMBO Rep* **16**: 863-878.

35. Maytin EV, Habener JF. 1998. Transcription factors C/EBP alpha, C/EBP beta, and CHOP (Gadd153) expressed during the differentiation program of keratinocytes in vitro and in vivo. *J Invest Dermatol* **110**: 238-246.

36. Eckert RL, Adhikary G, Young CA, Jans R, Crish JF, Xu W, Rorke EA. 2013. AP1 transcription factors in epidermal differentiation and skin cancer. *J Skin Cancer* **2013**: 537028.

37. Watt FM. 2016. Engineered Microenvironments to Direct Epidermal Stem Cell Behavior at Single-Cell Resolution. *Dev Cell* **38**: 601-609.

38. Cavazza A, Miccio A, Romano O, Petiti L, Malagoli Tagliazucchi G, Peano C, Severgnini M, Rizzi E, De Bellis G, Bicciato S et al. 2016. Dynamic Transcriptional and Epigenetic Regulation of Human Epidermal Keratinocyte Differentiation. *Stem Cell Reports* **6**: 618-632.

39. Carroll DK, Carroll JS, Leong CO, Cheng F, Brown M, Mills AA, Brugge JS, Ellisen LW. 2006. p63 regulates an adhesion programme and cell survival in epithelial cells. *Nat Cell Biol* **8**: 551-561.

40. Dorighi KM, Swigut T, Henriques T, Bhanu NV, Scruggs BS, Nady N, Still CD, 2nd, Garcia BA, Adelman K, Wysocka J. 2017. Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Mol Cell* **66**: 568-576 e564.

41. Simpson CL, Kojima S, Getsios S. 2010. RNA interference in keratinocytes and an organotypic model of human epidermis. *Methods Mol Biol* **585**: 127-146.

42. Yang A, Kaghad M, Wang Y, Gillett E, Fleming MD, Dotsch V, Andrews NC, Caput D, McKeon F. 1998. p63, a p53 homolog at 3q27-29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities. *Mol Cell* **2**: 305-316.

43. Mills AA, Zheng B, Wang XJ, Vogel H, Roop DR, Bradley A. 1999. p63 is a p53 homologue required for limb and epidermal morphogenesis. *Nature* **398**: 708-713.

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

45. Rinne T, Brunner HG, van Bokhoven H. 2007. p63-associated disorders. *Cell Cycle* **6**: 262-268.

46. Truong AB, Kretz M, Ridky TW, Kimmel R, Khavari PA. 2006. p63 regulates proliferation and differentiation of developmentally mature keratinocytes. *Genes Dev* **20**: 3185-3197.

47. Soares E, Zhou H. 2017. Master regulatory role of p63 in epidermal development and disease. *Cell Mol Life Sci*.

48. Kommagani R, Caserta TM, Kadakia MP. 2006. Identification of vitamin D receptor as a target of p63. *Oncogene* **25**: 3745-3751.

49. Kommagani R, Leonard MK, Lewis S, Romano RA, Sinha S, Kadakia MP. 2009. Regulation of VDR by deltaNp63alpha is associated with inhibition of cell invasion. *J Cell Sci* **122**: 2828-2835.

50. Avgustinova A, Benitah SA. 2016. Epigenetic control of adult stem cell function. *Nat Rev Mol Cell Biol* **17**: 643-658.

51. Blanpain C, Fuchs E. 2014. Stem cell plasticity. Plasticity of epithelial stem cells in tissue regeneration. *Science* **344**: 1242281.

52. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.

53. Dawkins JB, Wang J, Maniati E, Heward JA, Koniali L, Kocher HM, Martin SA, Chelala C, Balkwill FR, Fitzgibbon J et al. 2016. Reduced Expression of Histone Methyltransferases KMT2C and KMT2D Correlates with Improved Outcome in Pancreatic Ductal Adenocarcinoma. *Cancer Res* **76**: 4861-4871.

54. Toska E, Osmanbeyoglu HU, Castel P, Chan C, Hendrickson RC, Elkabets M, Dickler MN, Scaltriti M, Leslie CS, Armstrong SA et al. 2017. PI3K pathway regulates ER-dependent transcription in breast cancer through the epigenetic regulator KMT2D. *Science* **355**: 1324-1330.

55. Mo R, Rao SM, Zhu YJ. 2006. Identification of the MLL2 complex as a coactivator for estrogen receptor alpha. *J Biol Chem* **281**: 15714-15720.

56. Kim JH, Sharma A, Dhar SS, Lee SH, Gu B, Chan CH, Lin HK, Lee MG. 2014. UTX and MLL4 coordinately regulate transcriptional programs for cell proliferation and invasiveness in breast cancer cells. *Cancer Res* **74**: 1705-1717.

57. Uray IP, Dmitrovsky E, Brown PH. 2016. Retinoids and rexinoids in cancer prevention: from laboratory to clinic. *Semin Oncol* **43**: 49-64.

58. Cox J, Mann M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**: 1367-1372.

59. Capell BC, Drake AM, Zhu J, Shah PP, Dou Z, Dorsey J, Simola DF, Donahue G, Sammons M, Rai TS et al. 2016. MLL1 is essential for the senescence-associated secretory phenotype. *Genes Dev* **30**: 321-336.

60. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359.

61. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**: 576-589.

62. COSMIC, the Catalogue of Somatic Mutations in Cancer.

Chapter 4: Conclusions and Future Directions

4.1: Summary and main findings

A key role of transcription factor p63 in mediating epidermal commitment, development, and differentiation has been extensively demonstrated through mouse models and causative point mutations in human disease. In mice, knockout of p63 leads to developmental and morphological defects in the squamous epithelia and epidermis, causing abnormal craniofacial development, truncated limbs, and loss of salivary glands, hair follicles, and teeth[1,2]. In humans, germline missense mutations in p63 cause ectodermal dysplastic syndromes, which are characterized by orofacial clefting and limb malformations[3]. To date, how p63 engages and modifies chromatin to promote normal development during these early stages is still being elucidated.

Here we report a trans-differentiation model as a novel approach to determine the role of p63 in craniofacial development. We show that p63 can establish epithelial enhancers to regulate downstream genes and that this process is disrupted when p63 carries mutations derived from human disease. In particular, our results show a novel mechanism underlying pathogenic p63 mutations, in which failure to open and establish enhancers leads to transcriptional dysregulation. We integrate our epigenomic data with GWAS datasets derived from nsCL/P patients to identify novel CL/P candidate genes regulated by p63. We also uncover strong enrichment of nsCL/P-associated SNPs at enhancers established by p63, providing mechanistic insight into how these SNPs lead to human pathogenesis. Further, in a follow up study, we identify histone methyltransferase KMT2D as a key partner of p63 at epithelial enhancers. Our data demonstrates for the first time a critical role for KMT2D in epithelial homeostasis by maintaining the adhesion and proliferative capacity of the epidermal basement membrane through its interaction with p63 to

regulate critical enhancers for genes involved in both maintaining the basal epidermis as well as in the proper coordination of epithelial stratification.

In our first study[4] using an inducible trans-differentiation model converting fibroblasts into keratinocyte-like cells we demonstrated the following main findings:

1. <u>p63 establishes enhancers to upregulate epithelial and inflammation genes</u>: while previous studies had focused on the role of p63's engagement with chromatin during keratinocyte differentiation[5–7], *de novo* engagement of p63 with chromatin had not been extensively addressed due in part to a lack of tractable human models.

2. <u>Co-binding of p63 and KLF4 leads to robust establishment of enhancers that correspond remarkably well to *bone fide* enhancers in basal and differentiating keratinocytes</u>: based on published reports[8] we expanded our in vitro system to better recapitulate the epigenetic conditions of early development.

3. <u>Binding of p63 to chromatin is a pre-requisite for enhancer establishment:</u> using a patient derived p63 mutation in the DBD we show complete loss of p63 binding to chromatin and loss of enhancer establishment.

4. <u>Binding of p63 to chromatin is required but not sufficient for enhancer establishment:</u> an AEC patient derived substitution mutation in the SAM domain of p63 leads to a protein that retains binding to chromatin but shows defects in enhancer establishment. The main defect is in opening chromatin, pointing towards a potentially novel function of the p63 SAM domain in recruiting chromatin remodelers for enhancer establishment. Further, our results contrast a recent report[9] suggesting the main disease mechanism in AEC is aggregation of mutant p63, and posits a novel molecular link to explain disease phenotypes found in these patients.

5. <u>We identified potential candidate causative genes for cleft lip/palate regulated by p63 at and outside of known risk loci:</u> by integrating epigenomic data-sets from our trans-differentiation model with GWAS datasets we were able to gain novel insights into molecular mechanisms underlying nsCL/P. In particular, we were able to prioritize candidate genes within half of the known risk loci and also identify over 100 new candidate genes and risk loci for nsCL/P. Importantly, our analysis also has the potential to identify the enhancers that regulate them.

6. <u>Enhancers established by p63+KLF4 are enriched for SNPs associated with nsCL/P:</u> using 2 different co-localization analyses we identify enrichment of SNPs associated with nsCL/P only at newly established enhancers. This thus provides a novel molecular mechanism for how these SNPs may contribute to disease at a molecular level.

Taken together, our findings in our first study address how p63 binds and remodels chromatin with its first encounter, identifies and prioritizes nsCL/P candidate genes and the enhancers that regulate them, and provides new explanations for how p63 mutations and SNPs associated with nsCL/P underlie disease.

In our second study[7] we wanted to ask which proteins interact with p63 on chromatin to establish enhancer structure. We identified methyltransferase KMT2D interacting with p63 at a broad array of enhancers involved in epithelial development, adhesion, and differentiation. The main findings in this study were:

1. <u>KMT2D plays a key role in maintaining the hallmark gene expression program and proliferative capacity of epithelial progenitors</u>: KMT2D, one of the most frequently mutated genes in human cancer, has been extensively characterized. However, its role in epithelial cells and homeostais had not been previously explored. We show

depletion of KMT2D leads to reduced proliferation and reduced expression of epithelial development and adhesion genes, including *DLG2, RARG, ITGB4* and *LAMB2.*

2. <u>KMT2D is enriched genomewide at p63 target enhancers</u>: ChIP-seq analysis demonstrated that KMT2D and p63 are enriched at overlapping sites across the genome, with as high as 49% of p63 peaks in keratinocytes containing a directly overlapping KMT2D peak. Many of these included known p63 target enhancers, such as *MINK1, SERINC2, IRF6* and *TP63* itself.

3. <u>KMT2D and p63 interact on chromatin</u>: formaldehyde crosslinking followed by immuno-precipitation and mass spectrometry show p63 and KMT2D interact on chromatin. We also find over 1000 shared binding partners which could mediate this interaction.

4. <u>KMT2D is enriched at p63 targets and maintains the expression of key genes involved in epithelial homeostasis</u>: Knock down(KD) of KMT2D shows reduction of KMT2D at p63 target enhancers and reduction of gene expression of corresponding genes. Pathways affected include epidermal commitment and stratification, as well as cell junction and adhesion.

5. <u>KMT2D can drive expression of p63-target genes through both catalytic histone methylation activity, as well as catalytically independent mechanisms at enhancers</u>: Profiling of H3K4me1 (catalyzed by KMT2D) and H3K27ac (KMT2D recruits the acetyltransferase CBP/p300[10–12]) shows KD of KMT2D correlates with loss of H3K4me1 and H3K27ac at p63 regulated enhancers of 82 genes involved in development and differentiation. In another 20 loci we observe loss of gene

expression with no changes in H3K4me1 and H3K27ac, consistent with recent evidence demonstrating histone methyltransferase independent functions of KMT2D in regulating gene expression[13].

6. <u>KMT2D coordinates with p63 to maintain epidermal progenitor gene expression and proper epidermal differentiation and stratification:</u> Investigation of genes upregulated upon KD of KMT2D showed a striking enrichment of genes involved in keratinocyte differentiation, including canonical genes, such as *KLF4, GRHL1, GRHL3* and *ZNF750*. Using a three-dimensional human organotypic skin culture we observed keratinocytes treated with shKMT2D showed disrupted stratification dynamics, as well as premature cornification. Together with the above findings these results demonstrated a key role of KMT2D in maintaining epidermal progenitor gene expression and preventing terminal differentiation.

In conclusion, we provide the first description of KMT2D's role at epithelial enhancers, including its genome-wide interactions with the transcription factor p63 to coordinate key gene expression programs required for both epithelial cell identity and the maintenance of epithelial progenitors.

For an extended discussion on both studies, please refer to the discussion section of chapter 2 (pages 45-49) and chapter 3 (pages 80-81).

4.2: Future directions

The studies described in chapter 2 and 3 greatly increased our knowledge about p63's role in epithelial enhancer establishment and revealed a potential disease mechanism in orofacial clefting, connecting SNPs and enhancer regulation. Further, they provided insight into possible binding partners of p63 at enhancers, including histone methyltranferase KMT2D; and described

a previously unknown role of KMT2D in maintaining epithelial homeostasis. However, many questions remain. In the following sections I will describe open questions and possible follow up experiments.

4.2.1: Dissecting how specific SNPs may cause CL/P

In chapter 2 we found SNPs highly associated to CL/P are enriched at newly established enhancers by p63 and KLF4. We further discussed in detail two loci that are prominently linked to CL/P but where a molecular mechanism to explain disease has yet to be found: 8q24 locus and *MAFB* locus[14–17]. Our colocalization analyses identified in both cases newly established enhancers by p63 and KLF4 colocalize with SNPS highly associated to CL/P, however how these SNPs cause disease remains to be shown. In particular, in these 2 cases and in many other examples we find several enhancers (exhibiting open chromatin and de novo H3K27ac) established by p63 and KLF4 and it remains unclear if all are necessary for downstream regulation of gene expression. Further, in the case of the 8q24 locus and other known risk loci, the exact genes affected by the SNPs have not been determined. To start to answer these questions it would be interesting to test some of these enhancers using luciferase reporter systems to see if they are active in human keratinocytes and our trans-differentiation model, as well as during epidermal development and development of the lip and palate in murine models. SNPs could easily be integrated into these enhancer reporter assays to test their effect in gene expression. A more thorough, albeit technically challenging alternative is to use CRISPR systems to introduce specific SNPs either in our trans differentiation model or an epidermal differentiation model derived from pluripotent stem cells, to test how they affect downstream gene expression and differentiation.

A more high-throughput possibility to identify how newly established enhancers regulate downstream gene expression is using chromatin capture assays to study whether these newly established enhancers are looping to promoters. Looping to promoters could then be correlated with differential gene expression to filter key enhancers and corresponding SNPs to test.

Another observation from our studies is that many nsCL/P highly associated SNPs cluster around established enhancers but do not seem to be found within regulatory elements (based on chromatin signature). Whether these SNPs are simply in linkage disequilibrium with disease causative SNPs or actively contribute to disease through independent mechanisms remains to be elucidated. The experiments proposed above would start to shed light into these processes and greatly expand our knowledge around cleft lip/palate pathogenesis.

4.2.2: Understanding the role of lncRNAs in epidermal specification and differentiation

Another exciting result from our trans-differentiation experiments is that many lncRNAs become *de novo* expressed after establishment of enhancers by p63 and KLF4. In particular, in the 8q24 risk locus, which shows the strongest effect size across populations of Asian and European origin[18] we observe *de novo* expression of lncRNA *LINC00977*. How and whether this and other lncRNAs regulate epidermal specification and differentiation, as well as more complex processes such as lip and palate development is not understood. A recent study harnessed the versatility of gene editing tools to probe how lncRNAs could confer resistance to melanoma treatments[19]. By using an enzymatically inactive Cas9 (dCas9) version coupled to an activating VP64 protein complex targeted to over 10000 lncRNA promoter regions they were able to screen which ones conferred a resistance phenotype, suggesting a role for specific lncRNAs in this process. To begin to understand the role of lncRNAs in epidermal specification and differentiation, I propose a similar screen. We could target a large amount of lncRNAs in human keratinocytes with either activation (dCas9-VP64) or repression (dCas9-KRAB) while in parallel

inducing keratinocyte differentiation using calcium supplemented media. After three days of differentiation, keratinocytes can be sorted using basal vs differentiation surface markers. Comparison of sgRNA counts can then point us towards lncRNAs crucial for keratinocyte differentiation. We could then confirm their function using three-dimensional skin organotyoic cultures as described in chapter 3 and then test their role in development using murine models.

4.2.3: Identifying new SNPs associated to CL/P and CPO

A remarkable finding from the GWAS data on CL/P is that the 40 identified risk loci explain about 25% of the cases of non-syndromic cleft lip/palate[20,21]. This is particularly surprising given the small sample sizes profiled (between 100-1000 patients per cohort). Most of the data has been collected from patients of Asian and European descent. In order to better understand genetic effects and how non-coding regions can cause orofacial clefting, it is imperative to increase the number of patients sequenced, as well as expanding to diverse non-European populations. Further, while GWAS have provided important insight into CL/P, much more remains unknow about cleft palate only (CPO) pathogenesis. To date, only a single risk loci has been confidently identified to explain CPO[22]. A major challenge here is the lower incidence of cases, as well as lack of large samples of data collected in these populations.

4.2.4: Understanding discordant phenotypes in identical twins with p63 mutations

A key observation in patients with p63 substitution mutations is that all patients found so far are heterozygous for the mutation and that even patients with the same mutation show very different phenotypes. Given the severe phenotype observed in mice with homozygous deletion of *Trp63* (encoding for p63 in mice)[1,2] it is likely that complete loss of function of p63 in humans is

embryonically lethal. A recent study[23] reported the first case of monozygotic twins carrying a p63 mutation in the DBD. Interestingly, the twins exhibited discordant phenotypes, with one twin showing limb malformations and normal craniofacial development and the other showing normal limb development but CL/P[23]. Together with our studies and the fact that p63 acts as a tetramer, this suggests there is an important dose and spatiotemporal component to p63's epithelial enhancer establishment during development. Previous studies looking at heterogenous composition of tetramers in p53 found that tetramers composed of at least three mutant p53 proteins showed impairment of transcriptional activity[24]. It is possible that a similar effect is seen in patients carrying a substitution mutation in one of their p63 alleles. To test this hypothesis, it would be interesting to overexpress wildtype p63 in a cell line that does not endogenously express it. In parallel, mutant versions of p63 could be introduced under a doxycycline inducible promoter and effects on gene expression could be monitored. It would also be interesting to obtain keratinocytes from the monozygotic twins mentioned above and profile active enhancers to see if there are specific differences in enhancer structures. Given the stochastic nature of tetramer formation, it may be possible that defects on enhancer establishment and gene regulation may be lost within a population of cells. A better system would be to utilize a murine model in which a substitution mutation is introduced in a single allele of p63 followed by monitoring diverse phenotypes, more effectively recapitulating the spatiotemporal and dose dependent nature of p63's enhancer establishment in development.

4.2.5: Translation to treatment to prevent orofacial clefting

An important question is how a better understanding of p63's role in orofacial clefting can be translated into therapy, especially since CL/P is one of the most common congenital malformations reported, second only to heart congenital malformations[20]. Firstly, in the case of patients carrying a substitution mutation in p63, patients could possibly benefit from silencing of the mutant allele during development. This idea is bolstered by the observation that mice with

homozygous deletion of p63 show severe phenotypes that are lethal shortly after birth, while heterozygous mice with only 1 allele of p63 deleted show no difference compared to mice with two copies of WT p63[25]. How this would be achieved in human embryo development is presently unclear, a therapy involving allele specific siRNA or a small molecule inhibitor that only recognized mutant versions of p63 are a possibility but must be thoroughly tested for safety and efficiency.

Downstream pathways affected by p63 enhancer establishment may provide less disruptive treatment possibilities to prevent orofacial clefting. For example, we observe p63 establishes enhancers upstream of folic acid receptors *FOLR1* and *FOLR3* and that establishment of enhancers is followed by *de novo* expression of the respective genes in our trans-differentiation model. In mice it has been observed that KO of *folbp1* (encoding for the homolog of *FOLR1* in mice) leads to orofacial clefting[26]. This phenotype can be mitigated by supplementation of folic acid during embryo development in these mice. Similarly, it may be possible that defects in enhancer establishment in humans due to SNPs at enhancers of folic acid receptors lead to poor upregulation of these receptors and that risk of orofacial clefting might be mitigated through supplementation of folic acid in families that carry these SNPs.

4.2.6: Molecular mechanism behind birth defects observed with thalidomide treatment

Thalidomide, a drug used in the 1950s to treat morning sickness in pregnant women lead to over 10.000 infants with birth defects in Europe[27]. For almost 60 years, the molecular mechanisms behind the resulting birth malformations were poorly understood. A recent study showed degradation of *SALL4* as an important pathway in thalidomide induced birth defects[27]. Remarkably, the phenotypes observed in newborns, included limb and craniofacial malformations (e.g. CL/P) and closely mirrored the developmental phenotypes observed in patients carrying mutations in p63. Because of the high similarities in the phenotypes it is possible that thalidomide

113

interferes in epithelial lineage specification and enhancer establishment via p63. To test whether this is the case, it would be interesting to supplement cells in the fibroblast to keratinocyte-like trans-differentiation model with thalidomide at different doses and see if enhancer establishment is dysregulated. Presently, thalidomide continues to be used to treat a diverse array of diseases, from erythema nodosum leprosum to multiple myeloma. Treatment is correlated with significant adverse effects in the central nervous system, cardiovascular system and the dermis[28]. Understanding whether thalidomide interferes with enhancer establishment could provide important understanding of adverse effects associated with this drug.

4.2.7: p63's role in cancer and other disorders

While p63 initially garnered huge interest due to the high degree of sequence identity compared to p53, it was not shown to play a similar role in tumor suppression[25]. However, it has been observed that p63 is downregulated in some cancers undergoing epithelial to mesenchymal transition[29,30], suggesting a key role of p63 in protecting and maintaining epithelial identity. Intriguingly, in a subset of squamous carcinomas, p63 has been shown to be overexpressed, and overexpression has been connected to poor outcomes[31]. Based on the results of our studies, it would be interesting to study how enhancer landscape changes with loss or overexpression of p63 in cancer. Particularly, in cancers where p63 is overexpressed, specific enhancers might be aberrantly established to modulate a gene expression program conducive to tumor growth and maintenance. Some evidence for this was recently reported in a subtype of pancreatic ductal adenocarcinoma[31]; however, evidence in other cancer subtypes remains to be elucidated.

Another interesting phenotype observed in p63 KO mice and human patients carrying substitution mutations of p63 is T-cell lymphopenia[32,33]. It has been demonstrated that p63 plays an important role in the development of the thymic epithelium[32]. Importantly, in our first study we observed upregulation of genes involved in immune function when p63 was overexpressed by itself in dermal fibroblasts. Thus, our system might provide novel insight into T-cell lymphopenia and other diseases involved in T-cell development and maturation.

114

4.2.8: Mass spectrometry to dissect what parts of p63 recruit chromatin modifiers and other proteins to establish enhancers

Our experiments introducing substitution mutations in p63 revealed that the SAM domain of p63 is crucial for promoting open chromatin during enhancer establishment in epithelial lineage specification. Nevertheless, it remains unclear which proteins can engage and interact with the SAM domain of p63 to achieve this and it is also unknown how substitution mutations disrupt this process. To identify novel interacting partners, it would be interesting to clone and recombinantly purify a WT and a mutated version of the p63 SAM domain, followed by binding to a column and flow through of keratinocyte cell lysate. To the best of our knowledge, such a study has not been conducted yet. Interacting partners could then be isolated and probed using conventional bottom-up mass spectrometry methods, as shown in chapter 3. These could then be validated in our trans-differentiation model using shRNA knock downs to study their effects on chromatin remodeling and enhancer establishment.

4.2.9: Integration of GWAS data and epigenomics in other diseases

Around 85% of human DNA under evolutionary constraint corresponds to non-coding protein-sequences[34,35] a large fraction of which correspond to cis-regulatory elements. To date, many examples have been reported in which genetic variation at these sequences can cause disease[36–39]. At the same time, a plethora of GWAS studies have been carried out for all kinds of diseases, from rheumatoid arthritis to coronary heart disease and bipolar disorder. A main takeaway from these studies is that the vast majority of disease associated SNPs are not found within exons but rather at intronic and intergenic regions. While more recent studies have begun to combine gene expression data and expression quantitative trait loci (eQTLs) with GWAS data[40], they have garnered limited success. Particularly in the case of development, integration of

trans-differentiation and in vitro differentiation strategies similar to the one we show in chapter 2 could provide better insight into how specific SNPs may alter cis-regulatory elements to modulate gene expression.

4.2.10: KMT2D in cancer

The gene encoding the histone methyltransferase KMT2D is amongst the most frequently mutated genes in human cancer[41–45]. Previous studies have reported conflicting roles for KMT2D either as an oncogene or as a tumor suppressor in cancer, dependent on cellular context[42,46,47]. Our results in chapter 3 demonstrated that KMT2D depletion reduces keratinocyte proliferation, which is consistent with reports that KMT2D deficiency slows the proliferation of both breast and colon cancer cell lines[48,49], and correlates with improved outcomes in pancreatic cancer[50]. In contrast, our data suggesting a role for KMT2D in regulating genes involved in vitamin D and retinoid signaling would suggest a potential tumor suppressive role[51,52]. Indeed, deciphering these potential oncogenic versus tumor suppressive functions of KMT2D in epithelial cancers remains a critical question and important direction for future studies. Interestingly in pancreatic ductal adenocarcinoma reduced expression of KMT2D has been reported to correlate with better outcome[50], while overexpression of p63 has been demonstrated to lead to enhancer rewiring and poor outcome[31], pointing towards a role of KMT2D together with p63 in driving carcinogenesis in this type of cancer. More experiments must be conducted to understand how these proteins function together in the context of pancreatic cancer.

An additional intriguing observation from our mass spectrometry studies in chapter 3 is that KMT2D and wild type p53 can interact directly. A previous study from our lab showed that mutant p53 could interact with KMT2A, an H3K4 methyltransferase in the same family of proteins

as KMT2D, to drive cancer growth[53]. Given the key role of p53 in tumor suppression, it would be interesting to further characterize the interaction between these two proteins in the context of normal tumor suppression and in cancer.

4.3: References

1.   Yang, A. *et al.* p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. *Nature* 398, 714–718 (1999).

2.   Mills, A. A. *et al. p63* is a *p53* homologue required for limb and epidermal morphogenesis. *Nature* 398, 708–713 (1999).

3.   Bokhoven, H. van & McKeon, F. Mutations in the p53 homolog p63: allele-specific developmental syndromes in humans. *Trends Mol. Med.* 8, 133–139 (2002).

4.   Lin-Shiao, E. *et al.* p63 establishes epithelial enhancers at critical craniofacial development genes. *Sci. Adv.* In Press, (2019).

5.   Kouwenhoven, E. N. *et al.* Transcription factor p63 bookmarks and regulates dynamic enhancers during epidermal differentiation. *EMBO Rep.* 16, 863–878 (2015).

6.   Rapisarda, V. *et al.* p63 Transcription Factor Regulates Nuclear Shape and Expression of Nuclear Envelope-Associated Genes in Epidermal Keratinocytes. *J. Invest. Dermatol.* 137, 2157–2167 (2017).

7.   Lin-Shiao, E. *et al.* KMT2D regulates p63 target enhancers to coordinate epithelial homeostasis. *Genes Dev.* 32, 181–193 (2018).

8.   Chen, Y., Mistry, D. S. & Sen, G. L. Highly Rapid and Efficient Conversion of Human Fibroblasts to Keratinocyte-Like Cells. *J. Invest. Dermatol.* 134, 335–344 (2014).

9.   Russo, C. *et al.* Protein aggregation of the p63 transcription factor underlies severe skin fragility in AEC syndrome. *Proc. Natl. Acad. Sci.* 115, E906–E915 (2018).

10.   Wang, C. *et al.* Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. *Proc. Natl. Acad. Sci.* 113, 11871–11876 (2016).

11.   Lai, B. *et al.* MLL3/MLL4 are required for CBP/p300 binding on enhancers and super-enhancer formation in brown adipogenesis. *Nucleic Acids Res.* 45, 6388–6403 (2017).

12.   Wang, S.-P. *et al.* A UTX-MLL4-p300 Transcriptional Regulatory Network Coordinately Shapes Active Enhancer Landscapes for Eliciting Transcription. *Mol. Cell* 67, 308-321.e6 (2017).

13.   Dorighi, K. M. *et al.* Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Mol. Cell* 66, 568-576.e4 (2017).

14.   Beaty, T. H. *et al.* A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near *MAFB* and *ABCA4*. *Nat. Genet.* 42, 525–529 (2010).

15.   McDade, S. S. *et al.* Genome-wide analysis of p63 binding sites identifies AP-2 factors as co-regulators of epidermal differentiation. *Nucleic Acids Res.* 40, 7190–7206 (2012).

16.   Uslu, V. V. *et al.* Long-range enhancers regulating *Myc* expression are required for normal facial morphogenesis. *Nat. Genet.* 46, 753–758 (2014).

17.   Watanabe, K., Taskesen, E., Bochoven, A. van & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* 8, 1826 (2017).

18.   Grant, S. F. A. *et al.* A Genome-Wide Association Study Identifies a Locus for Nonsyndromic Cleft Lip with or without Cleft Palate on 8q24. *J. Pediatr.* 155, 909–913 (2009).

19.   Joung, J. *et al.* Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. *Nature* 548, 343–346 (2017).

20.   Ludwig, K. U. *et al.* Imputation of orofacial clefting data identifies novel risk loci and sheds light on the genetic background of cleft lip ± cleft palate and cleft palate only. *Hum. Mol. Genet.* 26, 829–842 (2017).

21.   Yu, Y. *et al.* Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. *Nat. Commun.* 8, 14364 (2017).

22.   Mangold, E. *et al.* Sequencing the GRHL3 Coding Region Reveals Rare Truncating Mutations and a Common Susceptibility Variant for Nonsyndromic Cleft Palate. *Am. J. Hum. Genet.* 98, 755–762 (2016).

23.   Wenger, T. *et al.* Expanding the phenotypic spectrum of TP63-related disorders including the first set of monozygotic twins. *Am. J. Med. Genet. A.* 176, 75–81 (2018).

24.   Chan, W. M., Siu, W. Y., Lau, A. & Poon, R. Y. C. How Many Mutant p53 Molecules Are Needed To Inactivate a Tetramer? *Mol. Cell. Biol.* 24, 3536–3551 (2004).

25.   Keyes, W. M. *et al.* p63 heterozygous mutant mice are not prone to spontaneous or chemically induced tumors. *Proc. Natl. Acad. Sci.* 103, 8435–8440 (2006).

26.   Piedrahita, J. A. *et al.* Mice lacking the folic acid-binding protein Folbp1 are defective in early embryonic development. *Nat. Genet.* 23, 228–232 (1999).

27.   Donovan, K. A. *et al.* Thalidomide promotes degradation of SALL4, a transcription factor implicated in Duane Radial Ray syndrome. *eLife* 7, e38430 (2018).

28.   Lilienfeld-Toal, M. V. *et al.* A systematic review of phase II trials of thalidomide/dexamethasone combination therapy in patients with relapsed or refractory multiple myeloma. *Eur. J. Haematol.* 81, 247–252 (2008).

29.   Lindsay, J., McDade, S. S., Pickard, A., McCloskey, K. D. & McCance, D. J. Role of ΔNp63γ in Epithelial to Mesenchymal Transition. *J. Biol. Chem.* 286, 3915–3924 (2011).

30.   Olsen, J. R. *et al.* p63 Attenuates Epithelial to Mesenchymal Potential in an Experimental Prostate Cell Model. *PLOS ONE* 8, e62547 (2013).

31.   Somerville, T. D. D. *et al.* TP63-Mediated Enhancer Reprogramming Drives the Squamous Subtype of Pancreatic Ductal Adenocarcinoma. *Cell Rep.* 25, 1741-1755.e7 (2018).

32.   Sun, L., Luo, H., Li, H. & Zhao, Y. Thymic epithelial cell development and differentiation: cellular and molecular regulation. *Protein Cell* 4, 342–355 (2013).

33.   Jiang, R. *et al.* Defects in limb, craniofacial, and thymic development in Jagged2 mutant mice. *Genes Dev.* 12, 1046–1057 (1998).

34.   Ward, L. D. & Kellis, M. Evidence of Abundant Purifying Selection in Humans for Recently Acquired Regulatory Functions. *Science* 337, 1675–1678 (2012).

35.   Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A. & Bejerano, G. Enhancers: five essential questions. *Nat. Rev. Genet.* 14, 288–295 (2013).
36.   Smemo, S. *et al.* Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Hum. Mol. Genet.* 21, 3255–3263 (2012).

37.   Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci.* 106, 9362–9367 (2009).

38.   Fakhouri, W. D. *et al.* An etiologic regulatory mutation in IRF6 with loss- and gain-of-function effects. *Hum. Mol. Genet.* 23, 2711–2720 (2014).

39.   Lettice, L. A., Hill, A. E., Devenney, P. S. & Hill, R. E. Point mutations in a distant sonic hedgehog cis-regulator generate a variable regulatory output responsible for preaxial polydactyly. *Hum. Mol. Genet.* 17, 978–985 (2008).

40.   Folkersen Lasse *et al.* Association of Genetic Risk Variants With Expression of Proximal Genes Identifies Novel Susceptibility Genes for Cardiovascular Disease. *Circ. Cardiovasc. Genet.* 3, 365–373 (2010).

41.   Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* 502, 333–339 (2013).

42.   Ford, D. J. & Dingwall, A. K. The cancer COMPASS: navigating the functions of MLL complexes in cancer. *Cancer Genet.* 208, 178–191 (2015).

43.   Rao, R. C. & Dou, Y. Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. *Nat. Rev. Cancer* 15, 334–346 (2015).

44.   Cerami, E. *et al.* The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov.* 2, 401–404 (2012).

45.   Sze, C. C. & Shilatifard, A. MLL3/MLL4/COMPASS Family on Epigenetic Regulation of Enhancer Function and Cancer. *Cold Spring Harb. Perspect. Med.* 6, a026427 (2016).

46.   Ortega-Molina, A. *et al.* The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development. *Nat. Med.* 21, 1199–1208 (2015).

47.   Toska, E. *et al.* PI3K pathway regulates ER-dependent transcription in breast cancer through the epigenetic regulator KMT2D. *Science* 355, 1324–1330 (2017).

48.   Mo, R., Rao, S. M. & Zhu, Y.-J. Identification of the MLL2 Complex as a Coactivator for Estrogen Receptor α. *J. Biol. Chem.* 281, 15714–15720 (2006).

49.   Kim, J.-H. *et al.* UTX and MLL4 Coordinately Regulate Transcriptional Programs for Cell Proliferation and Invasiveness in Breast Cancer Cells. *Cancer Res.* 74, 1705–1717 (2014).

50.   Dawkins, J. B. N. *et al.* Reduced Expression of Histone Methyltransferases KMT2C and KMT2D Correlates with Improved Outcome in Pancreatic Ductal Adenocarcinoma. *Cancer Res.* 76, 4861–4871 (2016).

51.   Bikle, D. D. Vitamin D and the skin: Physiology and pathophysiology. *Rev. Endocr. Metab. Disord.* 13, 3–19 (2012).

52.   Uray, I. P., Dmitrovsky, E. & Brown, P. H. Retinoids and rexinoids in cancer prevention: from laboratory to clinic. *Semin. Oncol.* 43, 49–64 (2016).

53.   Zhu, J. *et al.* Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth. *Nature* 525, 206–211 (2015).