ABSTRACT

POLKOFF, KATHRYN M. LGR5 in Biological Barriers: Roles in Skin and Lung Development and Homeostasis. (Under the direction of Dr. Jorge Piedrahita).

LGR5, a marker of adult and fetal stem cells in various tissues, has been studied extensively for the past decade. Cells expressing LGR5 have key roles in development and regeneration, and therefore having tools to track and study these cells in real time is key to tissue engineering and therapies. However, most of the research of these stem cell populations has been performed using mice models, which have significant anatomical, physiological, and molecular differences from the human. In this work, we present the development of a transgenic porcine model expressing a nuclear histone 2B fused with green fluorescent protein under the control of the endogenous *LGR5* promoter. Here, we dissect the roles of LGR5-expressing cells in the skin and lung. LGR5 expression in the skin marks hair follicle stem cells, is highly conserved across species, and the associated populations share a high degree of similarity in location, behavior, and gene pathways. We also show that they demonstrate fate plasticity in *in vitro* culture conditions. However, in the lung, the pig has drastically different LGR5-expressing populations from the mouse, and the porcine populations reflect those of the human. In this work, we have discovered two, independent, previously undescribed populations of LGR5-expressing cells in the human and porcine lung; one is a mesenchymal stem cell population defining airway stromal cells, and one is a transient, development-specific population. Overall, this work emphasizes the importance of adding additional animals to the functional study of organs and cell populations.

© Copyright 2022 by Kathryn M. Polkoff

All Rights Reserved

LGR5 in Biological Barriers: Roles in Skin and Lung Development and Homeostasis.

by
Kathryn M. Polkoff

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Comparative Biomedical Sciences: Cell Biology

Raleigh, North Carolina
2022

APPROVED BY:

_____	_____
Dr. Jorge Piedrahita Committee Chair	Dr. Kyle Mathews
_____	_____
Dr. Alon Greenbaum	Dr. Jason Haugh

BIOGRAPHY

Kathryn M. Polkoff started her academic career at University of Illinois at Urbana-Champaign, with a bachelor's and master's degree in Animal Sciences. Her undergraduate and Master's research in reproductive biology, including ovarian follicles and bovine embryo development, led her to a passion for studying stem cells early and throughout development. This, combined with her interest in large animal models, translational medicine, and genome editing tools, led her to pursue a PhD in Comparative Biomedical Sciences at North Carolina State University. Her research area of interest is toward understanding genetic and epigenetic regulation of cell fate, and her long-term goal is to develop therapies to modulate aberrant gene regulation.

ACKNOWLEDGMENTS

I suppose that now is perhaps not the best time to sit down and write a reflective letter to acknowledge all the things I wish to acknowledge; this thing is due tomorrow and it's approximately half in here, and half in various Word, Illustrator, Powerpoint, PDF, Google Doc, and Drive files… that may be the more pressing issue. But to submit it without fully acknowledging the depth of the experience and the people that made it possible would be perhaps the same as submitting with a full chapter missing.

I hope any of the (expected few) readers of this dissertation can grasp the amount of personal and scientific growth reflected in this document. Scientifically, as on any good journey, it took a bit of time to really gain traction. Reflecting on this degree from beginning to end, quite honestly, I went through a long period of addiction to the lab, a period of a healthy life balance, a pandemic (never thought I'd list *that* on here), a brief period of academic languishing (I still liked the work but couldn't find motivation), back to the lab addiction, and a final (feels long at the moment) period of wrangling loose ends, formatting them, and submitting them for publication, all of which can be summarized by my slowly and steadily increasing dosage of caffeine. I've heard over and over again that the last thing that someone wants to work on after their PhD is … more of the work from their PhD. But I can honestly say that at this moment, I have never been more excited about the research I've been doing (peep chapter 4). I know I have a long way to go from here as far as expectations for myself and the goals for the scientist I want to be, but after reflection on this work, it's something that I'm proud of. Although 75% of the experiments I did probably didn't end up in this document (mostly because they didn't work—or rather they worked but gave negative results, and I certainly wish to include them, but I don't

have time), and I probably will be haunted by the loose ends indefinitely, I still think the time feels right to move to the next.

With that, the first and absolute foremost acknowledgement I must make is to my very first two mentors: my parents. From the earliest days, although none of us knew it, they were training me will the skills and the mindset of a scientist. From my mom, it was instilling that constant sense of active wonder and curiosity. “That’s a great question,” she would say, “why do you think that is?” Or, even more so, “how do you think we can figure out the answer?” Time and time again, I will credit her with the framework for my intellectual independence, critical thinking, and problem solving. And to this day, I’m still trying to figure out how it never felt like a lesson—I was never force-fed information or held with my nose to a topic that did not interest me. Perhaps that’s just it—but I know for a fact that she knew exactly how to get me excited about topics that probably would not have been exciting to your average child (have you heard of phonics?). Most importantly, from her I get the creativity, the wonder, the excitement of learning something new. Without all those intrinsic motivators, there is no interest in science, no passion for discovery, no drive to figure it out. Thank you, Mom.

On the other side of that coin, you get the intentional, precise and thorough nature of my father. With my dad, if something is going to be done, it is going to be done right. Whether we were building a plywood stable for my model horses or making pizza dough, everything will be planned meticulously, measured, and executed according to the plan. It was from him that I was instilled with the attention to detail, the awareness, and the technical execution that makes laboratory work now seem so second-nature to me. Also from him—one thing I did not appreciate in adolescence but that I can appreciate now—there was always a very specific set of rules to follow, and however you operate within those rules was your freedom. This is really the

crux of scientific progress. I guess I spent most of my higher education experience figuring out the rules for biology, and my PhD figuring out exactly how to use that freedom. So, thank you Dad, for shaping me with the mindsets that makes it possible to design, think, and execute at a high level.

Thank you also to my brother Michael, whose gift for strategic endeavors forced me to develop quick and sound approaches to our logical (or sometimes illogical) arguments. This trained me to be prepared and thorough in my oral and written scientific logic and theses. The constant and high-level competition from our youth pays dividends now, when I need to perform well and stay clear headed under pressure. Thanks for always challenging me to be the best I can be.

There are so many people I need to thank, and each of them deserve their own paragraph but instead I'll just have to list names. Thank you to Dr. Matt Wheeler, who encouraged me to pursue research using gene editing and large animal models. Thank you to my peers and friends; you are the reason I retain some semblance of mental stability and life balance (especially you, Hanna (and Oligo), Greg, and Science Corner—you know who you are), and to the friends and family who live just one phone call away.

Thank you to my lab colleagues over the years (Laura, Katherine, Yanet, Bruce, Jaewook, Ross), who contributed to these findings significantly, and who know the chaos, excitement, and labor that comes with the all-hands-on-deck nature of working with pigs. Of course, thank you to the people who work with us in LAR and CPL, who also know that chaos. Thank you to the entire second floor of the Biomedical Partnership Center; from the potlucks to Halloween costume contests, to Friday afternoons at Raleigh Brewing, it's a fun place to work and a fun place to get things done. Thank you also to our collaborators especially in the

Greenbaum, Gonzalez, and Tata labs. I am so thankful for the experiences and connections I gained as a member and as a leader within the Comparative Medicine Institute—especially the Young Scholar Program and the undergraduates and collaborators I worked with. Amber D, Dylan, and Amber C, it has been an absolute joy to watch you grow and learn how to think and operate within the lab. It's always my goal for a trainee to become a colleague, and you all have absolutely blown away all expectations I could have had as such—I can't wait to call you all "Doctor" in a couple years yet.

Thank you especially to Nithin Gupta, originally a mentee from a CMI project but who quickly expanded his role to that of a lab partner. Reflecting on the time we spent working together over three years, I am so proud of how much you've grown; from a fledgling undergrad to an independent scientist, mentoring people and designing experiments on your own. From the insanely late nights in the tissue culture hood, or the pig pen, or the mouse room, to the "no way, no way!!" moments of excitement in the microscope room, to the crazy ideas and white board sessions, I count myself extremely lucky to have shared the scientific ups and downs with you. I can honestly say that without you, the amount and quality of research in this dissertation would have been significantly reduced.

And of course, thank you to Dr. Jorge Piedrahita. Thank you for always respecting me and my ideas, for giving me the opportunity to be heard by others, and for treating me as a teammate—I have come to realize that combination is actually pretty uncommon. Thank you for encouraging my scientific independence and collaboration, for training me in all aspects of academia and scientific leadership, and thank you for buying the occasional round on Friday afternoons.

TABLE OF CONTENTS

ABSTRACT i
BIOGRAPHY ii
ACKNOWLEDGMENTS iii
LIST OF TABLES ix
LIST OF FIGURES x
CHAPTER 1 1
The transformational impact of site-specific DNA modifiers on biomedicine and agriculture 1
CHAPTER 2 11
LGR5 is a conserved marker of hair follicle stem cells across species and is present early and throughout follicle morphogenesis 11
Abstract 11
Introduction 12
Results 13
Discussion 23
Materials and Methods 27
Supplementary figures 34
REFERENCES 38
Chapter 3 43
LGR5+ hair follicle stem cells exhibit self-renewal and multi-lineage differentiation in organoid culture 43
Abstract 43
Introduction 44
Discussion 60
METHODS 64
REFERENCES 70
Chapter 4 73
Two roles for LGR5 in lung development and homeostasis 73
Introduction 74
Results 75
Discussion 100

Methods..... 103
REFERENCES..... 109
Appendix A..... 118
In vitro validation of transgene expression in gene-edited pigs using CRIPSR Transcriptional Activators..... 118

LIST OF TABLES

Chapter 1

Table 1 Effect of site-specific DNA modifiers on multiple aspects of gene editing in mammals 3

Chapter 2

Table S1 Genes and primer sequences used in RT-qPCR 33

Chapter 3

Table S1 Organoid Complete Media 68
Table S2 Growth factor formulation for M1 68
Table S3 Primers used in this study 69

LIST OF FIGURES

Chapter 1

Figure 1 Gene editing outcomes using targeted modifiers 4

Chapter 2

Figure 1 LGR5 is expressed in the hair follicle outer root sheath in the lower bulge throughout anagen, catagen, and telogen in the pig and human 14

Figure 2 LGR5+ cells give rise to transient amplifying populations in hair matrix in anagen 15
Figure 3 LGR5+ epidermal cells express known stem cell and epidermal markers. 16
Figure 4 Human, Pig, and Mouse LGR5+ epidermal populations are enriched in extracellular matrix structure and organization pathways 20
Figure 5 LGR5 is expressed early and late throughout developing porcine hair follicles 23
Figure 6 LGR5 expression in rhesus and mouse hair follicle morphogenesis 25
Figure S1 Cross section of porcine hair follicle shows that LGR5 is expressed at a high level in the outer root sheath, and low level in the inner root sheath 34
Figure S2 Representative flow cytometry gating strategy and controls 35
Figure S3 Related to figure 4. Shared upregulated gene ontology pathways of upregulated genes in LGR5-high cells, compared pairwise across human, mouse and pig datasets 36
Figure S4 SHH expression in porcine LGR5+ D50 epidermis 37

Chapter 3

Figure 1 $LGR5^{HI}$ epidermis forms more organoids at a higher efficiency than $LGR5^{LO}$ or $LGR5^{NEG}$ cells 45
Figure 2 Growth factor array for organoid media formulation 47
Figure 3 Anagen bulge LGR5+ cells are the primary drivers of organoid formation 50
Figure 4 Generation of hair follicle organoids 52
Figure 5 Organoid derived from LGR5-high hair follicle stem cells produce organoids at higher efficiency and size than those from LGR5-low and LGR5-negative populations in M1 conditions 54
Figure 6 $LGR5^{LO}$ and $LGR5^{NEG}$ derived organoids demonstrate limited viability in long-term culture 56
Figure 7 Fully differentiated skin organoids have same markers of epidermis 58
Figure 8 Gene expression of organoids over time 59
Figure 9 Mouse organoids do not form under the same conditions. 61

Chapter 4

Figure 1 Two populations of LGR5 in development and homeostasis 76
Figure 2 Human lung distribution of *LGR5* matches the pig 78
Figure 3 Molecular mapping of LGR5 expression in the human lung confirms the presence of *LGR5* expression in multiple populations in development, and one at homeostasis 80
Figure 4 Single cell RNA sequencing of fetal porcine LGR5+ populations shows that fetal epithelial and mesenchymal populations are unrelated 81
Figure 5 LGR5 marks mesenchymal subset surrounding airways throughout the porcine postnatal lung..... 83
Figure 6 LGR5 expression in the mesenchymal population always occupies the sub-basal mesenchyme, but pattern and orientation depend on size of airway 84
Figure 7 Additional analysis of adult LGR5+ mesenchymal population..... 86
Figure 8 Fetal LGR5+ mesenchymal populations appear surrounding mature airways and resemble adult patterns of expression..... 87
Figure 9 Postnatal LGR5+ cells are mesenchymal stromal cells and support airway basal cell niche..... 89
Figure 10 LGR5+ epithelial expression becomes more restricted throughout development... 91
Figure 11 Molecular analysis of LGR5+ epithelial population..... 92
Figure 12 Characterization of fetal lung epithelial organoids..... 95
Figure 13 Transient LGR5 expression appears in adult epithelial organoids 96
Figure 14 LGR5 is expressed in organoids with a bronchioalveolar fate, but not a bronchiolar fate..... 99

CHAPTER 1

The transformational impact of site-specific DNA modifiers on biomedicine and agriculture

This article was published online in the journal Animal Reproductions on August 16th, 2018.

PMID: 34178139

doi: 10.21451/1984-3143-AR2018-0065

DOI: 10.21451/1984-3143-AR2018-0065

Proceedings of the 32nd Annual Meeting of the Brazilian Embryo Technology Society (SBTE); Florianopólis, SC, Brazil, August 16th to 18th, 2018.

The transformational impact of site-specific DNA modifiers on biomedicine and agriculture

Kathryn Polkoff[1,2], Jorge A. Piedrahita[1,2]

[1]Comparative Medicine Institute, College of Veterinary Medicine, North Carolina State University, Raleigh, NC, 27606, USA.

[2]Department of Molecular Biomedical Sciences, College of Veterinary Medicine, North Carolina State University, Raleigh, NC, 27606, USA.

Abstract

The development of genetically modified livestock has been dependent on incremental technological advances such as embryo transfer, homologous recombination, and somatic cell nuclear transfer (SCNT). This development rate has increased exponentially with the advent of targeted gene modifiers such as zinc finger nucleases, TAL-effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeats (CRISPR-Cas). CRISPR-Cas based systems in particular have broad applicability, and have low technical and economic barriers for their implementation. As a result, they are having, and will continue to have, a transformational impact in the field of gene editing in domestic animals. With these advances also comes the responsibility to properly apply this technology so it has a beneficial effect throughout all levels of society.

Keywords: domestic animals, gene editing.

Introduction

Embryo culture and embryo transfer, pronuclear injection, homologous recombination: technical advances that gradually made it possible to generate genetically modified large animals such as pigs, cattle and sheep. While each of these incremental advances have impacted the field, two have had such a large impact as to be properly defined as transformational. They are somatic cell nuclear transfer (SCNT), and site-specific gene editing via targeted gene modifiers, including zinc finger nucleases (Le Provost *et al*., 2010), TAL-effector nucleases (TALENs; Christian *et al*., 2010) and clustered regularly interspaced short palindromic repeats (CRISPR-Cas; Jinek *et al*., 2012).

Our previous inability to isolate and culture embryonic stem cells (ES) from domestic species to generate transgenic animals prevented the implementation of techniques such as homologous recombination (HR) (Gonçalves *et al*., 2014; Koh and Piedrahita, 2014). In spite of over thirty years of work in this area, no ES cell lines from domestic species have been isolated that allow the practical and efficient generation of transgenic animals (Koh and Piedrahita, 2014). Thus, while techniques such as HR using ES cells to create germ-line chimeras became the norm to generate transgenic mice, these approaches could not be used in domestic species; that is, until the advent of SCNT. From the initial observation of Keith Campbell and Ian Wilmut that sheep could be cloned from a somatic cell using SCNT (Campbell *et al*., 1996), multiple groups rapidly moved to genetic modification of somatic cells *in vitro* followed by SCNT. This led to the first reports of SCNT-generated transgenic sheep (Schnieke *et al*., 1997), pigs (Dai *et al*., 2002) and cattle (Cibelli *et al*., 1998). And these initial reports included application of HR in somatic cells before transfer (McCreath *et al*., 2000). As a result, there was tremendous excitement in the field and most, if not all, laboratories worldwide working in the area of genetic modification of domestic animals quickly moved to implement SCNT. While this transition was successful for many groups, gene targeting by HR remained a significant barrier. For reasons that are still not well understood, HR in somatic cells is extremely inefficient and in spite of significant efforts by many groups, only a few gene targeted animals were generated (reviewed by Prather *et al.*, 2008; Aigner *et al*., 2010; Piedrahita and Olby, 2011.

That all changed with the development of gene editing using targeted DNA endonucleases such as Zinc Finger Nucleases (ZFN), Tal Effector Nucleases (TALENs), and CRISPR-Cas9 nucleases (Sander and Joung, 2014). All three approaches make gene targeting in any cell, including somatic cells, more efficient by several orders of magnitude (Gaj *et al*., 2013). Using pigs as an example, we show in Table 1 that the impact of this technology on the efficiencies of generating a transgenic pig is indeed transformational. While all three approaches (ZNFs, TALENs, and CRISP-Cas) have been used to develop gene edited domestic species, this review will concentrate on the CRISPR-Cas based systems. This is due to the lower costs, ease of use, and expanding repertoire of modified enzymes that further increase the utility of the system. We will cover applications that focus on gene editing (genetic modifiers) as well as approaches that modify gene expression by acting on the epigenome (epigenetic modifiers). While these epigenetic modifiers have not yet been fully implement in domestic animals, we feel they have tremendous potential as models for clinical applications in humans.

Corresponding author: Jorge A. Piedrahita
japiedra@ncsu.edu
Received: May 4, 2018
Accepted: June 12, 2018

Table 1. Effect of site-specific DNA modifiers on multiple aspects of gene editing in mammals.

Before site-directed DNA modifiers*	After site-directed DNA modifiers	References
Homologous Recombination (HR)	Homologous recombination and targeted NHEJ	Smithies, 2001; Smithies *et al*., 1984; Smithies, 2008; Le Provost *et al*., 2010
Long homology arms	Short or no homology arms	Vazquez *et al*., 1998; Sander and Joung, 2014; Suzuki *et al*., 2016; Brown *et al*., 2016
Selectable markers	No selectable markers	Smithies, 2008; Gaj *et al*., 2013
Single gene	Multiple genes	Smithies, 2008; Piedrahita *et al*., 1992; Park *et al*., 2017
Single allele	Both alleles	Fu *et al*., 2013
Only cultured cells	Cultured cells and direct zygotic injection	Dai *et al*., 2002; Hai *et al*., 2014
Overall frequency of HR 1 in a million	Overall frequency of targeted gene editing 100%	Smithies, 2008
Only dividing cells	Dividing and non-dividing cells	Yao *et al*., 2017
Global but not targeted epigenetic modifications	Single and multi loci targeted epigenetic modifications	Ng and Bird, 1999; Zhou *et al*., 2018
Frequency too low for in vivo or ex vivo clinical applications	Frequency high enough that in vivo or ex vivo clinical applications can be developed	Wang et al., 2013; Hai *et al*., 2014

*Includes both genetic and epigenetic modifiers.

CRISPR-Cas editors

Originally derived from bacteria as a defense against bacteriophages, investigators have harnessed the ability of CRISPR-Cas to recognize a specific DNA sequence and create a double-stranded break. As depicted in Fig 1, CRISPR-Cas has two functional components: a guide RNA (gRNA) and a CRISPR Associated protein (Cas) nuclease. The gRNA is composed of an RNA sequence that recognizes the target DNA and an RNA region known as tracrRNA or transactivating CRISPR RNA. The Cas protein complexes with the gRNA and binds the target DNA.

In the first, and most commonly used, CRISPR-Cas system derived from *S. Pyogenes* (SpCas9), the gRNA contains a 20 nucleotide sequence complementary to a DNA sequence that is directly upstream of a protospacer adjacent motif (PAM) 5'-NGG. CRISPR-Cas systems from other bacteria such as *Staphylococcus aureus* (SaCas9) and Prevotella and Francisella (Cpf1) (Ran et al., 2015; Zetsche et al., 2015) have different gRNA sequence and PAM requirements, but all create a double-stranded break at the target site. This results in a system that can be targeted to specific regions of the genome using the gRNA followed by a double stranded DNA cleavage via the Cas9 endonuclease (Sander and Joung, 2014). The cell then senses this DNA damage and activates DNA repair pathways. It is this process of DNA repair that forms the basis for gene editing using CRISPR-Cas systems. As shown in Fig. 1, the DNA damage can be repaired by multiple mechanisms. The most frequently used, the non-homologous end joining (NHEJ) pathway, recruits cellular machinery to ligate the cleaved ends back together. However, this system is error-prone and creates random insertions and deletions (indels) at the damaged site. If the indels are located in the coding sequence of the gene, they can create a frame-shift mutation and therefore an abnormal or absent protein. Thus, NHEJ is often used to inactivate genes. This is such a highly efficient system that it can generate loss of function of one (heterozygous mutant) or both copies of the gene (homozygous mutant; Table 1). Prior to targeted endonucleases, the only way to obtain homozygous mutants was through breeding heterozygotes, or by performing two rounds of genetic modifications using sequential SCNT (Kuroiwa *et al*., 2004); neither of which are practical or easy to apply to domestic animals. This alone is transformational as, by avoiding the need for breeding, CRISPR-Cas induced NHEJ drastically reduces the time required to generate an animal or cell line devoid of a specific protein.

However, in some cases the goal is not to knock out a gene but to instead knock in or replace genes. This process, for instance, can be used for targeted insertion of a gene such as a fluorescent tag for cell or protein tracking, insertion of human genes, or addition of favorable agricultural traits. For targeted

homology directed repair, the double stranded break in the target region requires a donor DNA construct containing regions of homology to either side of the cut site. The cell will then repair the double stranded break by two competing mechanisms, the NHEJ described above, or by homology-directed repair (HDR) resulting in incorporation of the donor DNA into the target region. This process is analogous to conventional HR with the main difference being that in conventional HR there is no induced double stranded break, only the donor DNA. Both conventional HR and HDR require cell division as DNA replication is an integral component of the homologous recombination process. As described in Table 1, without the DNA break, HDR occurs at frequencies of 0.000001% (1 in 10E56) or lower. In contrast, with a targeted DNA break, HDR occurs at frequencies ranging from 10% to as high as 50%.

But there are differences in the composition of the donor DNA as well. Conventional HR requires that the donor DNA contains several kb of homology to the target gene, as well as positive and negative selectable markers to enrich for those are cells that had been modified (Vazquez *et al*., 1998). As a result, the donor DNA plasmids are difficult and expensive to develop, some requiring several months to complete. In contrast, donor DNA used for HDR requires regions of homology ranging from a total of 1 kb to less than 100 bp, does not require selectable markers, and can be rapidly and inexpensively generated. This allows the use of two types of donor DNA, small oligo that can be used to modify small regions of the DNA and larger DNA donors that can be used to replace or insert (knockin) a gene or gene fragment into the desired target region (Sander and Joung, 2014).

Recently, a new mechanism for gene insertion has been described for homology independent targeted integration (HITI; Brown *et al*., 2016; Suzuki *et al*., 2016). By a process little understood at present, the double stranded break created by the CRISPR-Cas9 is repaired by an NHEJ-driven mechanism, does not require DNA replication, and results in the insertion of a donor DNA in the absence of any homology to the target region. It does, however, require that the donor DNA plasmids are also cleaved by a CRISPR-Cas9. What is more surprising is that the frequency of targeted insertions is higher using HITI that using HDR (Suzuki *et al*., 2016). Since then, similar approaches using micro-homology arms (<50 bp) or homology arms of less than 1 kb of total homology have been described that also work in non-dividing cells (Yao *et al*., 2017), referred to as Micro-homology Mediated End Joining (MMEJ) and Homology-Mediated End Joining (HMEJ), respectively. Interestingly, the efficiency of the different integration methods differs drastically depending on cell type. In mouse ES cells, for instance, HDR and HMEJ occur at approximately the same rate, while in mouse embryos, HMEJ is 5-10 fold more effective than HDR (Yao *et al*., 2017). In summary, multiple approaches that have been or are being developed allow the modification or inactivation of essentially any gene in any cell type at high efficiency.

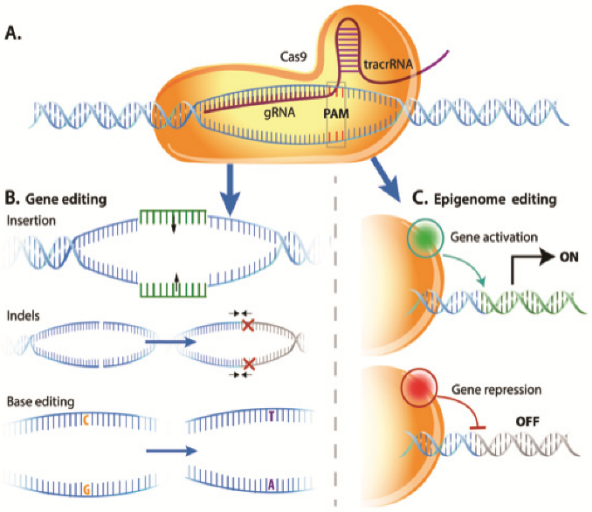


Figure 1. Gene editing outcomes using targeted gene modifiers. A) gRNA complexes with Cas9 protein to bind a specific 20 nucleotide sequence in the target DNA. B) Cas9 nuclease initiates cell-based repair mechanisms to create changes in DNA sequence: Homology directed repair or HITI for gene insertion, non-homologous end joining for indels, or base editing for site specific nucleotide changes. C) Catalytically inactive Cas9 protein fused with transcriptional modifiers leads to targeted gene activation or repression.

Base editors

While HDR and NHEJ are effective for knock-outs and knock-ins, the advent of base editors introduces a new paradigm for therapeutic gene editing. Base editors are CRISPR-based enzymes that can catalyze the conversion of specific bases within a specified target window without a double-stranded break. The first base editors consist of a Cas9 nickase fused with a cytidine deaminase (APOBEC-1 or AID), which is directed by a gRNA to elicit a targeted C*G to T*A conversion (Komor *et al*., 2016; Shimatani *et al*., 2017). The mechanism involves deamination of a cytidine, thereby converting it to a uridine, which pairs with an adenine upon cellular repair. Further iterations of base editors also allow for improved specificity of the target window to eliminate unintended conversion of cytosines neighboring the target base pair (Kim *et al*., 2017). More recent advances also allow the conversion of T*A pairs to C*G pairs by replacing the cytidine deaminase with an adenosine deaminase (Gaudelli *et al*., 2017). In addition, base editors have been fused to other targeted endonucleases such as Cpf1 (Li *et al*., 2018), allowing for targeting of sites with various PAM sequences.

The use of base editors has several advantages: there is no longer a need to simultaneously deliver a repair template with the endonuclease, the lack of double-stranded breaks diminishes the chance of unwanted indels, and the specific activity window reduces the number of potential off-target sites (discussed below). However, base editors are still restricted in that they only can catalyze conversions between C*G and T*A base pairs, and only can target a small window which must be approximately 15 base pairs upstream of the PAM sequence. Future base editors must be more flexible to allow for editing of clinically relevant sites that are currently out of reach due to lack of appropriate PAM location or that require different base conversions. Furthermore, microinjection of a base editor into mouse embryos showed that its nickase activity can still introduce indels at a relatively high frequency (Kim *et al*., 2017).

Approaches to generating gene edited offspring

Unlike conventional HR, where the efficiencies are so low that *in vivo* applications in embryos or somatic tissue are impractical, the increases in gene editing frequencies associated with systems such as CRIPS-Cas make *in vivo* gene editing possible. While gene editing was initially carried out in cells in culture, *in vivo* applications quickly developed. Initial reports showed that direct injection of CRISPR-Cas9 into the cytoplasm of one cell mouse embryos resulted in 50/56 (90%) of the offspring being modified via NHEJ (Wang *et al*., 2013). What was more surprising was the large number of offspring that had biallelic modifications 45/56 (80%). These initial reports were soon confirmed in other species including domestic animals such as pigs (Hai *et al*., 2014), sheep (Crispo *et al*., 2015), goats (Wang *et al*., 2015) and cattle (Bevacqua *et al*., 2016). It was also applied to multiple loci at the same time resulting in the generation of multi transgenic offspring (Park *et al*., 2017).

Initial reports of HDR by direct cytoplasmic injection were not as successful as NHEJ, suggesting that homologous recombination is not efficient in embryos, with frequencies ranging from 8 to 34% in mice (Yang *et al*., 2013). Approaches to enhance HDR were tested (Maruyama *et al*., 2015) with some success but it was not until NHEJ-dependent approaches were used that targeted insertion into zygotes became practical. The MHEJ system already described is highly efficient when used directly on embryos with over 25% of mouse embryos carrying the correct insertion and over 50% of non-human primate embryos generated by ICSI, followed by HMEJ, carrying the correct modification (Yao *et al*., 2017). In domestic species, HDR directly in embryos has been reported for pigs (Park *et al*., 2017) and goats (Niu *et al*., 2018) with efficiencies ranging from 15 to 50%. Of concern, however, is that in addition to the HDR-mediated insertion into one allele, the remaining allele was mutated via NHEJ. Thus, a large number of offspring need to be generated to create one carrying the desired insertion but without a mutation in the other allele.

This leads to the question: which system is better, gene editing directly in zygotes or gene editing in cultured cells followed by SCNT? There is no simple answer. It depends on the question being addressed, the technical capabilities available, and the regulatory environment in which one operates. The benefits of gene editing combined with that SCNT is that the donor cells can be extensively analyzed before SCNT so the genetics of the offspring are known. It also allows for complex gene edits and for sequential gene editing via multiple rounds of SCNT. This allows generation of multi-transgenic animals in a relatively short time, something that is crucial in species with longer generational intervals (cattle, goats, and pigs for instance). The drawbacks are that SCNT is technically complex, requires expensive specialized equipment, and can be unreliable. Zygotic injection, in contrast, is technically simple, can be carried out with less expensive equipment, and can be applied, in theory, to any mammalian species where zygotes are available; even those where SCNT is either impractical or has not been developed. The drawback is that the process is completely random so many of the offspring generated will have to be euthanized as they will not carry the desired gene edit. In addition, it cannot be used to generate sequential gene edits without breeding to produce new zygotes and that in species such as cattle can take years rather than months. However, if all that is needed is inactivation of 1-2 genes or modification of one loci, zygotic injection will produce the desired outcomes in a shorter period of time, even if some of the offspring do not carry the desired mutation and will need to be discarded. In an ideal system, having both SCNT and zygotic injection will give the greatest flexibility and provide the capabilities to tackle essentially any gene edit desired, regardless of complexity.

Epigenetic modifiers

Although gene targeting for knock-outs, knock-ins, and editing is very promising, there is also a need to address diseases that are a result of aberrant cellular regulation. In the past decades, modulation of gene expression has depended heavily on RNA interference, which focuses mostly on gene repression (Gemberling and Gersbach, 2018). The development of CRISPR-based gene regulators provides a powerful new strategy for targeted gene therapy. These epigenomic editors are composed of a catalytically inactive CRISPR protein, dead-Cas9 (dCas9), fused with an effector domain for transcriptional activation or suppression. These complexes are then paired with a gRNA and targeted to a specific site in the genome. With control of transcriptional activity, these editors can be used to suppress harmful genes, upregulate those that are deficient or silenced, or completely reprogram cell fate.

A series of dCas9 transcriptional activators have been developed, the first of which were dependent on the transcriptional activator VP64 (Gilbert *et al.*, 2013; Perez-Pinera *et al.*, 2013). Improved versions depend on addition of fused domains, protein scaffolds, or RNA scaffolds to recruit additional upregulating factors for improved efficiency (Chakraborty *et al.*, 2014; Konermann *et al.*, 2015; Chavez *et al.*, 2016). Other dCas9 activators rely on epigenomic modifiers such as histone acetyltransferase (Hilton *et al.*, 2015). As for transcriptional repressors, early versions relied upon dCas9 binding to interfere and block transcription initiation (Qi *et al.*, 2013). Soon after, dCas9 was used to recruit chromatin-modifying repressor complexes, such as the Kruppel-associated box (KRAB) domain, to effectively silence target gene expression (Gilbert *et al.*, 2013). Moreover, because specific effects of epigenetic elements on gene regulation are not well understood, targeted epigenetic modifications by DNA methyltransferase (Vojta *et al.*, 2016), or histone deacetylase (Kwon *et al.*, 2017) can be employed to better understand these phenomena. This type of screen for regulatory elements can also be performed in a high-throughput fashion with loss- and gain-of-function editors (Klann *et al.*, 2017).

Like their active-nuclease counterparts, dCas9 epigenome modifiers can also be delivered for therapeutic and fundamental purposes. Several studies have shown the ability of CRISPR activators to modify cell fate. For example, *in vitro* studies have demonstrated effective direct reprogramming of fibroblasts into neurons by targeted activation of three specific genes (Black *et al.*, 2016). These factors can also be delivered *in vivo* by the same approaches as the targeted nucleases, such as AAV. An impressive study by Liao *et al.* (2017) was the first to use CRISPR/Cas9 type systems to modify transcription for several purposes. They show the ability to increase muscle mass in a dystrophic mouse model by local injection into hindlimbs by upregulating utrophin, compensate for acute kidney injury by upregulating Klotho or IL-10, and completely reprogram liver cells into insulin producing cells to treat a mouse model of type 1 diabetes. This is the first of many future studies using *in vivo* transcriptional modifiers as therapeutics for disease and perhaps for production or reproductive traits in large animals.

Other uses for dCas9 delivery include reprogramming of astrocytes into neurons in transgenic mice by activation of multiple genes (Zhou *et al.*, 2018) or the ability to screen for potential oncogenes (Chow and Chen, 2018). Because they are so new, the *in vivo* delivery of targeted transcriptional regulators has thus far been limited to small animals, but as the therapies are translated to humans, we expect large animal models such as pigs to be important for scale-up and evaluation of physiological effects. Pigs have already been established as a model for epigenetic programming. For example, an Oct4-Enhanced GFP pig provides a valuable tool for the evaluation of reprogramming efficiency and pluripotency (Nowak-Imialek *et al.*, 2010). Even in a pre-targeting era, pigs have been useful for the study of epigenetic control of gene expression, silencing, or tissue specific control of transgenes (Archer *et al.*, 2003; Kues *et al.*, 2006).

Hurdles and challenges: off-target effects

One major limitation for the use of CRISPR is the potential for off-target effects. While each gRNA has been synthesized to target a specific genomic sequence, there is the possibility for binding and cleavage at closely related sequences elsewhere in the genome, resulting in unwanted indels. The presence of off-target effects from CRISPR-Cas was shown in human cells early on (Fu *et al.*, 2013), and hence there has been a push to develop methods for detection and prevention of off-target effects.

Initial efforts for safe and effective CRISPRs led to in silico design tools for gRNAs that score the probability of on- and off-target events (Hsu *et al.*, 2013; Heigwer *et al.*, 2014). While these are a good starting point and are free to use, in silico design tools are only moderately accurate for prediction of true off-target effects (Tsai *et al.*, 2015). To better understand the frequency and location of off-target sites, a handful of techniques have been established. GUIDE-seq (Tsai *et al.*, 2015), CIRCLE-seq (Tsai *et al.*, 2017), Digenome-seq (Kim *et al.*, 2015), and HTGTS (Frock *et al.*, 2015) are all examples of unbiased, sensitive tools that capture the double-stranded breaks created with *in vitro* or in situ following delivery of Cas9 and analyze based on sequence reads. However, these techniques are expensive and frequently require a full reference genome. It is of note that off-target sites detected by GUIDE-seq showed only modest overlap with in silico predictors especially because many actual off-target sites were excluded from consideration by the programs (Tsai *et al.*, 2015). As studies continue to elucidate the precise rules for CRISPR off-target binding (Boyle *et al.*, 2017), there is a need for a more accurate in silico

predictor tool. Likewise, off-target effects are highly characterized *in vitro*, but more work must be done to evaluate frequency of off-target events *in vivo*.

To combat these adverse effects, several strategies have been employed. Initial studies demonstrate that a shortened gRNA can increase specificity by eliminating ability to form bulges when binding (Fu *et al*., 2014). Further studies showed that the use of Cas9 nickases or paired nickases, which create only a single-stranded break, have fewer off-target effects (Shen *et al*., 2014; Frock *et al*., 2015). Additionally, modifications to the Cas9 protein for a high fidelity nuclease increase the specificity of the binding domain and decrease off-target effects and frequency (Kleinstiver *et al*., 2016). Finally, other CRISPR nucleases with less common PAM sequences or that are less tolerant of mismatches, such as Cpf1, have fewer off-target sites compared with Cas9 (Kim *et al*., 2016). The availability of new enzymes with higher fidelity and higher specificity combined with better in silico methods to design gRNA that will have single target specificity are likely to eventually lead to systems with undetectable off-target effects. For the present, however, it is important that off-target effects are taken into account when generating gene edited offspring, whether by SCNT or by zygotic injection.

Summary and conclusions

With the rapid adaptation of CRISPR-Cas and related gene editing technologies, the rate of applications to agriculture and biomedicine is growing exponentially. Previous methods of genetic modification of animals relied heavily on random insertion methods (pronuclear injection), use of genetically modified somatic cells followed by SCNT, or the use of viruses for transgene insertion; all methods with significant drawbacks. CRISPR-Cas and related systems not only do not suffer from these drawbacks but their implementation is both technically simpler and less costly. All these factors combined, and the high degree of plasticity of the procedure so it can be used to modify DNA as well as modify transcription, is transforming the field of gene editing of domestic animals.

However, as we continue to apply gene editors, whether it be for therapeutic delivery in medicine or disease resistance and growth traits in agriculture, we must be responsible and aware of our actions. The power of this technology is immense, and any misuse of it will decrease acceptance from the public who needs it the most. Nevertheless, proper use of these tools brings us the opportunity to cure disease, improve agricultural production to feed the growing population, and create a healthy future.

References

Aigner B, Renner S, Kessler B, Klymiuk N, Kurome M, Wünsch A, Wolf E. 2010. Transgenic pigs as models for translational biomedical research. *J Mol Med*, 88:653-664.

Archer GS, Dindot S, Friend TH, Walker S, Zaunbrecher G, Lawhorn B, Piedrahita JA. 2003. Hierarchical phenotypic and epigenetic variation in cloned swine. *Biol Reprod*, 69:430-436.

Bevacqua RJ, Fernandez-Martín R, Savy V, Canel NG, Gismondi MI, Kues WA, Carlson DF, Fahrenkrug SC, Niemann H, Taboga OA, Ferraris S, Salamone DF. 2016. Efficient edition of the bovine PRNP prion gene in somatic cells and IVF embryos using the CRISPR/Cas9 system. *Theriogenology*, 86:1886-1896.e1.

Black JB, Adler AF, Wang HG, D'Ippolito AM, Hutchinson HA, Reddy TE, Pitt GS, Leong KW, Gersbach CA. 2016. Targeted epigenetic remodeling of endogenous loci by CRISPR/Cas9-based transcriptional activators directly converts fibroblasts to neuronal cells. *Cell Stem Cell*, 19:406-414.

Boyle EA, Andreasson JOL, Chircus LM, Sternberg SH, Wu MJ, Guegler CK4, Doudna JA, Greenleaf WJ. 2017. High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proc Natl Acad Sci USA*, 114:5461-5466.

Brown A, Woods WS, Perez-Pinera P. 2016. Multiplexed targeted genome engineering using a universal nuclease-assisted vector integration system. *ACS Synth Biol*, 5:582-588.

Campbell KH, McWhir J, Ritchie WA, Wilmut I. 1996. Sheep cloned by nuclear transfer from a cultured cell line. *Nature*, 380(6569):64-66.

Chakraborty S, Ji H, Kabadi AM, Gersbach CA, Christoforou N, Leong KW. 2014. A CRISPR/Cas9-based system for reprogramming cell lineage specification. *Stem Cell Reports*, 3:940-947.

Chavez A, Tuttle M, Pruitt BW, Ewen-Campen B, Chari R, Ter-Ovanesyan D, Haque SJ, Cecchi RJ, Kowal EJK, Buchthal J, Housden BE, Perrimon N, Collins JJ, Church G. 2016. Comparison of Cas9 activators in multiple species. *Nat Methods*, 13:563-567.

Chow RD, Chen S. 2018. Cancer CRISPR screens in vivo. *Trends Cancer*, 4:349-358.

Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF. 2010. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, 186:757-761.

Cibelli JB, Stice SL, Golueke PJ, Kane JJ, Jerry J, Blackwell C, Ponce de León FA, Robl JM. 1998. Cloned transgenic calves produced from nonquiescent fetal fibroblasts. *Science*, 280(5367):1256-1258.

Crispo M, Mulet AP, Tesson L, Barrera N, Cuadro F, dos Santos-Neto PC, Nguyen TH, Crénéguy A, Brusselle L, Anegón I, Menchaca A. 2015. Efficient generation of myostatin knock-out sheep using CRISPR/Cas9 technology and microinjection into zygotes. *PLoS One*, 10(8):e0136690. doi: 10.1371 /journal.pone.0136690.

Dai Y, Vaught TD, Boone J, Chen SH, Phelps CJ, Ball S, Monahan JA, Jobst PM, McCreath KJ, Lamborn AE, Cowell-Lucero JL, Wells KD, Colman

A, Polejaeva IA, Ayares DL. 2002. Targeted disruption of the alpha1,3-galactosyltransferase gene in cloned pigs. *Nat Biotechnol*, 20:251-255.

Frock RL, Hu J, Meyers RM, Ho Y-J, Kii E, Alt FW. 2015. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat Biotechnol*, 33:179-186.

Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol*, 31:822-826.

Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol*, 32:279-284.

Gaj T, Gersbach CA, Barbas CF III. 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol*, 31:397-405.

Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, Liu DR. 2017. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*, 551(7681):464-471.

Gemberling M, Gersbach CA. 2018. Boosting, not breaking: CRISPR activators treat disease models. *Mol Ther*, 26:334-336.

Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS. 2013. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, 154:442-451.

Gonçalves NN, Ambrósio CE, Piedrahita JA. 2014. Stem cells and regenerative medicine in domestic and companion animals: a multispecies perspective. *Reprod Domest Anim*, 49:2-10.

Hai T, Teng F, Guo R, Li W, Zhou Q. 2014. One-step generation of knockout pigs by zygote injection of CRISPR/Cas system. *Cell Res*, 24:372-375.

Heigwer F, Kerr G, Boutros M. 2014. E-CRISP: fast CRISPR target site identification. *Nat Methods*, 11:122-123.

Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, Gersbach CA. 2015. Epigenome editing by a CRISPR/Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol*, 33:510-517.

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, Cradick TJ, Marraffini LA, Bao G, Zhang F. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*, 31:827-832.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096):816-821.

Kim D, Bae S, Park J, Kim E, Kim S, Yu HR, Hwang J, Kim JI, Kim JS. 2015. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat Methods*, 12:237-243- 1 p following 243.

Kim D, Kim J, Hur JK, Been KW, Yoon SH, Kim JS. 2016. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat Biotechnol*, 34:863-868.

Kim YB, Komor AC, Levy JM, Packer MS, Zhao KT, Liu DR. 2017. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat Biotechnol*, 35:371-376.

Klann TS, Black JB, Chellappan M, Safi A, Song L, Hilton IB, Crawford GE, Reddy TE, Gersbach CA. 2017. CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat Biotechnol*, 35:561-568.

Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, Joung JK. 2016. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, 529(7587):490-495.

Koh S, Piedrahita JA. 2014. From ES-like cells to induced pluripotent stem cells: a historical perspective in domestic animals. *Theriogenology*, 81:103-111.

Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, 533(7603):420-424.

Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F. 2015. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*, 517(7536):583-588.

Kues WA, Schwinzer R, Wirth D, Verhoeyen E, Lemme E, Herrmann D, Barg-Kues B, Hauser H, Wonigeit K, Niemann H. 2006. Epigenetic silencing and tissue independent expression of a novel tetracycline inducible system in double-transgenic pigs. *FASEB J*, 20:1200-1202.

Kuroiwa Y, Kasinathan P, Matsushita H, Sathiyaselan J, Sullivan EJ, Kakitani M, Tomizuka K, Ishida I, Robl JM. 2004. Sequential targeting of the genes encoding immunoglobulin-mu and prion protein in cattle. *Nat Genet*, 36:775-780.

Kwon DY, Zhao YT, Lamonica JM, Zhou Z. 2017. Locus-specific histone deacetylation using a synthetic CRISPR-Cas9-based HDAC. *Nat Commun*, 8:15315. **doi:** 10.1038/ncomms15315.

Le Provost F, Lillico S, Passet B, Young R, Whitelaw B, Vilotte J-L. 2010. Zinc finger nuclease technology heralds a new era in mammalian transgenesis. *Trends Biotechnol*, 28(3):134-141.

Li X, Wang Y, Liu Y, Yang B, Wang X, Wei J, Lu Z, Zhang Y, Wu J, Huang X, Yang L, Chen J. 2018. Base editing with a Cpf1-cytidine deaminase fusion. *Nat Biotechnol*, 36:324-327.

Liao HK, Hatanaka F, Araoka T, Reddy P, Wu MZ, Sui Y, Yamauchi T, Sakurai M, O'Keefe DD, Núñez-Delicado E, Guillen P, Campistol JM, Wu CJ, Lu LF, Esteban CR, Izpisua Belmonte JC. 2017. In vivo target gene activation via CRISPR/Cas9-mediated trans-epigenetic modulation. *Cell*, 171:1495-1507. e15.

Maruyama T, Dougan SK, Truttmann MC, Bilate AM, Ingram JR, Ploegh HL. 2015. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat Biotechnol*, 33:538-542.

McCreath KJ, Howcroft J, Campbell KH, Colman A, Schnieke AE, Kind AJ. 2000. Production of gene-targeted sheep by nuclear transfer from cultured somatic cells. *Nature*, 405(6790):1066-1069.

Ng HH, Bird A. 1999. DNA methylation and chromatin modification. *Curr Opin Genet Dev*, 9:158-163.

Niu Y, Zhao X, Zhou J, Li Y, Huang Y, Cai B, Liu Y, Ding Q, Zhou S, Zhao J, Zhou G, Ma B, Huang X, Wang X, Chen Y. 2018. Efficient generation of goats with defined point mutation (I397V) in GDF9 through CRISPR/Cas9. *Reprod Fertil Dev*, 30:307-312.

Nowak-Imialek M, Kues WA, Petersen B, Lucas-Hahn A, Herrmann D, Haridoss S, Oropeza M, Lemme E, Schöler HR, Carnwath JW, Niemann H. 2010. Oct4-enhanced green fluorescent protein transgenic pigs: a new large animal model for reprogramming studies. *Stem Cells Dev*, 20:1563-1575.

Park KE, Powell A, Sandmaier SE, Kim CM, Mileham A, Donovan DM, Telugu BP. 2017. Targeted gene knock-in by CRISPR/Cas ribonucleoproteins in porcine zygotes. *Sci Rep*, 7:42458. doi: 10.1038/srep42458.

Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, Thakore PI, Glass KA, Ousterout DG, Leong KW, Guilak F, Crawford GE, Reddy TE, Gersbach CA. 2013. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods*, 10:973-976.

Piedrahita JA, Zhang SH, Hagaman JR, Oliver PM, Maeda N. 1992. Generation of mice carrying a mutant apolipoprotein E gene inactivated by gene targeting in embryonic stem cells. *Proc Natl Acad Sci USA*, 89:4471-4475.

Piedrahita JA, Olby N. 2011. Perspectives on transgenic livestock in agriculture and biomedicine: an update. *Reprod Fertil Dev*, 23:56-63.

Prather RS, Shen M, Dai Y. 2008. Genetically modified pigs for medicine and agriculture. *Biotechnol Genet Eng* Rev, 25:245-266.

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 152:1173-1183.

Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, Koonin EV, Sharp PA, Zhang F. 2015. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature*, 520(7546):186-191.

Sander JD, Joung JK. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*, 32:347-355.

Schnieke AE, Kind AJ, Ritchie WA, Mycock K, Scott AR, Ritchie M, Wilmut I, Colman A, Campbell KH. 1997. Human factor IX transgenic sheep produced by transfer of nuclei from transfected fetal fibroblasts. *Science*, 278(5346):2130-2133.

Shen B, Zhang W, Zhang J, Zhou J, Wang J, Chen L, Wang L, Hodgkins A, Iyer V, Huang X, Skarnes WC. 2014. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Methods*, 11:399-402.

Shimatani Z, Kashojiya S, Takayama M, Terada R, Arazoe T, Ishii H, Teramura H, Yamamoto T, Komatsu H, Miura K, Ezura H, Nishida K, Ariizumi T, Kondo A. 2017. Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat Biotechnol*, 35:441-443.

Smithies O, Koralewski MA, Song KY, Kucherlapati RS. 1984. Homologous recombination with DNA introduced into mammalian cells. *Cold Spring Harb Symp Quant Biol*, 49:161-170.

Smithies O. 2001. Forty years with homologous recombination. *Nat Med*, 7:1083-1086.

Smithies O. 2008. Turning pages (Nobel lecture). *Chembiochem*, 9:1342-1359.

Suzuki K, Tsunekawa Y, Hernandez-Benitez R, Wu J, Zhu J, Kim EJ, Hatanaka F, Yamamoto M, Araoka T, Li Z, Kurita M, Hishida T, Li M, Aizawa E, Guo S, Chen S, Goebl A, Soligalla RD, Qu J, Jiang T, Fu X, Jafari M, Esteban CR, Berggren WT, Lajara J, Nuñez-Delicado E, Guillen P, Campistol JM, Matsuzaki F, Liu GH, Magistretti P, Zhang K, Callaway EM, Zhang K, Belmonte JC. 2016. In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature*, 540(7631):144-149.

Tsai SQ, Zheng Z, Nguyen NT, Liebers M, Topkar VV, Thapar V, Wyvekens N, Khayter C, Iafrate AJ, Le LP, Aryee MJ, Joung JK. 2015. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol*, 33:187-197.

Tsai SQ, Nguyen NT, Malagon-Lopez J, Topkar VV, Aryee MJ, Joung JK. 2017. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat Methods*, 14:607-614.

Vazquez JC, Nogues C, Rucker EB, Piedrahita JA. 1998. Factors affecting the efficiency of introducing precise genetic changes in ES cells by homologous recombination: tag-and-exchange versus the Cre-loxp system. *Transgenic Res*, 7:181-193.

Vojta A, Dobrinić P, Tadić V, Bočkor L, Korać P, Julg B, Klasić M, Zoldoš V. 2016. Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res*, 44:5615-5628.

Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R. 2013. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*, 153:910-918.

Wang X, Yu H, Lei A, Zhou J, Zeng W, Zhu H, Dong Z, Niu Y, Shi B, Cai B, Liu J, Huang S, Yan H, Zhao X, Zhou G, He X, Chen X, Yang Y, Jiang Y, Shi L, Tian X, Wang Y, Ma B, Huang X, Qu L, Chen Y. 2015. Generation of gene-modified goats targeting MSTN and FGF5 via zygote injection of CRISPR/Cas9

system. *Sci Rep*, 5:13878.

Yang H, Wang H, Shivalila CS, Cheng AW, Shi L, Jaenisch R. 2013. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell*, 154:1370-1379.

Yao X, Wang X, Hu X, Liu Z, Liu J, Zhou H, Shen X, Wei Y, Huang Z, Ying W, Wang Y, Nie YH, Zhang CC, Li S, Cheng L, Wang Q, Wu Y, Huang P, Sun Q, Shi L, Yang H. 2017. Homology-mediated end joining-based targeted integration using CRISPR/Cas9. *Cell Res*, 27:801-814.

Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A, Koonin EV, Zhang F. 2015. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, 163:759-771.

Zhou H, Liu J, Zhou C, Gao N, Rao Z, Li H, Hu X, Li C, Yao X, Shen X, Sun Y, Wei Y, Liu F, Ying W, Zhang J, Tang C, Zhang X, Xu H, Shi L, Cheng L, Huang P, Yang H. 2018. In vivo simultaneous transcriptional activation of multiple genes in the brain using CRISPR-dCas9- activator transgenic mice. *Nat Neurosci*, 21: 440-446.

CHAPTER 2

LGR5 is a conserved marker of hair follicle stem cells across species and is present early and throughout follicle morphogenesis

Abstract

Hair follicle stem cells are key for driving growth and homeostasis of the hair follicle niche and have remarkable regenerative capacity throughout hair cycling and display fate plasticity during cutaneous wound healing. Due to the need for a transgenic reporter, essentially all observations related to LGR5+ hair follicle stem cells have been generated using transgenic mice, which have significant differences in anatomy and physiology from the human. Using a transgenic pig model, a widely accepted model for human skin and human skin repair, we demonstrate that LGR5 is a marker of hair follicle stem cells across species in homeostasis and development. We also report the strong similarities and important differences in expression patterns, gene expression profiles, and developmental processes between species. This information is important for understanding the fundamental differences and similarities across species, and ultimately improving human hair follicle regeneration, cutaneous wound healing, and skin cancer treatment.

Introduction

The skin is the largest organ in the body and is responsible for maintaining homeostatic conditions such as thermoregulation, hydration, and protection from the environment [1]. To maintain these processes, the skin epidermis contains multiple populations of stem cells. One of these populations, hair follicle stem cells (HFSC), has been studied extensively in mice for understanding stem cell behavior in homeostasis and repair [2,3]. Although studies in mice have provided in depth mechanistic insight, translational studies using mouse models are limited due to differences in anatomy and physiology; mice have dense fur coats which undergo cyclic periods of growth and rest, thin dermis and epidermis, and loose skin attachment [1,4]. In contrast, humans have sparse hair coats, asynchronous follicle cycling, thick dermis and epidermis, and tight skin attachments; all these characteristics are shared with pigs, making them a widely accepted model for human skin and repair [5]. However, while several genes have been proposed as HFSC markers, they often vary across species [6,7], making it difficult to perform functional comparisons between species.

In mice, HFSC have been defined by the leucine-rich G protein coupled receptor-5 (LGR5) [8–10], a known potentiator of WNT signaling when bound to its ligand R-spondin [11]. Exciting findings from murine studies show that LGR5+ HFSC contribute to all regions of the hair follicle, including sebocytes [8], are vital for telogen to anagen transition during hair cycling [12–14], migrate from the hair follicle niche to contribute to re-epithelialization during wound healing [9,15], and are the cells responsible for hair regeneration in wound-induced follicle neogenesis [16,17]. While LGR5 is a widely used marker in mice, there have been no previous studies to our knowledge that evaluate the location of LGR5 expression in the skin of humans or non-murine models. This

is possibly because its study has been limited due to the lack of a reliable commercially available anti-LGR5 antibody [18,19].

To solve this issue, we developed a transgenic pig that expresses LGR5-driven H2B-GFP and here we use it to identify conserved pathways of gene expression, compare cell behavior, and study the role of LGR5+ HFSC in fetal and post-natal stages across species.

Results

Generation of a porcine LGR5-H2B-GFP model

To facilitate the study of LGR5+ cells in a non-murine species, we generated a transgenic porcine model expressing H2B-GFP under the control of the *LGR5* promoter using CRISPR/Cas9 mediated gene knock-in. We elected to use a nuclear H2B-GFP we and others have previously used [2,20,21]. The H2B-GFP sequence was inserted into exon 1 of one *LGR5* allele and after validation for accurate transgene insertion by PCR and sequencing (Figure 1A-C), clonal cell lines were used for somatic cell nuclear transfer. To examine whether H2B-GFP expression accurately reflects LGR5 mRNA expression, we overlaid RNA fluorescent *in situ* hybridization (RNA-FISH) of porcine *LGR5* and LGR5-H2B-GFP expression (Figure 1d). Based on colocalization, H2B-GFP in this model is faithfully representative of *LGR5* expression.

LGR5 is a marker of the outer root sheath and bulge cells in anagen, catagen, and telogen stages of the hair cycle in porcine and human skin, and is expressed at a low level in the inner root sheath.

Studies in mice have shown LGR5 expression in the lower bulge and outer-root sheath throughout anagen, catagen and telogen. The “bulge” niche is defined as the region in the hair

follicle below follicles throughout the hair cycle (Figure 1H-J). In addition, a closer examination of nuclear GFP fluorescence in the pig beyond the outer root sheath reveals a GFP-dim population in the inner root sheath (Figure S1a). This LGR5-low population is expressed in the inner root sheath above and at the bulge (Figure S1b-d).

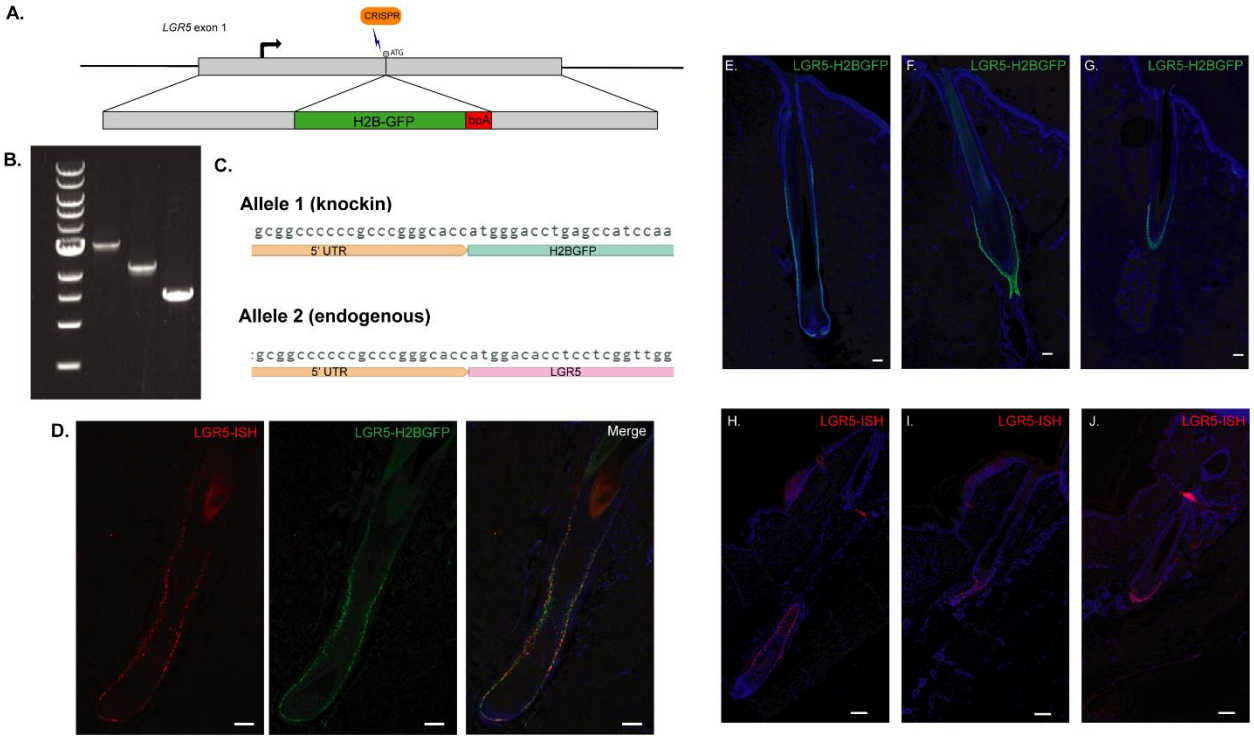


Figure 1. LGR5 is expressed in the hair follicle outer root sheath in the lower bulge throughout anagen, catagen, and telogen in the pig and human. A) Schematic of H2B-GFP knock-in at the porcine *LGR5* locus. 500 base pair regions of homology on either side of the CRISPR target site at the start codon were included in the homology directed repair template. B) PCR amplification of DNA from a colony containing the knock-in, from left: 200 bp DNA Ladder, 5' junction (allele #1), 3' junction (allele #1), unedited genomic allele (allele #2). C) Sequencing of *LGR5* allele #1 and allele #2 shows the endogenous locus of allele #1 remains unaltered and expresses the LGR5 protein, while allele #2 encodes H2B-GFP instead. D) *LGR5* transcript expression correlates with H2B-GFP expression in the skin of *LGR5*-H2B-GFP pigs, shown by endogenous *LGR5*-H2B-GFP and *LGR5* fluorescent *in situ* hybridization to a pig *LGR5* probe (*LGR5*-ISH). E-G) Expression of *LGR5-H2B-GFP* in anagen (E), catagen (F), and telogen (G) stage follicles. H-J) Detection of *LGR5* expression in human hair follicles by RNA *in situ* hybridization. As in pigs, *LGR5* is expressed in the outer root sheath of the lower bulge in anagen (H), catagen (I), and telogen (J).

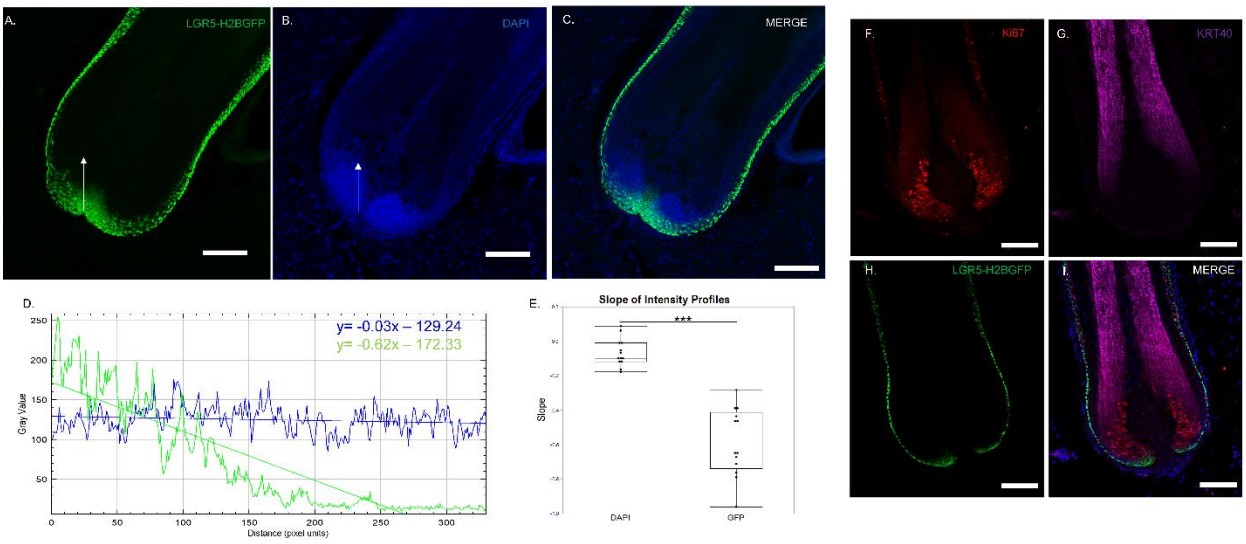


Figure 2. LGR5+ cells give rise to transient amplifying populations in hair matrix in anagen. A-C) 20x magnification confocal image of base of anagen hair follicle depicting H2BGFP expression as compared with DAPI. Intensity profile of each fluorescence along the white arrow is plotted in (D), which is summarized by the fit line. E) Quantification of slope from profile plots of GFP or DAPI profiles, with profile line drawn from base toward inner root sheath, showing that H2BGFP intensity is diluted significantly while DAPI remains relatively unchanged, n=12 follicles from 3 different pigs. IHC Staining to detect F) KI67 G) Keratin 40 or H) H2BGFP show LGR5+ cells giving rise to proliferating transient amplifying cells, followed by differentiation into the inner root sheath or hair shaft. I) Merge. Scale bar represents 100 μM.

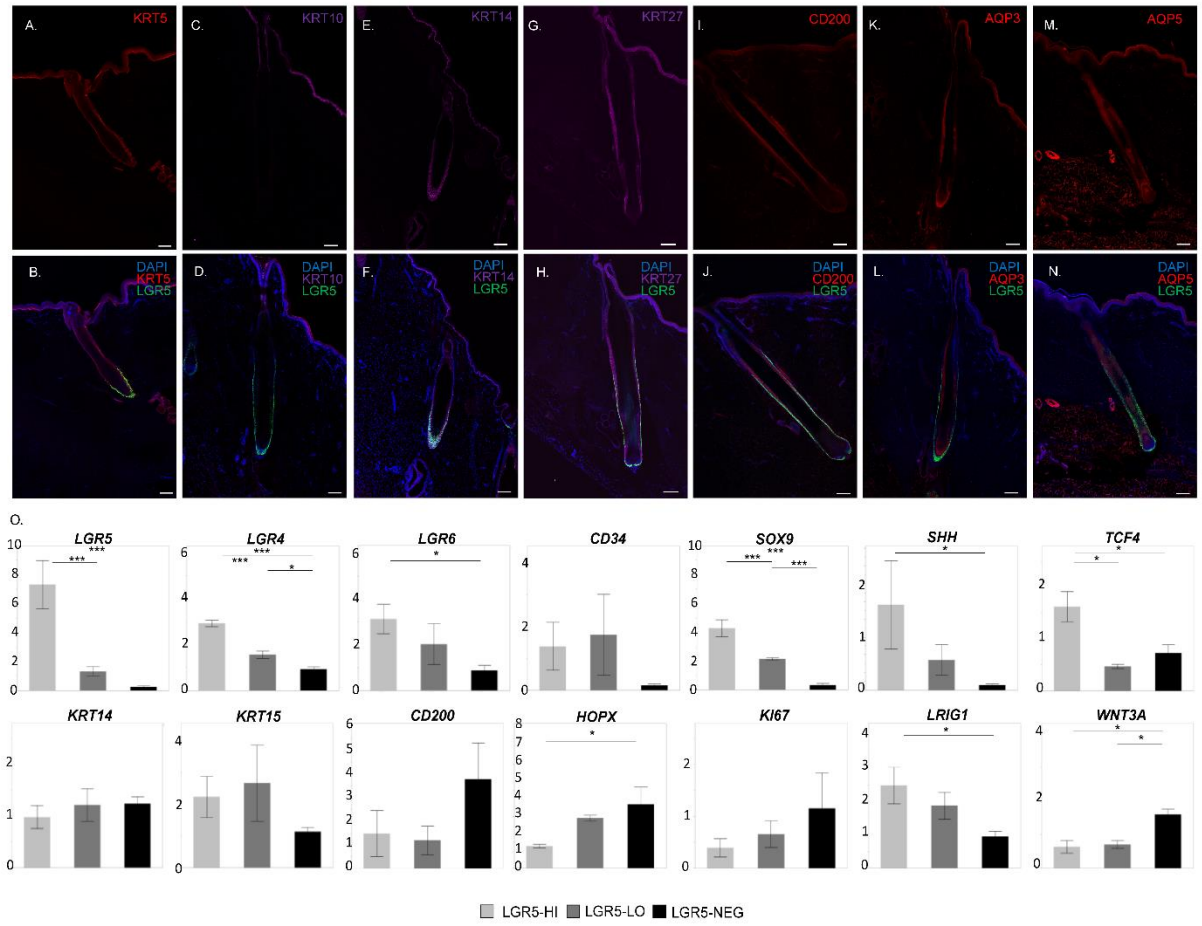


Figure 3. LGR5+ epidermal cells express known stem cell and epidermal markers. Antibody staining by fluorescent confocal microscopy of known skin markers (top A-M) and overlaid with DAPI and H2B-GFP (bottom B-N). A-B) Keratin 5, C-D) Keratin 10, E-F) Keratin 14, G-H) Keratin 27, I-J) CD200, K-L) Aquaporin 3, M-N) Aquaporin 5. Scale bar represents 200 μM. O) Relative expression of known stem cell markers based on delta-delta Ct analysis of RT-qPCR from LGR5-high (LGR5-HI, light gray), LGR5-low (LGR5-LO, dark gray), or LGR5-negative (LGR5-NEG, black) populations after flow sorting. Each sample is normalized to *ACTB* and *GAPDH* and then to unsorted epidermis, n=3 pigs. Student's T-test for each pair *** indicates P<0.005, * indicates P<0.05, error bars indicate SEM.

LGR5-expressing cells in the hair germ give rise to the inner root sheath and hair shaft

During anagen, HFSC are activated, exit the bulge region, proliferate downward in the outer root sheath to the hair matrix region. Here, they give rise to a transient amplifying population, which ultimately differentiates into 7 different lineages within the hair shaft and inner root sheath [23–25]. Thus, we next asked whether LGR5+ cells in the porcine hair follicle give rise to the hair and the inner-root sheath. The H2B-GFP marker in the nucleus of LGR5 expressing cells allows limited lineage tracing as the H2B-GFP signal drops in intensity with each cell division [2,26]. LGR5-expressing cells at the base of the hair follicle in contact with the basement membrane and adjacent to the dermal papilla, show bright GFP expression, which decreases in intensity as the cells divide into the inner root sheath and hair shaft (Figure 2A-C). The line profile drop in intensity was compared to DAPI, which remains constant along the same vector (Figure 2D-E). KI67 staining confirmed that indeed the LGR5+ cells are dividing as the GFP intensity decreases (Figure 2F), and that the transient amplifying cell populations terminally differentiate into the inner root sheath and hair shaft and cortex, as shown by hair keratin, Keratin 40 (Figure 2G-I).

Immunohistochemistry and gene expression analysis of LGR5+ cell localization with known markers of the epidermis

Next, we applied antibody staining to evaluate LGR5 localization in the context of well-known epidermal markers of mice and/or human skin. KRT5 and KRT14, which form an intracellular dimer [27], are present in the entire basal layer of the porcine epidermis, and co-localize with LGR5-H2B-GFP in the bulge region of the hair follicle (Figure 3A-B, E-F). KRT10, a marker of differentiated interfollicular epidermis cells, does not co-localize with LGR5-H2B-GFP, and was not detected in the follicle below the isthmus (Figure 3C, D). KRT27, a marker of the Henle,

Huxley and Cuticle regions of the inner root sheath [28], stains the LGR5-low populations in the inner root sheath, in addition to the entire interfollicular epidermis (Figure 3G-H). Unlike it's reported stem-cell specific affinity in humans [29], CD200 stained all layers of the epidermis, including LGR5-H2B-GFP cells in the bulge (Figure 3I-J). AQP3, a known marker of keratinocytes [30,31], shows minimal co-localization with LGR5-H2B-GFP, and instead stains the interfollicular epidermis and continues through the inner root sheath (Figure 3K, L). Aquaporin-5 (AQP5) has recently been suggested as an alternative to LGR5 to mark stem cells in the stomach [18], but is not specific in the skin (Figure 3M-N). From this we can confirm that in the pig, LGR5+ HFSC partially co-localize with markers of the basal layer stem cells including KRT5, KRT14, CD200 and AQP5, but not with more differentiated markers of the epidermis and keratinocytes such as KRT10, KRT27, and AQP3.

While LGR5 is recognized as a marker of HFSC in mice [8], other stem cell markers such as SOX9, TCF4, LRIG1, LGR4 and LGR6, have also been linked to stem cell populations of the epidermis in mice and humans [22,32]. To examine whether LGR5+ HFSCs also express other stem cell markers, we used fluorescence activated cell sorting (FACS) to separate the LGR5-high, low, and negative populations (Figure S2) and performed RT-qPCR separately on each population (Figure 3O). Our results show that LGR5+ cells are also enriched for *LGR4* and *LGR6* and that *SOX9, TCF4,* and *LRIG1* are significantly upregulated in the LGR5-high populations. Markers that have been used to identify and/or enrich for murine HFSC such as *KRT5, KRT15,* and *CD34* are not significantly enriched in the porcine LGR5+ HFSC.

Transcriptome analyses reveal extracellular matrix and structure organization as top conserved pathways across species

To examine the gene expression profile of LGR5+ cells in the pig at a deeper molecular level, we performed RNA sequencing (RNAseq) on RNA extracted from FACS sorted GFP high (LGR5-high) or GFP negative (LGR5-negative) cells as in Fig S4. Statistical analyses found 1619 significantly upregulated and 1161 significantly downregulated genes in the LGR5-high population compared with the LGR5-negative population after correction with Bonferroni adjustment for multiple comparisons. A heatmap of the 30 most differentially expressed genes shows significant differences in many epidermal, structural and stem cell genes (Figure 4A). Two of the top three most significantly down-regulated genes in the LGR5-high group were KRT10 and AQP3, which are confirmed by our immunostaining results (Figure 3D, L).

We next aimed to evaluate how the gene expression profile and signalling pathways of the LGR5+ cells in pigs compares with the human and the mouse. To do so, we used publicly available single cell RNAseq datasets from human [33] and mouse [34], and clustered the cells based on high or undetectable levels of *LGR5* expression. After quality control, we retained 3593 and 1422 human and mouse cells, respectively for further analysis. Clustering based on *LGR5* expression yielded 3400 negative and 59 positive human cells and 1278 negative and 82 positive mouse cells. In total, 836 common genes, as well as 1,329; 953; and 5,763 uniquely expressed genes detected in human; mouse; and pig, respectively (q-value $\leq$ 0.1). The gene ontology analysis revealed conserved pathways in comparison of species by pairs, with "extracellular matrix organization", "extracellular structure organization," and "cell-substrate adhesion" common between all three comparisons (Figure 4B, Fig S4A-C), which suggest a role for LGR5+ cells orchestrating the extracellular microenvironment. Comparative analysis using

IPA revealed a significant number of potential upstream regulators with the 20 most significant genes shown in Figure 4C, including those involved in growth (*TGFB family*, specifically*TGFB1*), transcription factors and cell signaling (*MYC*, *MYCN*, *CTNNB1*, *TNF*, *ESR1*, *ESR2*, *TP63*), tumor suppressor genes (*TP53*) cell differentiation (*SP1*, *GLI1*), and regulation of cell division (*HRAS*, *KRAS*). Overall, while many genes and pathways are shared across the species and cellular expression, there is also great variability when it comes to specific gene expression.

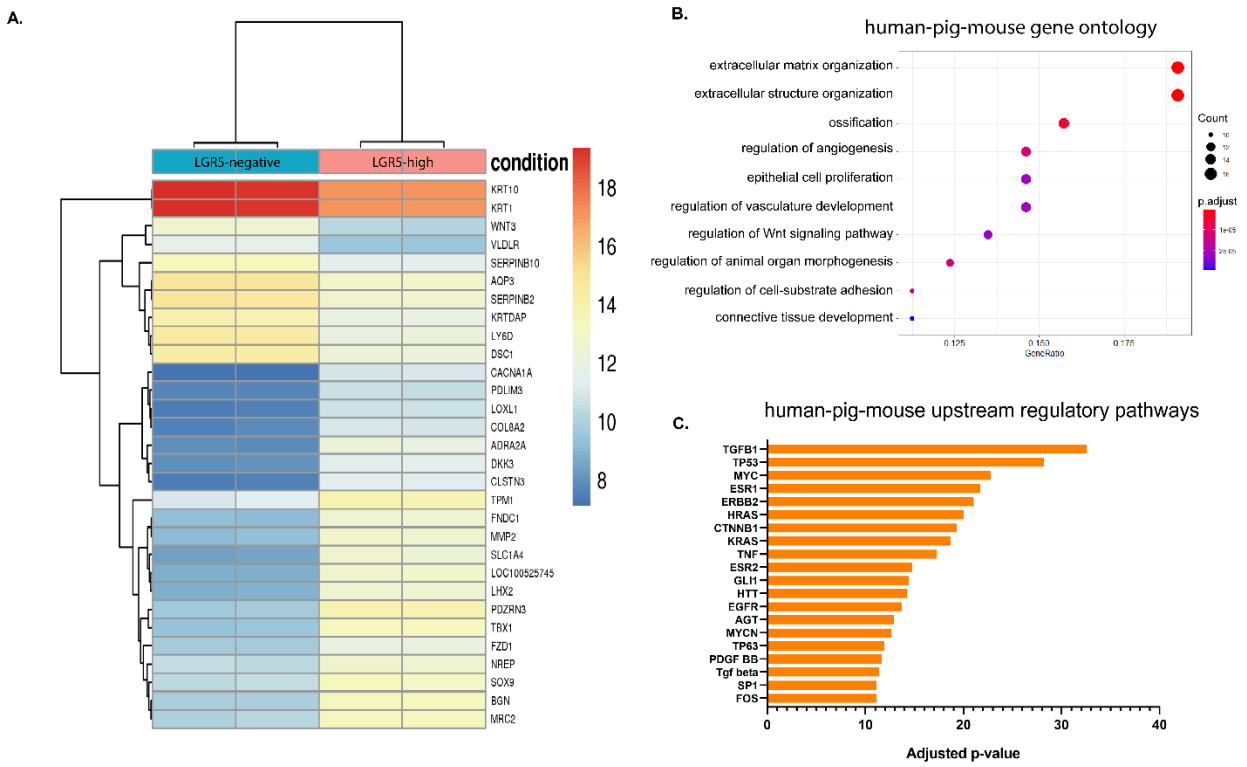


Figure 4. Human, Pig, and Mouse LGR5+ epidermal populations are enriched in extracellular matrix structure and organization pathways. A) Top 30 significantly differentially expressed genes between LGR5-high and LGR5-negative sorted porcine samples. B) Common gene ontology terms from genes that are upregulated in LGR5-high epidermis across all species. C) Upstream regulators that target a significant portion of the genes that are upregulated in LGR5+ cells from mice, human, and pig datasets.

LGR5 is expressed early in hair follicle development and throughout morphogenesis.

We next asked how LGR5 expression fits in fetal hair follicle neogenesis. In pigs we examined two key fetal stages, gestational day 50 (D50) when hair follicles are in the late placode/early hair germ stage, and gestational day D80 (D80) in the follicle stage of morphogenesis. At D50, whole mount dorsal fetal skin shows the evenly patterned distribution of LGR5 expression throughout the epidermis (Figure 5A, B), and LGR5+ cells are oriented perpendicular and directly adjacent to the dermal condensate marked by SOX2 (Figure 5C). SOX9, a known marker of hair follicle progenitor cells, occupies the suprabasal position and only dimly overlaps with LGR5 expression (Figure 5D-F). The majority of LGR5+ cells are not proliferating, with only a few LGR5+ cells in the suprabasal position co-expressing both KI67 and LGR5 (Figure 5G-I), indicating that at this stage, LGR5+ cells are slow cycling and that the primary proliferating population is in the suprabasal position. Previous reports suggest there is a group of basal cells that are Wnt-high, SHH-high, and are the slow cycling progenitors that give rise to the SOX9+ population [35]. To prove that this population is equivalent to the LGR5+ cells in pigs, we sorted LGR5+ or LGR5- cells from D50 skin and evaluated *SHH* expression by RT-qPCR (Figure S4A-B), and found that the LGR5+ population is significantly enriched in *SHH* expression (Figure S4C), confirming that LGR5 is expressed in the earliest stem cell progenitor population of the hair follicle.

At D80, during the follicle stage of morphogenesis, LGR5 expression is robust throughout the lower bulge (Figure 5J-K). Interestingly, the pattern of follicle development in dorsal skin emerges as a triplet (Figure 5J-K), with all developing follicles enveloping the dermal papillae, shown by SOX2 expression (Figure 5L). While LGR5 specifically marks the developing outer root sheath of the lower follicle, SOX9 is expressed in mostly all cells in the lower follicle

(Figure 5M-O). The majority of proliferating LGR5+ cells were found at the base of the hair follicle, close to the dermal papillae, similar to anagen (Figure 5P-R).

To ask if LGR5 expression is conserved throughout development, and since human fetal skin was difficult to obtain, we obtained age-matched fetal rhesus skin and queried *LGR5* expression by RNA-FISH. *LGR5* is expressed at a low level in the placode (arrowhead), and strongly expressed in the hair peg (arrow) (Figure 6A-B). SOX9 expression is absent from the placode, but is found throughout the hair peg (Figure 6C-D). Interestingly, we consistently found that none of the cells in the hair peg, perpendicular to the basement membrane and adjacent to the dermal papillae, were SOX9 positive, although LGR5 was detected in these cells. Finally, LGR5 expression is consistent and robust throughout the lower hair follicle throughout development, along with SOX9 (Figure 6E-H).

While studies in mice have reported that LGR5 is not detected in mice until E18.5 [8,36], cryosections from LGR5-eGFP mouse dorsal skin show that LGR5 is expressed as early as 15.5 (Figure 6I-P). Since follicles develop at different stages based on location, we collected tail or dorsal skin for placode or hair germ stages, respectively. LGR5 was undetectable at the placode stage (Figure 6I-J), although SOX9+ cells occupied the same suprabasal position as in other species (Figure 6K-L). In mice, we show that LGR5 is first detectable in the hair peg stage (Figure 6N-P), although the pattern of expression was more limited than the pig and rhesus skin and is not expressed in cells that contact the dermal papillae, and closely aligns with SOX9 expression. From these results, we can conclude that LGR5 is expressed early and throughout hair follicle development across species, although expression in rhesus monkey and pig hair follicle morphogenesis is earlier and more widespread than in the mouse.

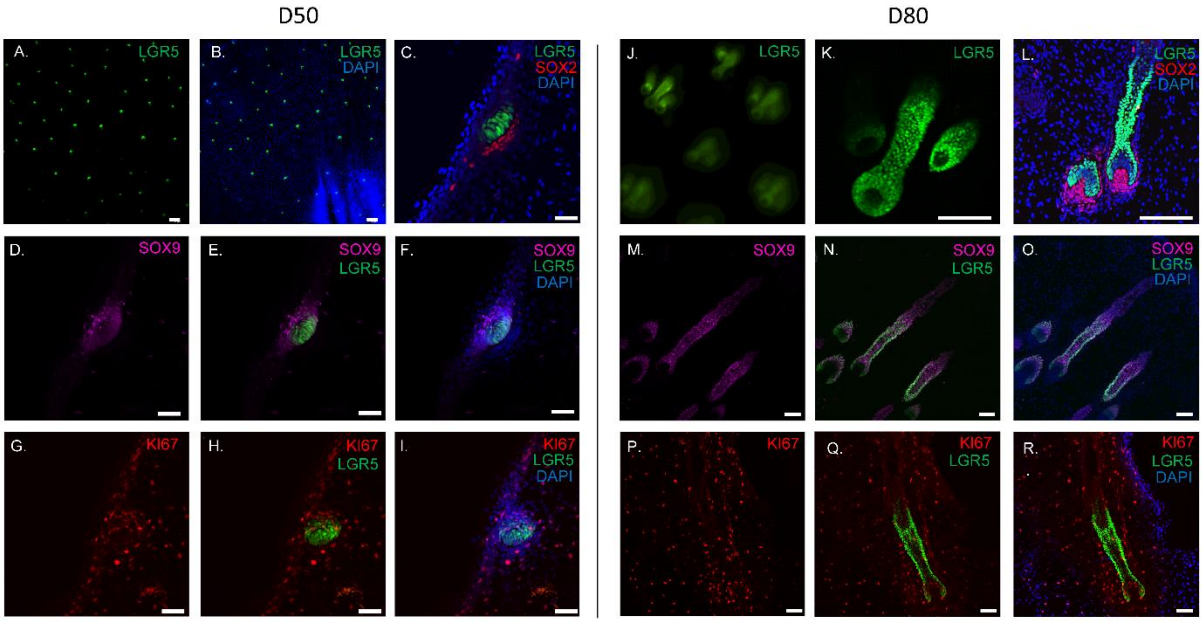


Figure 5. LGR5 is expressed early and late throughout developing porcine hair follicles. Fetal porcine skin at day 50 of gestation (A-K). Whole mount confocal imaging of showing patterning of LGR5 (A) and overlay with DAPI (B). Cryosectioned immunostaining show LGR5+ cells throughout the hair germ, perpendicular to the basement membrane and adjacent to the developing dermal papillae marked by SOX2 (D-F). Comparison of LGR5 or KI67 staining show little overlap (G-I), confirming our qPCR-data that the LGR5+ cells are a slow-cycling population. Day 80 porcine fetal skin (J-R). Whole mount staining shows triplet pattern of development. LGR5+ cells remain in close contact, surrounding the dermal papillae as they invaginate into the developing dermis (J-K). LGR5 expression is found in the outer root sheath of the lower developing follicle, partially overlapping with SOX9 expression which is not limited to the outer root sheath (J-R). KI67 shows that the majority of the proliferating LGR5+ cells are found at the base of the hair follicle, adjacent to the dermal papillae (P-R). Scale bar represents 200 μM (A-C), 50 μM (D-K), or 100 μM (L-T).

Discussion

Many studies have attempted to find markers for HFSC, but few of these markers are conserved across species. Some studies have suggested that CD200 may be a shared marker for HFSCs between mice and humans [37,38], however we and others show that CD200 in mice and pigs is expressed throughout the whole epidermis [6,7,39]. CD34, a functional marker of HFSC in mice, is not enriched in human HFSC [37] as our results confirm in pigs. KRT15 has also been used as a

marker of HFSC in mice, however it is also found in the entire basal layer of neonatal human skin, and in the rete ridges of adult human skin [40]. Despite the variation in patterns in many of the epidermal genes across species, in this work we show that in postnatal skin, LGR5 is a consistent marker of HFSC in the bulge across mice, pigs, and humans. In previous work in human cells, LGR5 was detected in human hair stem cell organoids [41], in the hair follicle region of haired men but not of bald men (alopecia) [42], and was identified via single cell RNAseq in human skin as a potential marker for the hair follicle lower bulge [33]. Nevertheless, this work is the first to spatially define the position of LGR5 expression in human hair follicles using RNA *in situ* hybridization.

The creation of the transgenic porcine model provides the opportunity to perform a cross-species examination of the transcriptome from the same cell type across multiple species. We recognize that our RNAseq analyses could not account for the differences in methods of collection from each dataset (especially related to depth of sequencing single cells vs bulk and number of replicates). However, it provides a starting point for examination of which genes and pathways are conserved, namely genes involved in extracellular matrix organization and extracellular matrix structure. More work needs to be done to understand the role of these stem cells in the modulation of the extracellular matrix and how that affects the cell fates within the niche, especially as hairs undergo cycling.

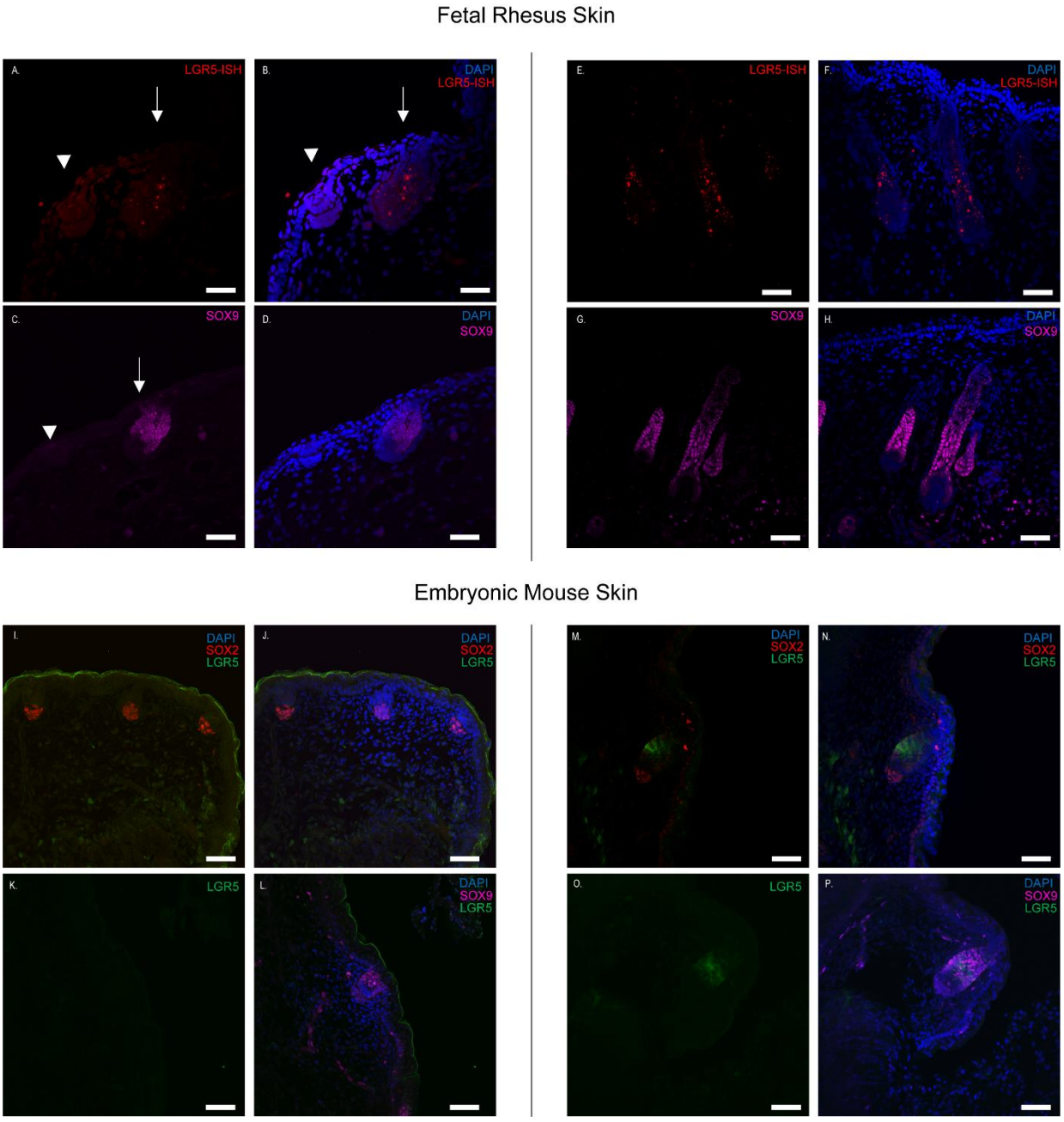


Figure 6. LGR5 expression in rhesus and mouse hair follicle morphogenesis. Gestational day 60 (A-D) or 90 (E-H) skin show rhesus skin also develops in a triplet pattern, with one more advanced developed follicle flanked by two lesser developed follicles. LGR5 is present, but expressed at a low level in the placode, while it is robustly expressed in the hair peg (A-B). SOX9 expression is absent in the placode, but present in the hair peg. The cells directly adjacent to the dermal papillae, with a perpendicular orientation to the basement membrane, are SOX9 negative (C-D). LGR5 expression is maintained throughout the lower hair follicle in later stages of development (E-F), as well as SOX9 (G-H). Embryonic day 15.5 mouse tail (I-L) or dorsal skin (M-P) show mouse placode or peg stage development, respectively. LGR5 is undetectable or absent at the placode stage (I-L), although SOX2 and SOX9 are present. In the hair peg stage, LGR5 is present, but limited to the mid-peg region and not in contact with the dermal papillae, similar to the pattern of SOX9 expression (M-P).

Beyond molecular comparisons of adult cells at homeostasis, we also can use this model to expand understanding of the role of LGR5 during follicle morphogenesis. The question has remained open as to what restricts canonical WNT signaling and polarizes the basal cells in epithelial bud formation in the placode [35]. Our results suggest that LGR5, a known potentiator of WNT signaling, could have a role in dictating which cells in the niche are WNT-responsive. Other studies have shown LGR5 to be involved with cell adhesion [43,44], suggesting a role for LGR5 in anchoring asymmetrically dividing progenitors to the WNT-hi niche. While this could be true for rhesus and pig, LGR5 is not detected until after the placode stage in the mouse. Furthermore, when it is first detected, the pattern of expression in the mouse is similar to that of SOX9, and is not expressed in the WNT-hi cells adjacent to the dermal papillae. While the exact role of LGR5 in follicle morphogenesis is unclear, these differences in early morphogenesis could point to slightly different signaling pathways in early development between mice and other species.

Overall, the depth of knowledge that can be gained from an additional model beyond the mouse can provide more clues toward the behavior of stem cells across species through stages of development, homeostasis, and disease. The development of this LGR5-H2B-GFP transgenic pig represents a translational milestone in which we are able to both confirm and expand knowledge gained from mouse models and develop it toward human medicine. Future experiments using this model will enable us to study the mechanisms of how the hair follicle stem cells contribute to wound healing, improve our understanding of hair disease such as alopecia, and better elucidate the complexities of the hair follicle stem cell niche and understand the utility of these cells for skin and hair regeneration.

Materials and Methods

Generation of LGR5-H2B-GFP pig

All experiments were performed in strict accordance with the approved Institutional Animal Care and Use Committee of North Carolina State University (IACUC protocol 17-028-B), in addition to the ARRIVE guidelines [45]. CRISPR/Cas9 nuclease was used to create a double-stranded break in the genomic DNA in exon 1 of the porcine LGR5 gene (gRNA sequence: ACCATGGACACCTCCTCGGT). A homology-directed repair template plasmid containing H2B-eGFP flanked by 1000bp homology arms flanking the cut-site was co-transfected with the Cas9 (Gift from Keith Joung, Addgene #72247) and gRNA (Gift from Keith Joung, Addgene #43860) plasmids, and cells were seeded at low density for colony outgrowth.

Porcine fetal fibroblasts isolated from day 42 fetuses were used for gene editing and somatic cell nuclear transfer. After transfection and low density seeding for colony formation, colonies were genotyped by PCR and sequencing to verify successful targeted transgene integration before somatic cell nuclear transfer. Somatic cell nuclear transfer was completed as previously described [46] and zygotes were surgically transferred into a surrogate and carried until term. Throughout this study, skin from 6 juvenile (2-4 months), 3 adult (>6 months), and 3 fetal day 50 and 3 fetal day 80 pigs were used. The results shown are consistent across offspring derived from somatic cell nuclear transfer (F0) in addition to their progeny (F1) and are representative of both sexes.

Human and Rhesus Samples

Human samples from adult male forearm skin were obtained from Accio Biobank Online and fixed in formalin within 24 hours of death. Samples were embedded in paraffin and sectioned at 7uM for further analyses.

Rhesus monkey fetal skin sections were kindly provided by Dr. Alice Tarantal, UC Davis. Two gestational ages were assessed: 60 days gestation (early second trimester) and 90 days gestation (late second trimester). Sections (5-6 um) were provided from formalin-fixed paraffin-embedded tissues and used for RNA-FISH and immunohistochemistry. Specimens were previously obtained under IACUC-approved protocols.

Immunofluorescence

Tissue was fixed with 4% paraformaldehyde then frozen in Optimal Cutting Temperature Compound (OCT) and sectioned at 20µM. Sections were blocked with IHC/ICC Blocking Buffer (Invitrogen) with 0.4% Triton X-100 (Sigma), incubated with primary and secondary (1:5000) antibodies and finally mounted in Prolong Gold Antifade Mount with DAPI (ThermoFisher). Antibodies and dilutions Anti-Ki67, 1:100 (Abcam ab15580); Anti-KRT14 1:200 (Thermo Fisher MA1-06323); Anti-CD200, 1:50 (ls-b11638); AQ3 1:200 (Abcam ab125219); Anti-KRT10 1:200 (Abcam ab9025); Anti-KRT40, 1:100 (Abcam ab16113); Anti-KRT5, 1:100 (Abcam ab64081). Immunostained samples were visualized by confocal microscopy (Olympus Fluoview FV3000 Confocal Microscope)**.** Line profile intensity was measured using ImageJ (NIH).

Single cell isolation and fluorescence activated cell sorting

Juvenile or adult porcine skin was cut into 5mm^2 pieces and incubated with 10mg/mL Dispase II (Sigma) in PBS without calcium and magnesium (Corning) for 1.5 hours at 37℃ or overnight at

4℃. Hair and epidermis were manually removed from dermis and incubated in 0.05% trypsin for 5-10 minutes at 37°C with shaking. Suspension was vortexed and strained with 70µm cell strainer (BD Falcon). Cells were resuspended in PBS with 10% fetal bovine serum (Corning) and 1% antibiotic-antimycotic (Corning). 250ng/mL propidium iodide (Biotium) was added to cells for live/dead detection and samples were sorted by Beckman Coulter MoFlo XPD**.** Cells were sorted for GFP Hi, Lo, and Neg and data were further analyzed using FlowJo™ (BD Biosciences).

Reverse transcriptase quantitative polymerase chain reaction (RT-qPCR)

Total RNA from skin tissue or sorted skin cells was isolated using Zymo Quick-RNA Microprep kit with on column DNase digest according to manufacturer's instructions. RNA was eluted into DNase/RNase free water and stored at -80℃ until further use. cDNA was synthesized with AffinityScript Multiple Temperature cDNA Synthesis Kit (Agilent), according to the manufacturer instructions. For RT-qPCR, iTaq Universal SYBR Green Supermix (BioRad) was used with cDNA template and forward and reverse primers were designed as listed in Table S1. For optimal conditions: 2min denaturation, 40 cycles of 95℃ denaturation for 5 s and 60℃ extension for 30 s, with final extension at 60℃ for 2 min. Primer sequences can be found in Figure S1. Each sample was amplified on a qTOWER^3 thermal cycler (Analytik Jena) with technical duplicates for three biological replicates with similar results. Each gene expression was normalized to *GAPDH* and *ACTB.*

RNA *in situ* hybridization

RNA *in situ* hybridization (RNA-FISH) was performed using RNAscope (Advanced Cell Diagnostics) according to the manufacturer's instructions. Briefly, paraffin embedded skin tissue was sectioned at 7um. Slides were deparaffinized with xylene, then heat treated followed by

protease digestion. The tissue was hybridized with a 10 ZZ probe targeting either the 560-1589 region of Homo sapiens *LGR5* mRNA, the 466-1464 region of Sus scrofa *LGR5* mRNA, or the 494-1423 region of Macacca mulatta *LGR5* mRNA. As controls, a positive control probe was used against porcine (1-642 region) or human (139-989 region) cyclophilin B, Rhesus peptidyprolyl isomerase B (119-916 region), or negative control probes targeting the bacterial gene dapB were used, followed by chromogenic development. Slides were washed, then mounted with Prolong Gold Antifade Mount with DAPI (ThermoFisher) and imaged by confocal microscopy.

Single-cell and bulk RNA-seq datasets and processing

We compiled two tissue-matched, single-cell epidermal RNA-seq datasets based on published samples for human, and mouse together with newly generated pig bulk RNA-seq samples. Aligned and processed sequencing data from single-cell human [accession numbers GSM3717037 [33]] and mouse [GSE67602 [34]] epidermal and hair follicle profiling studies were obtained from the Gene Expression Omnibus (Edgar, Domrachev, and Lash 2002). Single cell barcodes that had non-zero values for less than 500 genes or a high proportion of mitochondrial gene expression (> 5%) were excluded from further analysis. Bulk RNAseq samples were prepared from porcine cells, at least 500ng of RNA was extracted from sorted LGR5-GFP-high or LGR5-GFP-negative populations from 2 pigs, same as prepared for RT-qPCR. RNAseq was performed externally by GENEWIZ; library preparation with poly(A) selection was performed followed by paired end 150bp sequencing on Illumina HiSeq.

Clustering and analysis of differential gene expression

Single-cell populations were clustered based on *LGR5* gene expression into either *LGR5+* (expression > 1) or *LGR5-* (expression = 0), while bulk RNA-seq data was clustered based on fluorescent markers. A student's t-test was used to calculate p-values for each species followed by a Benjamini-Hochberg multiple test correction at a false discovery rate of 0.05 (Soneson and Robinson 2018). Further core and comparative analyses of the differentially expressed genes were conducted using Ingenuity Pathway Analysis (IPA) along with gene ontology (GO) analysis using the ClusterProfiler R package (QIAGEN 2020; Krämer et al. 2014; Yu et al. 2012; Oliveros 2007).

Data and code availability

The RNA-sequencing data reported in this study have been deposited in NCBI's Gene Expression Omnibus with the accession number GSE190069.

Acknowledgments

We are thankful for the support and hard work of the North Carolina State University Lab Animal Resources, Veterinary Services, and Flow Cytometry Core Facility. This work was supported by NIH F31AR077423 (KMP), NIH R01CA161608 (AJG), NIH T34GM131947 (DMW), and NIH R21OD019738 (JAP). We also thank Dr. Alice Tarantal, UC Davis, for providing rhesus monkey tissue sections (NIH U01-HL69748, R24-HL085794, and P51-OD011107).

Author Contributions:

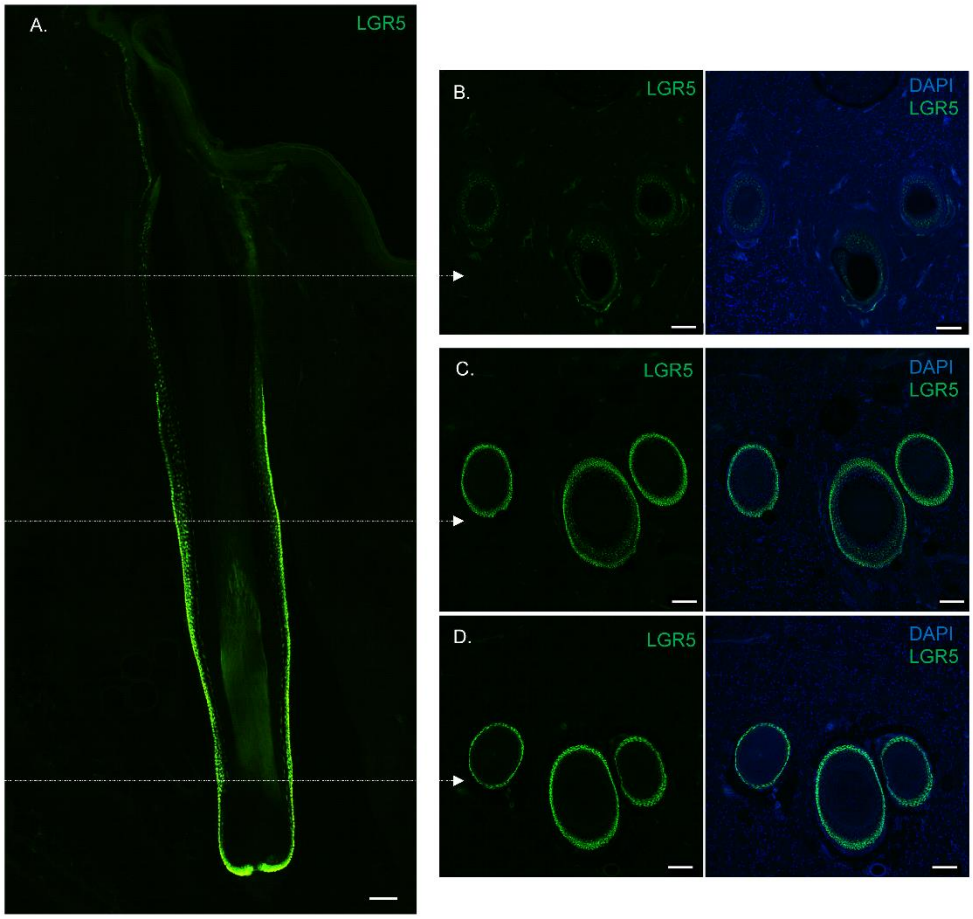
KMP, AJG and JAP designed research, KMP, NKG, AJG, YM, KLG, SGS, DMW, BC and JC performed experiments, KMP, NKG and JAP wrote the manuscript.

Conflict of Interest Statement: The authors have no conflicts of interest to declare.

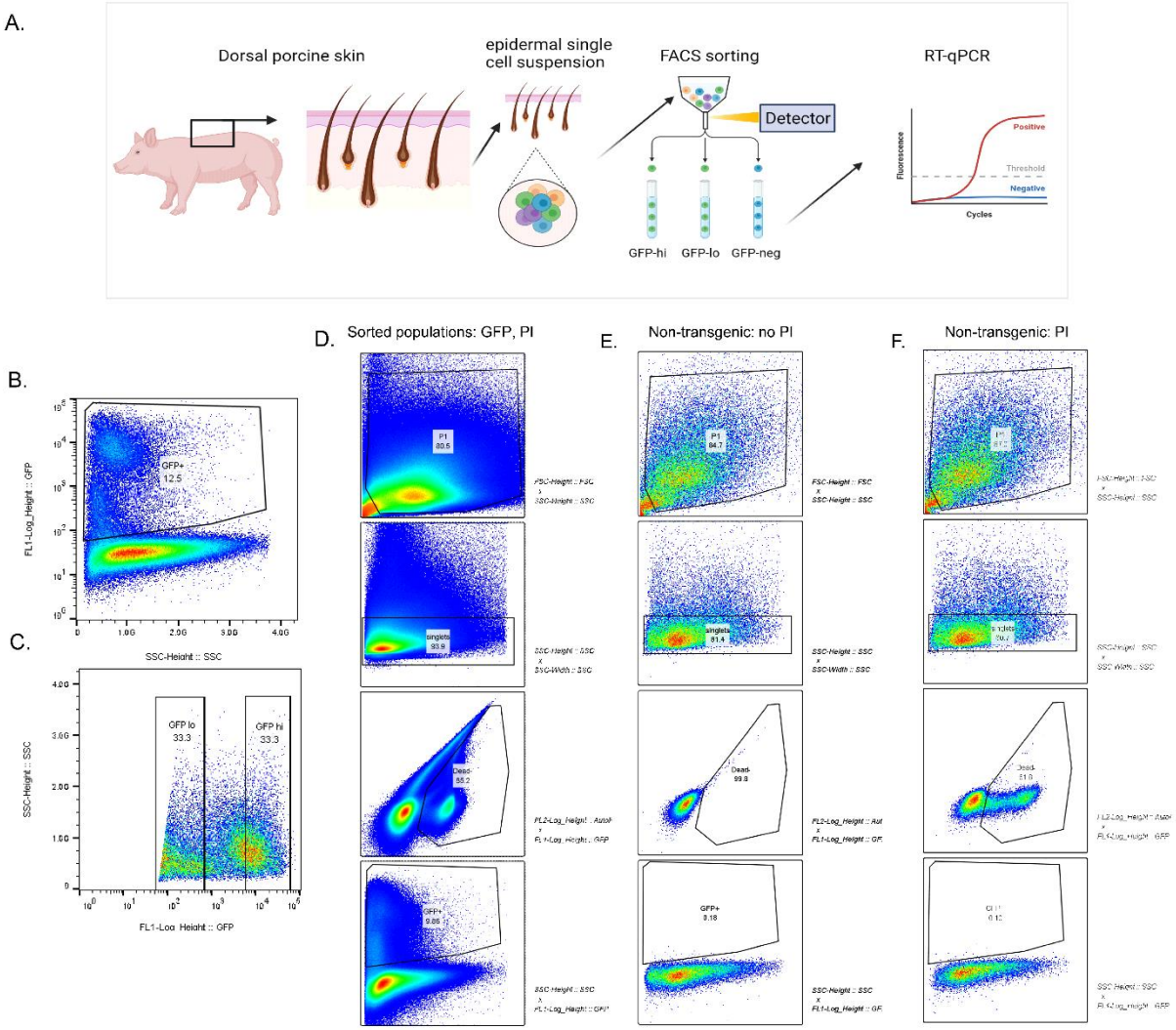
Table S1. Genes and primer sequences used in RT-qPCR analyses

Gene	F (5'-3')	R (5'-3')
ACTB	ACTGCCGCATCCTCTTCCTC	CTCCTGCTTGCTGATCCACATC
GAPDH	ATCCTGGGCTACACTGAGGAC	AAGTGGTCGTTGAGGGCAATG
LGR4	GACCGTCGGGTAGATTGCTC	CCAGCCAATCGTAGCTCCTC
LGR5	CCTTGGCCCTGAACAAAATA	ATTTCTTTCCCAGGGAGTGG
LGR6	CAGGAGGACGGCTTCATGC	GAGCTCCGTGAGGTTGTTCA
CD34	GGTATCTGCCTGGAGCGAAA	GGGTCTTCGCCCAGCCTTT
SOX9	CGGTTCGAGCAAGAATAAGC	GTAATCCGGGTGGTCCTTCT
KRT5	CGACAACGTCAAGAAGCAGT	GAGAGGGTGTTTGTGACGAC
KRT15	GCGAGATGGAGTGCCAGAAC	TCCACTGACTCCTCGACGTT
KRT14	GGAGGTGAAGATCCGCGAC	TCTGCAGCACGACATTAGCG
CD200	TGTTCCAAGTTACTAATCAGGCTGAA	AGCCCATTAGCAACATGATACTCTTT
SHH	CAGTTTATCCCCAACGTGGC	CCACTGGTTCATCACGGAGA
TCF4	TGCCTTAGGGACGGACAAAG	ATAGTTCCTGGACGGGCTTG
WNT3A	GCGACTTCCTCAAGGACAAG	GGTCACGTGTACCGAAGGAT
LRIG1	GACGCGGAGCCTAAACCTAA	CTCCACGCTGCGAATCCTAT
HOPX	GGAGGAGACCCAGAAATGGTT	TCTTGGTGGAAGGAAGCAGC
KI67	GGACCAGGCACAATGGATGG	CAGCTTTTGTCGAAGCGTCC

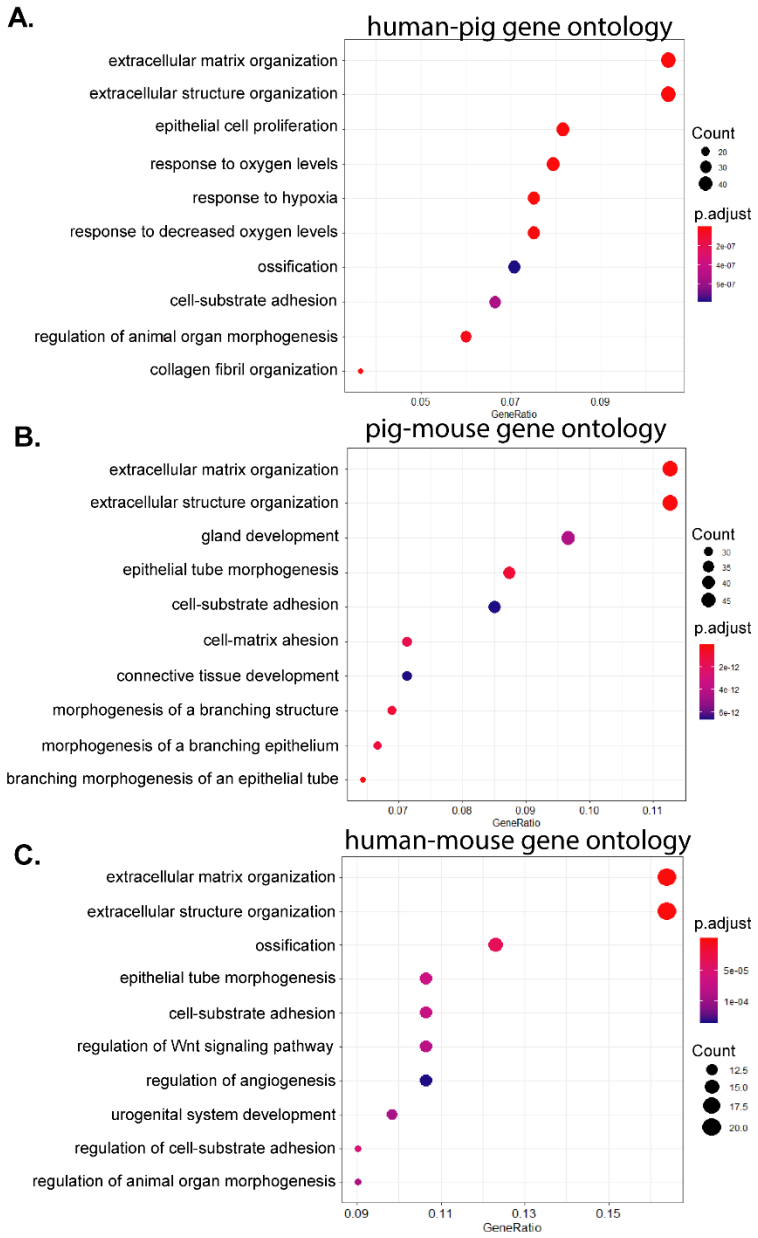
Supplementary figures



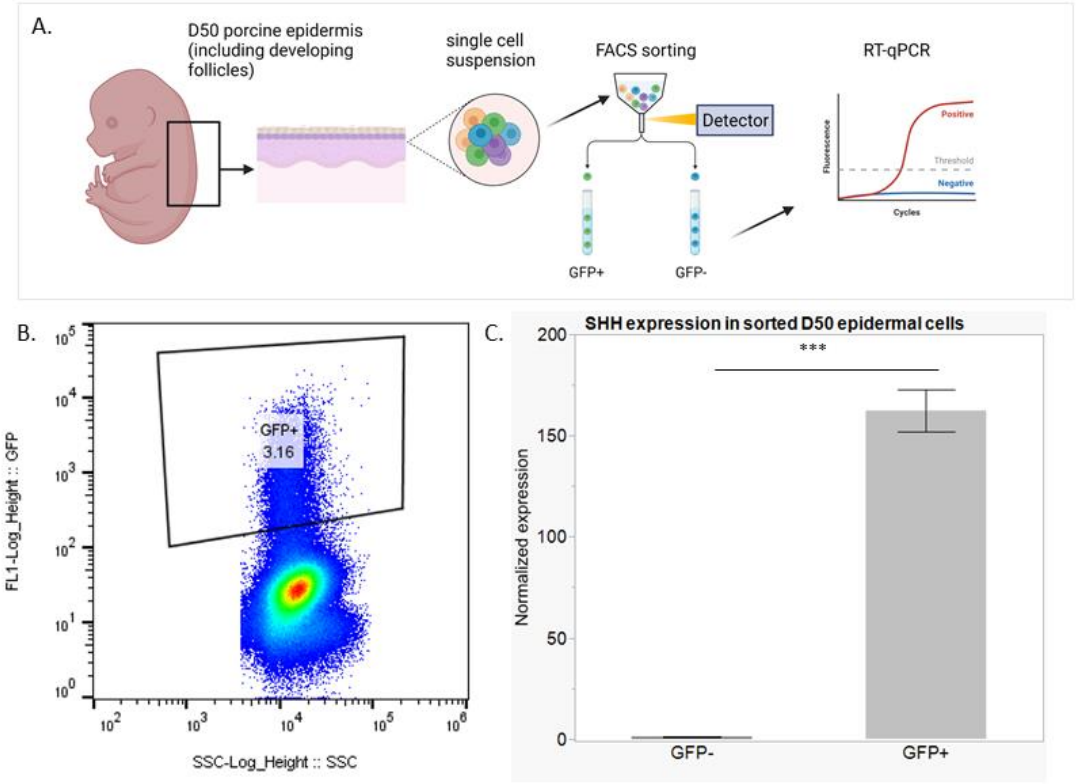
Supplementary Figure 1. Cross section of porcine hair follicle shows that LGR5 is expressed at a high level in the outer root sheath, and low level in the inner root sheath. H-J) Cross sections of hair follicles show distribution of GFP in correspondence with dashed white line. Scale bar represents 200 μM (A-C) or 100 μM (D-J).



Supplementary Figure 2. Representative flow cytometry gating strategy and controls. A) Schematic depicting process of cell isolation and fluorescence activated cell sorting (FACS), created with BioRender. GFP+ cells were split into GFP-high or GFP-low (B-C). Gating strategies determined as follows: D) Transgenic LGR5-H2B-GFP porcine epidermis stained with propidium iodide (PI) for live-dead, E) non-transgenic porcine epidermis with no PI, F) non-transgenic porcine epidermis with PI.



Supplementary Figure 3. Related to figure 4. Shared upregulated gene ontology pathways of upregulated genes in LGR5-high cells, compared pairwise across human, mouse and pig datasets.



Supplementary Figure 4: SHH expression in porcine LGR5+ D50 epidermis. A) Schematic depicting cell isolation, sorting, and RT-qPCR analysis processes, created with BioRender. B) Representative fluorescence activated cell sorting plot representing GFP+ population from D80 fetus. C) RT-qPCR relative expression of SHH of LGR5-GFP+ vs LGR5-GFP- sorted cells. Samples were normalized using a ddCT analysis to GAPDH and ACTB and then to the GFP- sample. Student's t-test *** indicates P=0.02, n=2 pigs.

REFERENCES

1. Di Meglio, P., Perera, G. K. & Nestle, F. O. The multitasking organ: recent insights into skin immune function. *Immunity* **35**, 857–869 (2011).
2. Fuchs, E. & Horsley, V. Ferreting out stem cells from their niches. *Nat. Cell Biol.* **13**, 513 (2011).
3. Jaks, V., Kasper, M. & Toftgård, R. The hair follicle—a stem cell zoo. *Exp. Cell Res.* **316**, 1422–1428 (2010).
4. Summerfield, A., Meurens, F. & Ricklin, M. E. The immunology of the porcine skin and its value as a model for human skin. *Mol. Immunol.* **66**, 14–21 (2015).
5. Sullivan, T. P., Eaglstein, W. H., Davis, S. C. & Mertz, P. The pig as a model for human wound healing. *Wound Repair Regen.* **9**, 66–76 (2001).
6. Joulai Veijouye, S. *et al.* Bulge Region as a Putative Hair Follicle Stem Cells Niche: A Brief Review. *Iran. J. Public Health* **46**, 1167–1175 (2017).
7. Eckert, R. L. *et al.* Biochemistry of epidermal stem cells. *Biochim. Biophys. Acta* **1830**, 2427–2434 (2013).
8. Jaks, V. *et al.* Lgr5 marks cycling, yet long-lived, hair follicle stem cells. *Nat. Genet.* **40**, 1291 (2008).
9. Joost, S. *et al.* Single-cell transcriptomics of traced epidermal and hair follicle stem cells reveals rapid adaptations during wound healing. *Cell Rep.* **25**, 585–597. e7 (2018).
10. Snippert, H. J. *et al.* Lgr6 marks stem cells in the hair follicle that generate all cell lineages of the skin. *Science* **327**, 1385–1389 (2010).
11. Lai, S., Cheng, R., Gao, D., Chen, Y.-G. & Deng, C. LGR5 constitutively activates NF-κB signaling to regulate the growth of intestinal crypts. *FASEB J.* **34**, 15605–15620 (2020).

12. Flores, A. *et al.* Lactate dehydrogenase activity drives hair follicle stem cell activation. *Nat. Cell Biol.* **19**, 1017–1026 (2017).
13. Smith, A. A. *et al.* Activating Hair Follicle Stem Cells via R-spondin2 to Stimulate Hair Growth. *J. Invest. Dermatol.* **136**, 1549–1558 (2016).
14. Hoeck, J. D. *et al.* Stem cell plasticity enables hair regeneration following Lgr5+ cell loss. *Nat. Cell Biol.* **19**, 666 (2017).
15. Mardaryev, A. N. *et al.* Lhx2 differentially regulates Sox9, Tcf4 and Lgr5 in hair follicle stem cells to promote epidermal regeneration after injury. *Development* **138**, 4843–4852 (2011).
16. Wier, E. M. & Garza, L. A. Through the lens of hair follicle neogenesis, a new focus on mechanisms of skin regeneration after wounding. *Semin. Cell Dev. Biol.* **100**, 122–129 (2020).
17. Wang, X. *et al.* Macrophages induce AKT/β-catenin-dependent Lgr5+ stem cell activation and hair follicle regeneration through TNF. *Nat. Commun.* **8**, 14091 (2017).
18. Tan, S. H. *et al.* AQP5 enriches for stem cells and cancer origins in the distal stomach. *Nature* **578**, 437–443 (2020).
19. Barker, N., Tan, S. & Clevers, H. Lgr proteins in epithelial stem cell biology. *Development* **140**, 2484–2494 (2013).
20. Sper, R. B. *et al.* Generation of a stable transgenic swine model expressing a porcine histone 2B-eGFP fusion protein for cell tracking and chromosome dynamics studies. *PLoS One* **12**, e0169242 (2017).
21. Chovatiya, G., Ghuwalewala, S., Walter, L. D., Cosgrove, B. D. & Tumbar, T. High-resolution single-cell transcriptomics reveals heterogeneity of self-renewing hair follicle

stem cells. *Exp. Dermatol.* (2020) doi:10.1111/exd.14262.

22. Purba, T. S. *et al.* Human epithelial hair follicle stem cells and their progeny: current state of knowledge, the widening gap in translational research and future challenges. *Bioessays* **36**, 513–525 (2014).
23. Fuchs, E. Skin stem cells: rising to the surface. *J. Cell Biol.* **180**, 273–284 (2008).
24. Hsu, Y.-C., Pasolli, H. A. & Fuchs, E. Dynamics between stem cells, niche, and progeny in the hair follicle. *Cell* **144**, 92–105 (2011).
25. Yang, H., Adam, R. C., Ge, Y., Hua, Z. L. & Fuchs, E. Epithelial-Mesenchymal Micro-niches Govern Stem Cell Lineage Choices. *Cell* **169**, 483–496.e13 (2017).
26. Kwak, M. & Ghazizadeh, S. Analysis of histone H2BGFP retention in mouse submandibular gland reveals actively dividing stem cell populations. *Stem Cells Dev.* **24**, 565–574 (2014).
27. Coulombe, P. A., Bernot, K. M. & Lee, C.-H. Keratins and the Skin. in *Encyclopedia of Biological Chemistry (Second Edition)* (eds. Lennarz, W. J. & Lane, M. D.) 665–671 (Academic Press, 2013).
28. Langbein, L. *et al.* K25 (K25irs1), K26 (K25irs2), K27 (K25irs3), and K28 (K25irs4) represent the type I inner root sheath keratins of the human hair follicle. *J. Invest. Dermatol.* **126**, 2377–2386 (2006).
29. Ohyama, M. *et al.* Characterization and isolation of stem cell-enriched human hair follicle bulge cells. *J. Clin. Invest.* **116**, 249–260 (2006).
30. Sougrat, R. *et al.* Functional expression of AQP3 in human skin epidermis and reconstructed epidermis. *J. Invest. Dermatol.* **118**, 678–685 (2002).
31. Hara-Chikuma, M. & Verkman, A. S. Roles of aquaporin-3 in the epidermis. *J. Invest. Dermatol.* **128**, 2145–2151 (2008).

32. Jimenez, F., Martínez, M. L., Escario, E. & Izeta, A. The hair follicle as a wound healing promoter and its application in clinical practice. *Wound Healing. Hoboken: John Wiley & Sons, Inc* 149–165 (2018).
33. Takahashi, R. *et al.* Defining Transcriptional Signatures of Human Hair Follicle Cell States. *J. Invest. Dermatol.* **140**, 764–773.e4 (2020).
34. Joost, S. *et al.* Single-Cell Transcriptomics Reveals that Differentiation and Spatial Signatures Shape Epidermal and Hair Follicle Heterogeneity. *Cell Syst* **3**, 221–237.e9 (2016).
35. Ouspenskaia, T., Matos, I., Mertz, A. F., Fiore, V. F. & Fuchs, E. WNT-SHH Antagonism Specifies and Expands Stem Cells prior to Niche Formation. *Cell* **164**, 156–169 (2016).
36. Xu, Z. *et al.* Embryonic attenuated Wnt/β-catenin signaling defines niche location and long-term stem cell fate in hair follicle. *Elife* **4**, e10567 (2015).
37. Cotsarelis, G. Gene expression profiling gets to the root of human hair follicle stem cells. *The Journal of clinical investigation* vol. 116 19–22 (2006).
38. Ohyama, M. *et al.* Characterization and isolation of stem cell–enriched human hair follicle bulge cells. *J. Clin. Invest.* **116**, 249–260 (2006).
39. Lin, S. *et al.* Comparison of the transcriptional landscapes between human and mouse tissues. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 17224–17229 (2014).
40. Webb, A., Li, A. & Kaur, P. Location and phenotype of human adult keratinocyte stem cells of the skin. *Differentiation* **72**, 387–395 (2004).
41. Yang, R. *et al.* Generation of folliculogenic human epithelial stem cells from induced pluripotent stem cells. *Nat. Commun.* **5**, 3071 (2014).
42. Garza, L. A. *et al.* Bald scalp in men with androgenetic alopecia retains hair follicle stem

cells but lacks CD200-rich and CD34-positive hair follicle progenitor cells. *J. Clin. Invest.* **121**, 613–622 (2011).

43. Carmon, K. S. *et al.* LGR5 receptor promotes cell–cell adhesion in stem cells and colon cancer cells via the IQGAP1–Rac1 pathway. *Journal of Biological Chemistry* vol. 292 14989–15001 (2017).
44. Leung, C., Tan, S. H. & Barker, N. Recent advances in lgr5+ stem cell research. *Trends Cell Biol.* **28**, 380–391 (2018).
45. Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M. & Altman, D. G. Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. *J. Pharmacol. Pharmacother.* **1**, 94–99 (2010).
46. Walker, S. C. *et al.* A highly efficient method for porcine cloning by nuclear transfer using in vitro-matured oocytes. *Cloning Stem Cells* **4**, 105–112 (2002).

Chapter 3

LGR5+ hair follicle stem cells exhibit self-renewal and multi-lineage differentiation in organoid culture

Abstract

The plasticity of epidermal stem cells has long been known to enable fate-switching between niches, suggesting that the stem cells throughout the skin are interchangeable depending on their microenvironment. Furthermore, culturing cells in 3D as organoids has emerged as an alternative to 2D culture, allowing a recapitulation of *in vivo* architecture. Using a transgenic porcine model, we have established culture conditions in which LGR5+ hair follicle stem cells form two types of epidermal organoids from a single cell. When cultured in expansion conditions, organoids maintain LGR5 expression, expand rapidly, and ultimately express hair follicle genes. When cultured in differentiation conditions, they switch to an interfollicular epidermis-like fate and the structure contains all layers of epidermis, from basal cells to stratum corneum. These results provide an *in vitro* model supporting results from *in vivo* experiments in which various stem cell populations within the epidermis exhibit dynamic fates depending on their microenvironment.

Introduction

The skin, the largest organ in the body, requires a high degree of stem cell and progenitor coordination for a rapid turnover of tissues. Within the skin, several compartments of stem cells have been noted, including those in the interfollicular epidermis (IFE), isthmus of the hair follicle, bulge region of the hair follicle, and sebaceous gland (Dekoninck and Blanpain 2019; Hsu, Pasolli, and Fuchs 2011; Yang et al. 2017). LGR5 has emerged as a marker of multiple epithelial stem cell populations throughout the body, including intestine, colon, mammary gland, stomach, and regenerating liver (Barker, Tan, and Clevers 2013; Guiu et al. 2019; Huch et al. 2013). In the skin, LGR5 is a marker of epidermal stem cells residing in the lower bulge of the hair follicle (Jaks et al. 2008; Joost et al. 2018), and while there is much variation in stem cell markers across species, LGR5 remains a constant marker of the hair follicle, as we have shown previously. While this LGR5+ hair follicle stem cell (HFSC) population resides exclusively in the hair follicle and governs hair follicle growth and homeostasis, upon a wound in the vicinity, these cells contribute to re-epithelialization repair (Joost et al. 2018; Kang et al. 2020; Page et al. 2013). This research has proven that stem population cells throughout the skin share a common set of genes, and further research suggests that they display dynamic stem cell fate between subpopulations within the epidermis, especially upon wounding. Furthermore, wounding has been shown to induce a "wound state" epigenetic signature, which is also induced by growing these cells in *in vitro* conditions (Ge et al. 2017).

Recently, organoids from single cells have emerged as a valuable tool to develop a "mini-niche" from single stem cell populations. To better understand the cell plasticity of HFSC, we asked if we could study the fate decisions of LGR5+ HFSC *in vitro* using a 3D environment to produce organoids. While much of the research in the skin has been performed using mice models, mouse

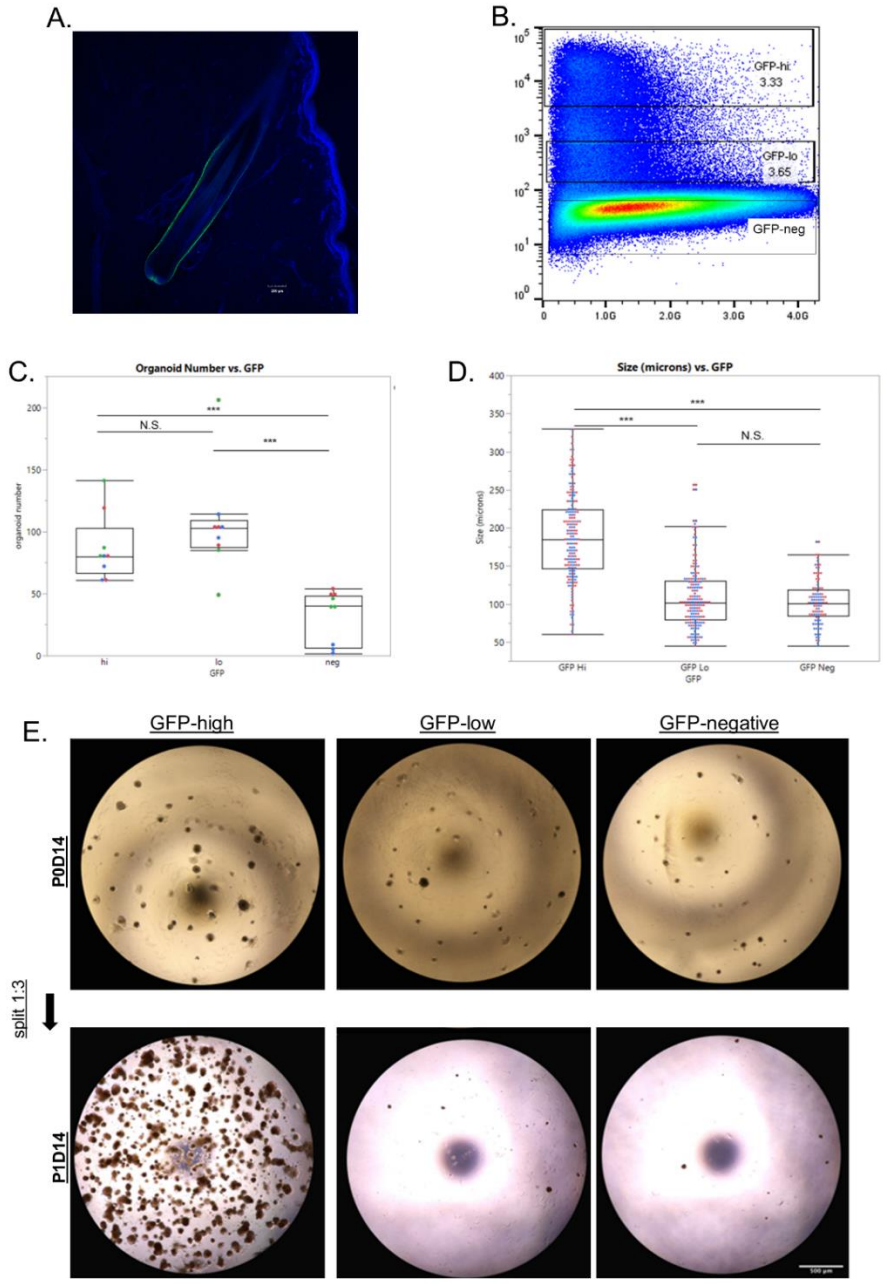


Figure 1. LGR5^{HI} epidermis forms more organoids at a higher efficiency than LGR5LO or LGR5NEG cells. A) Confocal image of cross section of porcine hair follicle showing distribution of LGR5-GFP throughout the outer-root-sheath of the hair follicle. B) Fluorescent activated cell sorting representative gating strategy based on GFP intensity (y-axis) for GFP-HI, -LO, and –NEG cells. C) Organoids formed per 5,000 cells seeded/well based on GFP fluorescent intensity. n=3 biological replicates, each with technical triplicates. D) Size (organoids from C) measured at D14. E) Brightfield images of whole Matrigel patty of organoids at day 14, or after being passaged and split 1:3 before growing to day 14.

skin has notable differences from human skin: thin epidermis and dermis, loose skin attachment, fur instead of sparse hair, and has been shown to spontaneously regenerate hair follicles within wounds--none of which apply to humans (Summerfield, Meurens, and Ricklin 2015; Di Meglio, Perera, and Nestle 2011). With this in mind, we elected to use pigs as a more physiologically relevant model to understand stem cell dynamics in the skin. To overcome the limitations of antibodies detecting LGR5, we used a transgenic pig expressing the marker H2B-GFP under the control of the endogenous LGR5 promoter to enable us to track and preferentially sort the hair follicle stem cell populations, as previously discussed in chapter 2.

In this research, we have developed methods to culture LGR5+ hair follicle stem cells as epidermal organoids from single cells. We show that these organoids can be expanded and maintain LGR5 expression, and depending on the growth conditions they can either differentiate into hair follicle organoids or switch to an interfollicular epidermis fate. This system is useful for studying the fate of primary HFSCs *in vitro* and can be used further for mechanistic studies of cell fate dynamics in healthy and diseased states.

Results

Defining and optimizing culture conditions for HFSC derived organoids

To focus on the hair follicle stem cells, we used skin from transgenic pigs expressing H2B-GFP under the control of the endogenous *LGR5* promoter, which is a conserved marker of HFSC across species (Figure 1A). Single cell epidermal suspensions were fluorescence activated cell sorted based on GFP-high ($LGR5^{HI}$), GFP-low ($LGR5^{LO}$), or GFP-negative ($LGR5^{NEG}$) expression (Figure 1B). When designing these experiments, there were many protocols for growing epithelial organoids, but very few focused on the skin. Therefore, we based our media

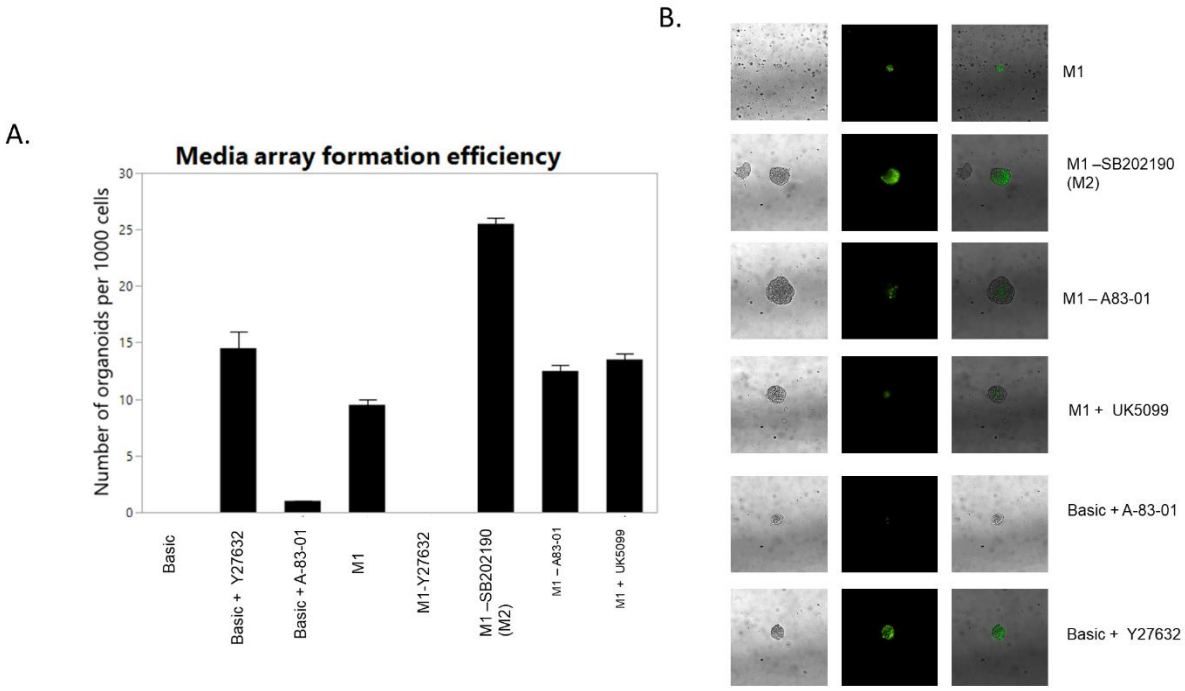


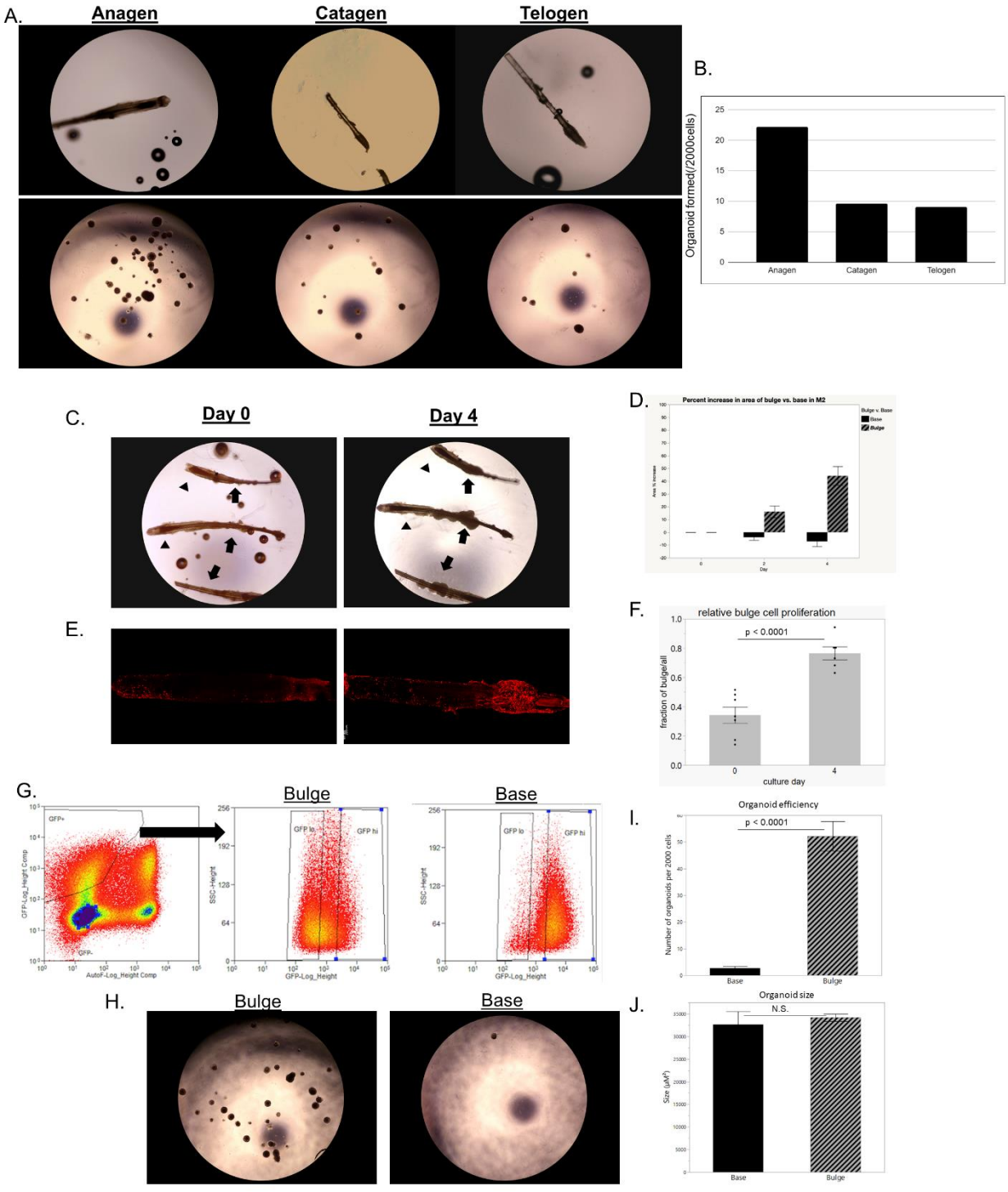
Figure 2. Growth factor array for organoid media formulation. A) Media formation efficiency or B) microscope images based on different growth factor conditions. All organoids were cultured in DMEM/F12 mixed with LWRN conditioned media and supplements as listed in Supplementary table 1. All media conditions are based off of M1 (Supplementary table 2) with or without the factors listed. SB202190 is a p38MAPK inhibitor, A83-01 is a TGF-beta inhibitor, UK5099 is a driver of lactate dehydrogenase activity, and Y27632 is a rho-kinase inhibitor.

formulation on a well-established porcine LGR5+ intestinal organoid media (Gonzalez et al. 2013), hypothesizing that the two epithelial stem cell populations of the intestine and the skin would share similar factors in the niche. To optimize conditions, we subjected LGR5-high expressing cells to a media array (Figure 1) in which we measured formation efficiency and presence of GFP. Once we determined the optimal conditions, termed HF growth media [M2], we seeded LGR5-hi, lo, and neg cells at 2000 cells/ well in a matrigel dome . LGR5-high and LGR5-low cells formed at a significantly higher efficiency than LGR5-negative, and $LGR5^{HI}$ derived organoids were significantly larger than low or negative organoids (Figure 1C, D). To further analyze how each population differed, we next asked if cells in the organoids remained potential for self renewal after passage. When organoids were dissociated and split 1:3, LGR5-high derived organoids showed a robust expansion of organoids, whereas there was almost no organoid formation in $LGR5^{LO}$ or $LGR5^{NEG}$ wells (Figure 1E). $LGR5^{HI}$ organoids were fast growing and maintain GFP expression before and after passage. From this we can conclude that in these growth conditions, the $LGR5^{HI}$ derived organoids undergo robust expansion and self-renewal, whereas $LGR5^{LO}$ and $LGR5^{NEG}$ do not.

Bulge-derived $LGR5^{HI}$ cells from anagen follicles drive organoid generation

While LGR5 is expressed consistently throughout the lower bulge of the hair follicle, we wanted to better understand which specific cells from the hair follicle were most efficient at organoid formation. We first asked how stage of hair cycle would impact organoid formation, hypothesizing that the actively expanding anagen $LGR5^{HI}$ cells would be more likely to form organoids than those from the quiescent catagen or telogen stages. To study this, we enzymatically isolated whole follicles from porcine skin and separated follicles based on stage of the hair cycle (Figure 3A).

Figure 3. Anagen bulge LGR5+ cells are the primary drivers of organoid formation. A) Whole follicles were enzymatically and manually isolated from the dermis and sorted based on stage of the hair cycle. Cells from each stage follicle were dissociated, sorted for HI LGR5-GFP expression, and plated at 2000 cells/ well. B) Quantification of organoid formation efficiency shows that cells from anagen follicles form organoids more efficiency than catagen or telogen, n=2 biological replicates each with technical triplicates. C) Whole anagen follicles were embedded in Matrigel for four days with the same growth factor conditions as organoids (M2). Arrows indicate identical locations in the bulge region for each follicle, area of which expands over 4 days. D) Quantification of bulge size by the percent increase in area over 4 days in culture, n=3 biological replicates with 10 total technical replicates. E) KI67 staining of D0 follicles or after 4 days in culture. F) Quantification of E represented as proportion: the number of proliferating cells (KI67+) in the bulge region divided by all the proliferating cells by day in culture. n=3 biological replicates and technical duplicates for each. G) Hair follicles were manually bisected at the mid-point between the bulge and the base of anagen follicles using a 22 gauge needle, and cells were enzymatically dissociated from either the bulge or base. Plots show FACS gating strategy for the groups based on GFP fluorescent intensity. All GFP positive, live cells were isolated and the top 50% of all GFP cells (from the first panel) was used to set the gates used to isolate GFP-hi from the bulge and the base populations; the base population had average higher fluorescent intensity. H) Cells from each group were plated at 2000 cells/well and efficiency (I) and size (J) were quantified. N=4 biological replicates and 3-8 technical replicates each. Error bars are indicative of SEM. Significance was determined by two-tailed Student's t-test for each comparison.


A.
Anagen
Catagen
Telogen
B.
Organoid formed(/2000cells)
25
20
15
10
5
0
Anagen
Catagen
Telogen
C.
Day 0
Day 4
D.
Percent increase in area of bulge vs. base in M2
Bulge v. Base
Base
Bulge
Day
E.
F.
relative bulge cell proliferation
p < 0.0001
fraction of bulge/all
1.0
0.8
0.6
0.4
0.2
0.0
0
4
culture day
G.
Bulge
Base
GFP+
GFP-
GFP-Log_Height Comp
AutoF-Log_Height Comp
SSC-Height
GFP lo
GFP hi
GFP-Log_Height
I.
Organoid efficiency
p < 0.0001
Number of organoids per 2000 cells
Base
Bulge
H.
Bulge
Base
J.
Organoid size
N.S.
Size (uM²)
Base
Bulge

Next, we isolated single cells from each stage of the cycle, sorted based on GFP intensity as in Figure 1, and seeded the cells from each stage in matrigel with HF growth media. As predicted, anagen-derived LGR5^{HI} cells form organoids at a higher efficiency than catagen or telogen (Figure 3A, B). Since the anagen stage hair follicle occurs when the dermal papillae and base of the follicle stretch deep into the dermis, further away from the bulge, we next asked whether there were differences in growth of cells in the bulge region vs the base region ex vivo. First, we embedded whole anagen stage follicles in matrigel and cultured them for 4 days (Figure 3C), where the bulge region increased in percent of total area over time (Figure 3D). To further quantify this, we stained the follicle explants with KI67 and compared the ratio of proliferating cells in the bulge vs the base at D0 vs D4, in which case the proportion of proliferating cells in the bulge significantly increased (Figure 3E,F). To ask if this is true of specifically the LGR5^{HI} cells, we next isolated single cells from the bulge or base epidermis. After setting a gate for the top 50% of GFP+ cells using all epidermis, we FACS sorted the bulge or the base cells for GFP^{HI} (Figure 3G). Interestingly, although on average, the HF base was brighter in GFP, LGR5^{HI} cells from the bulge region formed significantly more organoids than the base (Figure 3H-I). Although we showed in Figure 1 that LGR5^{HI} expression marks cells with best organoid forming potential, this suggests that after a certain level, higher intensity is not indicative of greater stemness. Overall, we can conclude that the LGR5^{HI} cells from the bulge region, not the base region, of the anagen follicle are the primary drivers of organoid formation.

Generation of hair follicle organoids.

After determining the source of the organoid forming cells, we next sought to better understand the properties of the organoids themselves. When examined at D7, most of the organoids were

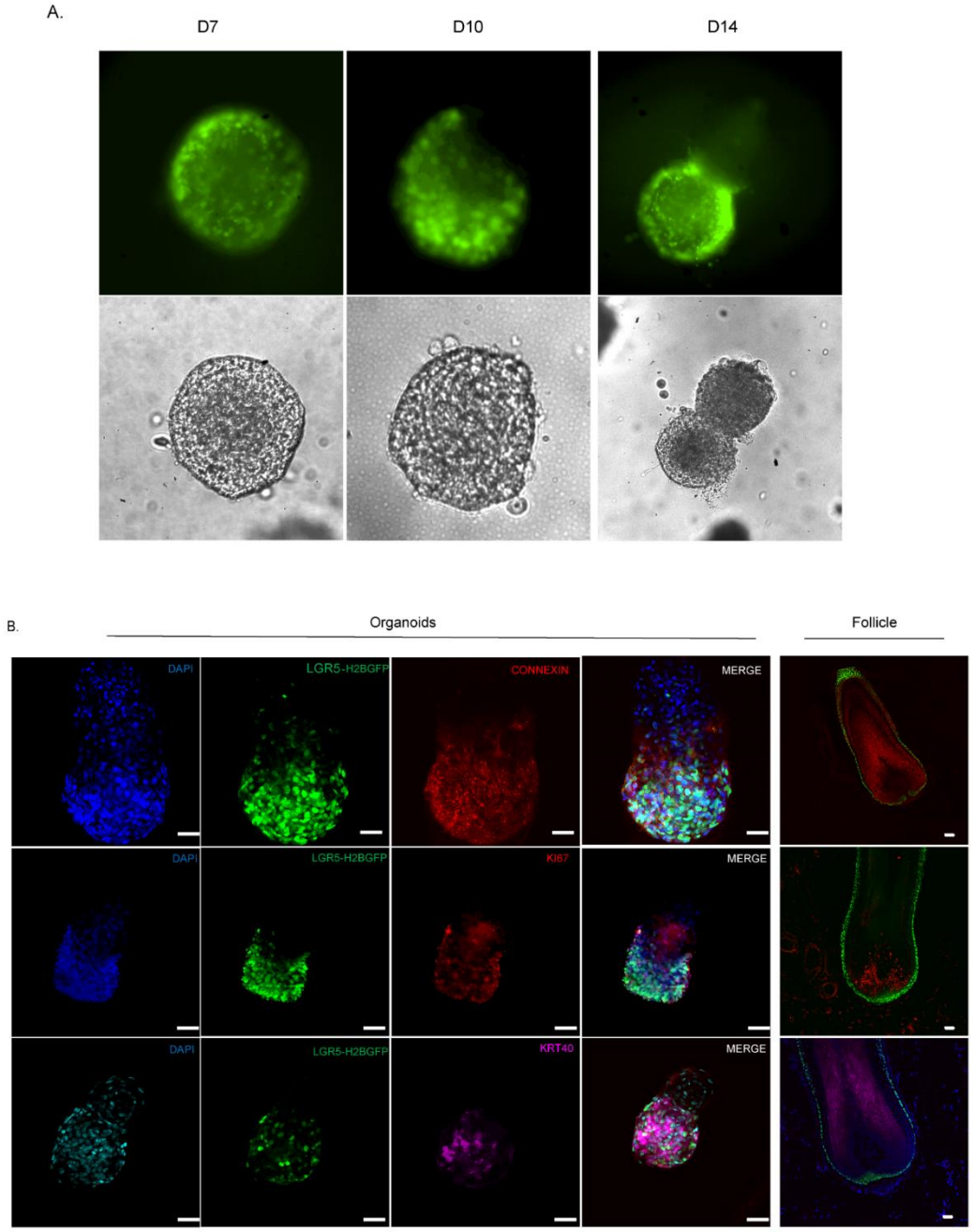


Figure 4. Generation of hair follicle organoids. A) Brightfield and GFP-fluorescent microscopy of organoids grown in M2 over 2 weeks of culture. Most organoids maintain bright GFP expression and undergo a round, elongated, and budding stage. B) Immunofluorescent detection of Connexin, KI67, and KRT40 in organoids or follicles shows similar pattern of expression in organoids and follicles. Auto-fluorescent properties of ECM can be detected in GFP channels (A) and red channels (B, KI67). Scale bars indicate 40 microns.

symmetric and spherical, with bright GFP expression throughout. We noticed that after 7-10 days of culture, many of the organoids became elongated, with asymmetric distribution of GFP to one side, and between day 10-14, they began to bud, while maintaining GFP expression on the opposite side of the bud (Figure 4A). This was reminiscent of the base of the hair follicle, where the LGR5+ cells proliferate upward to differentiate into the inner root sheath and hair shaft, while the cells in the outer-root sheath, in contact with the basement membrane (or matrigel *in vitro*), remain GFP+ (as shown in Chapter 2). To further understand the dynamics of organoid growth, we used whole mount organoids with immunofluorescent staining and compared patterns to that of the anagen hair follicle (Figure 4B). Immunofluorescent staining shows detection of similar patterns of connexin, KI67, and hair keratin KRT40 as found in the porcine hair follicle. It is also possible to see that in some of the more developed organoids, the extracellular matrix components give off an autofluorescent quality, which is a characteristic shared with the hair shaft in skin sections (4A, 4B-KI67). The pattern of expression of these markers suggests that the LGR5+ cells on the polarized end undergo a transient amplification while giving rise to cells that express markers of hair.

Despite multiple attempts, these hair follicle organoids were not capable of regenerating hair when grafted into nude mice. We hypothesize that this is because the organoids need consistent signaling from dermal papillae, which we have recapitulated with growth factors, but when we remove them from culture to graft them, they are not able to re-establish the correct signaling in time to continue the hair growth.

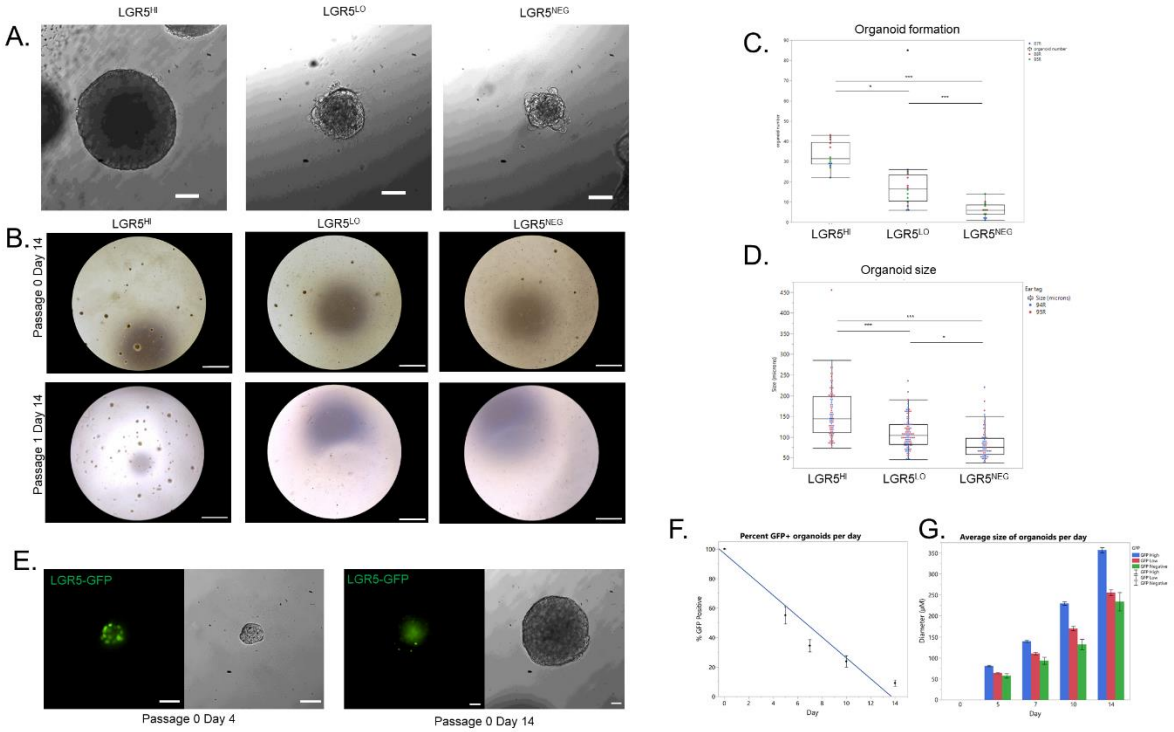


Figure 5. Organoid derived from LGR5-high hair follicle stem cells produce organoids at higher efficiency and size than those from LGR5-low and LGR5-negative populations in M1 conditions. A) Brightfield microscopy of LGR5HI, LO and NEG shows differences in morphology of organoids. B) LGR5HI cell derived organoids but not LGR5LO and LGR5NEG maintain the ability to reform after passage. C) LGR5HI cell derived organoids have significantly higher formation efficiency than LGR5LO and LGR5NEG organoids and (D) were significantly larger in diameter. E) Brightfield images of P0D4 and P0D14 organoids and (F) quantification of percent of organoids that maintain GFP+ cells by day in culture. G) Quantification of growth over time by cell type forming the organoids. Where appropriate bars represent mean and standard error, statistical significance shown by ANOVA and Tukey-Kramer HSD Comparison of Means *indicates P<0.05, *** indicates P<0.005. Scale bars are 1mm (B) and 50µM (E).

Generation of Interfollicular Epidermal organoids

While our initial media array had uncovered conditions that enabled growth and expansion of LGR5+ cells (Figure 2), we also noticed that cells grown in condition "M1" formed large spherical organoids over time in culture. To test the organoid growth in these altered conditions based on LGR5 expression, we enzymatically isolated the epidermis from porcine dorsal skin and sorted based on LGR5-H2B-GFP expression as in Figure 1B. As seen from the brightfield images in Figure 5A, LGR5-high organoids showed a large, smooth, round morphology when compared to LGR5-low or negative organoids which show minimal growth and frequent apoptosis. We next quantified the formation efficiency of $LGR5^{HI}$, $LGR5^{LO}$, or $LGR5^{NEG}$ organoids, and found that $LGR5^{HI}$ formed significantly more organoids and larger organoids than $LGR5^{LO}$, or $LGR5^{NEG}$, but interestingly that $LGR5^{LOG}$ formed significantly more and larger organoids than $LGR5^{NEG}$ (Figure 5B-D). Next, organoids from each group were dissociated into single cells and re-plated, and organoids from $LGR5^{HI}$ wells displayed a robust capability to re-establish organoids, whereas LGR5-low and negative derived populations did not (Figure 5B). Interestingly, while the LGR5-GFP expression was maintained initially in culture from the $LGR5^{HI}$ sorted population, expression gradually decreased over time (Figure 5E-F) and by day 14 in culture it was gone entirely from almost all organoids, suggesting that these cells are differentiating or exhibiting a switch of stem cell fates. Quantification of organoid size over time by GFP intensity reveals that $LGR5^{LO}$, or $LGR5^{NEG}$ cells have a slower growth trajectory (Figure 5G), and further analysis of $LGR5^{LO}$ and $LGR5^{NEG}$ organoids shows limited survival in prolonged culture (Figure 6), so for the remainder of the study we focused on $LGR5^{HI}$ derived organoids only.

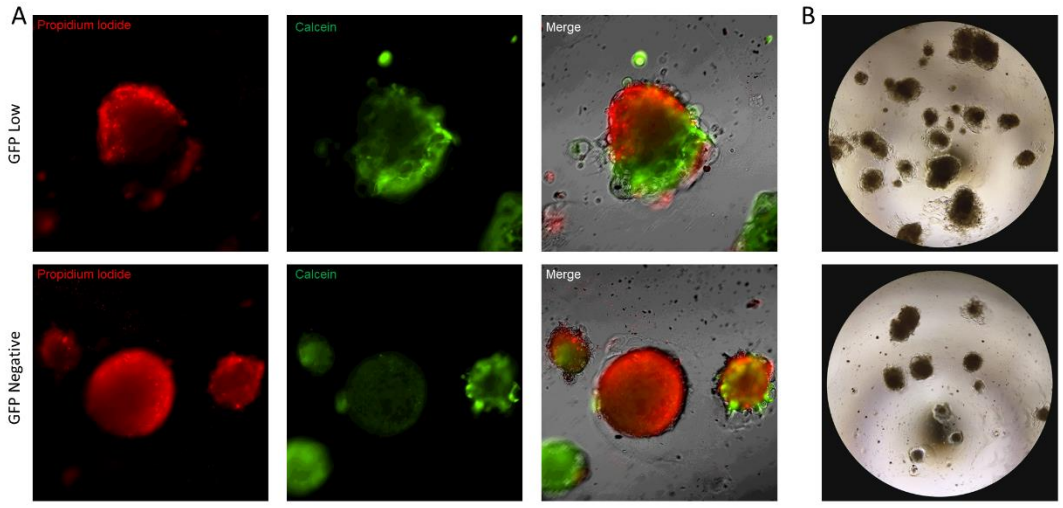


Figure 6. LGR5^{LO} and LGR5^{NEG} derived organoids demonstrate limited viability in long term culture. A) GFP-low and GFP-negative organoids stained with PI (dead), Calcein (live), with merge. B) Brightfield images of P0D42 GFP-low and GFP-negative organoids.

LGR5-hi organoids form all layers of epidermis.

Since the LGR5-hi sorted organoids displayed maintenance of stemness, but lost expression of the marker of the HFSC, LGR5, we hypothesized that they had undergone a fate-switch into an interfollicular epidermis-like fate. To evaluate this, we used immunofluorescence on sectioned organoids to compare markers and structure of the organoids with that of adult porcine epidermis. After 30 days of culture, epithelial organoids show a striking resemblance to postnatal epidermis, including similar expression and distribution of KRT85, actin, and CD200, which mark all layers of the cellular epidermis, and KRT14 a maker of the basal layer (Figure 7A). KRT10 expression in organoids was present but weak and inconsistent, suggesting that there still is some discrepancy between the organoids and the skin, potentially due to the organoids constantly being submerged in media

and not exposed to air. Immunohistochemistry and hematoxylin and eosin staining of the organoids and epidermis show similar distribution of layers, with a dark and cellular basal layer thick on the bottom, and the top consisting of an acellular stratum corneum in which cells have been completely cornified and stain in bright pink layers (Figure 7B,C). The innermost layer of cells in the organoids was cornified and stained positive for corneodesmin (CDSN), a marker exclusively staining the inner stratum corneum in the epidermis (Figure 7D). From this we can conclude that these organoids are indeed epidermal in nature, and exhibit all layers of the epidermis.

At day 14 of culture, early LGR5-hi derived organoids displayed a relatively homogenous expression of basal and epidermal markers (Figure 8A), expressing markers such as KRT14, CD200, and KRT85 in a bright but diffuse expression. By D30 and D42 of culture, organoids become more organized and expression of each marker is restricted to its respective region (Figure 7A, 8B). Finally, we aimed to support our immunofluorescent observations by using RT-qPCR to analyze the gene expression profile of the developing organoids, especially during the transition from hair follicle stem cells to an interfollicular epidermis state. Our results mirror the results from figures 7 and 8, in that markers of the stratum corneum such as CDSN, Involucrin (IVL) and S100A10 (Sehgal 2018) tended to increase in organoid culture as compared with LGR5. Furthermore, we confirmed that these organoids downregulate common markers of the hair follicle stem cells such as SOX9, LGR5, and LGR6. From this, we conclude that the $LGR5^{HI}$ hair follicle stem cells shift to an interfollicular epidermis state under these culture conditions and are able to generate organoids which recapitulate the architecture and gene expression of the adult porcine epidermis.

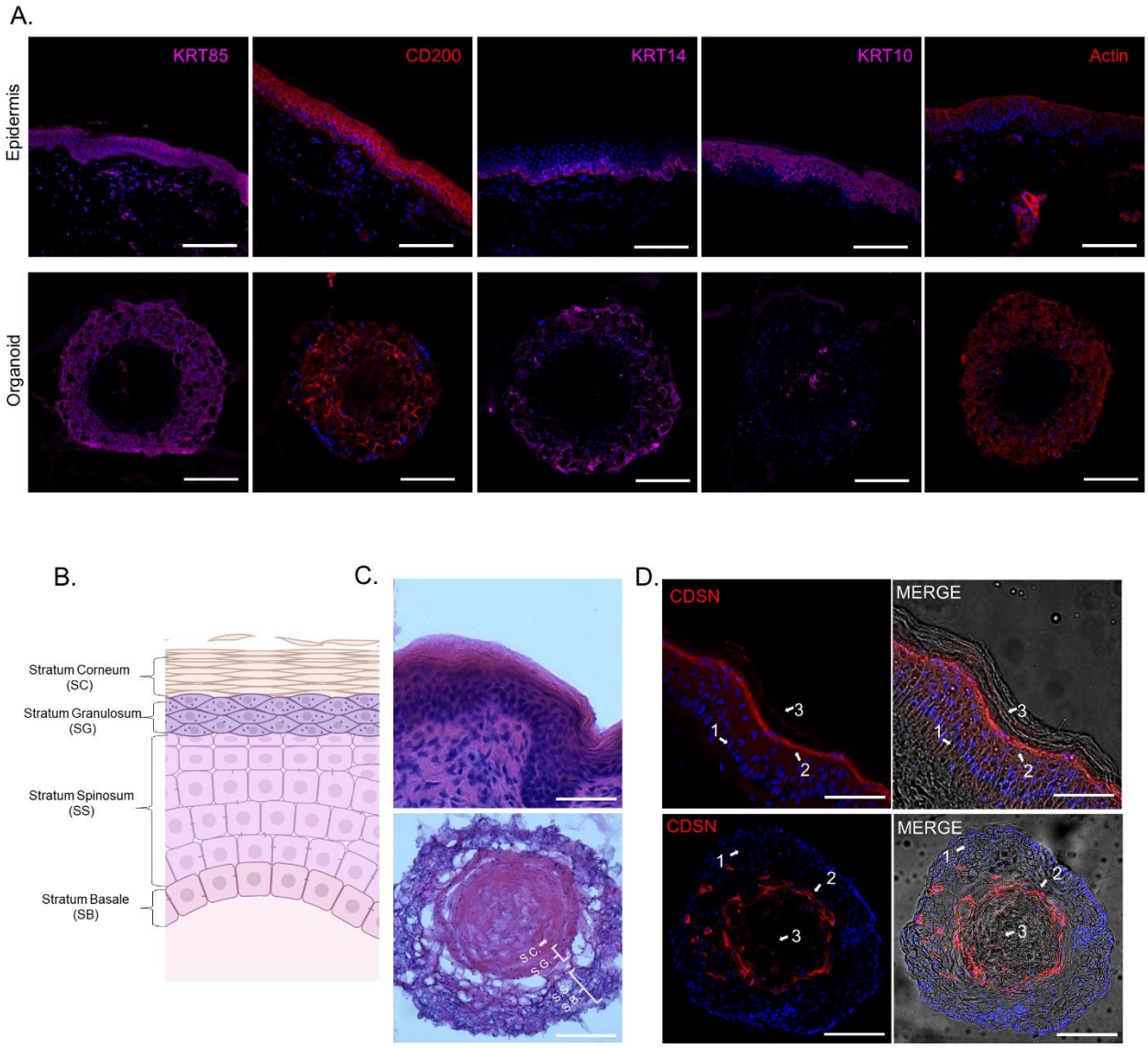


Figure 7. Fully differentiated skin organoids have same markers of epidermis. A) Epidermis and P0D30 GFP-high organoids stained with KRT85, CD200, KRT14, KRT10 and Actin (from left to right). B) Representative graphic of layers of the epidermis (figure made with BioRender), abbreviation which are mapped onto organoids in (C), hematoxylin and eosin staining of skin. D) CDSN is a marker of the stratified corneum in the epidermis, recapitulated in P0D30 LGR5HI organoids. Scale bars are 100µM.

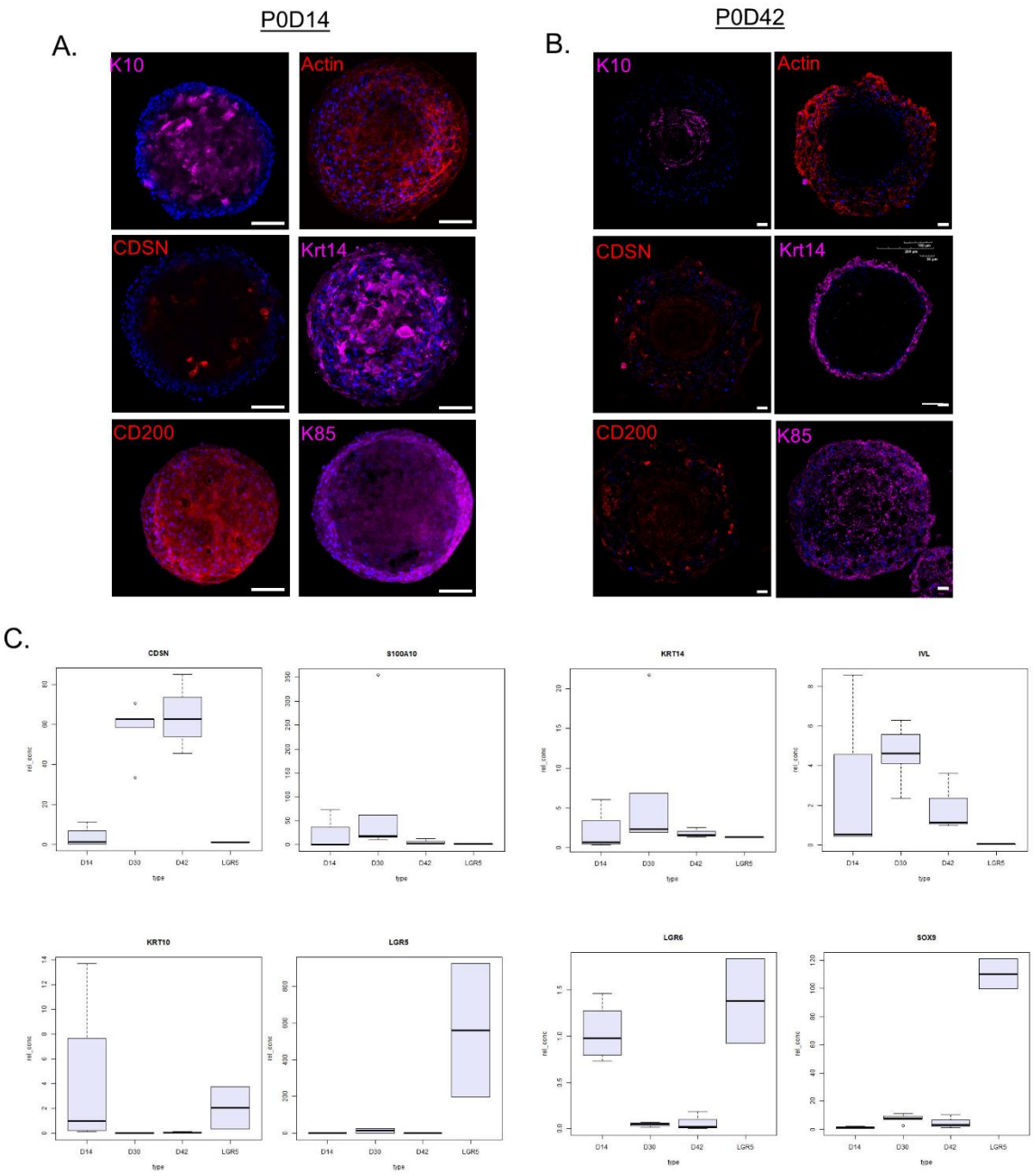


Figure 8. Gene expression of organoids over time. A) P0D14 or B) P0D42 sectioned organoids stained with same markers as figure 7 show how expression of each marker changes and becomes more restricted over time. C) Gene expression analysis of select skin and stem cell markers. based on time in culture.

While we are most interested in the translational applications based on the similarities between porcine and human skin, we did feel that it was critical to evaluate our results in the context of previous research conducted in murine tissues. As such, we used LGR5-IRES-GFP mice to evaluate the potential of mouse HFSC to maintain stemness and generate organoids resembling the IFE. Because the mouse marker is cytoplasmic (Figure 9A), we were not able to easily separate based on high or low GFP expression, and therefore just selected the LGR5-GFP positive or negative for the remaining experiments (Figure 9B). In order to be as thorough as possible, we tested our own media both as is, and supplemented with Forskolin, a molecule shown to be critical in the formation of organoids derived from mice (Boonekamp et al. 2019). As shown, the LGR5+ cells formed organoids at lower efficiency compared with negative (Figure 9C), and both conditions generated organoids at a much lower efficiency than what was observed in the pig. Furthermore, in brightfield images, the mouse organoids displayed a morphology similar to that of the pig GFP-negative organoids, with apparent cell death and lack of extracellular matrix stratification (Figure 9D). Using a live dead stain (Figure 9E), we confirmed that the morphology indicated unhealthy organoids. Thus, we concluded that either our culture methods are not compatible with murine LGR5+ hair follicle stem cells.

Discussion

Our results suggest that the organoids from $LGR5^{HI}$ stem cells retain higher levels of potency compared with $LGR5^{LO}$ or $LGR5^{NEG}$ sorted cells, which may have already differentiated terminally and do not regain stem cell properties upon *in vitro* culture. In this study, we have established a 3D culture system using multiple inhibitory small molecules and growth factors, expanding the LGR5+ hair follicle stem cells in a hair follicle forming fate, or switching to that

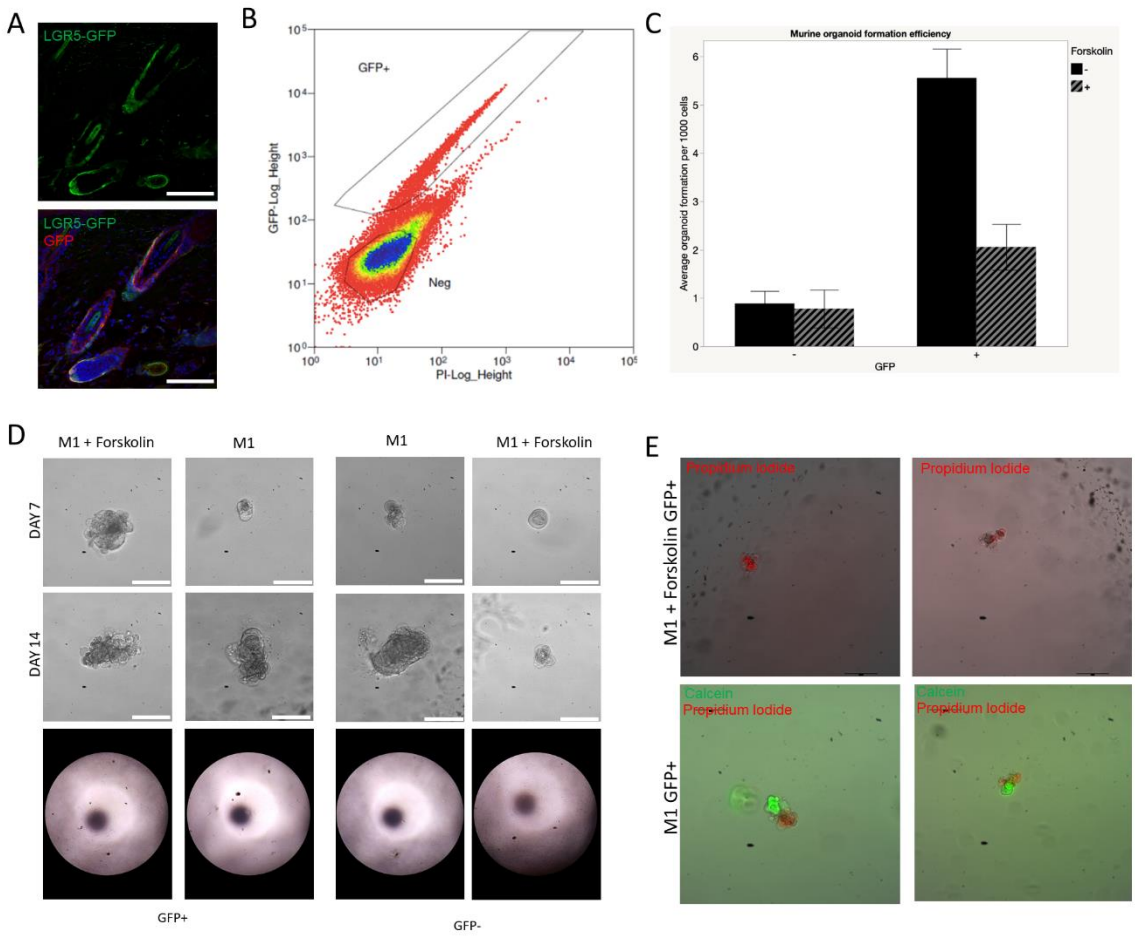


Figure 9. Mouse organoids do not form under same conditions. A) Cross section of LGR5-IRES-GFP mouse hair follicle (B) and gating scheme for isolation of GFP+ and GFP- cells. C) Formation efficiency of organoids with and without Forskolin and D) brightfield images at P0D14. Scale bars indicate 100µM. E) Live/dead staining with calcein and propidium iodide reflects poor health of mouse organoids.

of the interfollicular epidermal fate. Additionally, we demonstrate that while a subpopulation of the organoid maintains self renewal potential after passage, the organoids are able to generate the layers of the epidermis, recapitulating the architecture of adult skin. Therefore, these epidermal organoids could be used to study in depth the plastic nature of the LGR5+ stem cells and their ability to contribute to both the hair as well as the epidermis.

To our knowledge, this is the first report of LGR5+ hair follicle stem cells forming organoids with all layers of the interfollicular epidermis. While Clevers and colleagues were able to develop a method for long term expansion of epidermal stem cells as organoids, their research focused instead on the stem cell population marked by Sca-1 in the isthmus (Boonekamp et al. 2019). These results are among the first to prove the plastic fate of the stem cells within the epidermis *in vitro*. Our results show significantly increased size and formation of LGR5+ organoids as compared with LGR5-GFP-low or -negative. While the increased efficiency could potentially be accounted for by the large number of terminally differentiated keratinocytes in the LGR5-negative population, it cannot explain the lack of organization and poor growth dynamics after passage of those organoids that did form. One important aspect of our methods was that we cultured our organoids in low oxygen conditions (5%), and healthy organoids would not form in atmospheric oxygen conditions. More work must be done to understand the role of oxygen in epidermal cell maintenance and organoid differentiation, or wound repair.

In wound repair, the stress signaling from the wound leads to RAS/ERK signaling that initiates fate switching by activating “wound epicenters”--regulatory regions in chromatin domains that increase lineage plasticity in a wounding response on account of super-enhancer dynamics (Ge et

al. 2017). Furthermore *in vitro* culture of HFSC has been shown to activate wound signaling based on the similarities in super enhancer dynamics (Adam et al. 2015). Both of these studies further prove that after wound healing completion, the hair follicle stem cells are then locked into an epidermal fate.

While further work needs to be done in our organoids to evaluate the chromatin dynamics of the HFSC as they develop into organoids, our results suggest that they are emulating wound-like states in their transition from hair follicle to epidermis. Interestingly, our media formulation optimization suggests that the key culture ingredient in inducing epidermal differentiation from HFSC was the small molecule p38MAPK inhibitor, SB202190, which has also been shown to increase the RAS/ERK phosphorylation and signaling pathway in leukemia (Hirosawa et al. 2009). This could point to importance of the RAS/ERK signaling pathway in the cell fate switching of hair follicle stem cells. On the other hand, research in the context of keratinocytes and wound healing suggests that p38MAPK inhibition slows keratinocyte migration to a wound (Rötzer et al. 2016). More work must be done to understand how this pathway could initiate epidermal fate switching. Given that re-epithelialization in wounds is significantly slower without HFSC (Langton, Herrick, and Headon 2008; Zawacki and Jones 1967) and is potentially inhibited in chronic, non healing wounds (Stojadinovic et al. 2005), future studies can use these organoids as a model to further understand the mechanisms of fate switching of the hair follicle stem cells toward epidermal fates, and furthermore can endeavor to understand factors that may inhibit it.

METHODS

Animals

The transgenic porcine model expressing *LGR5* driven H2B-GFP was generated as previously reported (Chapter 2) and animals were bred in house. All experiments were performed in strict accordance with the approved Institutional Animal Care and Use Committee of North Carolina State University (IACUC 17-028-B).

Single cell isolation and flow sorting

Single cell isolation and flow sorting was performed as previously described (Chapter 2). Briefly, skin was cut into 5mm^2 pieces and incubated with Dispase II (Sigma) at 10mg/mL in PBS without calcium and magnesium (Corning) for 1 hour at 37℃ or overnight at 4℃. Hair and epidermis were manually removed from dermis and incubated in 0.05% trypsin for 5 minutes at 37°C with shaking. Suspension was vortexed and strained with 70µm cell strainer (BD Falcon). Cells were centrifuged at 1.2RPM for 5minutes and resuspended in PBS with 10% fetal bovine serum (Corning) and 1% antibiotic-antimycotic (Corning). 250ng/mL Propidium iodide (Biotium) was added to cells and sorting was performed with Beckman Coulter MoFlo XPD**.** Cells were sorted for LGR5-GFP HI, LO, and NEG and data was analyzed using FloJo^{TM} (BD Biosciences).

Organoid Culture

Sorted cells were washed with PBS, 10% FBS, 1% anti-anti, after sorting and counting. 2000 cells per well were then pelleted and resuspended in a mixture of Growth Factor Reduced Matrigel **(**Corning**)** and growth factor master mix. 25uL droplets of suspended cells were formed

on a prewarmed 48-well plate. After plating, Matrigel domes were allowed to polymerize for 20 minutes at 37℃. Following polymerization, 250uL of organoid complete media (see Supplementary table 2) was added to each well and plates were incubated at 5% CO_2 and 5% O_2 balanced with Nitrogen gas. Growth factors and media were replaced every other day. Growth factors conditions for M1 (IFE differentiation conditions) are listed in Supplentary table 2. M2 (hair follicle culture conditions) consisted of all factors in the table except SB202190.

For passaging, on day 14 of culture, media was aspirated and matrigel droplet was washed with PBS. After washing, ice cold Matrigel Recovery Solution (Corning) was used to break apart Matrigel and dissolve Matrigel by incubating for 20 minutes at 4C. Suspended organoids were then washed 2-3 times with ice cold PBS, and incubated again with ice cold Matrigel Recovery Solution for 10 minutes at 4C to remove residual Matrigel. Following 2-3 washes with ice cold PBS, organoids were incubated with 0.25% trypsin (Corning) for 15-25 minutes at 37C with shaking. Organoids were vortexed and pipetted to mechanically dissociate and PBS with 10% FBS was added to inactivate trypsin. Remaining organoids were removed using a 70uM cell strainer (Falcon) and single cells were then to pellet and washed with PBS. Single cells were then split 1:3 and resuspended in matrigel and growth factor mastermix, and overlaid with organoid complete media after polymerization.

Formation Efficiency, Size, and Loss of GFP Assays

Formation efficiency was measured by counting the number of organoids formed per well at day 14 and dividing by the total number of cells plated. To measure organoid size at each respective time point, representative images of each well were taken. ImageJ (NIH) was used to measure the diameter of each organoid in uM. Loss of GFP over time in LGR5-Hi organoids was

measured by taking a representative z-stack image of each well (Leica) and calculating the percent of organoids positive in each image. A GFP-positive organoid was defined as one with greater than 5 GFP-positive cells. To determine statistical differences, mean was calculated and then analyzed using one-way ANOVA with Tukey's post hoc test.

Histological Analysis

Pig and mouse skin was fixed overnight in 4% PFA at 4C, washed with PBS 3x, dehydrated overnight at 4C in 30% sucrose, and embedded in OCT. Organoids were washed with PBS 3x, fixed with 4% PFA for 5 minutes. For whole mount organoid staining, fixed organoids were washed with PBS 3x. For cryosectioning of organoids, organoids were washed after fixation and embedded in OCT. Hematoxylin & Eosin staining was performed on 20uM cryosections, and viewed with (LEICA). Sections or whole organoids were blocked with IHC/ICC Blocking Buffer with 0.4% Triton X-100 (Invitrogen, Sigma), incubated with primary antibodies; CD200 (ls-b11638, 1/50), KRT14 (MA1-06323, 1/200), K85 (ab192729, 1/200), CDSN (ab204235, 1/200), and KRT10 (ab9025, 1/200) and secondary antibodies (Thermofisher: A-11012, A-21235, and A-21450, all 1/500) in IHC/ICC Blocking Buffer with 0.4% Triton X-100 (Invitrogen, Sigma) diluted 1:1 with PBS (PBS, Corning). Slides were washed and mounted with Prolong Gold Antifade Mount with DAPI (Invitrogen, P36931). Immunostained samples were visualized with (Olympus Fluoview FV3000 Confocal Microscope)**.**

Reverse transcriptase quantitative polymerase chain reaction (RT-qPCR)

Total RNA from skin tissue or organoids was isolated using Zymo Quick-RNA Microprep kit with on column DNase digest according to manufacturer's instructions. RNA was eluted into

DNase/RNase free water and stored at -80℃ until further use. cDNA was synthesized with AffinityScript Multiple Temperature cDNA Synthesis Kit (Agilent), according to the manufacturer instructions. For RT-qPCR, iTaq Universal SYBR Green Supermix (BioRad) was used with cDNA template and forward and reverse primers were designed as listed in the table below for optimal conditions: 2min denaturation, 40 cycles of 95℃ denaturation for 5 s and 60℃ extension for 30 s, with final extension at 60℃ for 2 min. Each sample was amplified on a qTOWER^3 thermal cycler (Analytik Jena) with technical duplicates for three biological replicates with similar results. Each gene expression was normalized to *GAPDH* and *ACTB.*

Table S1. Organoid Complete Media

Reagent	Company/Cat. Number
25mL Advanced DMEM/F12	Thermofisher/12634010
25mL 100% L-WRN Conditioned Media	ATCC, according to manufacturer's instructions
10mM HEPES Buffer	Thermofisher/15630130
2mM Glutamax	Thermofisher/35050079
1X B27 Supplement minus Vitamin A	Thermofisher/12587010
1X N2 Supplement	Thermofisher/17502048
1X Antibiotic-antimycotic	Corning/30-004-CI

Table S2. Growth Factor formulation for M1

Reagent	Concentration	Company/Cat. Number
EGF	50ng/mL	Peprotech/AF-100-15
A-83-01	500nM	Peprotech/9094360
CHIR99021	2.5uM	Tocris/13122
SB202190	10uM	peprotech/1523072
Nicotinamide	1mM	Sigma-aldrich/N3376-100G
LY2157299	0.5uM	Selleckchem/S2230
Y-27632 Dihydrochloride	10uM	Peprotech/1293823

Table S3. Primers used in this study

Sus Scrofa Gene	Forward (5' -> 3')	Reverse (5' -> 3')
LGR5	CCTTGGCCCTGAACAAAATA	ATTTCTTTCCCAGGGAGTGG
LGR6	CAGGAGGACGGCTTCATGC	GTCCTGGGATGTGTGACAGG
SOX9	CGGTTCGAGCAAGAATAAGC	GTAATCCGGGTGGTCCTTCT
KRT14	GGAGGTGAAGATCCGCGAC	TCTGCAGCACGACATTAGCG
KRT10	TGGTACGACAAGCATGGCAA	GGCGCAGGGTTACCTCATTC
IVL	GGCCATCCATCCTACTGTGAG	ACTTGCTCCTGCTGGGTATC
S100A10	ACCTCACGAAAATGCCGTCT	GCCCAGCGATTAGCGAAAAG
CDSN	GGCCACACAGTCCAAGATGG	GGTCCTTGCAGGGGTCTAAG
GAPDH	ATCCTGGGCTACACTGAGGAC	AAGTGGTCGTTGAGGGCAATG

REFERENCES

Adam, R. C., Yang, H., Rockowitz, S., Larsen, S. B., Nikolova, M., Oristian, D. S., Polak, L., Kadaja, M., Asare, A., Zheng, D., & Fuchs, E. (2015). Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. *Nature*, *521*(7552), 366–370.

Barker, N., Tan, S., & Clevers, H. (2013). Lgr proteins in epithelial stem cell biology. *Development* , *140*(12), 2484–2494.

Boonekamp, K. E., Kretzschmar, K., Wiener, D. J., Asra, P., Derakhshan, S., Puschhof, J., López-Iglesias, C., Peters, P. J., Basak, O., & Clevers, H. (2019). Long-term expansion and differentiation of adult murine epidermal stem cells in 3D organoid cultures. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(29), 14630–14638.

Dekoninck, S., & Blanpain, C. (2019). Stem cell dynamics, migration and plasticity during wound healing. *Nature Cell Biology*, *21*(1), 18–24.

Di Meglio, P., Perera, G. K., & Nestle, F. O. (2011). The multitasking organ: recent insights into skin immune function. *Immunity*, *35*(6), 857–869.

Ge, Y., Gomez, N. C., Adam, R. C., Nikolova, M., Yang, H., Verma, A., Lu, C. P.-J., Polak, L., Yuan, S., Elemento, O., & Fuchs, E. (2017). Stem Cell Lineage Infidelity Drives Wound Repair and Cancer. *Cell*, *169*(4), 636–650.e14.

Gonzalez, L. M., Williamson, I., Piedrahita, J. A., Blikslager, A. T., & Magness, S. T. (2013). Cell lineage identification and stem cell culture in a porcine model for the study of intestinal epithelial regeneration. *PloS One*, *8*(6), e66465.

Guiu, J., Hannezo, E., Yui, S., Demharter, S., Ulyanchenko, S., Maimets, M., Jørgensen, A., Perlman, S., Lundvall, L., Mamsen, L. S., Larsen, A., Olesen, R. H., Andersen, C. Y.,

Thuesen, L. L., Hare, K. J., Pers, T. H., Khodosevich, K., Simons, B. D., & Jensen, K. B. (2019). Tracing the origin of adult intestinal stem cells. *Nature*, *570*(7759), 107–111.

Hirosawa, M., Nakahara, M., Otosaka, R., Imoto, A., Okazaki, T., & Takahashi, S. (2009). The p38 pathway inhibitor SB202190 activates MEK/MAPK to stimulate the growth of leukemia cells. *Leukemia Research*, *33*(5), 693–699.

Hsu, Y.-C., Pasolli, H. A., & Fuchs, E. (2011). Dynamics between stem cells, niche, and progeny in the hair follicle. *Cell*, *144*(1), 92–105.

Huch, M., Dorrell, C., Boj, S. F., van Es, J. H., Li, V. S. W., van de Wetering, M., Sato, T., Hamer, K., Sasaki, N., Finegold, M. J., Haft, A., Vries, R. G., Grompe, M., & Clevers, H. (2013). In vitro expansion of single Lgr5+ liver stem cells induced by Wnt-driven regeneration. *Nature*, *494*(7436), 247–250.

Jaks, V., Barker, N., Kasper, M., Van Es, J. H., Snippert, H. J., Clevers, H., & Toftgård, R. (2008). Lgr5 marks cycling, yet long-lived, hair follicle stem cells. *Nature Genetics*, *40*(11), 1291.

Joost, S., Jacob, T., Sun, X., Annusver, K., La Manno, G., Sur, I., & Kasper, M. (2018). Single-cell transcriptomics of traced epidermal and hair follicle stem cells reveals rapid adaptations during wound healing. *Cell Reports*, *25*(3), 585–597. e7.

Kang, S., Long, K., Wang, S., Sada, A., & Tumbar, T. (2020). Histone H3 K4/9/27 Trimethylation Levels Affect Wound Healing and Stem Cell Dynamics in Adult Skin. *Stem Cell Reports*, *14*(1), 34–48.

Langton, A. K., Herrick, S. E., & Headon, D. J. (2008). An Extended Epidermal Response Heals Cutaneous Wounds in the Absence of a Hair Follicle Stem Cell Contribution. *The Journal of Investigative Dermatology*, *128*(5), 1311–1318.

Page, M. E., Lombard, P., Ng, F., Göttgens, B., & Jensen, K. B. (2013). The epidermis comprises autonomous compartments maintained by distinct stem cell populations. *Cell Stem Cell*, *13*(4), 471–482.

Rötzer, V., Hartlieb, E., Winkler, J., Walter, E., Schlipp, A., Sardy, M., Spindler, V., & Waschke, J. (2016). Desmoglein 3-Dependent Signaling Regulates Keratinocyte Migration and Wound Healing. *The Journal of Investigative Dermatology*, *136*(1), 301–310.

Sehgal, V. (2018). *The Impact of Microanatomy and Changing Physiology of Stratum Corneum, the Skin Barrier, on Evolution of Atopic Dermatitis*. https://www.semanticscholar.org/paper/8e5800e992a3db04e9f5f9dfa89fcbc95ef4b0f8

Stojadinovic, O., Brem, H., Vouthounis, C., Lee, B., Fallon, J., Stallcup, M., Merchant, A., Galiano, R. D., & Tomic-Canic, M. (2005). Molecular pathogenesis of chronic wounds: the role of β-catenin and c-myc in the inhibition of epithelialization and wound healing. *The American Journal of Pathology*, *167*(1), 59–69.

Summerfield, A., Meurens, F., & Ricklin, M. E. (2015). The immunology of the porcine skin and its value as a model for human skin. *Molecular Immunology*, *66*(1), 14–21.

Yang, H., Adam, R. C., Ge, Y., Hua, Z. L., & Fuchs, E. (2017). Epithelial-Mesenchymal Micro-niches Govern Stem Cell Lineage Choices. *Cell*, *169*(3), 483–496.e13.

Zawacki, B. E., & Jones, R. J. (1967). Standard depth burns in the rat: the importance of the hair growth cycle. *British Journal of Plastic Surgery*, *20*, 347–354.

Chapter 4

Two roles for LGR5 in lung development and homeostasis

Introduction

LGR5, a surface receptor for R-spondin, is a known potentiator of Wnt signaling (Barker et al., 2013). LGR5 is often involved in developmental processes and regeneration, and generally marks key progenitor populations in organ development, such as in the mammary gland, or at homeostasis, such as in intestinal crypts, taste buds, and stomach (Barker et al., 2013). In the case of the liver, LGR5 was shown to mark hepatic progenitor cells in development (Prior et al., 2019) but was not expressed in the healthy adult liver at homeostasis. However, upon injury LGR5 was again expressed and the LGR5+ cells in the damaged liver were capable of organoid generation and organ regeneration (Huch et al., 2013). This shows the dynamic nature of LGR5 expression, from development to homeostasis to repair and regeneration.

One organ that has been largely unstudied regarding the role of LGR5 is the lung. WNT signaling has long been known to be vital to lung development (Hussain et al., 2017; Kadzik et al., 2014), in addition to repair and regeneration (reviewed here:(Raslan & Yoon, 2020)). Furthermore, RSPO2, the ligand for LGR5, was shown years ago to have a vital role in lung branching morphogenesis, and the drastically limited lung growth resulting from the knockout was correlated with reduced canonical WNT signaling (Bell et al., 2008). Despite these important studies, and the inclusion of RSPO2 growth factors in many human lung organoid growth media (Nikolić et al., 2017; Sachs et al., 2019), the expression of LGR5 in the developing lung has still not been mapped or studied.

In the postnatal lung, a recent study of LGR5 expression in mice beautifully describes a unique role for LGR5+ cells centered around the alveolar progenitor cells: orchestrating the alveolar

niche (Lee et al., 2017). However, this research has not been supported by data from human cells, and there are many physiological differences between the mouse and the human lungs. To begin to study the role of LGR5 and LGR5+ cells in the lung, we elected to use a porcine model, which represents a more similar anatomical and physiological model of the human lung (Benahmed et al., 2014; Judge et al., 2014; Lunney et al., 2021; Meurens et al., 2012). To accurately track the LGR5+ cells in real time, we generated a transgenic porcine model expressing nuclear GFP under the control of the endogenous LGR5 promoter (LGR5-H2B-GFP).

In this work, we show that pigs and humans have a drastically different pattern of LGR5 expression in the lung than the mouse, with LGR5 expression in postnatal lung found in a mesenchymal support niche surrounding the airways. This points to a role in the regulation of the airway niche by LGR5-expressing mesenchymal cells, which are multipotent stem cells. Interestingly, we find a very different role for LGR5 expression in fetal lung morphogenesis, in which LGR5 is expressed in developing airway epithelial cells and is tightly coupled with elongating bud tips throughout development. Together, these results identify a new lung stromal cell population in homeostasis which could have a role in orchestrating airway repair and contribution to fibrosis, and furthermore identify transient epithelial LGR5 expression as a pathway that could be used to forge lung repair and regeneration.

Results

LGR5 is expressed consistently in an airway mesenchymal population, and transiently in developing bud tips

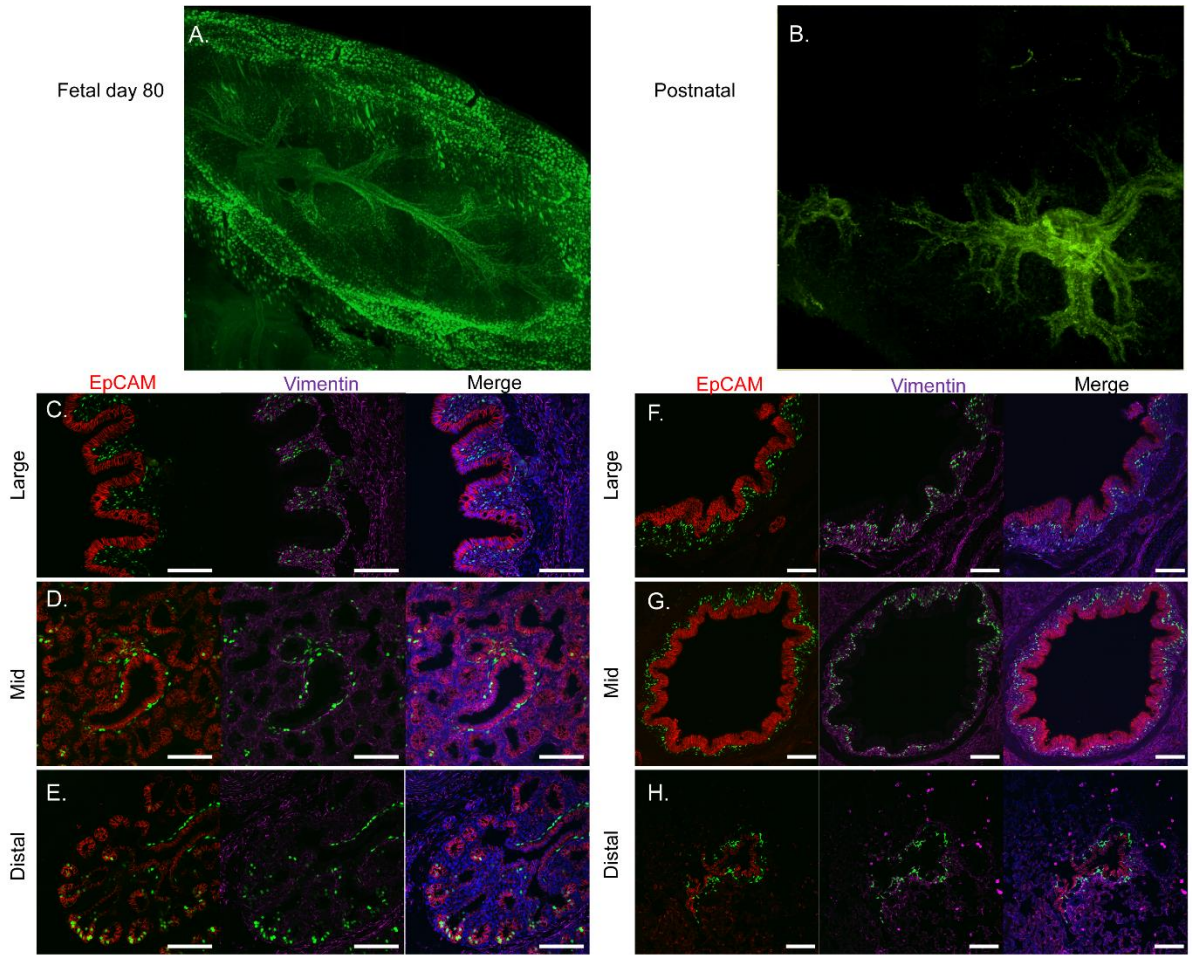


Figure 1. Two populations of LGR5 in development and homeostasis. In the fetal lung (A), LGR5 is expressed surrounding the airways in a mesenchymal population, and also in the developing bud tips. In the postnatal lung (B) LGR5 expressed only surrounding the airways. Examination of distribution of LGRR-GFP with EpCAM (red) or vimentin (purple) based on airway size show that the mesenchymal population always expresses vimentin and surrounds the airway epithelium of large, mid and distal airways (C-E), while LGR5+ cells in the bud tips are EpCAM+. F-H) Post-natal expression shows that the vimentin+ mesenchymal population is present in airways of all sizes, and occupies the sub-basal space.

To initially understand the distribution of LGR5 expression throughout development and homeostasis, we used a transgenic porcine model expressing H2B-GFP under the control of the endogenous LGR5 promoter, as reported previously (Chapter 2). For initial mapping of expression, we used the tissue clearing technique (Wang et al., 2019) to render the lungs transparent and used light sheet microscopy to acquire 3D images of fetal developmental day 80 (D80) and juvenile porcine lungs. In the fetus, LGR5 expression is robust surrounding both proximal and distal airways, in addition to the developing bud tips (Figure 1A), while the postnatal lung solely expresses the population surrounding the airways (Figure 1B). In cryosections imaged by confocal, co-staining fetal lungs with EpCAM for epithelial cells and vimentin for mesenchymal cells shows that the population surrounding the airways is mesenchymal, while the distal bud tips are epithelial (Figure 1C-E). Imaging of large, mid, and distal airways in juvenile tissue show that the LGR5+ population is present in airways of all sizes, and as in the fetus, the LGR5+ mesenchymal population occupies a sub-basal position, directly adjacent to the airway epithelium (Fig 1F-H). These initial data suggest that there are multiple previously undescribed roles for LGR5 in lung development and homeostasis, and that they differ largely from that of the mouse.

First, to determine whether the distribution of LGR5 in the developing and postnatal pig lungs is representative of the human, we employed fluorescent in situ hybridization (FISH) with a probe against human LGR5 mRNA transcripts in human fetal and adult lung tissue. Like the pig, LGR5 expression is robust in human developing lung bud tips (Figure 2A), with an almost identical pattern of expression as the pig (as in 1E). In addition, LGR5 expression was detected in the subepithelial population surrounding maturing airways in the developing lung, as shown with

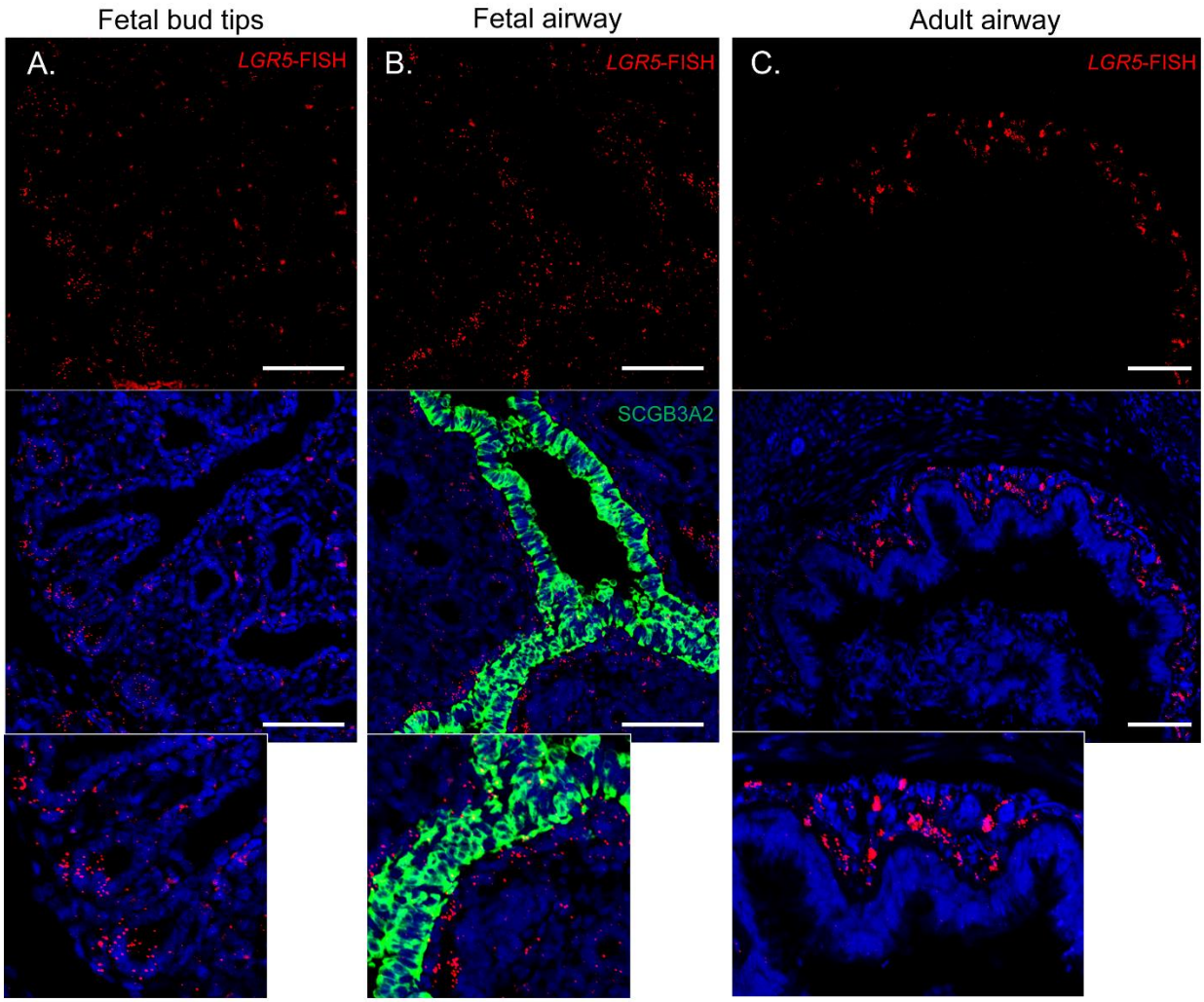


Figure 2. Human lung distribution of *LGR5* matches the pig. *In situ* hybridization with a probe detecting human *LGR5* transcripts in human fetal day 135 (A) or day 142 (B) shows the presence of epithelial *LGR5* expression in developing bud tips (A), and the presence of the mesenchymal *LGR5+* expression in cells surrounding the airways. C) Adult human lung shows robust *LGR5* expression in the mesenchymal population directly adjacent to the airway epithelium.

co-staining with SCGB3A2 (Figure 2B). Finally, robust detection of LGR5 is evident surrounding adult airways (Figure 2C). To further examine the molecular distribution of LGR5 expression in the human, we initially mapped LGR5 to previously established single-cell RNA sequencing (scRNAseq) datasets of lung non-epithelial cells (Vieira Braga et al., 2019), and see that LGR5 expression is highly upregulated and specific to lung fibroblasts (Figure 3A-C). Moreover, in developing human lungs (Q. Yu et al., 2021), LGR5 is detected in both epithelial and mesenchymal populations (Figure 3D). Combined, we can conclude that LGR5 expression in the human is aligned with expression in the pig, and furthermore that the pattern in the human and the pig is significantly different than previously described in the mouse.

For an initial high-resolution comparison between LGR5+ epithelial cells and LGR5+ mesenchymal cells, we performed scRNAseq on fluorescence-activated cell sorting (FACS) sorted LGR5+ cells in the developing lung. UMAP dimension reduction analysis identifies the two populations separated by EpCAM expression (Figure 4A), which can each be broken up to subclusters to yield 11 clusters (Figure 4B). Top cluster markers identify key genes that can be used to relate each population to established human developmental populations (Figure 4C). Separate clusters within the EpCAM+ population could suggest that one represents a more differentiated state as cells start to turn on HOPX (cluster 5), a known marker expressed in the bud tip-adjacent stalk region (CITE), and lower expression of bud tip marker SFTPB. Due to the dilution effect of H2B-GFP, cells that have begun to differentiate still retain a low-level GFP expression and could be detected by FACS. Within the EpCAM- mesenchymal population, there are several clusters, each with a relatively undescribed set of genes. Recent research (He et al., 2022) describes how the fetal lung allows for lineage analysis because both the mature,

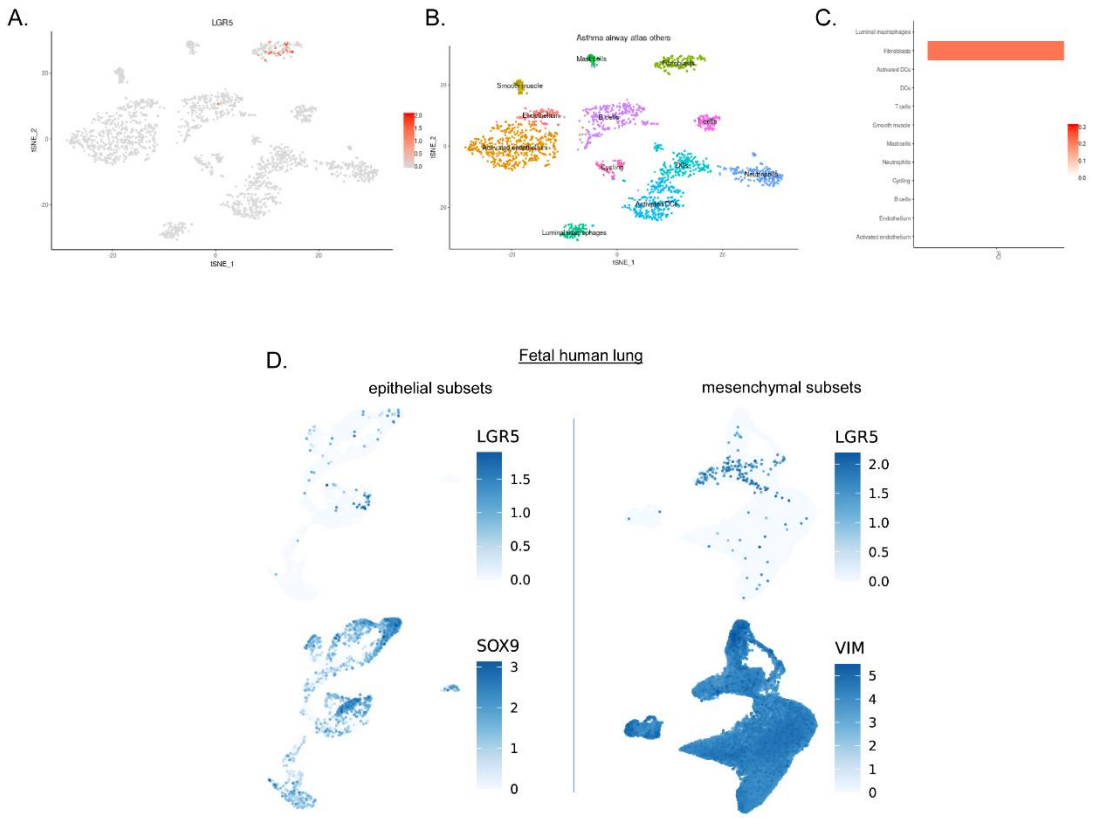


Figure 3. Molecular mapping of LGR5 expression in the human lung confirms the presence of *LGR5* expression in multiple populations in development, and one at homeostasis. A-C) The Sanger lung "Asthma Airway Atlas," asthma negative control population, identifies presence of LGR5 in postnatal lungs in the fibroblast populations. D) The developing human lung atlas webtool from Yu et al 2021 maps *LGR5* showing data from expression to both the developing fetal epithelium and mesenchymal populations.

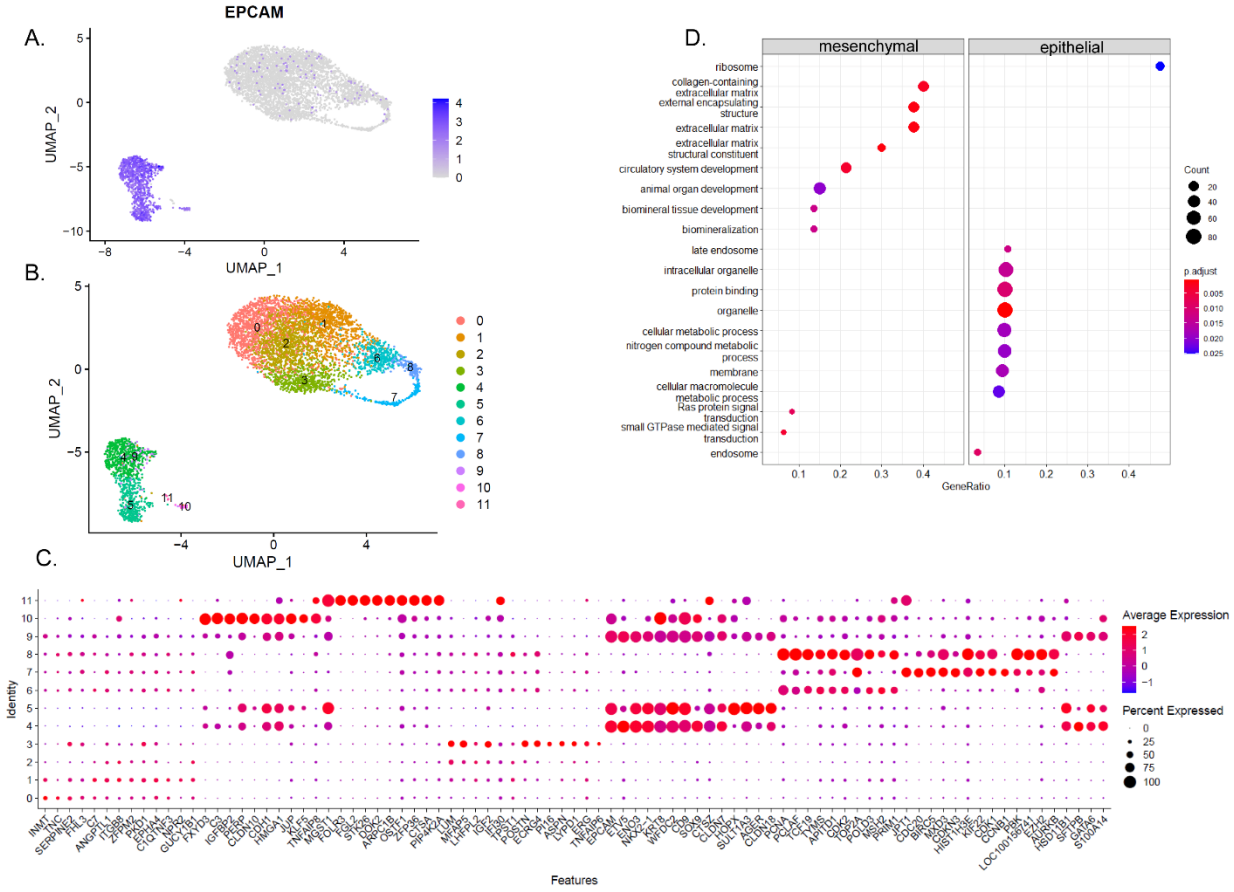


Figure 4. Single cell RNA sequencing of fetal porcine LGR5+ populations shows that fetal epithelial and mesenchymal populations are unrelated. All LGR5+ cells were FACS sorted from a porcine fetal d80 lung, and after quality assessment and filtering 5,284 cells were retained for analysis. A) UMAP analysis shows two main groups of cells marked by differential expression of EpCAM. Cell populations can be separated into 11 clusters (B), prestoMarkers analysis determines the mostly highly expressed markers for each cluster. D) Gene ontology analysis of the EpCAM+ vs EpCAM- populations from plot in (A) shows most significantly upregulated pathways in each population.

differentiated airways and developing distal bud tips exist at once. This suggests that each population in the mesenchymal cluster could represent the developmental state of the airway they surround. More research must be done to identify which genes change through the development of the airways and to identify the spatial distribution of each cluster. Gene Ontology analysis on differentially expressed genes between the mesenchymal and epithelial population shows that mesenchymal subsets are highly enriched in the extracellular matrix and matrisome terms (Figure 4D).

LGR5+ mesenchymal population is a stromal cell and airway niche population

To begin to dissect the role of the LGR5+ mesenchymal population, we focused on postnatal lungs. Further 3-dimensional analysis revealed that the LGR5+ mesenchymal cells are aligned length-wise down the airways detected by nuclear shape, and form a striated pattern as they follow the folds of the airways, remaining in close contact with the basement membrane and the epithelial cells (Figure 5A). Confocal imaging of whole mount lung allows examination of cross-hatch patterns in the larger airways, with a small population of LGR5+ cells in a perpendicular orientation (Figure 6A-B). However, in the respiratory bronchioles or distal airways, the pattern is more randomly distributed, but nevertheless occupies the same sub-basal position (Figure 6C). Closer analysis of other known mesenchymal markers reveals that all LGR5+ cells express vimentin (Figure 5B), a subset of PDGFRa+ fibroblasts express LGR5 (Figure 5C), LGR5+ cells are directly adjacent to the EpCAM+ epithelial cells, and LGR5 expression is independent of alpha-smooth muscle actin (αSMA) (Figure 5E). Since smooth muscle populations have been repeatedly described surrounding airways, we decided to define the relationship between LGR5+ and αSMA+ fibroblasts from various sized airways (Figure 6D-F). Cryosections reveal that in proximal and mid-sized airways, the αSMA forms a band of smooth muscle, and LGR5+ cells

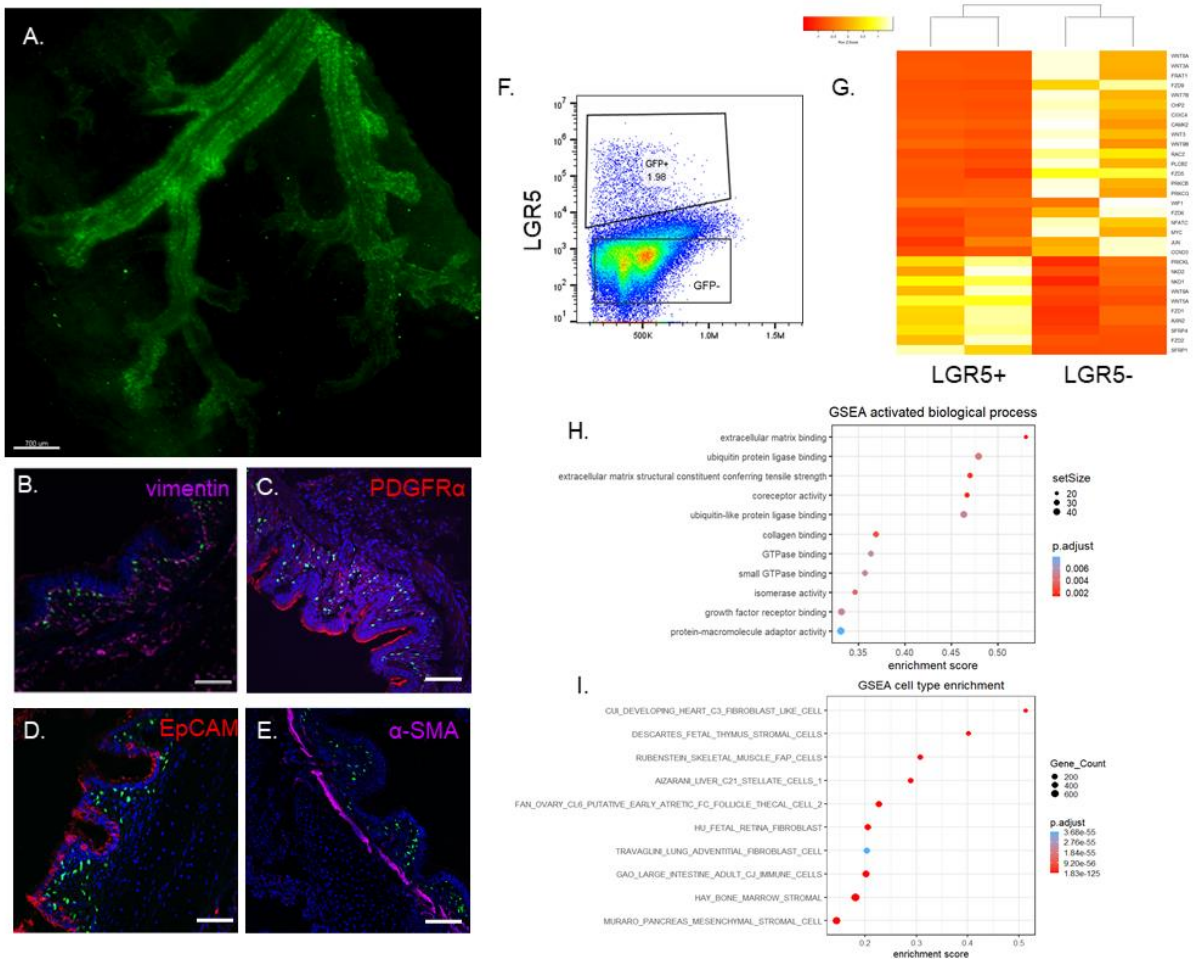


Figure 5. LGR5 marks mesenchymal subset surrounding airways throughout the porcine postnatal lung. A) Tissue clearing and light sheet microscopy show volumetric rendering of LGR5 distribution throughout airways of all sizes. B-E) Confocal imaging of bronchiolar cross-sections, shows co-expression of LGR5 with known fibroblast markers vimentin and PDGFRα, but not with myofibroblast and smooth muscle marker αSMA or epithelial cell marker EpCAM. Scale bar indicates 100 μM unless marked otherwise. F) Fluorescence activated cell sorting gates to identify LGR5-GFP+ population from postnatal porcine lungs. G) Differential gene expression from adult LGR5- vs LGR5+ cell RNA sequencing comparison show significant differentially expressed genes from the KEGG WNT pathway. H) Gene ontology biological processes enriched in LGR5+ cells. I) Gene set enrichment analysis for "cell type" shows significant enrichment in genes associated with human cell type datasets.

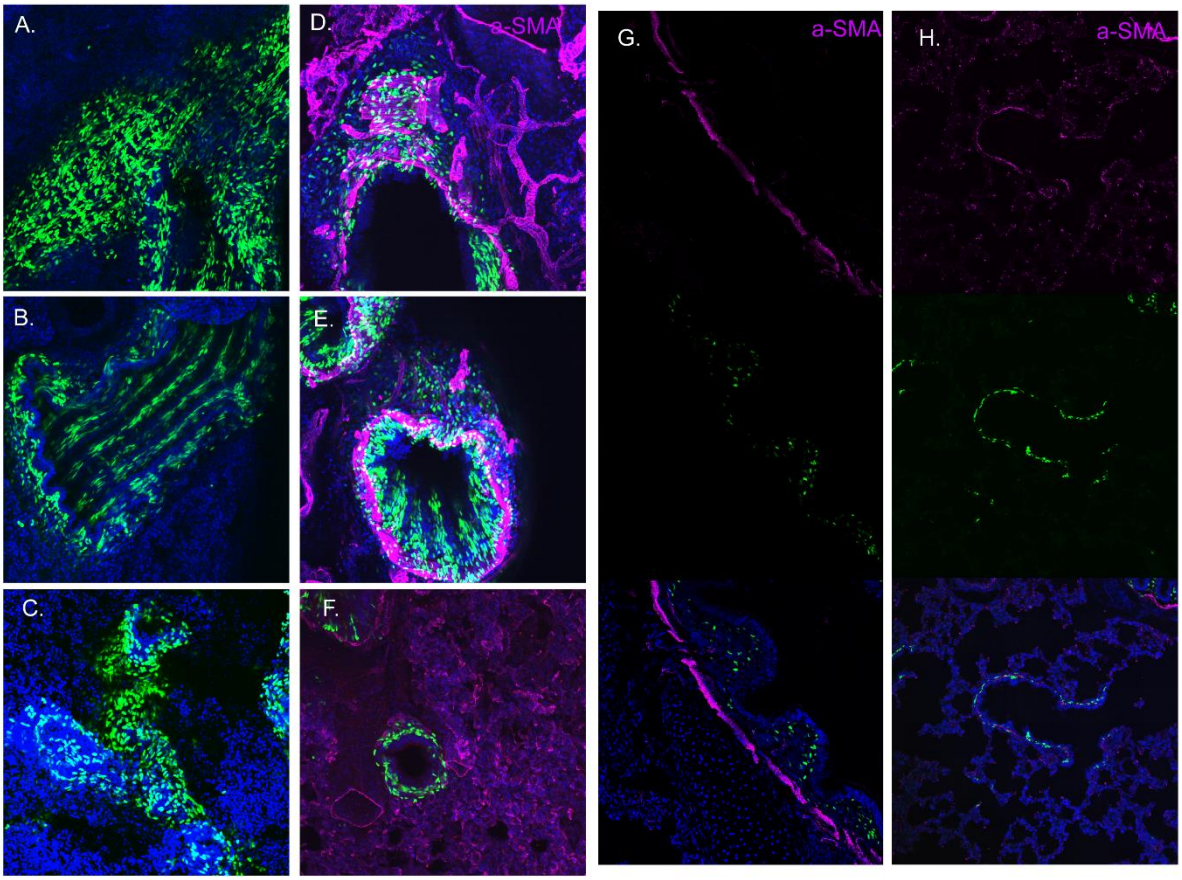


Figure 6. LGR5 expression in the mesenchymal population always occupies the sub-basal mesenchyme, but pattern and orientation depend on size of airway. A) Whole mount confocal imaging of large (A,D,G), mid (B,E), or distal (C,F,H) airways without (A-C) or with (D-H) α-SMA co-staining shows changes in LGR5+ cell pattern depending on size and shape of its respective airway. G,H) Confocal imaging of cross-sections show that LGR5 and α-SMA are not overlapping.

reside in the space between the αSMA and the airway epithelium (Figure 6G). In distal or respiratory airways, LGR5+ and aSMA populations remain separate, and instead are intercalated in the sub-basal space separating the airway from the surrounding interstitia (Figure 6H). This suggests separate but co-existing roles for αSMA+ and LGR5+ populations throughout airways of all sizes.

For further molecular analysis, we next sorted by LGR5-H2B-GFP, and used RNA sequencing to evaluate the differential gene expression between the FACS sorted LGR5+ and LGR5- populations from postnatal porcine lung (Figure 5F). Since LGR5 is a known potentiator of WNT signaling, we first asked how the WNT pathway genes (KEGG) were differentially regulated in the LGR5+ population. Results showed significant upregulation of a select few, such as WNT5A and WNT9A, but primarily significant downregulation of WNT genes. Importantly, many of the significantly upregulated genes in the LGR5+ population from the KEGG WNT pathway are actually negative regulators of WNT signaling, such as SFRP1, SFRP4, PRICKLE1, NKD1, and NKD2 (Figure 5G). This suggests a key role for the LGR5+ mesenchymal cells in modulating WNT signaling in their niche.

Significantly enriched genes were evaluated by Gene Set Enrichment Analysis (GSEA, Broad Institute); the most significantly enriched gene ontology biological process pathways include several extracellular matrix binding pathways and several protein co-receptor activities (Figure 5H). Comparisons to annotated gene sets in the database reveal the ten most significantly enriched subsets, seven of which are stromal cell subsets (Figure 5I). Further analysis of one specific stromal cell dataset detailing upregulated genes found in adipose-derived mesenchymal

stem cells (Boquest et al., 2006) shows that of the significantly differentially expressed genes the LGR5+ population shared with the adipose derived stem cells, the majority are significantly upregulated (Figure 7).

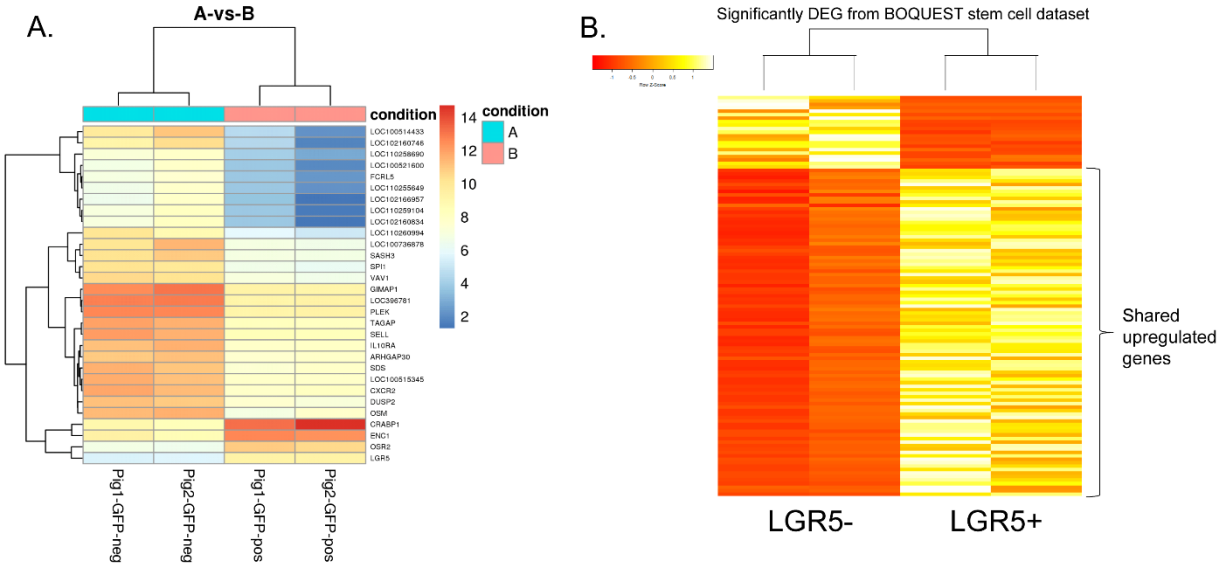


Figure 7. Additional analysis of adult LGR5+ mesenchymal population. A) Heatmap of top expressed genes in LGR5+ or LGR5- populations. DEG heatmap of genes in "BOQUEST stem cell upregulated genes" dataset.

To understand the distribution of LGR5 expression surrounding airways in the fetus, and compare it with the adult, we performed confocal-whole mount imaging and see that the pattern remains largely the same (Figure 8A-C). By molecular analysis, we determined that while many genes are the same, among the most significantly upregulated genes in the fetus were AGTR2 (a gene known to be widely expressed in fetal but not adult tissue), FBN3, and TMEM213, suggesting slightly different roles due to a developing microenvironment (Figure 8D).

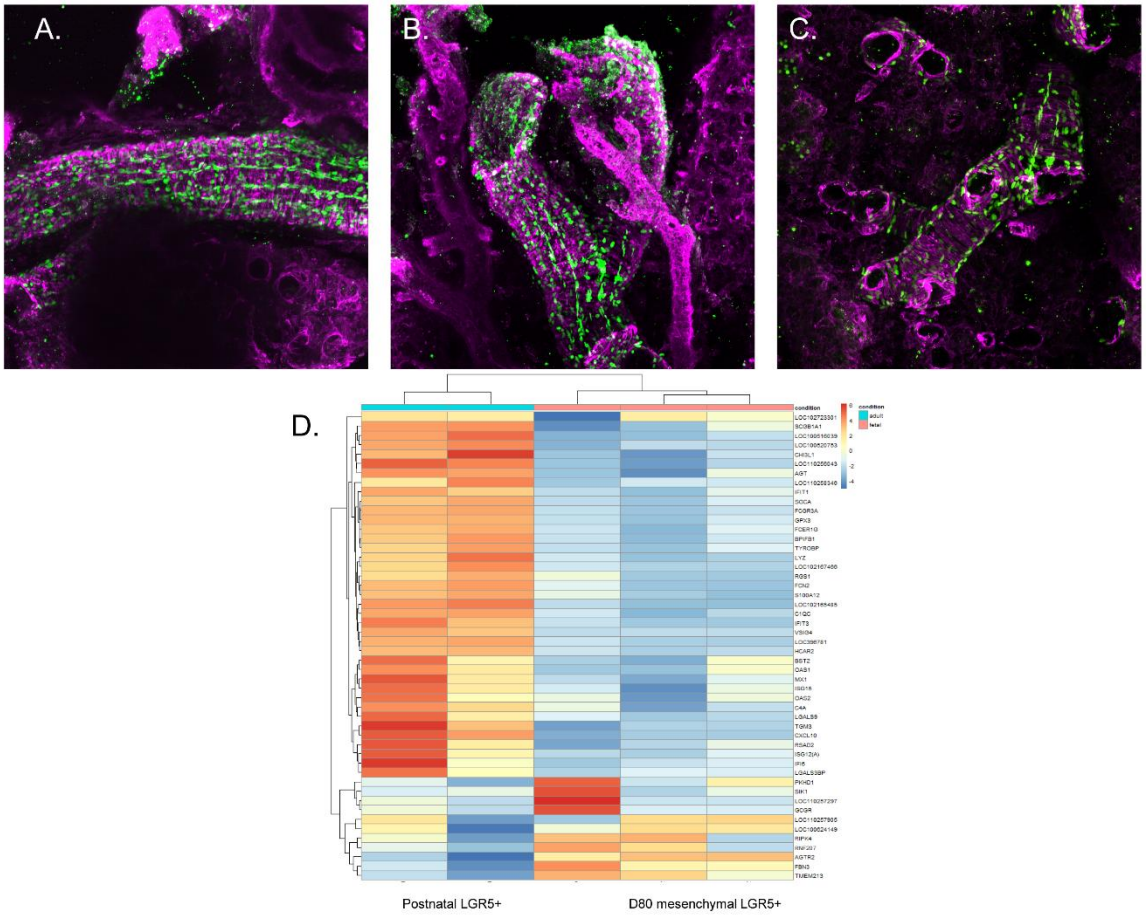


Figure 8. Fetal LGR5+ mesenchymal populations appear surrounding mature airways and resemble adult patterns of expression. A-C) Whole mount staining of D80 lungs with a-SMA show similar pattern of LGR5 expression to adult in large, mid and distal airways. D) RNAseq differential gene expression comparison of LGR5+ mesenchymal populations between fetal and adult.

Since the postnatal LGR5+ lung population shared a gene signature with many stromal cell and mesenchymal stem cell populations, we next asked whether the LGR5+ lung cells present functional properties of mesenchymal stromal and niche cells. When cultured in vitro in expansion followed by induction media, LGR5+ mesenchymal lung cells indeed differentiated into osteocytes, adipocytes, and chondrocytes, and therefore we concluded that LGR5+ mesenchymal cells share the pluripotency that defines mesenchymal stem or stromal cells (MSC) (Figure 9A-D). Furthermore, since stromal cells generally define a support niche for other stem cell subsets, we hypothesized that these LGR5+ MSC would designate the support niche for airway epithelium. To test this, we isolated EpCAM+ cells from lung bronchioles and co-cultured at a 1:1 ratio with LGR5+ MSC in a matrigel dome with or without growth factors (Figure 9H). Remarkably, the LGR5+ cells alone were capable of supporting organoid formation with no additional growth factors (Figure 9E-F). Furthermore, we noticed that the organoids in the LGR5+ wells took on a round, hollow phenotype, whereas the conditions with growth factors represented an irregular, filled shape. To quantify the differences, we used keratin (KRT14), a marker of airway basal cells, and pro-surfactant protein C (SFTPC), a marker of AECII alveolar cells to distinguish between the round bronchiolar type and the irregular bronchioalveolar type, similar to previously described methods (Kathiriya et al., 2022). Bronchiolar organoids were classified as those that expressed KRT14 evenly in most cells throughout the round shape, and did not express SFTPC. Bronchioalveolar organoids express both SFTPC and KRT14, each distributed partially throughout the organoids (Figure 9G). The prevalence of each type was quantified for each culture condition, and results show that the presence of the growth factors induced a bronchioalveolar fate, but LGR5+ MSC alone in basic media supported the growth of bronchiolar organoids, and more specifically KRT14+ basal cells (Figure 9H).

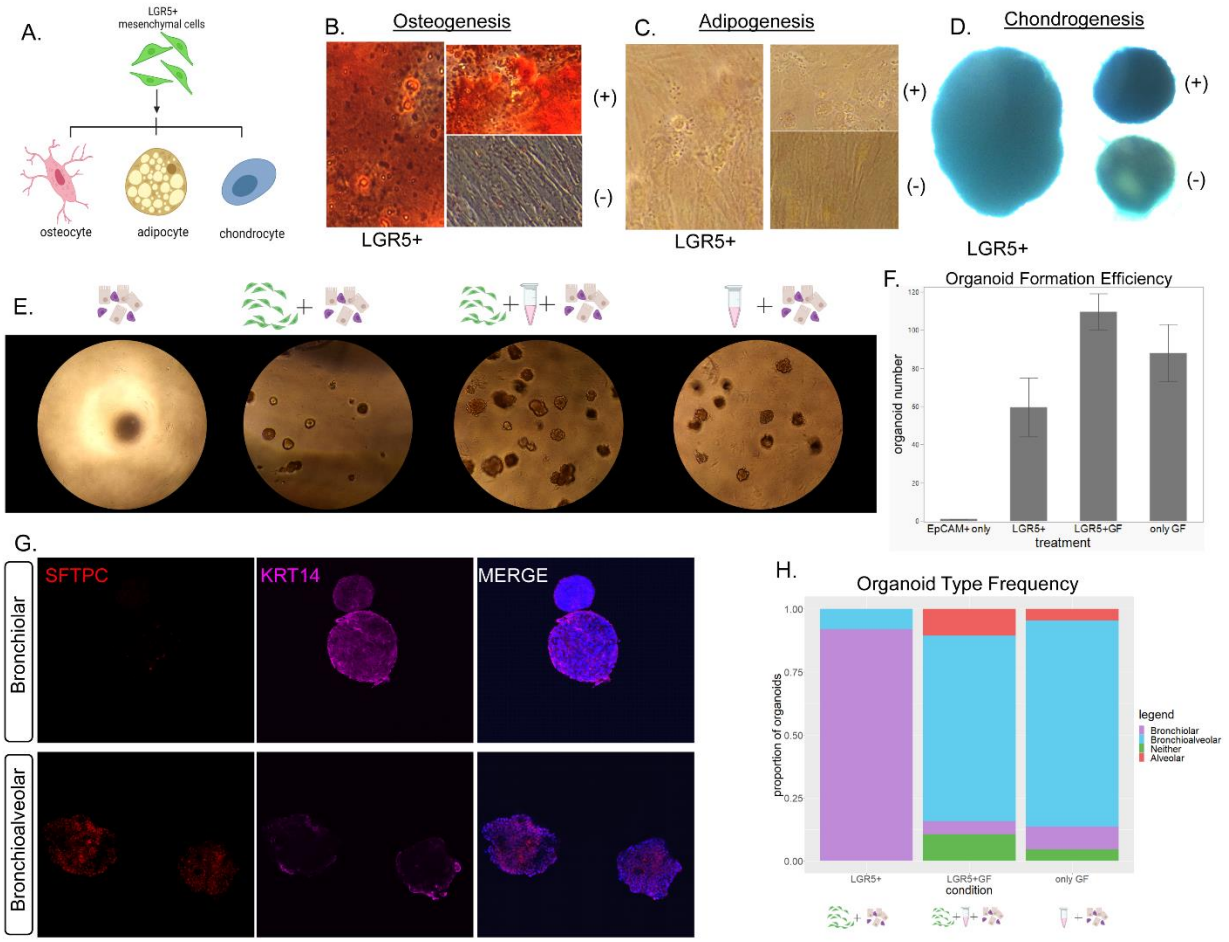


Figure 9. Postnatal LGR5+ cells are mesenchymal stromal cells and support airway basal cell niche. A) LGR5+ mesenchymal cells were cultured in osteocyte, adipocyte or chondrocyte differentiation media. B-D) LGR5+ cells show ability to differentiate into osteocytes, adipocytes, and chondrocytes. Porcine bone marrow mesenchymal stem cells were used as a positive control, and porcine dermal fibroblasts were the negative control. E) Airway epithelial cells were seeded in growth factor reduced Matrigel alone, with LGR5+ mesenchymal cells only, with LGR5+ mesenchymal cells and a growth factor cocktail, or a growth factor cocktail alone. F) Quantification of organoid formation efficiency depending on growth conditions. G) Bronchiolar organoids were characterized by a spherical shape, robust expression of KRT14, and absence of SFTPC. Bronchioalveolar organoids had an irregular shape and expressed both SFTPC and KRT14. H) Organoids from epithelial cells co-cultured with LGR5+ cells only were almost entirely bronchiolar, while conditions containing LGR5+ cells and growth factors or growth factors alone yielded primarily bronchioalveolar organoids.

These data support that this LGR5+ population is a multipotent mesenchymal niche population supporting the bronchiolar basal cells. This suggests that LGR5+ MSC not only have a role in providing the niche signals to the basal cell population, but the fact that added growth factors still induce an alveolar fate suggests that the LGR5+ cells may play an active role in blocking ambient WNT signaling from lung interstitia from reaching the airway. Combined with the RNAseq analysis this suggests a role for LGR5 in the airway fibroblasts: the LGR5-driven potentiation of the WNT receptors could allow this population to "soak up" (canonical) WNT ligands, such as WNT2, that are expressed highly throughout the lung. Ongoing experiments in the lab will test if this transition is solely dependent on WNT signaling.

Transient expression of LGR5 in fetal epithelial population marks developing bud tips

We next asked what is the role of LGR5+ cells in the developing epithelial population. From our scRNAseq data, we know that the epithelial and mesenchymal populations are significantly different. Close analysis of the bud tip region in a D80 fetus shows that the mesenchymal population (arrow) extends from the developed airways throughout the stalk, but is not present surrounding the bud tip population (arrowhead) (Figure 10A). To further analyze the epithelial LGR5+ population, we stained the fetal D50 and D80 with known markers of developing epithelium, such as EpCAM, SFTPC, KI67 (Figure 10B-G), along with SOX2 and SOX9. Since SOX9 represents a well-established marker of bud tips, and SOX2 marks the stalk and more developed airways, we used these as a basis for comparison to ask how the expression of LGR5 changes throughout development. For each time point, we quantified the overlap of epithelial LGR5+ populations with SOX2 or SOX9, representative images are shown (Figure 10H-M). From quantification of the SOX2 and LGR5, we see that in the pseudoglandular stage (D50)

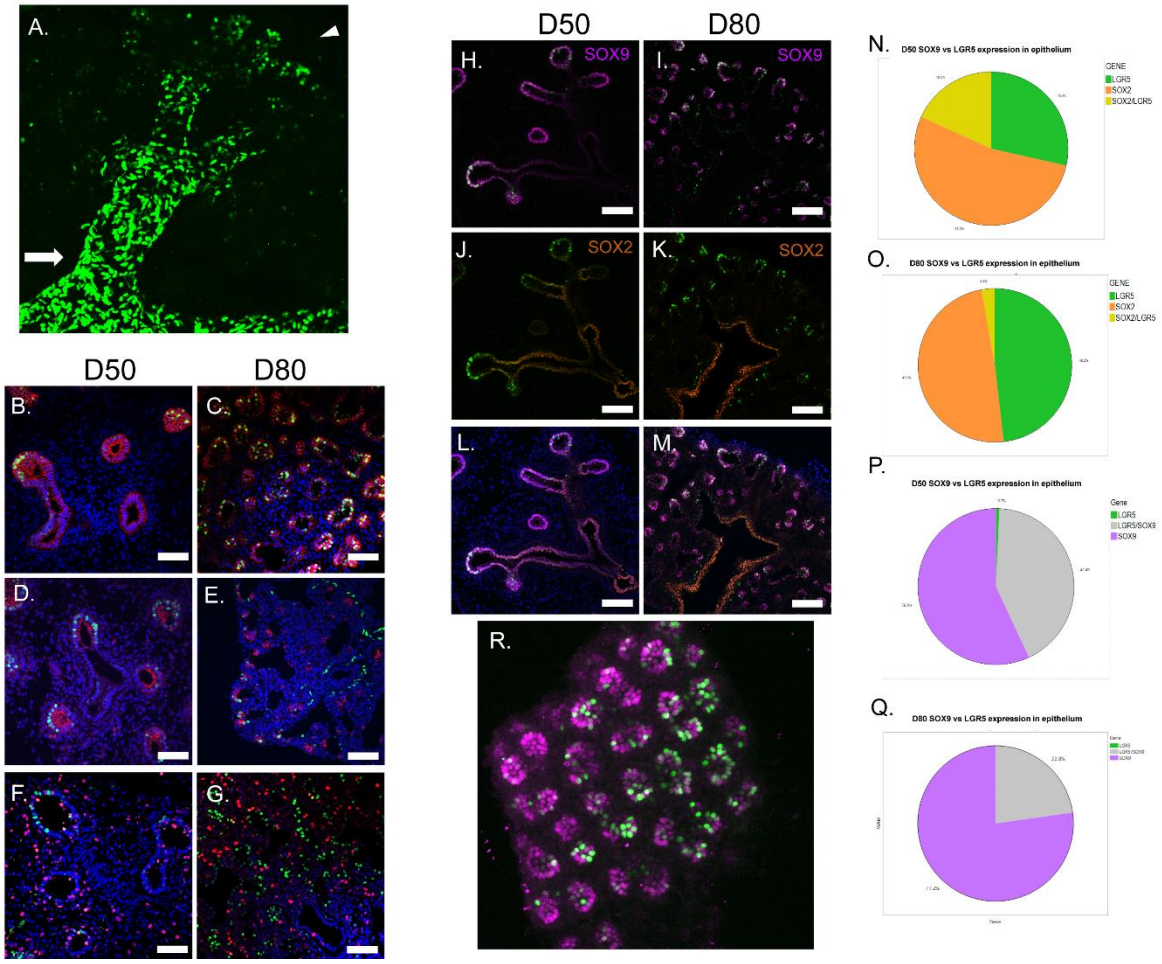


Figure 10. LGR5+ epithelial expression becomes more restricted throughout development. A) LGR5+ mesenchymal population (arrow) surrounds distal airways, but not the bud tips (arrowhead), where LGR5 is expressed in epithelial population. B-G) Co-detection of LGR5 with known markers of development including EpCAM, SFTPC, and Ki67 at fetal D50 and D80 shows consistent staining with EpCAM and co-localization with SFTPC, and occasional co-expression with Ki67. H-M) Representative images of relationship of LGR5 with SOX9, a known marker of developing bud tips, and SOX2, a marker of more differentiated airways. N-Q) Quantification of co-expression of LGR5+ epithelial cells with SOX2 at D50 or D80 shows that LGR5 is occasionally co-expressed with SOX2 at D80, but rarely at D80. With respect to SOX9, LGR5 is expressed in almost half as many SOX9 positive cells at D80 compared with D50. Each chart is representative of 10 image from 2 different fetal lungs. R) 3D visualization of LGR5 and SOX9 expression at D80 via whole-mount confocal imaging.

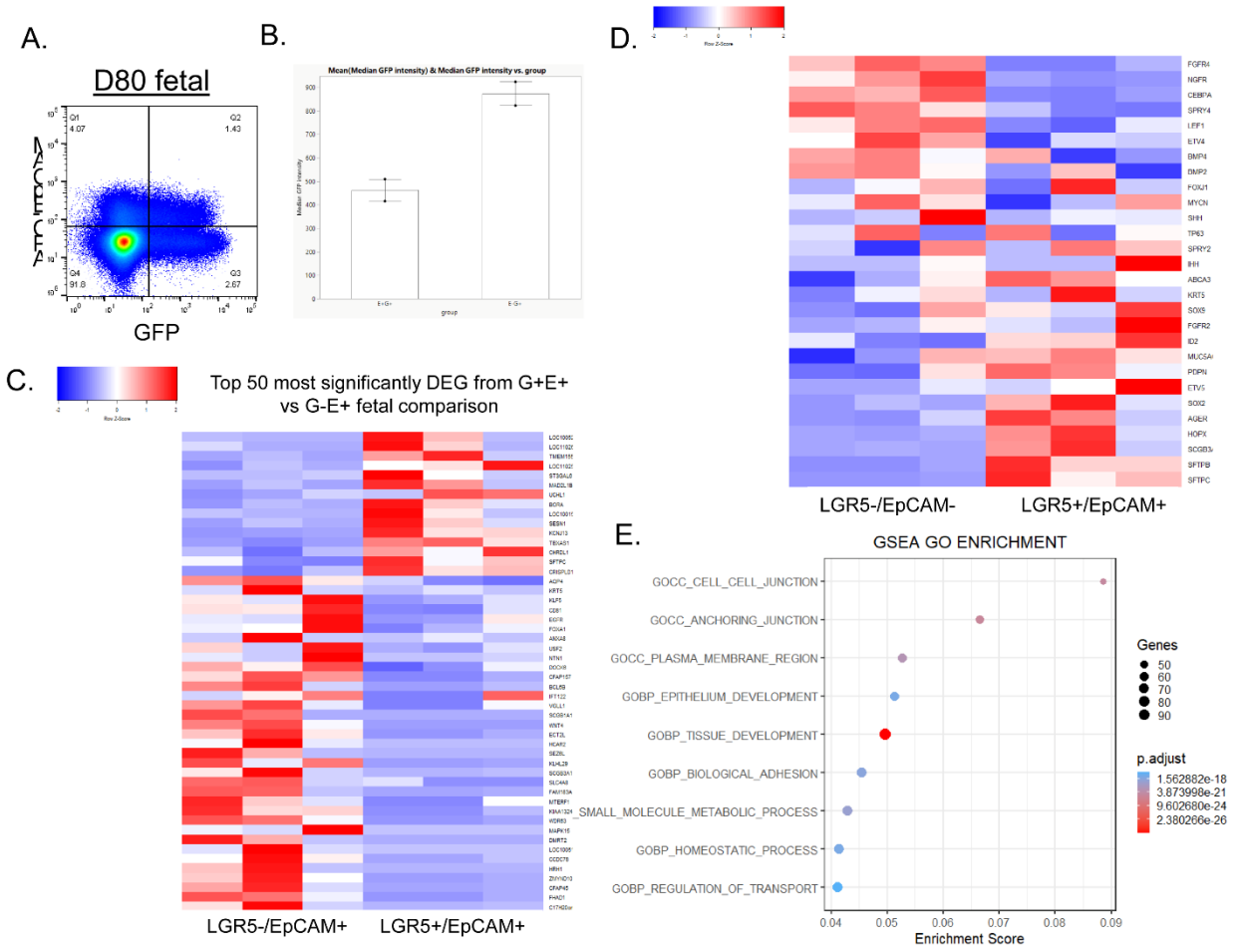


Figure 11. Molecular analysis of LGR5+ epithelial population. A) Fluorescence activated cell sorting by LGR5-GFP or APC-EpCAM shows four distinct populations. B) Median GFP intensity from LGR5+/EpCAM+ population compared with LGR5+/EpCAM- populations shows epithelial population has a significantly lower intensity of LGR5-GFP expression (*n*= two fetuses). C) Top 50 most significantly differentially expressed genes between the LGR5+/EpCAM+ populations and LGR5-/EpCAM+ population (*n*=3 different fetal lungs). D) LGR5-/EpCAM- vs LGR5+/EpCAM+ differential gene expression using a subset of previously described genes upregulated in the developing human lung bud tip and stalk. E) Gene ontology terms enriched in LGR5+/EpCAM+ cells.

that LGR5 occasionally overlaps with SOX2, as has been shown in human tissue with SOX2 and SOX9 in early development (Figure 10N). However, at fetal D80 in the canalicular stage, there is almost no overlap between the two populations (Figure 10O). In the case of SOX9, almost 100% of LGR5+ epithelial cells also express SOX9, at both D50 and D80 (Figure 10P-Q). However, while at D50, approximately 42% of SOX9+ bud tip cells also expressed LGR5, by D80, that number is cut almost in half to 22.2%. 3D imaging of bud tips co-stained with SOX9 shows how the pattern varies by individual tip, with most tips containing at least 2-3 LGR5+ cells, but some containing many more (Figure 10R). Together, these results show that LGR5 expression becomes more restricted throughout development as the fate of the developing airway cells becomes more defined.

To analyze the transcriptome of fetal cells in more detail and compare with other epithelial or parenchymal populations, we sorted D80 fetal lung cells based on EpCAM and/or GFP (Figure 11A). We also noticed that the GFP fluorescent intensity of the epithelial population based on MFI of the epithelial population was significantly lower than the mesenchymal population (n=2), which could be a result of the transient nature of LGR5 expression vs permanent (Figure 11B). When comparing the significantly differentially expressed genes from the LGR5+/EpCAM+ population and the LGR5-/EpCAM+ population, the top significantly differentially expressed genes include upregulation of markers of the tip, like SFTPC, and downregulation of more differentiated airway basal and club cell markers like KRT5 and SCGB1A1 (Figure 11C). A heat map of differential gene expression based on a defined set of bud tip genes shows that compared with LGR5-/EpCAM- cells, LGR5+/EpCAM+ cells are highly enriched in markers of bud tips (SFTPC, SFTPB, FGFR2), and again significantly downregulate differentiated airways markers

(NGFR) (Figure 11D). GSEA GO terms for LGR5+/EpCAM+ include cell-cell junction, anchoring junctions, biological adhesion, and homeostatic process, which combined could imply a role of this population in being anchored into the niche while maintaining homeostasis, not differentiating.

LGR5 is expressed in vitro in a transient state between basal cells and alveoli.

To further explore the role of LGR5 in the epithelium, we used FACS to sort the LGR5+/EpCAM+ or LGR5-/EpCAM+ population from the developing lung. Cells from each group were placed in matrigel with growth media and growth factors. Interestingly, at D50, the LGR5+ epithelial cells formed at a significantly higher efficiency than the LGR5+/EpCAM+, although both populations form big budding organoids that expressed LGR5 (FIgure 12A-B). The D50 cell derived organoids expressed SOX2, SOX9, and SFTPC, often in the same pattern with LGR5, with minimal expression of mature basal cell marker KRT14 (Figure 12C-D). Moreover, organoids from both d50 populations formed at significantly higher efficiency than LGR5+/EpCAM+ cells at D80, although neither group from D80 formed many organoids (Figure 12E). Overall this suggests significant changes in cell state through development, where the LGR5+ bud tip cells are receptive to different signals depending on the stage of development, and potentially reflects the restriction in LGR5 expression we see between D50 D80 tips. Additionally, it could mean that the LGR5+ cells from D50 organoids have higher lineage potential than LGR5- epithelial cells.

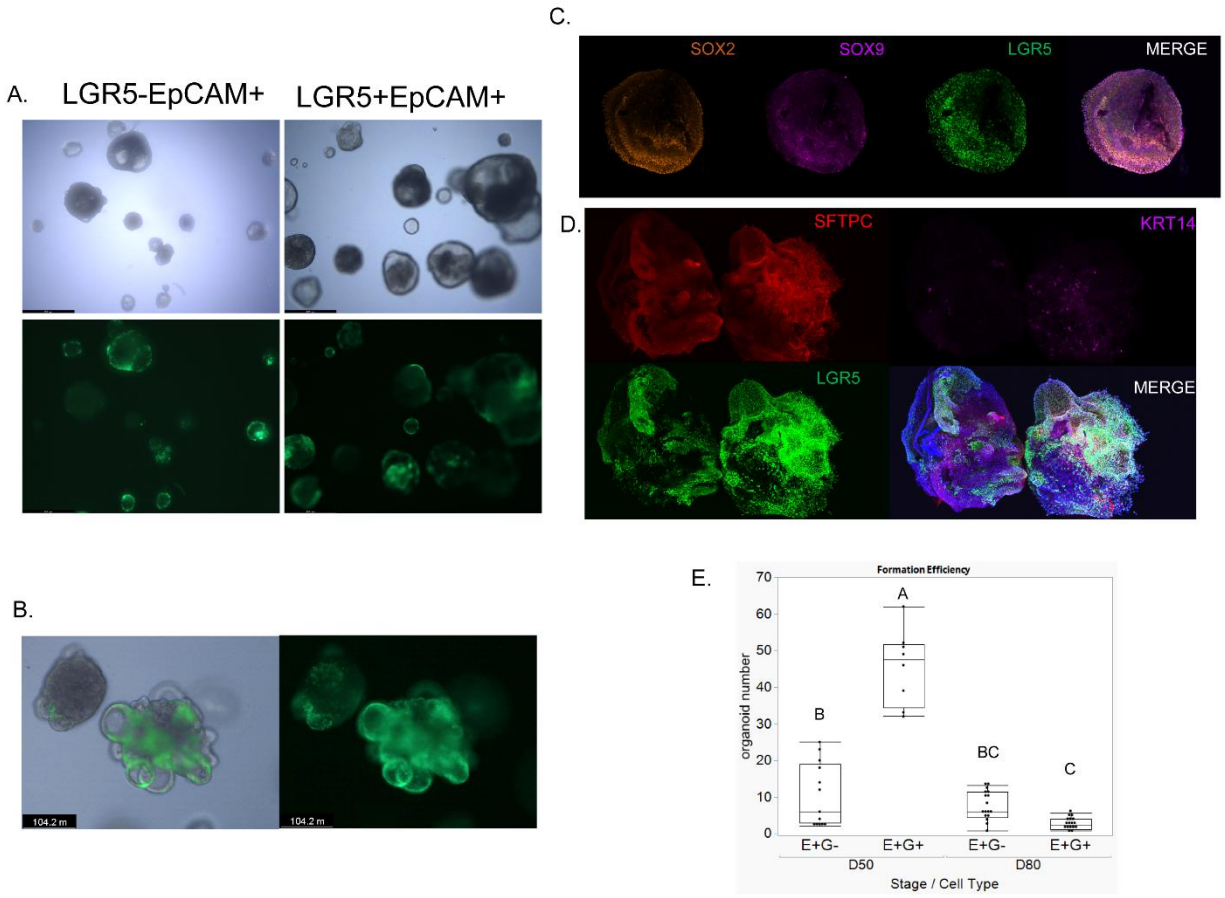


Figure 12. Characterization of fetal lung epithelial organoids. A) Brightfield and fluorescent microscopy images of organoids from D50 fetal epithelium, LGR5- or LGR5+. Both groups show large organoids with bright GFP expression, but LGR5- organoids appear to be slightly smaller after 14 days in culture. B) By day 21, EpCAM+/GFP+ organoids are GFP bright and have multiple buds. C) Immunofluorescent staining of LGR5+EpCAM+ organoids show correlation of LGR5+ cells with SOX2 and SOX9 expression. D) D50 epithelial organoids show bright expression of SFTPC, dim expression of KRT14, and bright expression of LGR5. E) Organoid formation efficiency as number of organoids formed per 2 x 10^3 cells seed after sorting by presence or absence of EpCAM ("E") and LGR5-GFP ("G"). Comparison of all pairs was performed using a one-way ANOVA by group, and comparison of means was performed by Tukey-Kramer HSD for all pairs, n=2 biological and 10 technical replicates for each group.

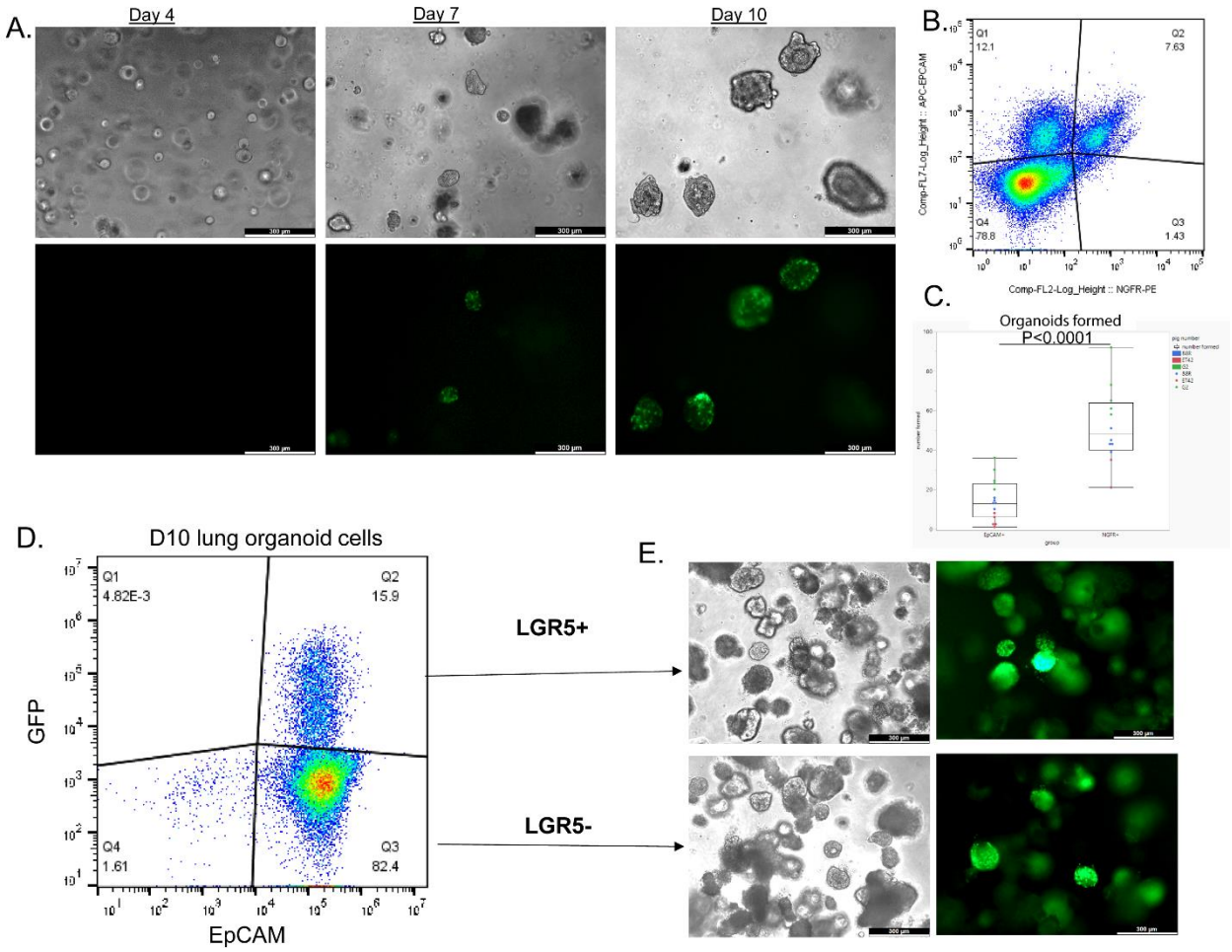


Figure 13. Transient LGR5 expression appears in adult epithelial organoids. A) All EpCAM+ (LGR5-) cells from adult lungs were seeded in Matrigel with a growth factor cocktail. While organoids remained dark after 4 days of culture, by 5-7 days, some organoids began to turn on LGR5 expression. By day 10, most organoids were bright with GFP. B) To determine differences between all EpCAM+ cells and specifically airway basal cells marked by NGFR, each group was FACS sorted, and organoid formation efficiency was quantified (C), showing that while both basal cells and other epithelial cells form organoids, the NGFR+ basal cells do so more efficiently. D) After 10 days in culture, organoids were dissociated and LGR5+ cells were sorted from LGR5- cells by FACS. E) Brightfield and immunofluorescent images of lung organoids formed from sorted D10 LGR5+ or LGR5- organoid cells.

The plastic nature of the stem cells in the lung has long been known, with airway basal cells at the top of the hierarchy, capable of differentiating a variety of cells, including secretory or club cells, a transition that is reversible in some circumstances (Davis & Wypych, 2021; Tata & Rajagopal, 2017). For example, secretary cells, after basal cell ablation in vivo or when put into basal cell growth conditions in vitro, de-differentiate back into basal cells (Tata et al., 2013). Club cells, in turn, are capable of self renewal, or generation of alveolar epithelial type 2 cells (Davis & Wypych, 2021). There is discussion in the literature of whether a bronchioalveolar stem cell (BASC) state exists in humans, a cell state between AT2 cells and club cells, with potential for generation of either. However, there has been some difficulty tracing this cell state in humans due to lack of specific markers (Chen & Fine, 2016). Recent research in our collaborating lab suggests there is an AT0 state in humans, which would be the equivalent of a BASC, that displays both secretary cell and alveolar markers.

Initially as a baseline for comparison to the fetal organoids, we seeded adult epithelial cells in matrigel with the same growth factors and growth medium. Surprisingly, after 5-7 days in culture, the majority of the organoids displayed at least some level of LGR5 expression (Figure 13A), and by day ten, approximately 15% of all organoid cells were LGR5+ (Figure 13D). To ask more specifically which cells were turning on LGR5 expression and forming organoids, we sorted lung cells based on NGFR, a known basal cell marker, or EpCAM only (Figure 13B). The NGFR+ basal cell population formed significantly more organoids per well at the same concentration (Figure 13C), but there were no significant differences between the phenotype of the organoids, and both expressed LGR5, suggesting that the basal cells or secretory cells had the same fate in these conditions. We next asked if there were any differences between the behavior

of the LGR5+ or LGR5- cells in the organoids, and hence FACS sorted and replated; there were minimal differences between the two populations, and organoids from the LGR5- population began to express LGR5 (Figure 13E). These two things indicate that in vitro, expression of LGR5 is transient, and that basal cells give rise to a cell population that expresses LGR5. We hypothesized that the organoid media conditions represented a regenerative state, and therefore we next asked what cell types were being generated. As defined above by expression of KRT14 and SFTPC, these organoids had a bronchioalveolar phenotype (Figure 14A). When co-staining with SFTPC, we noticed that LGR5 expression seemed to be correlated with SFTPC expression. This led us to ask what was the cell state of the LGR5+ cells, and FACS sorted organoid cells based on GFP expression (as in 13D) and subjected each population to RNAseq analysis. Among the top significantly upregulated genes were SFTPB, SFTPC, SCGB1A1, and SCGB3A2, all markers of club/secretory cells and AT2 cells, which are also co-expressed in LGR5+ cells in bud tip development. Among the downregulated genes were AGER (AT1 cell marker), IL33 and IGFBP4 (both basal cell markers) (He et al., 2022) (14B, C). Since we detected markers of AT2 and AT1 cells by immuno-fluorescence and by RNAseq after starting with basal cells or mixed basal and club cells, and since LGR5+ cells were downregulated in markers of AT1 and basal cells, we hypothesized that LGR5 was expressed by secretory cells undergoing a transition to an AT2 cell. If this is true, then LGR5 would not be expressed in basal cells or mixed basal and club cells that maintain a basal cell lineage in culture. To test this, we altered the media conditions to favor a basal cell phenotype by removing WNT and WNT agonist CHIR signaling. Indeed, after ten days in culture, none of the bronchiolar organoids expressed LGR5, while the bronchioalveolar controls were bright with GFP expression (Figure 14D). While more experiments need to be done in vivo using lineage tracing after injury and

during AT2 regeneration from club cells, LGR5 could potentially represent the marker for a transient state between a multi-potential basal stem cell and committed lineage progenitor.

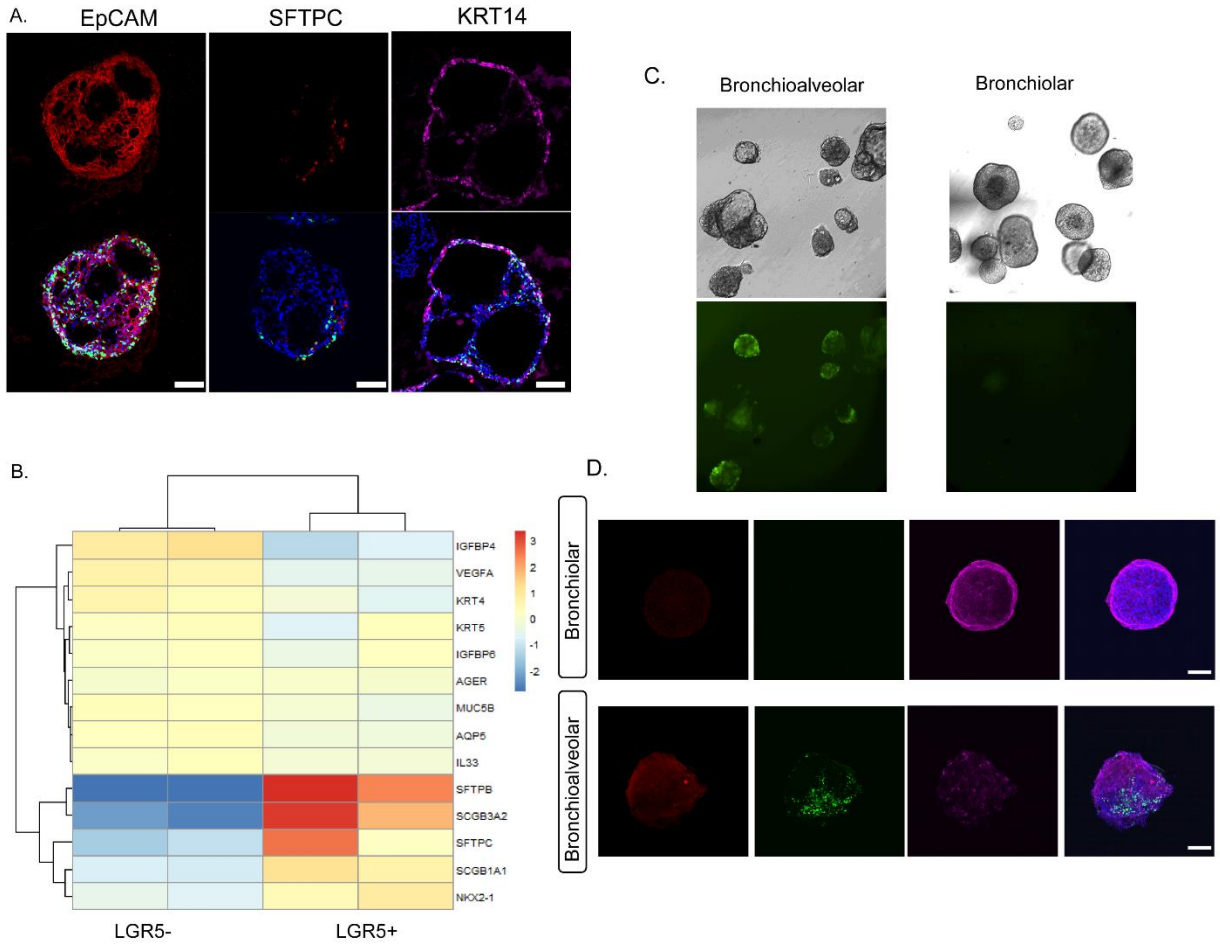


Figure 14. LGR5 is expressed in organoids with a bronchioalveolar fate, but not a bronchiolar fate. A) Initial immunofluorescent analysis of cryosectioned LGR5+ adult lung organoids imaged by confocal microscopy shows distribution of EpCAM, SFTPC, KRT14 and colocalization with LGR5. B) Differential gene expression between LGR5+ and LGR5- organoid cells (*n*=2 biological replicates) of select basal, secretory, AT2 and AT1 markers. C) Removal of WNT signaling pathway growth factors from the media shows morphological differences between bronchioalveolar and bronchiolar conditions, and absence of LGR5 expression in bronchiolar organoids. D) Immunofluorescent whole mount confocal imaging comparison of SFTPC, LGR5, and KRT14 expression between bronchioalveolar and bronchiolar organoids. Scale bar represents 100 μM.

Discussion

In this work, we have thoroughly identified novel roles for LGR5-expressing cells in development and homeostasis. While the two described populations are consistent in humans and pigs, it is unclear why the results are not consistent with mice. While LGR5 is reported in fibroblasts in mice, the roles of the murine LGR6+ cells seem to be more akin to the LGR5+ human and pig populations (Lee et al., 2017), given that they surround and support the airway niche, although the LGR6+ population overlaps significantly with the aSMA+ population in the mouse whereas the two populations are mutually exclusive in the pig. It is difficult to determine if the differences between the species are due to the differences in anatomy and associated physiology, or if the redundant function of LGRs allows for interchangeable gene function across species. However, since RSPO and WNT signaling have been studied extensively in murine lung development, it is quite interesting that LGR expression has not been reported in murine bud tip development. If the role of LGR5 in development is to help bud tip progenitors maintain self-renewing potential throughout development, perhaps the shortened developmental time frame and small lung size do not require such protective measures for its progenitor cells, although in that case there is a question as to what is the receptor for the necessary RSPO. Or perhaps LGR5 transcript is expressed at a comparatively low level, especially in the transient state as we have shown, and cytoplasmic resolution of murine models has made it difficult to elucidate.

Nevertheless, we report the role of LGR5+ mesenchymal cells in the lung as a stromal cell population supporting airways basal cells, in addition to their multi-lineage potential. There have been few reports of airway fibroblasts throughout the literature, but the majority have been concerned with the myofibroblast populations or maintenance of myofibroblasts populations

(Moiseenko et al., 2020; E. J. Walker et al., 2019), which we know to be separate from the LGR5+ population. Recent discussion after the report of a fetal airway fibroblast population has suggested that no airway fibroblast population or signaling niche has yet been reported in the adult (He et al., 2022). To our knowledge, we are among the first to identify an airway specific fibroblast population separate from myofibroblasts that is present at homeostasis, correlate this population with LGR5 expression, and define their role as a support niche cell and mesenchymal stromal cell.

A growing body of evidence has identified lung tissue resident mesenchymal stem cells (tr-MSC) (Lama et al., 2007; Rolandsson Enes et al., 2016; Sveiven & Nordgren, 2020). Recent publications have identified the lung tr-MSCs as key contributors to pulmonary fibrosis as they differentiate into accumulating myofibroblasts (Cao et al., 2018; Lemos & Duffield, 2018; Popova et al., 2010). Paradoxically, these populations have been used extensively in anti-fibrotic therapeutics and treatment of other lung diseases (Foronjy & Majka, 2012; Klein, 2021). Despite broad characterization and identification of key tr-MSC markers such as SFRP (Rolandsson Enes et al., 2016), which is expressed in the LGR5+ MSC, to our knowledge there has been very little information about where these tr-MSC reside (Klein, 2021), and there not been a connection until now identified between the tr-MSC and airway fibroblast populations. Our results suggest that if indeed tr-MSC have a role in contributing to fibrosis, LGR5+ MSC may represent major contributors, and furthermore LGR5 may represent a preventative or therapeutic pathway for restoring dysregulated cells.

On the epithelial side, we have identified a transient expression of LGR5 present in development and during the transition from basal cells to AT2 cells in vitro. Bronchioalveolar stem cells (BASC) were first identified as a novel population in mice that expresses both AT2 cell markers (SftpC) and club cell markers (Sgcb1a1), and have the capability to contribute to both an airway lineage and an alveolar population (Kim et al., 2005). While they have a minimal role in a healthy lung, they are highly active in repair. Their capability for self renewal has made them a hotly pursued topic in the name of regeneration, but also has been cited explicitly as a mechanism for adenocarcinoma formation (Jones et al., 2019; Jones-Freeman & Starkey, 2020). A major limitation for the study of these cells, cited repeatedly, has been the lack of genetic tools or markers that specifically identify the BASC population (Jones et al., 2019; Jones-Freeman & Starkey, 2020; Liu et al., 2019; Salwig et al., 2019). For this reason, there has been much debate about the existence of this population in vivo, since both putative markers are expressed consistently by other populations and cells are restricted geographically (Rawlins et al., 2009). Recently, two labs have generated complex lineage tracing strategies to track these populations in mice in vivo (Liu et al., 2019; Salwig et al., 2019) and have identified an active role in repair. However, to date, BASC have not been identified in human lungs. Since BASC have minimal role at homeostasis, and only become active upon injury, is it not possible that the BASC state may be transient? After all, it is known that AT2 cells express transient markers as they differentiate from AT2 to AT1 (Jones-Freeman & Starkey, 2020; Wu & Tang, 2021), suggesting that basal or club cells transitioning to AT2 cells (or maintaining multiple lineage potentials) may also take on a transition state. If this transient regenerative state mimics development, as has been suggested (Jones et al., 2019; Jones-Freeman & Starkey, 2020), then we would not expect murine BASC to express LGR5, which is why it has not been reported henceforth. However,

human and porcine both express LGR5 in the transient state in development, and that is repeated in in vitro regeneration, and thus our in vitro data supports that LGR5 should be considered as a potential marker of BASC in human lungs. Further supporting this, both LGR5 and BASC have separately reported roles in adenocarcinoma (Tammela et al, 2017; Zhang et al., 2016), which could suggest that LGR5 is upregulated in the case of an injury to maintain lineage potential in repair, but these same traits make this population a prime source of cancer cells. Further work needs to be done in human and porcine injured airways to evaluate whether LGR5 is expressed in BASC in the case of an injury. If so, then we have identified a novel pathway for regeneration, and a potential therapeutic target for cancer.

Methods

Transgenic porcine tissue

All experiments were carried out in strict accordance with the Institutional Animal Care and Use Committee of North Carolina State University (IACUC protocol 17-028-B), and followed the established ARRIVE guidelines (Kilkenny et al., 2010). LGR5-H2B-GFP transgenic pigs were generated by somatic cell nuclear transfer as previously described (S. C. Walker et al., 2002). Male and female transgenic offspring of the clones were used in these experiments, with no detectable differences. Experiments in this study are performed using lung tissue from 6 juvenile or adult pigs, 4 fetal D80 fetuses, and 3 fetal D50 fetuses, and data are representative of both sexes.

Immunofluorescence

Tissue was fixed in 4% paraformaldehyde and dehydrated in 30% sucrose before being embedded in OCT. Organoids were fixed in 2-4% paraformaldehyde and embedded in OCT or stained and imaged as whole mounts. Tissue or organoids were sectioned using a cryostat set to

20uM. For immunohistochemistry, sections were blocked with 0.4% Triton X-100 and IHC/ICC blocking buffer (Invitrogen) followed by incubation of primary antibody in 0.2% Triton-X in a 1:1 ratio of blocking buffer and PBS, for 1.5 hours at room temperature or 4℃ overnight. Primary antibodies were applied at a 1:200 ratio; anti-KI67(Invitrogen #MA5-14520), anti-EpCAM (Abcam #ab71916), anti-SOX9 (eBioscience #14-9765-82), anti-SOX2 (Abcam #ab97959), anti-Alpha-Smooth Muscle Actin (Invitrogen #14-9760-82), anti-Vimentin (Santa Cruz #sc-6260), anti-PDGFRa (Bioss #BSM-52829R), anti-Prosurfactant Protein C (Abcam #ab90716). Sections were stained with secondary antibodies (1:500) for 1 hour at room temperature and mounted with Prolong Diamond Antifade Mount with DAPI. For whole mount imaging, organoids or tissue sections were incubated with block, antibodies, and washes sequentially in eppendorf tubes and mounted on a slide immediately before analysis, and the addition of incubation with NucBlue (Invitrogen) with the secondary antibody. Slides were imaged with an Olympus Fluoview 3000 upright confocal microscope.

RNAscope in situ hybridization

RNAscope was performed according to the manufacturer's instructions (ACD Bio). Briefly, paraffin embedded lung tissue from human fetus or adult was sectioned at 7um. Slides were deparaffinized with xylene, then heat treated followed by protease digestion. The tissue was hybridized with a 10 ZZ probe targeting the 560-1589 region of Homo sapiens LGR5 mRNA. A positive control probe against human (139-989 region) cyclophilin B or as a negative control, a probe targeting the bacterial gene dapB were tested. After chromogenic development, slides were washed and imaged by confocal microscopy.

Tissue Clearing

Here, we used the BoneClear procedure for clearing the porcine lungs; therefore, we will only outline differences between our implementation versus the original protocol. Due to the size and complexity of pigs, perfusion step was not conducted. We skipped the decalcification step. Also, we increased the immunostaining time from 3 to 7 days and skipped the agarose embedding step. We then used the custom light-sheet microscopy (Moatti et al., 2020) to image the cleared lungs. The sample was mounted inside a custom chamber (aluminum and glass) to be immersed in 100% DBE. A silicone membrane was employed in front of the objective lens (10×/numerical aperture (NA) 0.6, Olympus; XLPLN10XSVMP-2) to minimize the potential damage to the objective lens when immersed in 100% DBE. The working distance of the detection objective and the field of view was 8 mm and 1.77 mm2, respectively.

Cell Isolation and Cell Sorting

Cells were isolated from lungs of adult or fetal LGR5 pigs. 2cm2 pieces of lung were collected in cold PBS and moved into a sterile hood. Lung tissue was then minced using scissors or razor blade and then washed twice with cold PBS to remove blood and immune cells. Minced tissues were then incubated for 1 hour in a roller incubator at 37℃ in an enzymatic digest buffer containing 10mg/mL Dipsase (Invitrogen), 2mg/mL DNase I (STEMCELL Technologies), and 5mg/mL (adult) or 2mg/mL (fetal) Collagenase Type II (Sigma). The digestion mixture was then strained using a 70μM cell strainer and washed with wash buffer (PBS, 5% FBS, 1% antibiotic-antimycotic) and centrifuged. For adult tissue, pellets were incubated in Red Blood Cell Lysis Buffer for 10 minutes at room temperature with gentle inversion. Following washing of cells, the pellet was resuspended in freezing media (65% a-MEM, 30% FBS, and 5% DMSO) at a concentration of no more than 5 x 106 cells/mL. Cells were then frozen overnight in -80C then

transferred to liquid nitrogen for long term preservation. For fluorescent activated cell sorting, cells were suspended in MACS buffer (Miltenyi) and incubated with 10µl/ml anti-EpCAM-APC (Thermo Scientific 17-5791-83) and/or anti-NGFR-PE (eBioscience #12-9400-41). Propidium iodide was added for dead cell exclusion immediately before fluorescence activated cell sorting (Beckman Coulter MoFlo XPD).

Organoid Culture

For co-culture conditions, LGR5+ mesenchymal cells were mixed at a 1:1 ratio (2,000 cells each) of EpCAM+ airway cells and resuspended in a 25 ul droplet of reduced growth factor Matrigel (Corning). Droplets were allowed to solidify upside down at 37℃ for 25 minutes and then were overlaid with 250uL of organoid media. Basic growth media consisted of DMEM/F12 (Life Technologies #12634-010), Glutamax (Life Technologies #35050-061), Hepes (Life Technologies #15630-106), N2 (Life Technologies #17502-048), B27 (Life Technologies #12587-010), 2% fetal bovine serum and 1% antibiotic-antimycotic (Gibco #15240-096). For conditions with growth factors, EpCAM+ cells were seeded at 2,000-5,000 cells per well. For bronchioalveolar conditions, LWRN-conditioned media was mixed with basic media at a 1:1 ratio, and a growth factor master mix was added for final concentrations of CHIR99021 (PeproTech #2520691) (5 µM), EGF (PeproTech AF-100-15) (100 ng/ mL), Nicotinamide (Life Technologies #17502-048) (2mM), LY2157299 (Cayman Chemical Company #15312) (1 µM), A-83-01 (2 nM), Y27632 (20uM), and FGF10 (16.6ng/mL). For bronchiolar conditions, no LWRN was added and the mastermix contained added R-Spondin (R&D Systems # 4645RS/CF) (2 µg/mL), reduced CHIR99021(1 µM), EGF (100 ng/ mL), (Nicotinamide (2mM), LY2157299(1µM), A-83-01 (2nM), Y27632 (20µM), and FGF10 (16.6ng/mL). Media was changed every other day, by adding 250 µL of media and 3 µL of growth factor master mix. For

dissociation for FACS, Matrigel dome containing organoids was disrupted using pipette tip and resuspended in 250μL of ice cold Matrigel Recovery Solution, then incubated for 20 minutes at 4℃. Once free of Matrigel, organoids were resuspended in TrypLE Select and incubated with shaking for 10-15 minutes at 37℃. Wash buffer or warm basic media was added to stop enzymatic dissociation and dissociated organoids were filtered through a 40 μM filter to remove any clumps before further use.

RNA sequencing analysis

For bulk RNAsequencing, each group was Ultra-low input RNA-seq

Library Preparation, and HiSeq Sequencing RNA library preparations and sequencing reactions were conducted at GENEWIZ, LLC. (South Plainfield, NJ, USA). For fetal samples, but not for postnatal or organoids, SMART-Seq v4 Ultra Low Input Kit for Sequencing was used for full-length cDNA synthesis and amplification (Clontech, Mountain View, CA), and Illumina Nextera XT library was used for sequencing library preparation. The final library was assessed with Agilent TapeStation. The sequencing libraries were multiplexed and clustered on a flowcell. After clustering, the flowcell was loaded on the Illumina HiSeq instrument according to the manufacturer's instructions. The samples were sequenced using a 2x150 Paired End (PE) configuration. Image analysis and base calling were conducted by the HiSeq Control Software (HCS). Raw sequence data (.bcl files) generated from Illumina HiSeq was converted into fastq files and de-multiplexed using Illumina's bcl2fastq 2.17 software. Sequence reads were trimmed to remove possible adapter sequences and nucleotides with poor quality using Trimmomatic v.0.36. The trimmed reads were mapped to the Sus scrofa reference genome available on ENSEMBL using the STAR aligner v.2.5.2b. The STAR aligner is a splice aligner that detects splice junctions and incorporates them to help align the entire read sequences. BAM files were

generated as a result of this step. Unique gene hit counts were calculated by using feature Counts from the Subread package v.1.5.2. Only unique reads that fell within exon regions were counted. After extraction of gene hit counts, the gene hit counts table was used for downstream differential expression analysis. Using DESeq2, a comparison of gene expression between the groups of samples was performed. The Wald test was used to generate p-values and Log2 fold changes. Genes with adjusted p-values < 0.05 and absolute log2 fold changes > 1 were called as differentially expressed genes for each comparison. Gene set enrichment analysis was performed using GSEA (Broad Institute) (Mootha et al., 2003; Subramanian et al., 2005) and ClusterProfiler (G. Yu et al., 2012).

Single cell RNAsequencing

Single cell suspensions from D80 fetuses were loaded into a Chromium Controller (10x Genomics) for gel beads in emulsion. After scRNA-seq library preparation with the Chromium SIngle Cell 3' Library & Gel Bead Kit, quality analysis was performed using a bioanalyzer. Libraries were sequenced using the NextSeq500 (~400M repair pairs). Reads were mapped using the Sus scrofa 11.1 assembly, and raw counts were imported into R for analysis with the Seurat package. After stringent quality thresholds based on reads per cell, UMI identifiers were removed and 52384 cells were retained for analysis. Cell types were clustered by nearest neighbor and visualized using UMAP analysis in Seurat. PretsoMarkers was used to identify the most significant markers for each cluster. Mapping LGR5 expression using single cell RNAsequencing of established human datasets was performed using previously acquired datasets in the lung (Vieira Braga et al., 2019)(Q. Yu et al., 2021), and figures were generated with the associated webtools (https://www.lungcellatlas.org/) and (https://github.com/Camp-Lab/GutTubeR/).

REFERENCES

Barker, N., Tan, S., & Clevers, H. (2013). Lgr proteins in epithelial stem cell biology. *Development* , *140*(12), 2484–2494.

Bell, S. M., Schreiner, C. M., Wert, S. E., Mucenski, M. L., Scott, W. J., & Whitsett, J. A. (2008). R-spondin 2 is required for normal laryngeal-tracheal, lung and limb morphogenesis. *Development* , *135*(6), 1049–1058.

Benahmed, M. A., Elbayed, K., Daubeuf, F., Santelmo, N., Frossard, N., & Namer, I. J. (2014). NMR HRMAS spectroscopy of lung biopsy samples: Comparison study between human, pig, rat, and mouse metabolomics. *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, *71*(1), 35–43.

Boquest, A. C., Shahdadfar, A., Brinchmann, J. E., & Collas, P. (2006). Isolation of stromal stem cells from human adipose tissue. *Methods in Molecular Biology* , *325*, 35–46.

Cao, H., Wang, C., Chen, X., Hou, J., Xiang, Z., Shen, Y., & Han, X. (2018). Inhibition of Wnt/β-catenin signaling suppresses myofibroblast differentiation of lung resident mesenchymal stem cells and pulmonary fibrosis. *Scientific Reports*, *8*(1), 13644.

Chen, F., & Fine, A. (2016). Stem Cells in Lung Injury and Repair. *The American Journal of Pathology*, *186*(10), 2544–2550.

Davis, J. D., & Wypych, T. P. (2021). Cellular and functional heterogeneity of the airway epithelium. *Mucosal Immunology*, *14*(5), 978–990.

Foronjy, R. F., & Majka, S. M. (2012). The potential for resident lung mesenchymal stem cells to promote functional tissue regeneration: understanding microenvironmental cues. *Cells* , *1*(4), 874.

He, P., Lim, K., Sun, D., Pett, J. P., Jeng, Q., Polanski, K., Dong, Z., Bolt, L., Richardson, L., Mamanova, L., Dabrowska, M., Wilbrey-Clark, A., Madissoon, E., Tuong, Z. K., Dann, E., Suo, C., Kai'En, I. G., He, X., Barker, R., … Rawlins, E. L. (2022). A human fetal lung cell atlas uncovers proximal-distal gradients of differentiation and key regulators of epithelial fates. In *bioRxiv* (p. 2022.01.11.474933). https://doi.org/10.1101/2022.01.11.474933

Huch, M., Dorrell, C., Boj, S. F., van Es, J. H., Li, V. S. W., van de Wetering, M., Sato, T., Hamer, K., Sasaki, N., Finegold, M. J., Haft, A., Vries, R. G., Grompe, M., & Clevers, H. (2013). In vitro expansion of single Lgr5+ liver stem cells induced by Wnt-driven regeneration. *Nature*, *494*(7436), 247–250.

Hussain, M., Xu, C., Lu, M., Wu, X., Tang, L., & Wu, X. (2017). Wnt/β-catenin signaling links embryonic lung development and asthmatic airway remodeling. *Biochimica et Biophysica Acta, Molecular Basis of Disease*, *1863*(12), 3226–3242.

Jones-Freeman, B., & Starkey, M. R. (2020). Bronchioalveolar stem cells in lung repair, regeneration and disease. *The Journal of Pathology*, *252*(3), 219–226.

Jones, M., Zhang, J.-S., & Bellusci, S. (2019). Bronchioalveolar stem cells vindicated! *Biotarget*, *3*, 4–4.

Judge, E. P., Hughes, J. M. L., Egan, J. J., Maguire, M., Molloy, E. L., & O'Dea, S. (2014). Anatomy and bronchoscopy of the porcine lung. A model for translational respiratory medicine. *American Journal of Respiratory Cell and Molecular Biology*, *51*(3), 334–343.

Kadzik, R. S., Cohen, E. D., Morley, M. P., Stewart, K. M., Lu, M. M., & Morrisey, E. E. (2014). Wnt ligand/Frizzled 2 receptor signaling regulates tube shape and branch-point formation in the lung through control of epithelial cell shape. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(34), 12444–12449.

Kathiriya, J. J., Wang, C., Zhou, M., Brumwell, A., Cassandras, M., Le Saux, C. J., Cohen, M., Alysandratos, K.-D., Wang, B., Wolters, P., Matthay, M., Kotton, D. N., Chapman, H. A., & Peng, T. (2022). Human alveolar type 2 epithelium transdifferentiates into metaplastic KRT5+ basal cells. *Nature Cell Biology*, *24*(1), 10–23.

Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M., & Altman, D. G. (2010). Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. *Journal of Pharmacology & Pharmacotherapeutics*, *1*(2), 94–99.

Kim, C. F. B., Jackson, E. L., Woolfenden, A. E., Lawrence, S., Babar, I., Vogel, S., Crowley, D., Bronson, R. T., & Jacks, T. (2005). Identification of bronchioalveolar stem cells in normal lung and lung cancer. *Cell*, *121*(6), 823–835.

Klein, D. (2021). Lung Multipotent Stem Cells of Mesenchymal Nature: Cellular Basis, Clinical Relevance, and Implications for Stem Cell Therapy. *Antioxidants & Redox Signaling*, *35*(3), 204–216.

Lama, V. N., Smith, L., Badri, L., Flint, A., Andrei, A.-C., Murray, S., Wang, Z., Liao, H., Toews, G. B., Krebsbach, P. H., Peters-Golden, M., Pinsky, D. J., Martinez, F. J., & Thannickal, V. J. (2007). Evidence for tissue-resident mesenchymal stem cells in human adult lung from studies of transplanted allografts. *The Journal of Clinical Investigation*, *117*(4), 989–996.

Lee, J.-H., Tammela, T., Hofree, M., Choi, J., Marjanovic, N. D., Han, S., Canner, D., Wu, K., Paschini, M., & Bhang, D. H. (2017). Anatomically and functionally distinct lung mesenchymal populations marked by Lgr5 and Lgr6. *Cell*, *170*(6), 1149–1163. e12.

Lemos, D. R., & Duffield, J. S. (2018). Tissue-resident mesenchymal stromal cells: Implications for tissue-specific antifibrotic therapies. *Science Translational Medicine*, *10*(426). https://doi.org/10.1126/scitranslmed.aan5174

Liu, Q., Liu, K., Cui, G., Huang, X., Yao, S., Guo, W., Qin, Z., Li, Y., Yang, R., Pu, W., Zhang, L., He, L., Zhao, H., Yu, W., Tang, M., Tian, X., Cai, D., Nie, Y., Hu, S., … Zhou, B. (2019). Lung regeneration by multipotent stem cells residing at the bronchioalveolar-duct junction. *Nature Genetics*, *51*(4), 728–738.

Lunney, J. K., Van Goor, A., Walker, K. E., Hailstock, T., Franklin, J., & Dai, C. (2021). Importance of the pig as a human biomedical model. *Science Translational Medicine*, *13*(621), eabd5758.

Meurens, F., Summerfield, A., Nauwynck, H., Saif, L., & Gerdts, V. (2012). The pig: a model for human infectious diseases. *Trends in Microbiology*, *20*(1), 50–57.

Moatti, A., Cai, Y., Li, C., Sattler, T., Edwards, L., Piedrahita, J., Ligler, F. S., & Greenbaum, A. (2020). Three-dimensional imaging of intact porcine cochlea using tissue clearing and custom-built light-sheet microscopy. *Biomedical Optics Express*, *11*(11), 6181–6196.

Moiseenko, A., Vazquez-Armendariz, A. I., Kheirollahi, V., Chu, X., Tata, A., Rivetti, S., Günther, S., Lebrigand, K., Herold, S., Braun, T., Mari, B., De Langhe, S., Kwapiszewska, G., Günther, A., Chen, C., Seeger, W., Tata, P. R., Zhang, J.-S., Bellusci, S., & El Agha, E. (2020). Identification of a Repair-Supportive Mesenchymal Cell Population during Airway Epithelial Regeneration. *Cell Reports*, *33*(12), 108549.

Mootha, V. K., Lindgren, C. M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., Houstis, N., Daly, M. J., Patterson, N., Mesirov, J. P., Golub, T. R., Tamayo, P., Spiegelman, B., Lander, E. S.,

Hirschhorn, J. N., … Groop, L. C. (2003). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics*, *34*(3), 267–273.

Nikolić, M. Z., Caritg, O., Jeng, Q., Johnson, J.-A., Sun, D., Howell, K. J., Brady, J. L., Laresgoiti, U., Allen, G., Butler, R., Zilbauer, M., Giangreco, A., & Rawlins, E. L. (2017). Human embryonic lung epithelial tips are multipotent progenitors that can be expanded in vitro as long-term self-renewing organoids. *eLife*, *6*. https://doi.org/10.7554/eLife.26575

Popova, A. P., Bozyk, P. D., Goldsmith, A. M., Linn, M. J., Lei, J., Bentley, J. K., & Hershenson, M. B. (2010). Autocrine production of TGF-β1 promotes myofibroblastic differentiation of neonatal lung mesenchymal stem cells. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, *298*(6), L735–L743.

Prior, N., Hindley, C. J., Rost, F., Meléndez, E., Lau, W. W. Y., Göttgens, B., Rulands, S., Simons, B. D., & Huch, M. (2019). Lgr5+ stem and progenitor cells reside at the apex of a heterogeneous embryonic hepatoblast pool. *Development* , *146*(12). https://doi.org/10.1242/dev.174557

Raslan, A. A., & Yoon, J. K. (2020). WNT Signaling in Lung Repair and Regeneration. *Molecules and Cells*, *43*(9), 774–783.

Rawlins, E. L., Okubo, T., Xue, Y., Brass, D. M., Auten, R. L., Hasegawa, H., Wang, F., & Hogan, B. L. M. (2009). The Role of Scgb1a1 Clara Cells in the Long-Term Maintenance and Repair of Lung Airway, but Not Alveolar, Epithelium. In *Cell Stem Cell* (Vol. 4, Issue 6, pp. 525–534). https://doi.org/10.1016/j.stem.2009.04.002

Rolandsson Enes, S., Andersson Sjöland, A., Skog, I., Hansson, L., Larsson, H., Le Blanc, K., Eriksson, L., Bjermer, L., Scheding, S., & Westergren-Thorsson, G. (2016). MSC from fetal

and adult lungs possess lung-specific properties compared to bone marrow-derived MSC. *Scientific Reports*, *6*, 29160.

Sachs, N., Papaspyropoulos, A., Zomer-van Ommen, D. D., Heo, I., Böttinger, L., Klay, D., Weeber, F., Huelsz-Prince, G., Iakobachvili, N., Amatngalim, G. D., de Ligt, J., van Hoeck, A., Proost, N., Viveen, M. C., Lyubimova, A., Teeven, L., Derakhshan, S., Korving, J., Begthel, H., … Clevers, H. (2019). Long-term expanding human airway organoids for disease modeling. *The EMBO Journal*, *38*(4). https://doi.org/10.15252/embj.2018100300

Salwig, I., Spitznagel, B., Vazquez-Armendariz, A. I., Khalooghi, K., Guenther, S., Herold, S., Szibor, M., & Braun, T. (2019). Bronchioalveolar stem cells are a main source for regeneration of distal lung epithelia in vivo. *The EMBO Journal*, *38*(12). https://doi.org/10.15252/embj.2019102099

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43), 15545–15550.

Sveiven, S. N., & Nordgren, T. M. (2020). Lung-resident mesenchymal stromal cells are tissue-specific regulators of lung homeostasis. *American Journal of Physiology. Lung Cellular and Molecular Physiology*, *319*(2), L197–L210.

Tammela, T., Sanchez-Rivera, F. J., Cetinbas, N. M., Wu, K., Joshi, N. S., Helenius, K., Park, Y., Azimi, R., Kerper, N. R., Wesselhoeft, R. A., Gu, X., Schmidt, L., Cornwall-Brady, M., Yilmaz, Ö. H., Xue, W., Katajisto, P., Bhutkar, A., & Jacks, T. (2017). A Wnt-producing

niche drives proliferative potential and progression in lung adenocarcinoma. *Nature*, *545*(7654), 355–359.

Tata, P. R., Mou, H., Pardo-Saganta, A., Zhao, R., Prabhu, M., Law, B. M., Vinarsky, V., Cho, J. L., Breton, S., Sahay, A., Medoff, B. D., & Rajagopal, J. (2013). Dedifferentiation of committed epithelial cells into stem cells in vivo. *Nature*, *503*(7475), 218–223.

Tata, P. R., & Rajagopal, J. (2017). Plasticity in the lung: making and breaking cell identity. *Development* , *144*(5), 755–766.

Vieira Braga, F. A., Kar, G., Berg, M., Carpaij, O. A., Polanski, K., Simon, L. M., Brouwer, S., Gomes, T., Hesse, L., Jiang, J., Fasouli, E. S., Efremova, M., Vento-Tormo, R., Talavera-López, C., Jonker, M. R., Affleck, K., Palit, S., Strzelecka, P. M., Firth, H. V., … Teichmann, S. A. (2019). A cellular census of human lungs identifies novel cell states in health and in asthma. *Nature Medicine*, *25*(7), 1153–1163.

Walker, E. J., Heydet, D., Veldre, T., & Ghildyal, R. (2019). Transcriptomic changes during TGF-β-mediated differentiation of airway fibroblasts to myofibroblasts. *Scientific Reports*, *9*(1), 20377.

Walker, S. C., Shin, T., Zaunbrecher, G. M., Romano, J. E., Johnson, G. A., Bazer, F. W., & Piedrahita, J. A. (2002). A highly efficient method for porcine cloning by nuclear transfer using in vitro-matured oocytes. *Cloning and Stem Cells*, *4*(2), 105–112.

Wang, Q., Liu, K., Yang, L., Wang, H., & Yang, J. (2019). BoneClear: whole-tissue immunolabeling of the intact mouse bones for 3D imaging of neural anatomy and pathology. *Cell Research*, *29*(10), 870–872.

Wu, H., & Tang, N. (2021). Stem cells in pulmonary alveolar regeneration. *Development* , *148*(2). https://doi.org/10.1242/dev.193458

Yu, G., Wang, L.-G., Han, Y., & He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: A Journal of Integrative Biology*, *16*(5), 284–287.

Yu, Q., Kilik, U., Holloway, E. M., Tsai, Y.-H., Harmel, C., Wu, A., Wu, J. H., Czerwinski, M., Childs, C. J., He, Z., Capeling, M. M., Huang, S., Glass, I. A., Higgins, P. D. R., Treutlein, B., Spence, J. R., & Camp, J. G. (2021). Charting human development using a multi-endodermal organ atlas and organoid models. *Cell*, *184*(12), 3281–3298.e22.

Zhang, X., Xu, M., Su, S., Zhou, Z., Yang, H., Zhao, S., Zeng, D., Yang, K., Liu, Y., Wang, L., & Li, J. (2016). Lgr5-positive cells in the lung and their clinical significance in patients with lung adenocarcinoma. *Molecular and Clinical Oncology*, *5*(2), 283–288.

APPENDICES

Appendix A

In vitro validation of transgene expression in gene-edited pigs using CRIPSR Transcriptional Activators

This article was originally published in The CRISPR Journal on October 20th, 2020

https://doi.org/10.1089/crispr.2020.0037

The CRISPR Journal
Volume 3, Number 5, 2020
© Mary Ann Liebert, Inc.
DOI: 10.1089/crispr.2020.0037

The CRISPR Journal

RESEARCH ARTICLE

In Vitro Validation of Transgene Expression in Gene-Edited Pigs Using CRISPR Transcriptional Activators

Kathryn M. Polkoff,[1,2] Jaewook Chung,[1,2] Sean G. Simpson,[3,4] Katherine Gleason,[1,2] and Jorge A. Piedrahita[1,2,*]

Abstract

The use of CRISPR-Cas and RNA-guided endonucleases has drastically changed research strategies for understanding and exploiting gene function, particularly for the generation of gene-edited animal models. This has resulted in an explosion in the number of gene-edited species, including highly biomedically relevant pig models. However, even with error-free DNA insertion or deletion, edited genes are occasionally not expressed and/or translated as expected. Therefore, there is a need to validate the expression outcomes gene modifications *in vitro* before investing in the costly generation of a gene-edited animal. Unfortunately, many gene targets are tissue specific and/or not expressed in cultured primary cells, making validation difficult without generating an animal. In this study, using pigs as a proof of concept, we show that CRISPR-dCas9 transcriptional activators can be used to validate functional transgene insertion in nonexpressing easily cultured cells such as fibroblasts. This is a tool that can be used across disciplines and animal species to save time and resources by verifying expected outcomes of gene edits before generating live animals.

Downloaded by 207.244.86.173 from www.liebertpub.com at 02/07/22. For personal use only.

Introduction

The development of CRISPR-based gene editing technologies enabled complex and efficient editing of the genome: it is now possible to add, edit, and/or delete DNA sequences of living cells and gametes of multiple species and phyla. In animals, this has been applied to genetically modified model organisms so as to better understand the function of genes, or to develop animals with improved agricultural production traits.[1–3] Gene editing with tools such as CRISPR, TALENs, and other RNA-guided endonucleases has unlocked the ability to perform germline gene knockouts, knock-ins, and base editing at high efficiency in almost any species.[4]

All RNA-guided endonucleases rely on a site-specific DNA break to initiate genomic repair. This break is repaired by homology-directed repair or nonhomologous end-joining. In gene-editing processes, the homology-directed repair mechanism of the cell can be hijacked by supplying the cell with an alternate repair template containing an altered DNA sequence or inserted gene, resulting in a knock-in. Alternatively, the nonhomologous end-joining approach is applied when the error-prone cellular machinery repairs the DNA with bases inserted or missing, which usually causes frameshift mutations thought to result in gene knockouts.[5]

Although in theory, insertion-deletion edits cause frameshift mutations and a premature stop codon,[6] many laboratories have reported of alternate downstream start codons, active truncated variants, alternative splicing, or in frame exon-skipping that interfere with the complete knockout phenotype depending on the location of the indel.[7–17] This results in unwanted and complex phenotypes that are costly and time consuming to identify and resolve.

In large animals such as nonhuman primates, pigs, cattle, and sheep, this represents a large monetary and time investment due to high husbandry costs and long gestation lengths (such as almost 5.5 months in macaques, 4 months in a pig, and 9 months in a cow). There is,

[1]*Comparative Medicine Institute, North Carolina State University, Raleigh, North Carolina, USA;* [2]*Department of Molecular Biomedical Sciences, College of Veterinary Medicine, North Carolina State University, Raleigh, North Carolina, USA;* [3]*Deparment of Animal and Avian Sciences, University of Maryland, College Park, Maryland, USA; and* [4]*RenOVAte Biosciences, Inc., Reisterstown, Maryland, USA.*

*Address correspondence to: *Jorge A. Piedrahita, PhD, Department of Molecular Biomedical Sciences, College of Veterinary Medicine, North Carolina State University, Biomedical Partnership Center, Room 51, 1060 William Moore Drive, Raleigh, NC 27607, USA,* E-mail: japiedra@ncsu.edu

Downloaded by 207.244.86.173 from www.liebertpub.com at 02/07/22. For personal use only.

therefore, a need to develop a method to test the gene expression profile of a gene-edited organisms before generating offspring or performing experiments. At present, one of the most commonly used methods for generating precise gene knockouts or knock-ins in species such as pigs and cattle is by the use of somatic cell nuclear transfer (SCNT) where somatic cells are modified *in vitro* before being used for SCNT to generate offspring.[18–20] Having an edited cell line before SCNT provides the opportunity to examine the effect of the DNA edit on gene expression before generating a live animal. Unfortunately, most of the target genes of interest are not expressed in fibroblasts, one of the most commonly used SCNT donors. As proof of concept, we use two DNA-edited pig cell lines designed to express a cell lineage marker (Histone 2B-GFP or cytoplasmic GFP) under the control of a stem cell-specific promoter (*LGR5*). LGR5 is a marker of stem cells in a variety of tissues, including skin, gastrointestinal track, lung, and others.[21] SCNT offspring from one line did not express the transgene properly for unknown reasons, requiring us to generate a second line that demonstrated the desired phenotypic change.

In this study, we describe a system using CRISPR-dCas9 transcriptional activators (TAs) as a method to transcriptionally activate transgene expression in a non-expressing cell type. Our results show that TAs used *in vitro* can recapitulate the transgene expression profile of the gene of interest from *in vivo* tissue of transgenic offspring. Based on this, we propose that TAs could be a useful screening tool to confirmed desired expression profile of an intentional genomic alteration, such as a knockout or knock-in, before generation of a gene-edited animal, which is especially useful for laboratories investing time and resources on generating gene-edited large animals.

Materials and Methods

Animal welfare

This study was performed within strict accordance of the Institutional Animal Care and Use Committee at North Carolina State University (approved IACUC protocols 14-067-B and 17-082-B). The animals used in this study were generated by SCNT using surrogate gilts from the university herd. Euthanasia was performed by sedation followed by intravenous injection of sodium pentobarbital, which meets the recommended guidelines of the American Veterinary Medical Association for porcine euthanasia.

Generation of gene-edited pigs by SCNT

Two gene-edited pig models were generated. For model I, IRES-GFP, TALENs were designed to elicit a double-stranded break in the 3′ UTR of the porcine *LGR5* gene (TALEN sites: 5′ TATAATTTGTTCCGCTAC and 3′ AAATCCGAATGGACTTAG), roughly 20 bp downstream from the stop codon. A homology directed repair template containing the internal ribosomal entry site and green fluorescent protein (IRES-GFP) sequence and regions of homology surrounding the cut site was used to mediate gene knock-in (Fig. 1A).

For model II, H2B-GFP, CRISPR-Cas9 nuclease was used to create a double-stranded break in the genomic in exon 1 of the porcine *LGR5* gene (CRISPR target: ACCATGGACACCTCCTCGGT). A homology-directed repair template plasmid containing histone 2b and green fluorescent protein (H2B-GFP) flanked by 1000 bp homology arms flanking the cut site was co-transfected with the Cas9 (Addgene #72247) and gRNA (Addgene #43860) plasmids, and cells were seeded at low density for colony outgrowth (Fig. 1F).

Pig fetal fibroblasts isolated from embryonic day 42 fetuses were used for gene editing and SCNT. After transfection and single cell cloning, colonies were genotyped by polymerase chain reaction (PCR) and sequencing to verify successful targeted transgene integration before SCNT. SCNT was completed as previously described[22] and zygotes were surgically transferred into a surrogate and carried until term.

IHC and cell analysis of SCNT offspring

Skin and intestine were isolated immediately after euthanasia and fixed overnight in 4% paraformaldehyde. Tissues were dehydrated using 10%, 20%, and 30% sucrose solutions and then embedded in optimal cutting temperature compound. Cryosections were stained and mounted with Prolong Antifade Mountant with DAPI (ThermoFisher) and analyzed by fluorescent microscopy (Olympus).

TA gRNA design

Within the 1000 bp region directly upstream of the *LGR5* transcriptional start site, 5 gRNA were designed using Benchling Biology Software, targeting 20 nucleotides on either the forward or reverse strand with the protospacer adjacent motif (PAM) sequence of NGG. CRISPR target sites include guide 1: ACAAGATTTGCTCCTC ACTG, guide 2: GTCCCGCATTGTTCTACTAG, guide 3: CTCCAATGCTGTCTAACCCA, guide 4: CGGACA CAAGCAGACGCACA, and guide 5: TTTCTCCACT CCGCGGCTGG. Complementary oligonucleotides were ordered from IDT, annealed, and cloned into plasmid MLM3636 (Gift from Keith Joung; Addgene #43860).[23] In brief, 1 μg of MLM3636 backbone was digested with *Bsm*bI enzyme. Complementary oligos with 4 bp overhangs matching overhangs from digested backbone were

Downloaded by 207.244.86.173 from www.liebertpub.com at 02/07/22. For personal use only.

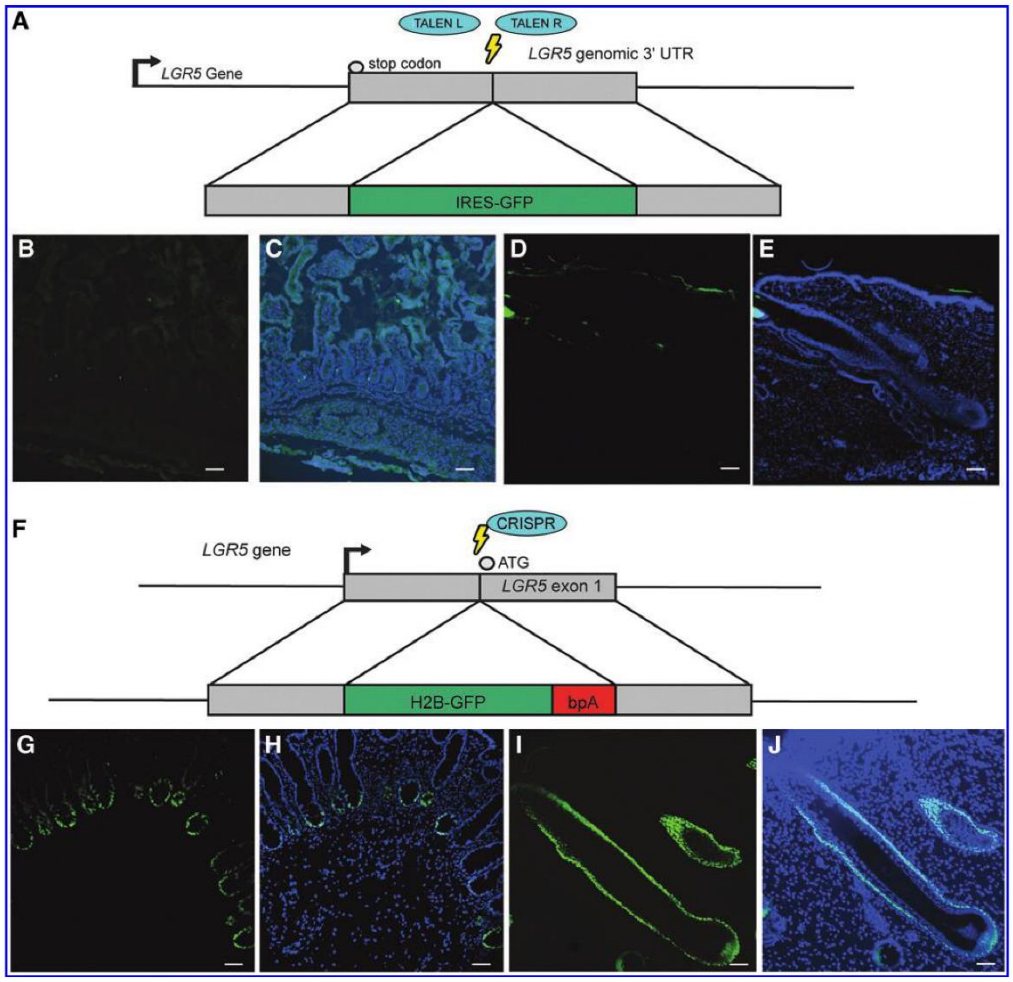


FIG. 1. GFP detection in model I and model II. GFP is not detected in adult stem cell populations in IRES-GFP model. **(A)** Schematic of knock-in strategy for model I. Cryosections of intestine **(B, C)** or skin **(D, E)** were evaluated for GFP expression **(B, D)** or co-stained with DAPI **(C, E)**. GFP is expressed in expected adult stem cell population in H2B-GFP model. **(F)** Schematic of knock-in strategy for model II. Cryosections sections of intestine **(G, H)** or skin **(I, J)** were evaluated for natural GFP expression **(G, I)** or co-stained with DAPI **(H, J)**. Scale bar indicates 100 μm. H2B-GFP, histone 2b and green fluorescent protein; IRES-GFP, internal ribosomal entry site and green fluorescent protein.

annealed for a 10 μM concentration, combined with 20–50 ng of backbone, and incubated overnight at 16°C with T4 DNA Ligase (NEB). Product was subsequently cloned in NEB-5α competent cells, cultured, and plasmids were extracted and sequenced to verify proper sequence.

Cell culture and transfection

Pig fibroblasts were cultured at 37°C and 5% CO_2 and passaged with 0.05% Trypsin/EDTA at 80% confluence. Culture media consisted of Dulbecco's Modified Eagle Medium (Gibco) supplemented with 15% fetal calf serum

Downloaded by 207.244.86.173 from www.liebertpub.com at 02/07/22. For personal use only.

(Gemini) and 1% penicillin/streptomycin. Cells were transfected by nucleofection (Lonza) using the U-023 setting. For each transfection, 5×10^5 cells of each group were transfected with 1 μg total of DNA: 500 ng of 5 gRNA plasmids (100 ng each) generated as described earlier, in combination with 500 ng VP64-dCas9 (Gift from Keith Joung; Addgene #47107). As a positive control, 1 μg of pmaxGFP (Lonza) was transfected for cytoplasmic GFP expression. Detection of GFP was analyzed by epifluorescent microscopy (Olympus) followed by flow cytometry; at least 10,000 events were recorded for each treatment (Cytoflex; Beckman Coulter).

Reverse transcription quantitative polymerase chain reaction analysis

RNA was extracted from fresh cell pellets with Quick-RNA Microprep Kit (Zymo). cDNA was synthesized (AffinityScript Multiple Temperature cDNA Synthesis Kit) according to the manufacturer's instruction. For quantitative polymerase chain reaction (qPCR) amplification, cDNA and primers were added to SYBR green mastermix (Bio-Rad) and amplified according to the manufacturer's instructions using a two-step amplification, denaturing at 98°C and annealing/extending at 60°C. Listed from 5′ to 3′, primers for *LGR5* were forward: CCTTGGCCCTGAAC AAAATA, reverse: ATTTCTTTCCCAGGGAGTGG. For GAPDH forward: ATCCTGGGCTACACTGAG GAC, reverse: AAGTGGTCGTTGAGGGCAATG. All reactions were performed in duplicate. Amplification was quantified using a delta-delta CT analysis, normalizing to GAPDH and nontransfected wild-type cells.

Statistical analysis

All results are shown as mean with standard error. Statistical analysis was performed using JMP Pro 14 (SAS Institute). When appropriate, a one-way paired Student's *t*-test was used to test for significance.

Results

Model I: 3′-LGR5-IRES-GFP. GFP is not detected in cells/tissues from gene-edited pig

We first attempted to tag cells that express *LGR5* with a cytoplasmic GFP marker. We began by integrating a DNA construct containing an IRES-GFP into the 3′ untranslated region of the endogenous pig *LGR5* locus, directly downstream of the final exon. To do this, we designed TALE Nucleases to create a double-stranded break in the DNA at the desired locus, and simultaneously transfected the TALENs with a homology directed repair template into pig fetal fibroblasts (Fig. 1A). Colonies selected and screened by PCR and sequencing analysis showed the successful insertion of the transgene (Supplementary Fig. S1A), and cells were subsequently used for SCNT. However, GFP fluorescence was not detected in the expected locations of fluorescence, including the skin and gut of tissue from piglets harboring the transgene (Fig. 1B, E).

Model II: 5′-LGR5-H2B-GFP. H2B-GFP is detected in cells/tissues from gene-edited pig

Given the lack of desired phenotype and GFP expression from the first model, we decided to tag *LGR5*-expressing cells with a pig H2B-GFP fusion protein by introducing an H2B-GFP sequence, without the IRES, into one of the *LGR5* alleles. To do this, we targeted CRISPR-Cas9 to the *LGR5* locus immediately downstream of the start codon, and co-transfected with a plasmid containing the H2B-GFP coding sequence and a bovine polyA signal using a homology-directed repair template (Fig. 1F). Cells containing the desired gene edits were then screened by PCR (Supplementary Fig. S1B) and used for SCNT. Tissues from piglets born from the second genotype were then analyzed for transgene expression. Unlike model I, the phenotype in model II piglets accurately reflected the genotype and H2B-GFP was detected in tissues that express *LGR5* (Fig. 1F, J). Furthermore, the *LGR5* expression correlated with the GFP mean fluorescence intensity (Fig. 2A, B) and, therefore, we concluded that the phenotype accurately reflected the intended outcome for these pigs.

Design and test of TAs

Based on this and previous experiences by us and others,[7–17] we aspired to generate a method for testing the RNA and protein expression outcomes of a DNA edit before spending the resources to generate the piglets. Because *LGR5* is not expressed in cells such as fibroblasts cultured *in vitro*, we were not able to validate the desired phenotype until we had already invested considerable time and resources into generating and validating the pigs by SCNT. This led us to ask whether it would be possible to test the phenotype in cells before SCNT by upregulating transgene expression using a catalytically inactive dCas9 fused to the TA VP64 for targeted gene upregulation. We subsequently designed five gRNA targeted to the 1 kb region directly upstream of the transcriptional start site (Fig. 2C). All five gRNA were encoded in plasmids and co-transfected with a plasmid encoding dCas9-VP64 into fibroblasts from wild-type, model I, and model II genotypes. Cell lines showed significant upregulation of *LGR5* mRNA after transfection with the TAs as detected by qPCR (Fig. 2D).

GFP expression profile using TAs

After verifying that the transcription was upregulated, we next evaluated the GFP phenotype of *LGR5* expressing

Downloaded by 207.244.86.173 from www.liebertpub.com at 02/07/22. For personal use only.

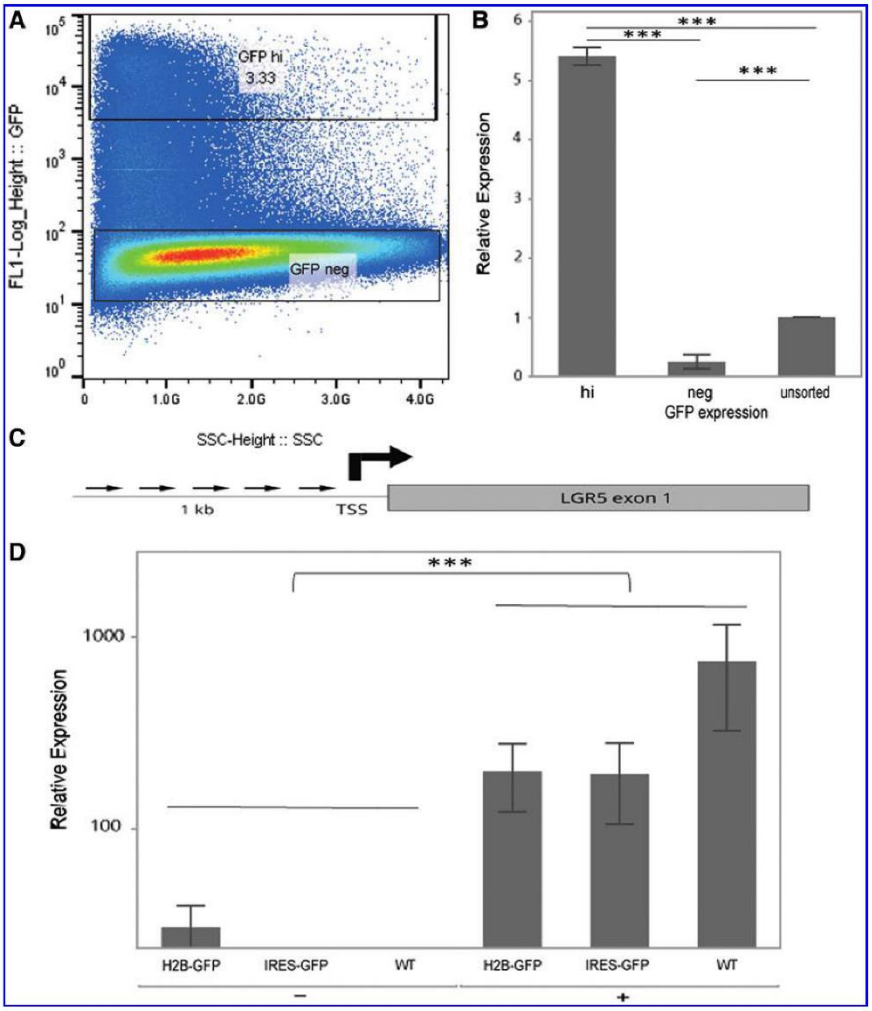


FIG. 2. *LGR5* mRNA expression in model II cells sorted by GFP expression and in all genotypes after transcriptional activation. **(A)** GFP protein expression correlates with *LGR5* mRNA expression. Epidermal cells extracted from model II—H2B-GFP were isolated and sorted based on GFP expression. **(B)** RT-qPCR analysis of relative *LGR5* expression from cells sorted in **(A)** indicates that GFP-hi cells from the model II—H2B-GFP are significantly enriched for *LGR5* expression, whereas GFP-neg cells do not express *LGR5*. Data representative of mean of two biological replicates performed in technical duplicate. **(C)** Schematic of *LGR5*-targeting TA design. **(D)** RT-qPCR analysis shows that *LGR5* expression is significantly upregulated in cell lines from model I, model II and unmodified upregulate after transfection with TAs ($n = 3$). A delta-delta CT analysis was used to normalize *LGR5* expression to GAPDH relative to nontransfected wild-type cells. ***Indicates $p < 0.05$. Error bars indicate SEM. RT-qPCR, reverse transcription quantitative polymerase chain reaction; SEM, standard error of the mean; TAs, transcriptional activators.

cells for each genotype *in vitro* after transfection with dCas9-VP64 TAs. First, we examined the presence of GFP by fluorescent microscopy before and after transcriptional activation (Fig. 3A). As shown, after TA nucleofection, the wild-type and the model I cells did not contain detectable GFP, whereas the model II cells expressed nuclear GFP as expected. We then quantified this upregulation by flow cytometry and showed that GFP fluorescence was not detected in any cell line without TA treatment (Fig. 3B). After transfection with TAs for *LGR5*, GFP was detected in an average of 35.1%±(14.3, standard deviation) of the LGR5-H2B-GFP cell line but not in the wild-type or IRES-GFP cell lines (Fig. 3C, Supplementary Fig. S2A). From this we concluded that cells

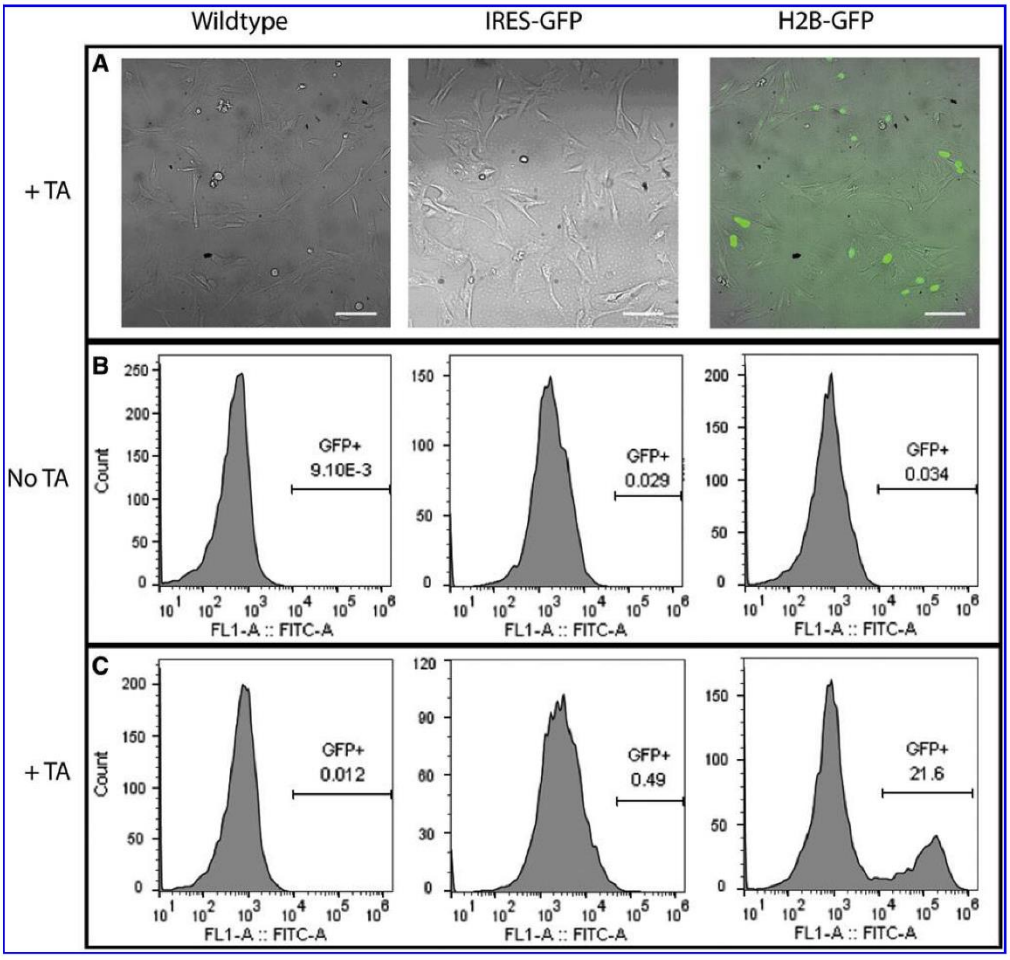


FIG. 3. *LGR5*-targeting TAs upregulate GFP expression in model II—H2B-GFP cells, but not wild-type, model I—IRES-GFP, or nontransfected cells. **(A)** Fluorescent-brightfield overlay showing nuclear GFP expression in model II—H2B-GFP cells treated with TA but not in wild-type or model I—IRES-GFP. **(B)** Flow cytometry analysis of GFP expression in each cell type without TA treatment. **(C)** Flow cytometry analysis of GFP expression in each cell type treated with *LGR5*-TAs. Scale bar indicates 100 μm.

Downloaded by 207.244.86.173 from www.liebertpub.com at 02/07/22. For personal use only.

transfected with TAs targeted to the 1 kb region upstream of the *LGR5* gene locus *in vitro* successfully predicted the *in vivo* GFP expression profile of *LGR5*-expressing cells.

TAs can be used to screen pooled cells for presence of transgene

Although our proof-of-concept study has shown that the *in vitro* phenotype of fibroblast cell lines from gene-edited pigs faithfully recapitulates the *in vivo* phenotype, we next asked whether this principle could best be applied *before* the generation of a transgenic animal. To test this, we transfected wild-type fetal fibroblasts with CRISPR nuclease, gRNA, and an HDR template as in the generation of model II. We grew out the transfected cells in a pool of mixed genotypes until recovered, and then re-transfected them with the TAs. Presence of DNA insertions was evaluated by PCR of the pooled cell DNA (Fig. 4B, C), and as shown by the flow cytometry and microscopy analysis, GFP was expressed in about 1% of the population of the cells (Fig. 4A, Supplementary Fig. S2B). This suggests that about 3–4% of the pooled cells contains the knock-in, of which only about 25–30% were upregulated. From these results, we conclude that flow cytometry and TAs can be used to screen a pool of edited cells for cells that contain the transgene.

Discussion

In this study, we employed CRISPR-based methods for evaluating outcomes of gene editing events. As a proof of concept, we used two different gene-edited genotypes in pigs, which were both generated with the goal of adding a GFP tag to the *LGR5* expressing cells. Our results show that model I, IRES-GFP animals did not show any GFP signals in any tissue examined. Although the DNA sequence of the transgene was intact based on PCR (Supplementary Fig. S1A) and sequencing, it is possible that the DNA repair mechanisms that facilitated the transgene integration also elicited mutations or deletions

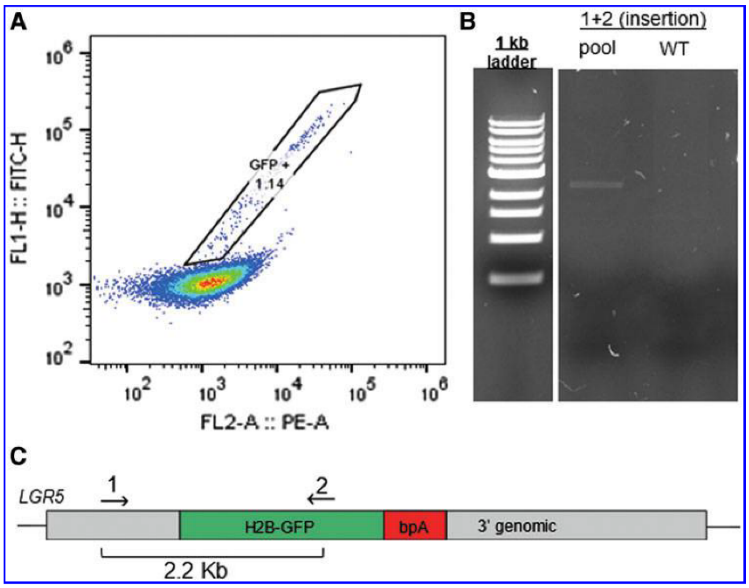


FIG. 4. TAs can induce transgene expression in a pool of transfected cells. The LGR5-H2B-GFP knock-in plasmid was transfected with CRISPR-Cas9 as described for the generation of the LGR5-H2B-GFP model. Cells were recovered and expanded for 1 week before pooled cells were transfected with *LGR5* TAs and evaluated for H2B-GFP expression by flow cytometry. **(A)** Detection of a population of GFP+ cells by flow cytometry. **(B)** PCR analysis of genomic DNA for H2B-GFP transgene inserted into *LGR5* locus or the endogenous *LGR5* allele around the CRISPR cut-site. Band from primers 1 + 2 shows that population of cells within the pool of edited cells contain the H2B-GFP insert, whereas WT cells do not. **(C)** Diagram of primers used in **(B)**, amplifying the 5′ junction of the insert. PCR, polymerase chain reaction; WT, wild-type.

Downloaded by 207.244.86.173 from www.liebertpub.com at 02/07/22. For personal use only.

upstream or downstream of the intended target site. It is also possible that the transgene interfered with enhancers or cis-regulatory elements that affected gene expression. Further experiments are necessary to determine why the transgene was not expressed as expected. In contrast, model II, generated from a CRISPR-mediated knock-in of the H2B-GFP gene into the 5′ of the *LGR5* gene, showed the expected nuclear GFP signal in *LGR5*-expressing cells. Using CRISPR-TAs, we were able to distinguish *in vitro* which genotype yielded the desired GFP signal profile.

Using CRISPR-TAs has advantages over traditional methods of transgene testing, such as cloning the transgene into an expression vector, for several reasons. First, the design and synthesis of gRNAs using a plasmid-based method is straightforward and can be completed much more quickly than the synthesis of an expression construct. Second, by upregulating the gene in its genomic location, it is possible to account for unexpected outcomes; for example, in model I, we detected the presence of the IRES-driven GFP using an expression vector and in gene-edited cells, but, nevertheless, GFP was not detected in the animal or the cells.

Notably, CRISPR-Cas systems are known to have off-target effects, which occur when there are mismatches between the gRNA sequence and the genomic DNA it binds,[24,25] and always must be considered when making changes to DNA sequences. However, since TAs represent a transient state of activation,[26,27] and generally more than one binding site is needed for effective activation, this suggests that there is little to no risk of misinterpreting results due to off-target binding of CRISPR-TAs. Nevertheless, CRISPR-Cas systems will continue to be developed to reduce off-target binding,[28] and thus the techniques described in this study can be updated with higher fidelity enzymes as they become available.

The application of CRISPR-TAs *in vitro* before generating the transgenic animals in this research would have saved our laboratory significant costs in labor, animal fees, time, and resources. As shown by us and others,[7–17] even with known DNA sequences in somatic cells, it is difficult to know the exact phenotypic outcome. For example, alternate start codons, hypomorphic proteins, alternative splicing, or undetected mutations upstream or downstream of the cut site or insert could all contribute to undesired phenotype, representing wasted resources, time, increased number of experimental animals, and misinterpretation of results. The CRISPR-TAs are a tool that all laboratories performing gene editing, especially large animal gene editing, should consider for validations of gene editing outcomes before investing in the creation of the animal. In Figure 5, we have summarized the process by which gene-editing strategies and outcomes can be efficiently and effectively analyzed for knockout or knock-in gene edits, using SCNT or microinjection strategies. Although microinjection does not use somatic cells from culture, the same knockout strategy is often applied in somatic cells to optimize CRISPR cutting efficiency and targeting strategy,[29–31] which yield the same edited sequence as the target of microinjection. In this case, those cells could be used with TAs to ensure that the modified sequence elicits gene expression as desired before proceeding with microinjection and embryo transfer. In gene knock-ins, this analysis includes proper synthesis, function, and localization of a protein. For CRISPR-based knockouts, this system can be used to ensure the absence of the protein and the nonsense-mediated decay of mRNA.[17] Although our study shows

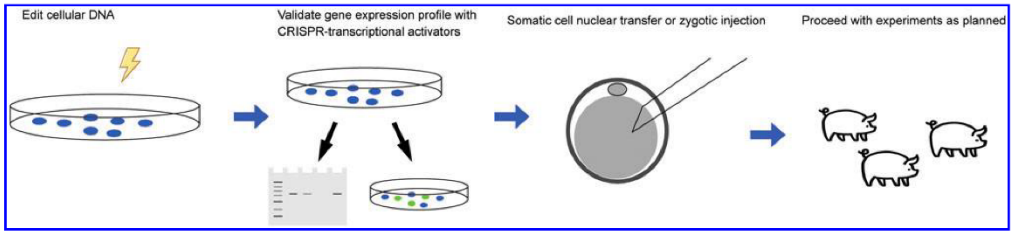


FIG. 5. CRISPR-TAs applied to gene-edited cells *in vitro* can validate intended phenotype of gene edits before generation of offspring. Suggested process for applying TAs: (1) Gene editing by site-specific RNA-guided endonucleases. (2) Transfect gene-edited cells with TAs and evaluate for desired outcomes. (3) Generate gene editing offspring using the cells or same editing strategy applied previously. (4) Proceed with experiments as planned.

Downloaded by 207.244.86.173 from www.liebertpub.com at 02/07/22. For personal use only.

knock-in detection from a pool, analysis of knockouts may require clonal expansion before analysis.

Furthermore, our results suggest that this novel method can be used to save time by enriching for cells that contain a properly inserted transgene from a pool of edited cells. Although this proof-of-concept study used a fluorescent marker, other methods can be used to screen for proper protein expression, such as antibody-based flow cytometry. In addition, TAs can also be used with cultured cells to develop controls for Western blot or flow cytometry or for testing target-specific reverse transcription quantitative polymerase chain reaction analysis (qRT-PCR) probes and primers.

Conclusion

In conclusion, CRISPR-Cas9 and RNA-guided endonuclease-mediated gene edits often elicit unpredictable outcomes, and in this research we applied CRISPR-TAs as a tool to evaluate outcomes *in vitro*. As proof of concept, we demonstrated that this technique can be successfully used to validate transgene expression in non-expressing cell types, and that it could be coupled with flow sorting to enrich a pool of cells containing the desired genotype. We suggest that this technology presents animal researchers with an opportunity for *in vitro* validation of the expected expression profile of site-specific DNA modifications before the generation of live gene-edited animals.

Author Confirmation Statement

K.P. and J.P. designed experiments. K.P., J.C., S.G., and K.G. performed experiments. K.P. and J.P. interpreted results and wrote the article. All co-authors have reviewed and approved this manuscript. This manuscript has been submitted solely to this journal and is not published, in press, or submitted elsewhere.

Author Disclosure Statement

No competing financial interests exist.

Funding Information

We gratefully acknowledge funding from National Institutes of Health R21OD019738 and R01OD023138 to J.P.

Supplementary Material

Supplementary Figure S1
Supplementary Figure S2

References

1. Van Eenennaam AL. Application of genome editing in farm animals: cattle. *Transgenic Res*. 2019;28:93–100. DOI: 10.1007/s11248-019-00141-6.
2. Whitelaw CBA, Sheets TP, Lillico SG, et al. Engineering large animal models of human disease. *J Pathol*. 2016;238:247–256. DOI: 10.1002/path.4648.
3. Polkoff K, Piedrahita JA. The transformational impact of site-specific DNA modifiers on biomedicine and agriculture. *Anim Reprod*. 2018;15:171–179. DOI: 10.21451/1984-3143-ar2018-0065.
4. Gilles AF, Averof M. Functional genetics for all: engineered nucleases, CRISPR and the gene editing revolution. *Evodevo*. 2014;5:43. DOI: 10.1186/2041-9139-5-43.
5. Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*. 2014;32:347–355. DOI: 10.1038/nbt.2842.
6. Ran FA, Hsu PD, Wright J, et al. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013;8:2281–2308. DOI: 10.1038/nprot.2013.143.
7. Miao J, Ying B, Li R, et al. Characterization of an N-terminal non-core domain of RAG1 gene disrupted Syrian Hamster model generated by CRISPR Cas9. *Viruses*. 2018;10:243. DOI: 10.3390/v10050243.
8. Pereira FJC, Teixeira A, Kong J, et al. Resistance of mRNAs with AUG-proximal nonsense mutations to nonsense-mediated decay reflects variables of mRNA structure and translational activity. *Nucleic Acids Res*. 2015;43:6528–6544. DOI: 10.1093/nar/gkv588.
9. Munoz DM, Cassiani PJ, Li L, et al. CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions. *Cancer Discov*. 2016;6:900–913. DOI: 10.1158/2159-8290.CD-16-0178.
10. Lalonde S, Stone OA, Lessard S, Lavertu A, Desjardins J, Beaudoin M, et al. Frameshift indels introduced by genome editing can lead to in-frame exon skipping. *PLoS One*. 2017;12:e0178700. DOI: 10.1371/journal.pone.0178700.
11. Kapahnke M, Banning A, Tikkanen R. Random splicing of several exons caused by a single base change in the target exon of CRISPR/Cas9 mediated gene knockout. *Cells*. 2016;5:45. DOI: 10.3390/cells5040045.
12. Chiavacci E, Kirchgeorg L, Felker A, et al. Early frameshift alleles of zebrafish tbx5a that fail to develop the heartstrings phenotype. *Matters* 2017. DOI: 10.19185/matters.201703000011.
13. Hosur V, Low BE, Li D, et al. Genes adapt to outsmart gene targeting strategies in mutant mouse strains by skipping exons to reinitiate transcription and translation. *Genome Biol* 2020;21:168. DOI: 10.1101/2020.04.22.041087.
14. Mou H, Smith JL, Peng L, et al. CRISPR/Cas9-mediated genome editing induces exon skipping by alternative splicing or exon deletion. *Genome Biol*. 2017;18:108. DOI: 10.1186/s13059-017-1237-8.
15. Huin V, Buée L, Behal H, et al. Alternative promoter usage generates novel shorter MAPT mRNA transcripts in Alzheimer's disease and progressive supranuclear palsy brains. *Sci Rep* 2017;7:12589. DOI: 10.1038/s41598-017-12955-7.
16. Makino S, Fukumura R, Gondo Y. Illegitimate translation causes unexpected gene expression from on-target out-of-frame alleles created by CRISPR-Cas9. *Sci Rep*. 2016;6:39608. DOI: 10.1038/srep39608.
17. Smits AH, Ziebell F, Joberty G, et al. Biological plasticity rescues target activity in CRISPR knock outs. *Nat Methods*. 2019;16:1087–1093. DOI: 10.1038/s41592-019-0614-5.
18. Tan W, Proudfoot C, Lillico SG, et al. Gene targeting, genome editing: from Dolly to editors. *Transgenic Res*. 2016;25:273–287. DOI: 10.1007/s11248-016-9932-x.
19. Ryu J, Prather RS, Lee K. Use of gene-editing technology to introduce targeted modifications in pigs. *J Anim Sci Biotechnol*. 2018;9:5. DOI: 10.1186/s40104-017-0228-7.
20. Mclean Z, Oback B, Laible G. Embryo-mediated genome editing for accelerated genetic improvement of livestock. *Front Agric Sci Eng*. 2020;7:148–160. DOI: 10.15302/J-FASE-2019305.
21. Barker N, Tan S, Clevers H. Lgr proteins in epithelial stem cell biology. *Development*. 2013;140:2484–2494. DOI: 10.1242/dev.083113.
22. Walker SC, Shin T, Zaunbrecher GM, et al. A highly efficient method for porcine cloning by nuclear transfer using in vitro-matured oocytes. *Cloning Stem Cells*. 2002;4:105–112. DOI: 10.1089/153623002320253283.
23. Maeder ML, Linder SJ, Cascio VM, et al. CRISPR RNA-guided activation of endogenous human genes. *Nat Methods*. 2013;10:977–979. DOI: 10.1038/nmeth.2598.

Downloaded by 207.244.86.173 from www.liebertpub.com at 02/07/22. For personal use only.

24. Tsai SQ, Nguyen NT, Malagon-Lopez J, et al. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR—Cas9 nuclease off-targets. *Nat Methods*. 2017;14:607. DOI: 10.1038/nmeth.4278.
25. Manghwar H, Li B, Ding X, et al. CRISPR/Cas systems in genome editing: methodologies and tools for sgRNA design, off-target evaluation, and strategies to mitigate off-target effects. *Adv Sci*. 2020;7:1902312. DOI: 10.1002/advs.201902312.
26. Gilbert LA, Larson MH, Morsut L, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*. 2013;154:442–451. DOI: 10.1016/j.cell.2013.06.044.
27. Liao H-K, Hatanaka F, Araoka T, et al. In vivo target gene activation via CRISPR/Cas9-mediated trans-epigenetic modulation. *Cell*. 2017;171:1495–1507.e15. DOI: 10.1016/j.cell.2017.10.025.
28. Kleinstiver BP, Pattanayak V, Prew MS, et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*. 2016;529:490–495. DOI: 10.1038/nature16526.
29. Wang X, Zhou J, Cao C, et al. Efficient CRISPR/Cas9-mediated biallelic gene disruption and site-specific knockin after rapid selection of highly active sgRNAs in pigs. *Sci Rep*. 2015;5:13348. DOI: 10.1038/srep13348.
30. Wang X, Yu H, Lei A, et al. Generation of gene-modified goats targeting MSTN and FGF5 via zygote injection of CRISPR/Cas9 system. *Sci Rep*. 2015;5:13878. DOI: 10.1038/srep13878.
31. Zhou W, Wan Y, Guo R, et al. Generation of beta-lactoglobulin knock-out goats using CRISPR/Cas9. *PLoS One*. 2017;12:e0186056. DOI: 10.1371/journal.pone.0186056.

Downloaded by 207.244.86.173 from www.liebertpub.com at 02/07/22. For personal use only.