ABSTRACT

Title of Thesis: OPTIMIZING GENE TARGETING METHODOLOGIES USING CRISPR/CAS RIBONUCLEOPROTEINS IN PORCINE SOMATIC CELLS

Chan Mi Kim, Master of Sciences, 2016

Thesis Directed By: Assistant Professor, Dr. Bhanu Telugu, Department of Animal Avian Sciences

The pig is one of the most valuable animal models for agricultural and biomedical research. In this thesis, we sought to standardize procedures for routine gene targeting and demonstrate that it is possible to create site-specific knockins of short DNA sequences in a locus of our interest, specifically downstream of a pig insulin gene and a ubiquitously expressed *COL1A* gene in pigs. Our study successfully established and optimized protocol for CRISPR/Cas mediated knockin in porcine somatic cells, while simultaneously creating a stable platform for future transgene integration and stepping forward to perform more efficient porcine transgenesis in the future.

OPTIMIZING GENE TARGETING METHODOLOGIES USING CRISPR/CAS RIBONUCLEOPROTEINS IN PORCINE SOMATIC CELLS

by

Chan Mi Kim

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Sciences
2016

Advisory Committee:
Professor Bhanu Telugu, Chair
Dr. Carol Keefer
Dr. Seong Ho Lee

Acknowledgements

Thank you to Dr. Bhanu Telugu who gave me the opportunity to pursue my master's study and trained me to become a more mature scientist. I would never have been able to complete this program without your guidance and support.

To Dr. Kieun Park and Anne Powell who helped me at the bench, constantly motivated me with words of encouragements, and guided me to make incremental progress through my project.

To my mom and dad who have gave me tremendous supports throughout my studies. I deeply appreciate all of your prayers of day and night.

Lastly, to my dearest husband, Carl Eunro Chu, thank you for your unfailing love, support, and encouragement.

Table of Contents

List of Tables

List of Figures

List of Illustrations

List of Abbreviations

SCNT: Somatic cell nuclear transfer

COL1A: Collagen 1A locus

PBCs: Pancreatic Beta Cells

GFP: Green Fluorescent Protein

DM: Diabetes Mellitus

rtTA: Reverse Tetracycline-Dependent Transactivator

IR: Insulin Receptor

HNF1A: hepatocyte nuclear factor-1α

GIPR: glucose-dependent insulinotropic polypeptide receptor

ADA: American Diabetes Association

TALENs: Transcription Activator-Like Effector Nuclease

TALEs: transcription activator-like effectors

CRISPR/Cas: clustered regularly interspaced short palindromic repeats and CRISPR-associated (Cas) nuclease system

DSB: double strand break

RVDs: repeat-variable di-residues

gRNA: guide RNA

PAM: Prototospacer Adjacent Motif

NHEJ: Non-Homologous End-Joining

HR: Homologous Recombination

HA: Hemagglutinin epitope

NLS: nuclear localization signals for nuclear localization

Chapter 1: Literature Review

Introduction

The use of animal models in scientific research remains essential for drug discovery and improvement of animal and human health. Among the animal models, the mouse is coveted for its small size and short generation interval. The laboratory mouse has been crucial in identifying both cellular and molecular components of the human diseases. Although mice are powerful research models, they are limited in translational research due to significant differences in anatomy and physiology between mice and men. Also, the mouse model precludes long-term investigations because of its small size and short lifespan. In order to translate findings from basic science to enhance human health, a large animal model such as pig that can more closely resemble human anatomy and pathological responses is highly preferred (Wolf, Braun-Reichhart et al. 2014).

Comparative genetic and protein analysis (http://www.ensembl.org/Sus_scrofa/Info/Index) has proven that the pig is evolutionarily more closer to human than mouse (Wernersson, Schierup et al. 2005), and resembles many human anatomical and physiological characteristics (Swindle, Makin et al. 2012). The pigs are attractive model organisms for biomedical and genetic engineering applications because of relatively early sexual maturity, short generation interval, and a litter bearing species producing multiple offspring in one pregnancy (Aigner, Renner et al. 2010).

The first pig was made transgenic for growth hormone in 1985, when foreign DNA was microinjected into zygotic pronuclei (Hammer, Pursel et al. 1985). Since then, the uses of porcine model in transgenesis have expanded. The major impacts of transgenic pigs took place in the field of medicine where pigs were used as models of human diseases (Whitelaw, Sheets et al. 2015). In the past, the pig model has been used to study neurodegenerative disease (Yang, Wang et al. 2010), cardiovascular disease (Carlson, Tan et al. 2012), diabetes (Renner, Fehlings et al. 2010, Wolf, Braun-Reichhart et al. 2014), and cystic fibrosis (Rogers, Stoltz et al. 2008). Many techniques that were established using the pig model have been transferred directly to clinics from both surgical and diagnostic perspective (Swindle, Makin et al. 2012).

Table 1: Applications of pig and minipig models in biomedical research. Below table summarizes the use of pig and minipig models in biomedicine and pharmaceutical research (Garrels, Holler et al. 2014)

MODEL	COMMENT	REFERENCE
Xenotransplantation	knockout of alpha-galactosyltransferase	Lai et al., 2002; Dai et al., 2002
Xenotransplantation	expression of tumour necrosis factor ligand	Klose et al., 2005
Xenotransplantation	expression of human leukocyte antigen	Weiss et al., 2009
Xenotransplantation	PERV-knock down	Dieckhoff et al., 2008
Xenotransplantation	expression of human thrombomodulin	Petersen et al., 2009
Xenotransplantation	expression of human A20 (anti-apoptotic gene)	Oropeza et al., 2009
Cystic fibrosis pig	knockout of cystic fibrosis transmembrane conductance receptor	Rogers et al., 2008
Diabetes model	expression of mutated hepatocyte nuclear factor-1	Umeyama et al., 2009
Diabetes model	expression of mutated insulin 2	Renner et al., 2010
Immunodeficient pig	knockout of light chain	Ramsoondar et al., 2010
Immunodeficient pig	knockout of joining gene cluster	Mendicino et al., 2010
Huntington model	expression of mutated huntingtin with polyglutamine tract	Yang et al., 2010
Alzheimer model	expression of mutated human amyloid precursor protein	Kragh et al., 2010
Breast cancer	knockout of BRCA1 gene	Luo et al., 2011
Macular degeneration	introduced deletion in ELOVL4 gene	Sommer et al., 2011

Even though, the pig has several advantages over murine models, and is considered a preferred model, the use of pigs in biomedical research has not been widespread because of the genetic engineering technologies that lagged the mouse model. Despite many attempts to establish embryonic stem cell (ESC) lines in porcine model, the porcine ESC (pESC) lines are currently unavailable for genetic modifications in pig. There are two prominent means for generating genetically modified livestock animals: the pronuclear injection and the somatic cell nuclear transfer (SCNT). The pronuclear injection is typically used for generating transgenic animals, where the transgene is injected into pronucleus. However, this procedure suffers from random integration of transgene, insertional mutagenesis (the transgene inserts into an existing gene potentially disrupting the endogenous gene's expression or function), lack of control over transgene copy number, silencing or aberrant expression of transgenes in non-target tissues based on the site of integration (positional variegation), and random assortment and segregation in subsequent generations (Park and Telugu 2013, Park, Park et al. 2016), to name few. A preferred and more widely used method for generating a genetically modified pigs is somatic cell nuclear transfer (SCNT). In this method, somatic cells, typically fetal fibroblasts were utilized to make the intended genetic modification. The modified fibroblasts are then used as a nuclear donor for reconstituting oocytes and generation of modified offspring.

As discussed above, introduction of transgene either via pronuclear injection or by transfection into somatic cells results in random integration of transgenes, which suffers from lack of control over integration sites, number of copies of inserted transgene and reliable expression of transgene to name a few. As such, targeting the transgene to a specific locus, called as gene targeting is highly desired. That said, the efficiency of gene targeting is relatively less (1 in 10^6 cells) in somatic cells, and has not been successfully tested in embryos. The use of site-specific nucleases that engineer a double strand break (DSB) at the locus of interest increases the efficiency of gene targeting by three orders of magnitude. An overarching goal of this thesis is to standardize procedures for routine gene targeting and demonstrate that it is possible to create site-specific knockins of short or long DNA sequences in a gene locus of our interest, specifically downstream of ubiquitously expressed *COL1A* gene or insulin gene, respectively.

There are 2 parts to this thesis:

Part 1: Targeted Knockin of pseudo attP sites downstream of COL1A gene in porcine somatic cells using CRISPR/Cas System

A) Designing and assembling *COL1A* locus targeting vector and site-specific nucleases (CRISPR/Cas)
B) Cell transfection and SCNT of the ssDNA oligonucleotide (containing two pseudo attP sites) along with site-specific nucleases

Part 2: Assembling and testing insulin locus targeting vectors and site-specific nucleases (either TALENs or CRISPR/Cas)

A) Designing and assembling insulin locus targeting vector and site-specific nucleases (TALENs and/or CRISPR/Cas)
B) Cell transfection and SCNT of the insulin-targeting vector along with the site-specific nucleases

Significance/ Rationale

As previously mentioned, a preferred means for generating genetically engineered porcine model is SCNT, where somatic cells typically fetal fibroblasts are modified to include the intended genetic modification and used as nuclear donors for generating genetically modified offspring (Park, Park et al. 2016). Gene targeting (knockin), requires homologous recombination, a type of genetic recombination in which nucleotide sequences are exchanged between two identical molecules of DNA, often suffers from poor efficiencies in somatic cells (1 in 10^6-10^7 cells) (Park, Park et al. 2016). Additionally, the use of most commonly used somatic cells (fetal fibroblasts) has limited viability in culture for screening recombinants. In this regard, site specific nucleases such as ZFNs (zinc finger nucleases), TALENs (transcription activator-like effector nucleases), and CRISPR (clustered regulated interspaced short palindromic repeat and CRISPR-associated (Cas) nuclease system (CRISPR/Cas) that engineer a double strand break (DSB) at the target site and promote gene targeting via homologous recombination can improve targeting efficiencies by nearly 1000 fold and thus could offer a solution (Moehle, Rock et al. 2007).

Among these genome editors, the CRISPR/Cas system has emerged as a tool of choice because of the ease of design, assembly, delivery, and a high degree of reliable gene modifications (Park, Park et al. 2016). In the past, the CRISPR/Cas

system has been utilized to successfully generate pigs and other large domestic animals (Park, Park et al. 2016). In this Thesis study, a mammalian codon optimized Type III Cas 9 from *Streptococcus pyogenes* along with a chimeric synthetic single-guide RNA (sgRNA) containing Cas9 binding sites and a 20 nucleotide guide sequences specific to the target site has been used to introduce DSBs (Zhou, Xin et al. 2015). Once the DSB is generated by CRISPR, the DSB is expected to be repaired by the cells own repairing mechanisms: either by non-homologous end-joining (NHEJ) pathway or by the homology directed repair recombination (HDR) pathway. The predominant DNA repair pathway is the error-prone NHEJ pathway, where the ends of DNA are simply ligated back together. Alternatively, the HDR pathway can occur in the presence of repair template that has the homology arms flanking the site of DSB. In other words, for generating gene ablation models, NHEJ is the preferred pathway, whereas for introducing gene knockins and point mutations, HDR pathway is preferred.

One of the goals of this study is to focus on utilization of CRISPR/Cas system to engineer targeted knockin of pseudo attP sites (50 bp) downstream of an ubiquitously expressed *COL1A* gene in porcine somatic cells. In order to achieve HDR at high frequency, several published options will be tested, including the use of a single stranded oligo as a targeting template, small molecules or inhibitors of DNA ligase IV such as SCR7 (Chu, Weber et al. 2015, Maruyama, Dougan et al. 2015), transient incubation of cells at low temperature (Doyon, Choi et al. 2010, Brown, Lee et al. 2011, Miller, Tan et al. 2011), and the choice of CRISPR reagents. CRISPR reagents can be delivered as DNA expression vector, RNA preps, and a complex of

Cas9 protein and sgRNA (Park, Park et al. 2016). Through this study, we outlined and examined different reagents and experimental conditions in a head to head comparison to standardize procedures for routine gene targeting experiments.

Additionally, by knocking-in pseudo attP sites (50 bp) downstream of a ubiquitously expressed *COL1A* locus, subsequent phiC31 integrase-mediated introduction of functional transgenes will be investigated. The phiC31 integrase, a serine integrase, that causes recombination between two 50 bp recognition sequences of attP and attB sites allow us to integrate functional transgenes like GFP using phiC31 integrase at the knockin site (Smith, Brown et al. 2010, Brown, Lee et al. 2011). When such integration takes place, we expected that the transgene is not only specifically integrated downstream of an ubiquitously expressed *COL1A* gene, but also protected from any potential limitations that are stated above including but are not limited to insertional mutagenesis, lack of control over transgene copy number, positional variegation, random assortment and segregation in subsequent generations.

By utilizing the gene targeting protocols optimized with CRISPR/Cas mediated knockin in porcine somatic cells; a second goal of this study is to generate insulin-GFP reporter pigs. The insulin reporter pigs will allow us to study biomedical applications, specifically type 1diabetes mellitus (DM). Diabetes Mellitus (DM), often referred to as diabetes, is a chronic metabolic disease in which there is uncontrollably elevated blood glucose levels for over a prolonged period (American Diabetes Association. 2013). Globally, DM is responsible for 4.6 million deaths per year (WHO report). DM affects about 346 million people and the numbers are expected to double by 2030 (WHO report). These staggering statistics prompted

United Nations to declare DM as a "global threat" (Raducanu and Lickert 2012). Insufficient PBC (pancreatic β cell) mass due to the autoimmune destruction of endogenous insulin-producing PBC results in Type I DM (Nir, Melton et al. 2007). There are two major medical therapies available for Type 1 DM patients: administration of synthetic insulin or transplantation of pancreatic sections or primary (1°) PBC. Although these treatments have advanced diabetic care greatly, they are inconvenient (often require immune-suppression), costly, and limited in supply. Consequently, alternative therapies and research in an animal model that best mimics humans in physiology, organ size is highly desired. Recent studies have suggested a potential new diabetic treatment option that involves inhibiting autoimmunity and restoring PBC to recover its function to produce insulin (Nir, Melton et al. 2007). As a result, the study of dynamics and regenerative capacity of PBC has become important for diabetic research. A significant majority of previous studies have used transgenic mouse models to specifically and conditionally ablate PBC to study the dynamics of the β cell function (Nir, Melton et al. 2007). Although mice are good model for studying β cell mass fluctuation and regulation mechanisms, the use of mice in the study of clinical and more practical implications of diabetes is highly limited because the mice do not accurately represent complex human physiology. Even though, the pig has several advantages over murine models, and is considered a "preferred" model, the use of pigs in diabetic research has not been widespread because of the genetic engineering technologies that lagged the mouse model. That being said, the successful adaptations of CRISPR/Cas system and other site-specific

nuclease to the porcine systems have greatly advanced our ability to genetically tailor the porcine models to replicate human diseases.

The domestic pig offers several advantages over other animal models for studying human diseases. This is because the pig is a monogastric animal, has a long lifespan, and resembles humans in size, anatomy, and physiology (Swindle, Makin et al. 2012). Specifically relevant to this thesis, the pig has similar morphology of the pancreas and overall metabolic status to those of humans. The swine pancreas is located posterior to the stomach and is involved in both exocrine and endocrine functions as in human (Swindle, Makin et al. 2012). From hormonal perspective, the pig's pancreas is similar to that of humans because it is responsible for secretion of insulin and glucagon in response to change in blood glucose levels (Renner, Blutke et al. 2015), and many more characteristics that we will discuss later in this thesis that suggest the pig is a promising model to study human diseases, particularly DM.

In part two of this thesis study, we hoped to tag the porcine insulin locus with green fluorescent protein (GFP) to generate insulin reporter pigs. Here, we utilized the optimized gene-targeting techniques that we have established in first part of this thesis study. In the future, we expected to use the insulin reporter pigs to further test potential cellular therapy, validate potential medications, and perform reprograming studies. Ultimately, we want to generate a swine model of diabetes. A good model to study PBC dynamics and regenerative capacity would allow us to not only specifically ablate PBC, but also conditionally kill PBC at a desired time (Nir, Melton et al. 2007). In the future, utilizing this system, we can plan to perform conditional gene ablation by using tetracycline-controlled transcriptional activation system, a

method of inducible gene expression where transcription is reversibly turned on (tet-On system) or off (tet-Off system) in the presence of doxycycline. In order to generate the pig model of diabetes as we further optimize our porcine insulin gene targeting methodology and demonstrate that we could potentially modify the porcine insulin locus. We hope that our efforts allow us to generate insulin reporter pigs and pig models of diabetes and further study the dynamics of PBC, its capacity for spontaneous regeneration, the cellular origins of new PBC, and the determinants of PBC (Nir, Melton et al. 2007). All of these studies on reprogramming potentials of porcine PBC may provide an alternative treatment option for human type-1 diabetic patients. The primary sources of islet from porcine tissues might correct insulin dependence upon transplantation. Moreover, the clinical implications that we could have not studied using the mouse model of diabetes could be studied more effectively using our pig model of diabetes (e.g. new diabetic medications could be tested using the pig model).

CRISPR/Cas-mediated Gene Targeting

The CRISPR (clustered regularly interspaced short palindromic repeats) /Cas system, a gene targeting system that is widely used in biological sciences today, is based on the natural system used by archaea to protect themselves from infectious viruses (Cong, Ran et al. 2013). In the presence of viral DNA, archaea were found to produce two types of short RNAs, one of which contains the sequence that matches that of invading virus (Cong, Ran et al. 2013). These two RNAs form an interfering complex with the protein called Cas 9, a nuclease that cuts the DNA of the virus when the matching sequence, known as guide RNAs, finds its target sequence on the

viral genome (Cong, Ran et al. 2013). Researchers have utilized this system such that we can use it to precisely manipulate the gene of our interest. The main components of CRISPR/Cas gene editing system are guide RNA (gRNA) and Cas9 endonuclease. When gRNA and Cas9 are expressed in a cell, the gRNA-Cas9 complex gets formed and is recruited to the target sequence that is located directly upstream of the PAM motif. The PAM motif, an abbreviation for prototospacer adjacent motif, refers to the DNA sequences (5'-NGG-3') that are located at about 20 nucleotides downstream of the target sequence and is required for gRNA-Cas9 to bind. Once the complex is localized to the target DNA, Cas9 cuts the desired locus with extreme accuracy resulting in a DSB. DSB is then repaired by one of two processes, NHEJ, which may result in small gene mutation, or HR, which may result in gene targeting event.

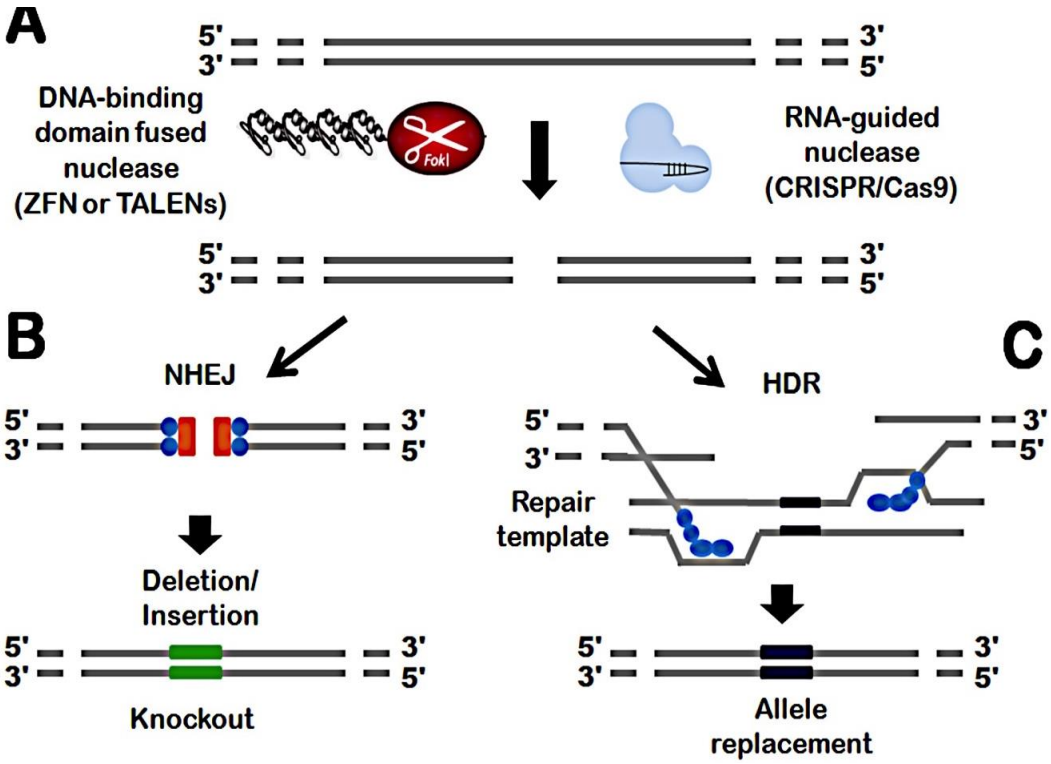


Figure 1: CRISPR/ Cas9 genome editing technique. A) Schematic representing the genome of interest recognized by the guide RNA and Cas9 complex undergo DSB. B) Once the DSB occurs, it is repaired by cells own reparing mechanisms: either NHEJ or HDR. NHEJ, where the ends of DNA are simply ligated back together, occur in the absence of a repair template. Alternatively, the HR can occur in the presence of repair template.

TALEN-mediated Gene Targeting

TALENs (Transcription Activator-Like Effector Nuclease) are novel and valuable tools for precisely manipulating mammalian genome. TALENs are generated by fusing transcription activator-like effectors (TALEs), a class of DNA binding protein produced by plant pathogenic bacteria in the genus *Xanthomonas*, to the catalytic domains of Fok1 nucleases (Cermak, Doyle et al. 2011). Because Folk1 nucleases act as a dimer, TAL effector nucleases also work in a pair, binding opposing targets across a spacer over which the FokI domains come together to create the double strand break (DSB). The targeting specificity of TALENs is determined by customizable arrays of polymorphic amino acid repeats called repeat-variable di-residues (RVDs) in the TAL effectors. Each RVD sequence that binds to a certain DNA nucleotide (NI = A, NK = G HD = C, NG = T) can be customized to specifically bind to the locus of interest and induce DSB (Cermak, Doyle et al. 2011). Once the DSB is created, it can be repaired by the cells own repairing mechanisms: either by the non-homologous end-joining (NHEJ) pathway or by the homologous recombination (HR) pathway. The NHEJ pathway, where the ends of DNA are simply ligated back together, is preferred in the absence of a repair template. Alternatively, the HR can occur in the presence of repair template that has the homology arms flanking the site of the DSB.

General Pancreatic Anatomy

The pancreas is located in the upper abdominal cavity behind the stomach, is an organ that is involved in both exocrine and endocrine functions. The human pancreas is about 12 to 15 cm long J- shaped organ and it weighs about 80g (Pandol

2010). The pancreas is organized into 4 general regions: the head, neck, body and tail. The head region, the widest part of the pancreas, lies in the loop of the duodenum (Pandol 2010). The body of the pancreas is located in the posterior region to the distal portion of the stomach between the tail and the neck (Pandol 2010). The tail of the pancreas lies near the hilum of the spleen (Pandol 2010). The anatomical pancreas is composed of two parts: an exocrine part (acinar and duct tissues), and an endocrine part (islets of Langerhans). The exocrine pancreas, comprises 85% of the mass of pancreas, and is responsible for secretion of digestive enzymes, ions and water into the duodenum of the gastrointestinal tract (Pandol 2010). The endocrine pancreas is responsible for secretion of insulin and glucagon into the blood stream to regulate blood glucose concentrations. Two major arteries, the celiac and superior mesenteric arteries are responsible for supplying blood to the pancreas (Pandol 2010). The two major veins, the splenic vein and the superior mesenteric vein are responsible for draining the blood from the pancreas (Pandol 2010).

Pancreatic Development

After approximately 5 weeks of gestation, the human pancreas arises from the embryonic digestive tube (Pandol 2010). The pancreas appears as two outpouchings of the endodermal lining of the duodenum (Pandol 2010). The two outpouchings refer to the ventral and the dorsal pancreatic bud (Pandol 2010). The ventral pancreatic bud, located close to the junction of the hepatic duct, begins to migrate posteriorly to the dorsal pancreatic bud as the stomach begins to dilate. These two buds eventually fuse and about 90% of the fused pancreatic bud population disintegrates from duodenum

(Gilbert 2000). The remaining 10% of the fused pancreatic bud population persists and gives rise to the main pancreatic duct that connects pancreas with duodenum (Gilbert 2000).

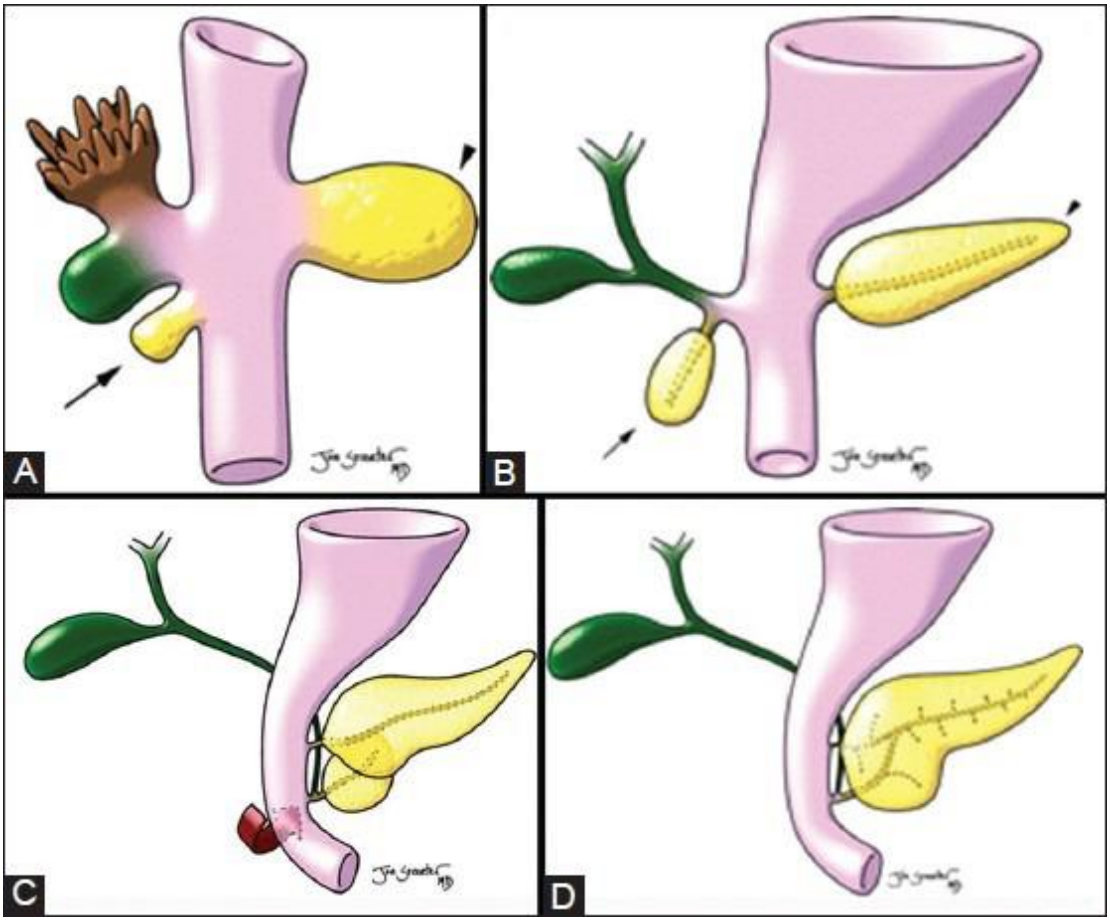


Illustration 1: Drawings of the embryonic development of the pancreas. The ventral and dorsal pancreatic buds arose from embryonic digestive tissue. The ventral pancreatic bud rotates clockwise around the caudal part of the foregut and fuses with the dorsal pancreas (Thakur, Jhobta et al. 2014).

The development of pancreas is tightly controlled by the notochord. The signaling pathways, directed by notochord are absolutely necessary to specify pancreas. The dorsal pancreas, because of its contact with the notochord expresses pancreatic specific genes, *the Pdx1* (Gittes 2009). There are two signaling factors expressed in notochord during pancreatic specification, activin-β and fibroblast growth factor 2 (FGF2) (Hebrok, Kim et al. 1998). These two factors are known to suppress Sonic hedgehog signaling system in dorsal pancreatic endoderm and thereby allow the marker gene of pancreas *Pdx1* to be expressed (Hebrok, Kim et al. 1998). On the contrary, the ventral pancreas, which is not in direct contact with the notochord, is primarily influenced by the signals from cardiogenic mesenchyme (e.g FGF and BMP) (Hebrok, Kim et al. 1998). In the absence of activin-β and FGF2 signals, Sonic hedgehog signaling system in ventral pancreas is activated (Gittes 2009). As a result, *Pdx1* is not expressed in the ventral region of the pancreas. Such influence makes this region of the pancreas to specifically differentiate into the major pancreatic ducts. Generally, the dorsal pancreatic bud is responsible for forming the head, body and tail of the pancreas, while the ventral pancreatic bud is responsible for forming the uncinate process (Gilbert 2000).

Pancreatic development involves formation of two distinct exocrine and endocrine tissue types from one simple epithelium (Gittes 2009). Previous studies have concluded that the fate of pancreatic epithelium is determined by the proximity of pancreatic epithelial cells to mesenchyme, the mesoderm that accumulates around pancreatic epithelial buds (Gittes 2009). When the epithelial cells are close to mesenchyme, proexocrine factor in mesenchyme interact with the epithelial cells in

cell-contact mediated fashion and exclusively differentiate the epithelial cells into exocrine acinar cells (Gittes 2009). On the contrary, in the absence of direct cell-contact mediated mesenchymal effect, epithelium exclusively differentiates into endocrine cells (Gittes 2009). There are many mesenchymal signals that are involved in the process of pancreatic cell specification (e.g. FGF10, Notch, TGFβ family, and Wnts) (Gittes 2009). Each of these signaling pathways is complicated and is beyond the scope of this thesis. In general, the pro-exocrine factors are mostly cell-contact mediated, while the pro-endocrine factors are diffusible (Gittes 2009). Relatively speaking, the early mesenchyme signals favor endocrine development, while late mesenchyme signals favor exocrine development (Gittes 2009). The notch-signaling pathway, for example, is a signaling pathway that promotes exocrine cell differentiation, is mostly expressed in late mesenchyme, and interacts with pancreatic epithelial cells via cell-mediated fashion (Kim, Shin et al. 2010).

The topology analysis of pancreas has revealed that the pancreatic differentiation involves signaling events that cumulatively activate transcriptional program (Gittes 2009). As mentioned previously, the notch-signaling pathway promotes exocrine cell differentiation. This is done so because the notch signal continues to stimulate the expression of *Ptf1α* (another transcription factor present in pancreatic progenitor cell like *Pdx1*) in exocrine progenitor cell. The endocrine pancreatic cells convert to non-epithelial cells as they slightly down regulate *Pdx1* and *Ptf1α* due to the absence of notch-signals (Kim, Shin et al. 2010). As a result, the transcription factor called neurogenin 3 (*Ngn3*) in endocrine progenitor cells gets expressed. The expression of *Ngn3* in the endocrine progenitor cells allows the islet

cell population to arise (Kim, Shin et al. 2010). Additional signals (ex. NeuroD1, Nkx2.2, Pax4, Nkx6.1 etc) are required for further specification of islet progenitor cells to give rise to different cell types (ex. α, β, δ, PP and ε) (Kim, Shin et al. 2010).

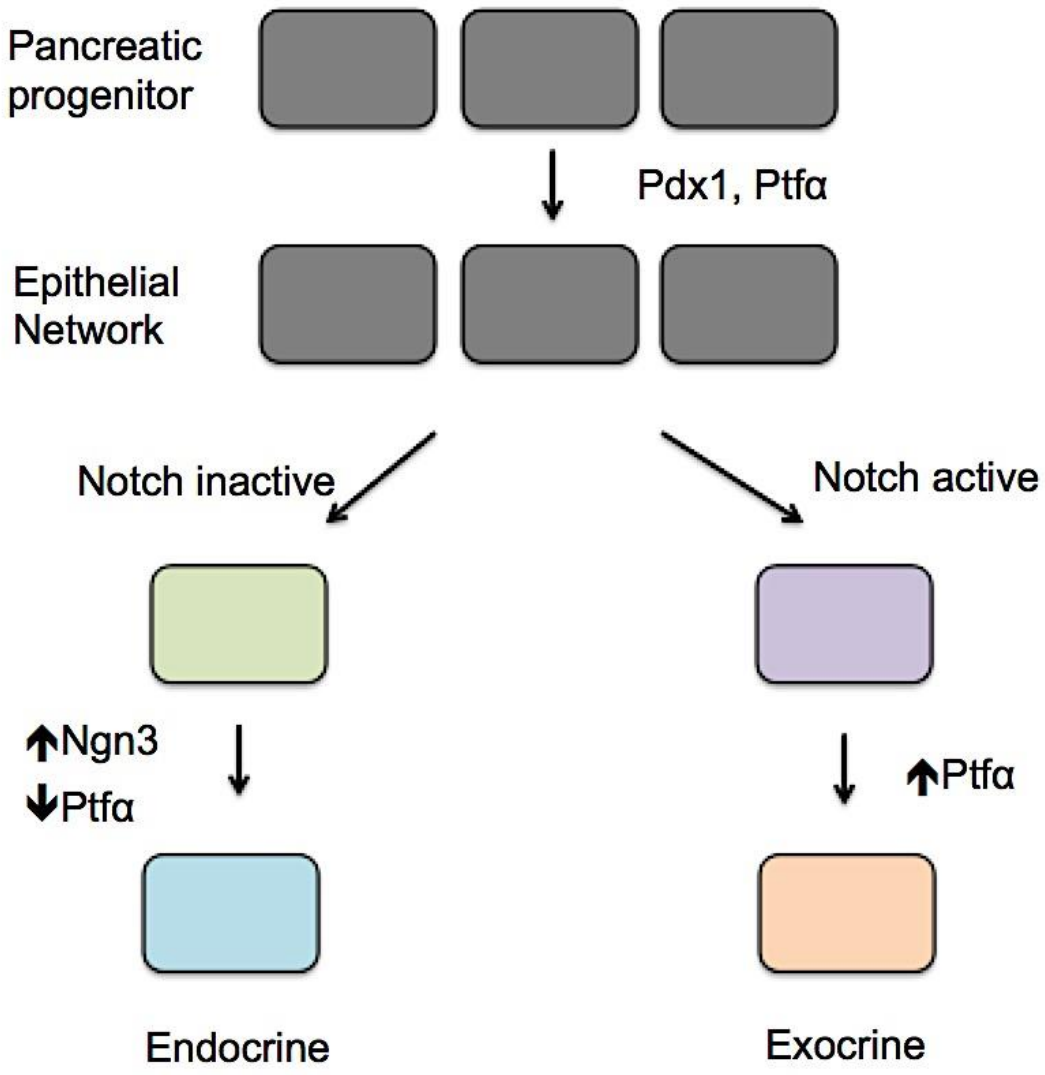


Illustration 2: Pancreatic cell lineage and the role of transcription factors during islet cell type development. Pancreatic cell types develop from pancreatic progenitor cells that express *Pdx1* and *Ptf1α*. The epithelial cell network, tightly regulated by the notch pathway, then gets converted into either endocrine or exocrine islet cell type.

Exocrine Pancreas

The exocrine pancreas plays a central role in the digestive process because its secretions provide digestion of a food into components that are available for absorption by the intestine. The exocrine pancreatic ducts are arranged in clusters of acini (from Latin term meaning "berries in a cluster"). Each unit of these acini (called acinus) is bound to each other by specialized intercellular connections known as tight junctions and gap junctions (Pandol 2010). The functional unit of the exocrine pancreas refers to an acinus and its draining ductule called intercalated ducts (Pandol 2010). The basolateral membranes of pancreatic acinus, consisting of many rough endoplasmic reticulum, is capable of synthesizing and modifying proteins (Pandol 2010). Once the pancreatic secretions are synthesized from acinus, they are secreted to the lumen via exocytosis. The secretions are then drained and accumulated in the intercalated duct that provides pathways for the secretions to be transported to the main pancreatic duct, the intralobular duct (Pandol 2010). The intralobular duct is important because it joins the pancreas to the common bile duct of gall bladder that supplies bile to the enzymes that are secreted by the pancreas (Pandol 2010). Together, the pancreatic secretions and bile get transported to the duodenum, where it starts to break down foods into small components.

The hormones secreted by cells in the stomach and duodenum (ex. gastrin, cholecystokinin, and secretin) control the synthesis of exocrine pancreatic secretion (Pandol 2010). The exocrine pancreas secretion contains proteolytic, amylolytic, and lipolytic digestive enzymes (e.g. trypsinogen, chymotrypsin, proelatase, and mesotrypsin). Most of these enzymes, because they are capable of autodegradating

pancreatic cells, are released in inactive precursor called zymogens (or proenzymes). Once the zymogens reach the surface of the duodenal lumen where enterokinase is located, are cleaved and activated (Pandol 2010). The pancreatic secretions, containing bicarbonate ions and enzymes, aid breaking down of carbohydrates, proteins and lipids in the chyme. As a result, nutrients from the chyme can readily be absorbed into the lumen of small intestine for body to utilize for energy.

Endocrine Pancreas

Glucose, a hexose sugar, is a physiologically important molecule because it is an absolute nutritional requirement for all cells, particularly for brain and neuron cells to generate energy and survive. The regulation of blood glucose concentration is crucial because there are many clinical implications that can potentially be entailed from both hyperglycemic (or high-blood sugar) conditions and hypoglycemic (or low-blood sugar) conditions. The endocrine pancreas, in partnership with cells of the liver, plays a major role in homeostatic regulation of the blood glucose level in the body.

The endocrine pancreas, surrounded by exocrine acini, constitute approximately 1-1.5% of the total pancreatic mass (Pandol 2010). The islet of Langerhans, the functional unit of an endocrine pancreas, consists of five different cell types: alpha (α), beta (β), delta (δ), gamma (γ), and epsilon (ε)-cells (Pandol 2010). The histology of the endocrine pancreas suggests that insulin producing PBCs are located in the central part of islets, whereas the other cells (α, δ, γ, and ε- cells) are located more toward the peripheral side of the islets (Pandol 2010). The cells of islets of Langerhans are involved in synthesizing peptide hormones like insulin (β-cells), glucagon (α-cells), pancreatic polypeptide (γ-cells) and somatostatin (δ-cells)

(Pandol 2010). For the purpose of this thesis, we will focus our discussion on insulin and glucagon, the ones that are most prominently involved in regulating blood glucose metabolism.

Insulin is a dipeptide hormone that consists of two polypeptide chains (each chain contains 21 to 30 amino acids) interlinked by hydrophobic interactions and three disulfide bonds (Blundell, Cutfield et al. 1971). When insulin, structurally homologous to insulin-like growth factors 1 and 2 (IGF-1 and 2), first gets synthesized, it is made as a single polypeptide molecule called preproinsulin. As it gets processed over from ER to Golgi apparatus, two preproinsulins get combined together and mature to an active form, the insulin. In the presence of high blood glucose (approximately 10 mmol/l), glucose from plasma enters the insulin-producing PBCs via glucose transporters (GLUT 2), a group of membrane proteins that facilitate the glucose transport over a plasma membrane. When glucose enters the PBC, it triggers glycolysis and the Krebs cycle to oxidize glucose, and more importantly, to produce ATPs. The change in ATP/ADP ratio is sensed by sulfonylurea receptor (SUR), a membrane protein that is known to sense intracellular levels of ATP and ADP and facilitate its associated potassium channel to promote insulin release from PBC (Campbell, Sansom et al. 2003). As a result of the change in ATP/ADP ratio, membrane depolarization and calcium influx occur. An increase in intercellular calcium ion concentration leads to the transcription of new insulin and the release of previously made insulin via secretory vesicles.

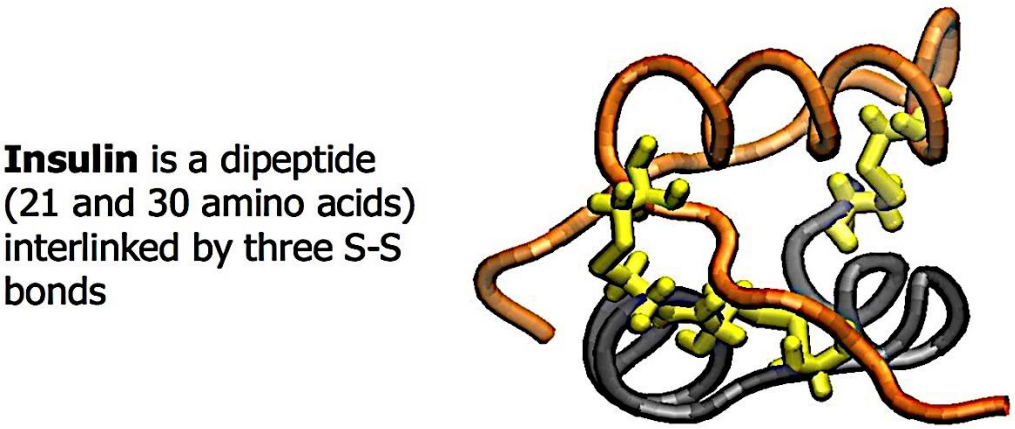


Illustration 3: 3D structure of an insulin. The backbone image of the X-ray crystallography image of an insulin protein is shown above. Note that the insulin protein is composed of two peptide chains interlinked by 3 disulfide bonds (Lehninger, Nelson et al. 2013)).

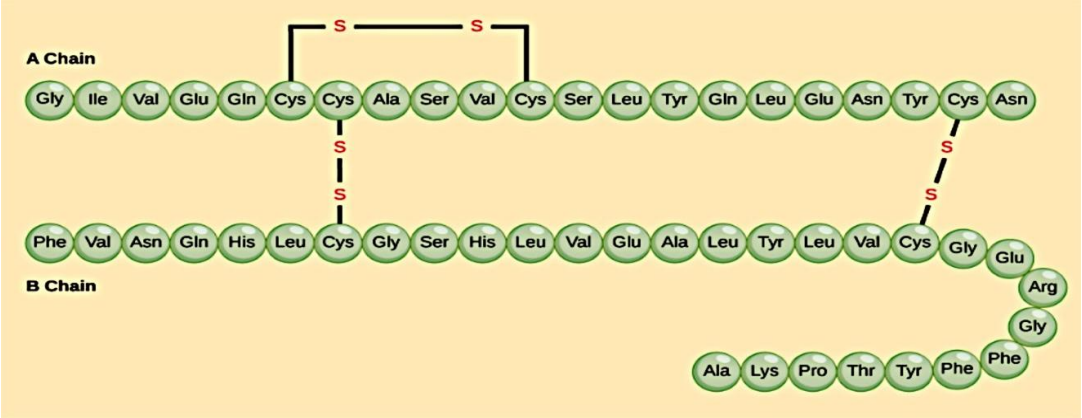


Illustration 4: Amino acid sequences of insulin. The each and every amino acid sequences of insulin gene is shown above. The thiol side chains in cysteine residues undergo oxidation to form disulfide bonds with each other (Lehninger, Nelson et al. 2013).

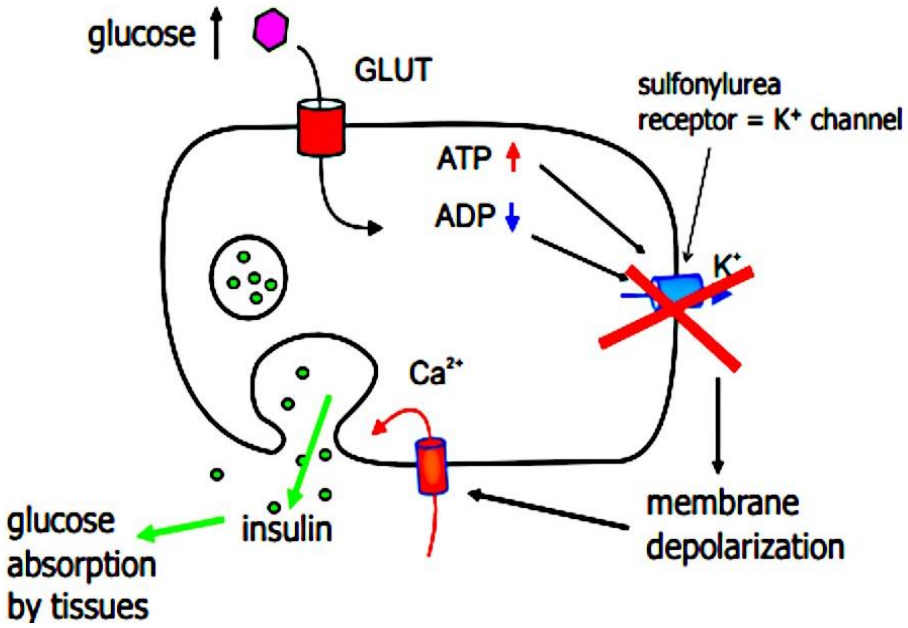


Illustration 5: Mechanisms of glucose-stimulated insulin secretion in PBC. The insulin-producing PBCs detect the raised blood glucose level via GLUT 2. Once this is perceived, PBCs undergo glycolysis and the Krebs cycle to oxidize glucose and produce ATPs. The change in ATP/ADP ratio is sensed by sulfonylurea receptor (SUR), which eventually leads to membrane depolarization and calcium influx. An increase in intercellular calcium ion concentration leads to the transcription of new insulin and the release of insulin.

Once insulin is released from PBC, insulin receptors (IR) on tissues (particularly on muscles and liver cells) detect the increased levels of circulating insulin. The IR belongs to the receptor tyrosine kinase family (phosphoinositide 3-kinase, PI3K) (Inoue 2015). The binding of insulin to IR induces the conformational changes within the IR that lead to the phosphorylation of the tyrosine residues within the IR domain. This phosphorylation recruits other adaptor proteins that eventually induce downstream processes involving blood glucose homeostasis (Kiselyov, Versteyhe et al. 2009). High levels of circulating insulin and its signal transduction cascades allow special transporter proteins on cell membranes of tissues (GLUT 4) to absorb glucose from the circulation to utilize it for energy. In addition to inducing glucose uptake, insulin triggers many metabolic changes. Insulin acts as a secondary messenger within the cell to stimulate the transcription of insulin-regulated genes such as *insulin receptor*, *IRS-1*, *p85αPI3K*, *p110αPI3K* (Ducluzeau, Perretti et al. 2001). Additionally, insulin triggers glycogen synthesis, lipid synthesis, and esterification of fatty acids to store glucose (Dimitriadis, Mitrou et al. 2011). On the contrary, insulin downregulates the pathways that counteract the effects of glucose storing pathways such as gluconeogenesis, lipolysis, and proteolysis.

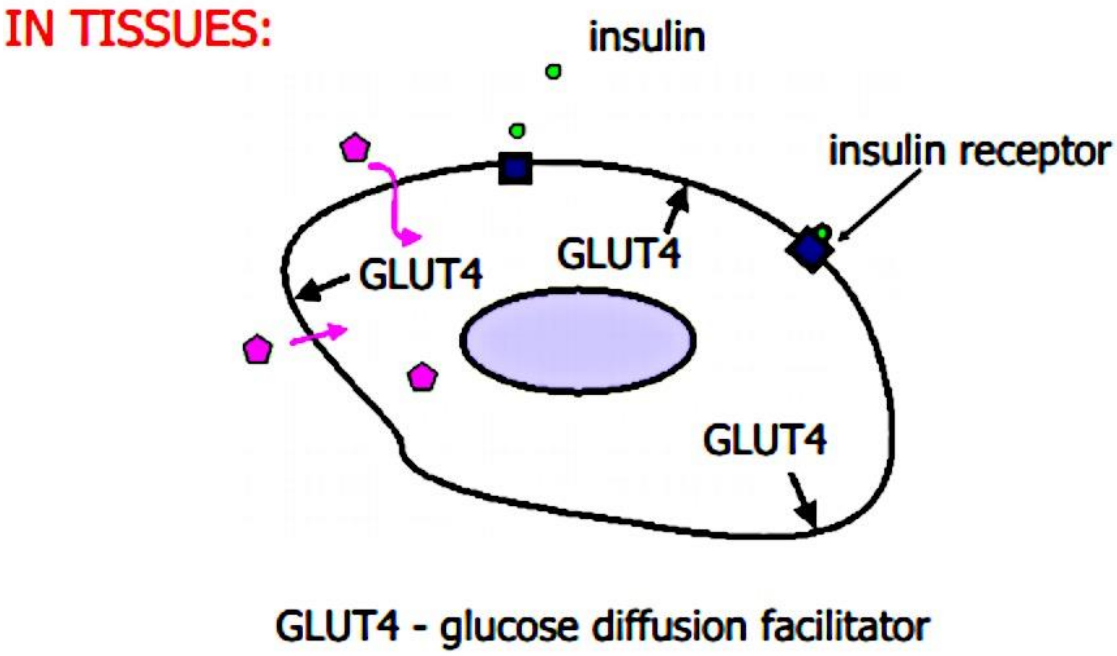


Illustration 6: A general actions of insulin on tissue cells. When insulin binds to an IR, IR changes its conformation and gets phosphorylated. This phosphorylation recruits other adaptor proteins that ultimately induce downstream processes involving blood glucose homeostasis (Lehninger, Nelson et al. 2013).

Glucagon, a peptide hormone produced by α-cells of the islets of pancreas, has opposing actions to those of insulin. Glucagon is a hormone consisting of 29 amino acids that folds to form a short α-helix (Nussey and Whitehead 2001). Similar to insulin, glucagon is synthesized as prepro-hormone (approximately 160 amino acids), which gets proteolytically cleaved to form active glucagon as it matures through Golgi apparatus in pancreatic α-cells (PAC) (Quesada, Tuduri et al. 2008). Glucagon is secreted when the PAC sense the absence or the low levels of glucose (4mmol/L). PACs have a specific set of channels that produce action potentials of sodium and calcium in the absence or at low levels of glucose (Quesada, Tuduri et al. 2008). Once the depolarization occurs in PAC, the change in electrical gradient stimulates chains of reactions that eventually lead to the secretion of glucagon from PAC via emiocytosis (Gosmanov, Gosmanov et al. 2000). The regulation of glucagon gene expression is not fully elucidated, although transcription factors like Arx, Pax6 and Nkx6.1 are necessary for PAC differentiation (Gosmain, Cheyssac et al. 2011).

Once glucagon is released from PAC, it binds to glucagon receptors, primarily located on the cells of liver and kidney. Glucagon receptors detect glucagon level and modulate the expression of several enzymes involved in glucose metabolism. The glucagon receptor, also known as seven-transmembrane domain receptors, belongs to G protein-coupled receptors (Mayo, Miller et al. 2003). The G-protein coupled receptor, when bound to its ligand like glucagon, undergoes a conformational change that triggers nearby GTP binding protein, or the G-protein, to be activated. The G-protein is made of trimeric subunits (α, β, and γ). To its alpha unit, G-protein is

bound by the guanosine diphosphate (GDP) subunit. When the glucagon receptor undergoes a conformational change, GDP bound to the alpha unit is replaced with guanosine triphosphate (GTP). The activated alpha subunit then dissociates from β and γ subunit to activate an effector molecule (e.g. adenylyl cyclase) that catalyzes the production of secondary messengers (e.g. cAMPs). The secondary messengers, cAMPs, are known to mediate the actions of glucagon on glycogenolysis and gluconeogenesis (Gosmanov, Gosmanov et al. 2000). The increased cAMP level triggers the intercellular kinases (e.g. protein kinase A (PKA)). PKA results in phosphorylation of phosphorylase that activates glycogen breakdown and deactivates glycogen formation (Gosmanov, Gosmanov et al. 2000).

Diabetes Mellitus

Diabetes mellitus describes a group of metabolic diseases that is characterized by hyperglycemia, or abnormally elevated glucose level in the blood stream. According to American Diabetes Association (ADA), a high blood glucose level with symptoms of hyperglycemia refers to a random plasma glucose level ≥ 200mg/dL, whereas a blood glucose level for individuals with no hyperglycemia have plasma glucose level ≤ 140mg/dL within two hours after a meal and between 70 to 99mg/dL after eight hours after a meal. The control of diabetes is important because the long-term hyperglycemia entails many clinical complications that result in damage or failure of different organs. These long-term complications of diabetes include retinopathy with potential loss of vision, nephropathy leading to renal failure,

peripheral neuropathy with risk of foot ulcer, and autonomic neuropathy causing cardiovascular symptoms (American Diabetes Association. 2013). There are four classes of diabetes: Class I, Class II, Class III, and gestational diabetes. For the purpose of this thesis, only Class I diabetes will be discussed in detail. Briefly, Class II diabetes, accounting for 90~95% of total DM cases, occurs when body tissues become insulin resistant (Wolf, Braun-Reichhart et al. 2014). Class III diabetes refers to DM that occur as a result of genetic defects and infections that interfere with insulin production in PBC (American Diabetes Association. 2013). Gestational diabetes, also called diabetes during pregnancy, is caused by excessive rise in blood glucose level due to increase in placental hormones (American Diabetes Association. 2013).

Class I diabetes (type 1 diabetes), often referred to as juvenile-onset diabetes or insulin-dependent diabetes, accounts for approximately 5 to 10% of all diabetes (American Diabetes Association. 2013). Class I diabetes is primarily caused by autoimmune destruction of the pancreatic β cell (PBC), the cell that is responsible for producing insulin. Because type 1 diabetic patients are not able to produce functional insulin, the tissues that have insulin receptors, particularly muscles and liver cells, are unable to effectively utilize blood glucose. As a result of this, approximately 85 to 90% of Class 1 diabetic patients exhibit hyperglycemic blood glucose levels even when they are fasted (American Diabetes Association. 2013).

The explicit cause of autoimmune destruction of PBC is not yet understood. However, there are some suggestions that multiple genetic predispositions and environmental factors play important roles in autoimmune destruction of PBC

(American Diabetes Association. 2013). The two major treatments available for type-1 diabetic patients are administration of synthetic insulin and transplantation of pancreatic sections or 1° PBC. Although these treatments have advanced diabetic care greatly, they are still inconvenient (often require immune-suppression), costly, and limited in supply. Consequently, there are still needs to not only enhance current diabetic treatments but also to explore next generation treatments.

Pig as a model to study DM

The pig has been suggested as an ideal model to study DM. The swine pancreas, like that of humans', is located posterior to the stomach (Ferrer, Scott et al. 2008, Swindle, Makin et al. 2012). There are three lobes of the pig pancreas: the splenic lobe, the duodenal lobe, and the connect lobe (Murakami, Hitomi et al. 1997). These three lobes relatively correspond to the tail/body, the head, and the uncinate process of the human pancreas respectively (Ferrer, Scott et al. 2008). Although Ferrer et al. and Morel et al. assert that the pancreatic ductal system vary significantly between different breeds, the pig pancreas in general is related to the proximal duodenum with a single pancreatic duct entering the duodenal lumen (Morel, Kaufmann et al. 1991, Ferrer, Scott et al. 2008, Swindle, Makin et al. 2012). The pancreatic ductal system and the distribution of the pancreatic lobes, the anatomical and morphological structures of the normal representation of the porcine model pancreas (the normal representation means that 97% of the 65 Landrace pigs exhibit the phenotype) resemble the pancreas of the human (Ferrer, Scott et al. 2008). As in human, this pancreatic duct is separate from the bile duct of gall bladder that supplies bile to the pancreatic enzymes (Pandol 2010). The vascular anatomy of the pig

pancreas is less variable than the ductal anatomy, and is also comparable to that of humans (Ferrer, Scott et al. 2008). Previous study conducted by Ferrer, Scott et al. has shown that the pig pancreas, especially that of the Landrace strain (the one that is mostly used for xenotransplantation), has analogous anatomical location of celiac trunk and main branches (splenic artery, hepatic artery and left gastric artery) to that of humans.

From a hormonal perspective, the glucose control mechanisms are very similar in pigs and humans (Eventov-Friedman, Tchorsh et al. 2006). In both species, the blood glucose level ranges between 70 and 115 mg/dl (Wolf, Braun-Reichhart et al. 2014) and its level tends to increase with the age for the same reasons in both species (Larsen, Rolin et al. 2001). The pig's pancreas, similar to that of humans, is involved in both exocrine and endocrine functions and is responsible for secretion of insulin and glucagon in response to changes in blood glucose levels (Renner, Blutke et al. 2015). Specifically relevant to this thesis, the proportion, distribution, and cell types of porcine Langerhans of islet are also very similar to that of humans (Steiner, Kim et al. 2010). As in human, porcine PBCs do not have sufficient regenerative capability (Bonner-Weir 2000), and their functionalities highly correlate with insulin levels in the body (Butler, Janson et al. 2003). Although porcine insulin and human insulin differ by one amino acid at position 30 of the B-chain (Aigner, Renner et al. 2010), a comparison of the effects of insulin analysis has proven that there is no clinically relevant difference between porcine insulin and human insulin (Richter and Neises 2003, Steiner, Kim et al. 2010). In fact, humans have used the porcine insulin

for decades to treat type 1 diabetic patients (Eventov-Friedman, Tchorsh et al. 2006, Aigner, Renner et al. 2010).

Lastly, our lab has successfully adapted means for genetically manipulating the pig genome (Whitelaw, Sheets et al. 2015). We have previously designed two TALEN sets for targeting prion-protein precursor (*PRNP*), a safe harbor locus using Golden Gate assembly platform (Addgene). We have tested these TALEN sets in both primary fetal fibroblasts and porcine embryos. Additionally, CRISPR/Cas system has been successfully adapted to modify gene loci of interest. We have bi-allelically knocked out one or more genes simultaneously by injecting in-vitro transcribed Cas9 mRNA and a target sgRNA into 1-cell embryo. For this experiment, Dr. Kieun Park, a colleague in the laboratory, has injected Cas9 mRNA alongside chimeric CRISPR sgRNA into 1-cell stage in vitro fertilized porcine embryo. The injected embryos were then cultured in vitro for 6 days, and the number of embryos progressing to the blastocyst stage was determined. The genomic DNAs from the blastocysts were then harvested, amplified using locus specific primers bordering the cut site, and sequenced to determine gene modifications. Our experiment has shown that the developmental rates of the injected embryos to blastocyst stage in vitro reached >30% (much higher than those of TALEN system), indicating that the CRISPR/Cas9 system is relatively non-toxic and very efficient.

There have been many studies where the pigs have made transgenic to study DM (Wolf, Braun-Reichhart et al. 2014). These pigs include transgenic pigs that down regulate GIPR (glucose-dependent insulinotropic polypeptide receptor) (Cheng, Wang et al. 2015), transgenic pigs that have defects in insulin-secreting gene named

HNF1A (hepatocyte nuclear factor-1α) (Wolf, Braun-Reichhart et al. 2014), and transgenic pigs that represent neonatal diabetes (Liu, Hodish et al. 2010). All of these studies have proven that the pig is a great model species that can be used to study human diseases, particularly DM.

Chapter 2: Materials and Methods

COL1A targeting plasmid construction and production of sgRNA

Expression plasmid for Cas9 nuclease (pMJ920) was a gift from Jennifer Doudna (Addgene plasmid # 42234) (Jinek, East et al. 2013). Targeting guide RNAs were designed based on the software available from MIT (Available online: http://www.genome-engineering.org/crispr/). Two complementary sgRNA oligo DNAs (22 nucleotides in length) were commercially synthesized (IDT DNA Technologies, Coralville, IA, USA), annealed to form double-strand DNA and cloned into a Bsa1 restriction enzyme digested in-house vector to yield a U6 promoter driven sgRNA expression cassette. The cloned fragments were DNA sequenced to confirm their fidelity. Confirmed sgRNA expressing vector was in vitro transcribed using MEGAshortscript T7 kit (Life Technologies, Carlsbad, CA, USA) to generate chimeric sgRNA. sgRNA was purified with MEGAclear kit (Life Technologies, CA, USA) for complexing with Cas9 protein prior to nucleofection.

Cell Nucleofection and Knockin Experiments

Cas9 plasmid + sgRNA DNA + oligo or Cas9 protein + sgRNA mRNA + oligo were prepared to a final volume of less than 5 μL. The Cas9 protein mixture was incubated

for 10 min at room temperature to allow ribonucleoprotein complex formation as per manufacturer's recommendation (Thermo Fisher Scientific, Waltham, MA, USA). Approximately, 1×10^6 cultured fetal fibroblast cells (culture conditions described below as for sorted cells) were harvested, washed once in PBS, and resuspended in nucleofection buffer (Lonza, Basel, Switzerland). About 5 µL of Cas9 plasmid mixture or Cas9 protein mixture, and cell suspension were combined in a Lonza 4D strip nucleocuvette. Reaction mixtures were electroporated using DO113 setting and immediately plated into one well of a 6-well plate at high density to facilitate recovery. Electroporated cells were incubated (with 10 µM/mL SCR7 or without SCR7) 30 °C for 72 h or 38.5 °C for 18 hand disassociated with 0.05% trypsin for sorting.

Single Cell Sorting and Culture For Colony Screening

Nucleofected cells were sorted for GFP expression on a BD Flowcytometer (San Jose, CA, USA) at the University of Maryland Flowcytometry Core at 1 cell/well density into 96-well plates, which were gelatinized by adding 100 µL of 0.1% gelatin solution for 1–3 h to facilitate attachment. Fifty microliters of 40% FCS High Glucose DMEM which was conditioned (CM) by incubating 20 mL with 2.5×10^6 irradiated CF1 mouse embryonic feeder cells/T75 flask. The CM was supplemented with 5 ng/µL bFGF and filtered before adding to 96 well plate. All wells were fed with 50 µL 10% FCS CM + bFGF immediately after sort, and 50 µL 20% FCS CM after 18 h. The sorted plates were incubated in 5% CO_2 + 5% O_2 38.5 °C for 7–10 days. Colonies that were 80%–100% confluent were split with 0.05% Trypsin, with one well split into 2 wells of 48 well (1:4; this is passage (P1)) and further incubated

at 38.5 °C for 3–5 days when 1 well was collected for DNA and the 2nd split into 1 well of 12 well (1:4) for further propagation (P2) or frozen in 92% FCS and 8% DMSO.

Somatic cell nuclear transfer (SCNT)

All animal work was performed as per the approved guidelines of Beltsville ARS, Institutional Animal Care and Use Committee. SCNT was performed as described in previous study (Jeong, Park et al. 2013). Cumulus-oocyte complexes (COCs) were purchased from a commercial supplier (De Soto Biosciences, Seymour, TN, USA). Briefly, matured oocytes were enucleated by aspirating the polar body and MII chromosomes with an enucleation pipette (Humagen, Charlottesville, VA, USA). After enucleation, a donor cell was introduced into the perivitelline space of an enucleated oocyte. Fusion of injected oocytes was induced by DC pulse (2.0 kV/cm for 30 μs using a BTX-Cell Manipulator 2001 (BTX, San Diego, CA, USA)). After fusion, the reconstructed oocytes were activated by an electric pulse (1.0 kV/cm for 60 μs), followed by 4 h of incubation in PZM3 medium containing 2 mM 6-dimethylaminopurine. Approximately 120–130 reconstructed oocytes were surgically transferred into the oviducts of naturally cycling gilts on the first day of standing estrus. Following transfer, pregnancies were confirmed on Day 30 by ultrasound. Fetuses were harvested from Day 45 pregnant euthanized sow. Tissues from four of the 11 fetuses were trypsin digested and incubated in T175 plates (P0) for subsequent culture and cryostorage.

Genotyping of Fibroblast Cell Colonies, and Edited Nuclear Transfer Fetuses

Single cell derived colonies cultured for 10–15 days were washed three times with PBS-PVA (p 7.4) medium. About 2–3 µL of colony suspension was transferred into 18 µL of colony lysis buffer (50 mM KCl, 1.5 mM $MgCl_2$, 10 mM Tris pH 8.0, 0.5% NP-40, 0.5% Tween-20 and 100 µg/mL proteinase K) and incubated for 1 h at 65 °C. The digestion was terminated by heating the mixture at 95 °C for 10 min, and 2 µL of supernatant was used as a PCR template. Tissue biopsies from fetuses were digested in a tissue lysis buffer (50 mM Tris pH 8.0, 0.1 M NaCl, 20 mM EDTA, 1% SDS, 50 µg/mL RNase A, 100 µg/mL proteinase K) overnight at 65 °C. Following overnight digest, the genomic DNA of the sample was extracted from the tissue lysate using phenol-chloroform, and recovered by resuspension in 100 µL of 10 mM Tris-HCl, pH 7.4 buffer following ethanol precipitation. Purified genomic DNA was amplified using PCR (primers in Appendix A), cloned into PCR2.1 vectors (Thermo Fisher Scientific, Waltham, MA, USA) and transformed into E. coli DH5-α maximum competent cells (Thermo Fisher Scientific). Five to ten colonies were picked, cultured, plasmid DNA extracted and sequenced (Macrogen, Rockville, MD, USA). Sequences were aligned by Bio-Edit software (Available online: http://www.mbio.ncsu.edu/BioEdit/bioedit.html) for comparison with wild-type and two attP sites per allele.

Targeted phiC31 Integrase Mediated Integration of GFP Transgene into Pseudo attP Sites in COL1A Locus

pDB2 and pCMV-Integrase plasmids were a gift from Michele Calos (Addgene plasmids # 18954 and 18935 respectively) (Groth, Olivares et al. 2000, Keravala,

Portlock et al. 2006). Porcine fetal fibroblasts containing pseudo attP sites downstream of COL1A site were nucleofected with a previously published GFP transgene (pDB2; 0.7 µg) containing consensus attB site for phiC31 integrase and CMV promoter driven integrase gene (1.3 µg) (Luo, Wang et al. 2015). A control nucleofection was performed with GFP transgene (pDB2) without the CMV-Integrase plasmid. Following nucleofection, the cells were selected for stable integration by selecting with 500 ng/µL of G418 for 7–10 days followed by flow cytometry. A PCR screen was performed with transgene specific and flanking COL1A sequence to show integration of the plasmid at the target site (primers in Appendix A).

Assembling and validating insulin locus targeting vectors

As in human, the porcine insulin gene consists of three coding sequences (exons) separated by two non-coding sequences (introns) (Han and Tuch 2001). Both the amino acid and the gene sequences of porcine insulin are available from the BLAST database (http://www.ncbi.nlm.nih.gov/gene/397415). We designed two sets of primers using the BLAST database (Appendix A). An approximately 1.2 KB of genomic fragment containing the region around the exon 3 of the insulin gene was PCR amplified and was cloned into PCR 2.1 with T4 ligase, both of which were commercially purchased from Life Technology, CA. This amplicon, essentially the homology arms of the targeting vector, was then sequentially amplified by overlap extension PCR amplification technique, a method that allowed us to introduce enzyme restriction sites (Sal1, Mlu1, and BsiW1) and delete some sequences (ex. stop codon, TAG) at the specific locus of our interest. Once we had the enzyme sites generated, we cloned in previously published selection cassettes (2A-GFP-PGK-Neo)

or reporter genes (2A-GFP) by restriction enzyme based cloning (Addgene plasmid #31938). The fused 2A-peptide refers to a 20 amino acid long self-cleaving peptide that originates from silkworm (Wang, Wang et al. 2015). The 2A-peptide has a cleavage site between the last G (glycine) and the first P (proline) residue at its C-terminus (ex. −LLNFDLLKLAGDVESNPG↓P-), allowing the nascent peptide chain to be released from ribosome as its peptide backbone undergoes hydrolysis of the peptidyl–$tRNA^{Gly}$ ester bond (Sharma, Yan et al. 2012, Wang, Wang et al. 2015). Subsequently, the 2A-peptide lets the ribosome to reinitiate translation on the same codon without any presence of a new promoter, allowing simultaneous expression of two individual proteins (ex. Insulin and GFP) from one open reading frame (Sharma, Yan et al. 2012).

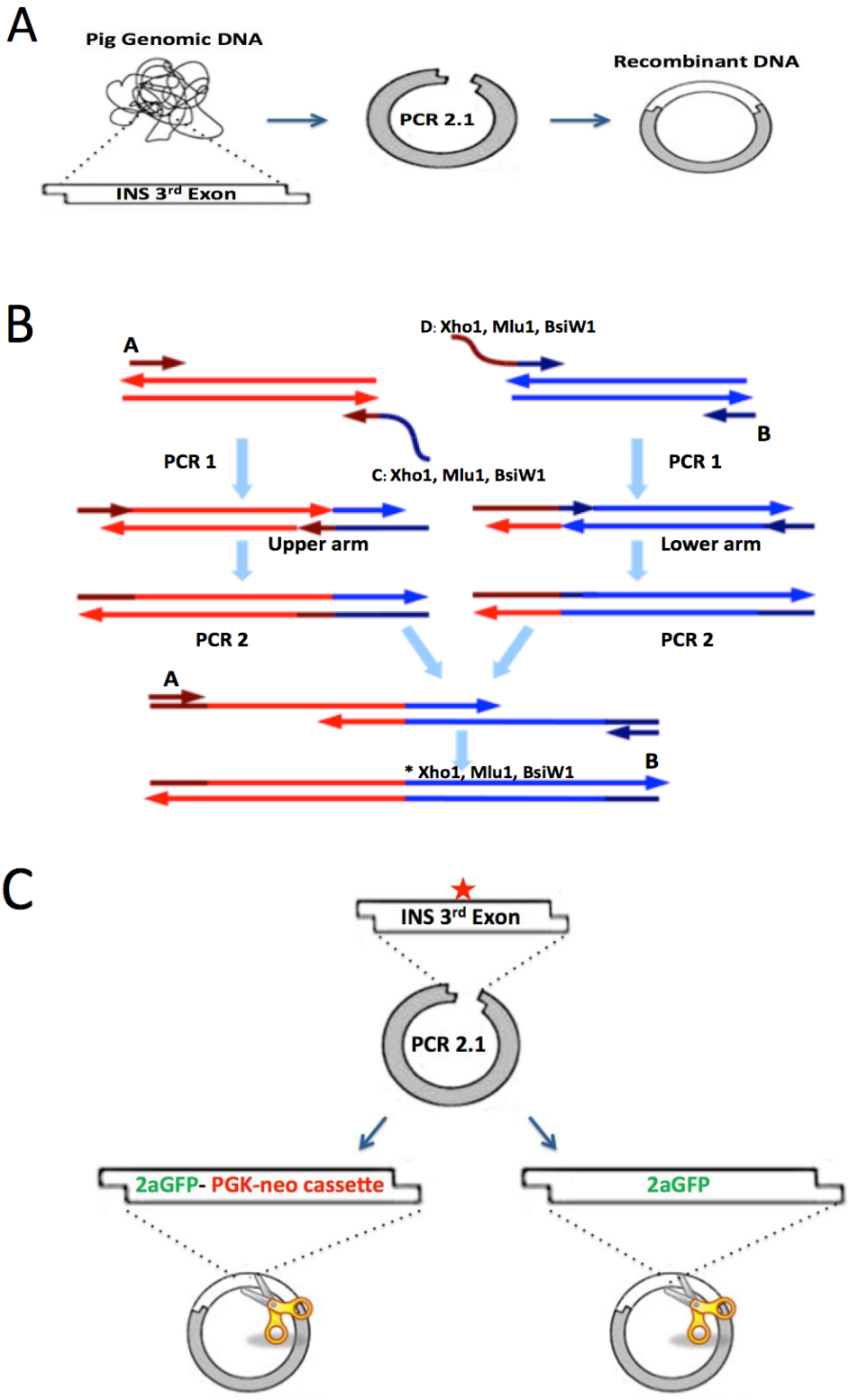


Figure 2: Assembling and validating insulin locus targeting vectors. *A)*

Cloning insulin exon 3 into PCR 2.1. An approximately 1.2 KB of gene region around the exon 3 of the insulin gene was amplified from porcine genome and cloned into PCR 2.1 vector**. *B) Overlap Extension PCR Amplification.***

The overlap extension PCR involves designing two primers in the opposite direction where one end of the primer (the 3' end) has complementary sequences of the template strand and the other end of the primer (the 5' end) has the mutation or deletion to be introduced. We designed two primers (indicated as primer C and D) in opposite direction to primers that contain the complementary sequences of the exon 3 insulin gene, and about 20 base pair sequences of restriction enzymes (Sal1, Mlu1, and BsiW1) to be introduced. We ran two separate PCR reactions (indicated as PCR 1) using the primers that we designed (primer C and D) in combination with the primers that we used to amplify the whole 1.2 KB fragment of the exon 3 insulin gene (indicated as A and B). We then obtained two fragments of the insulin exon 3 (indicated as upper arm and lower arm) that contain overlapping modified regions on one end of each of these fragment. We then ran another subsequent PCR with primers A and B, along with the two fragments of the insulin exon 3, such that the upper arm and lower arm fragment not only serve as a template but also serve as primers for each other. ***C) Targeting vector with selection cassettes.*** The resulting fragment with restriction enzyme sites added right in between the gene was then cloned into the PCR 2.1 vector. Once we had the enzyme sites generated, we cloned in selection cassettes (2A-GFP-PGK-Neo) or reporter genes (2A-GFP) by restriction enzyme based cloning.

Assembling and validating TALENs

TALENs were constructed using Golden Gate TALEN and TAL Effector Kit 2.0 from addgene. The targeting specificity of TALENs was determined by customizable arrays of repeat-variable di-residues (RVDs) in the TAL effectors. The Addgene kit offered RVDs that can be customized and bind to a certain DNA nucleotide (NI = A, NK = G HD = C, NG = T). The kit also provided TALEN backbone vector that contains TALE DNA binding domain (a site where we clone in our designed RVDs) fused to Fok 1 nucleases. We designed our TALENs using the TALEN Targeter software from Cornell University. When we provided the sequences of insulin exon 3, the software provided potential cut sites and corresponding RVD sequences based on naturally established guidelines from nine different genes from plants, animals, and protists (Cermak, Doyle et al. 2011). In order for the TALENs to be most efficient, the potential cut sites must be preceded by the nucleotide T (Cermak, Doyle et al. 2011). For this experiment, we chose the cut site that lies right before the termination codon of insulin exon 3 (↓TAG). We then assembled TALENs following the Golden Gate Kit protocol provided by Addgene. **Figure 3** shows Folk1 nucleases bound to opposing targets across a spacer over which the Fok1 domains come together to create the DSB. The TALENs specifically bind around the stop codon of insulin exon 3 and induce DSB. Once the DSB is created, we expected it to be repaired by HR with a targeting vector that has either GFP (**Fig 3. C**) or PGK-Neo cassettes (not shown).

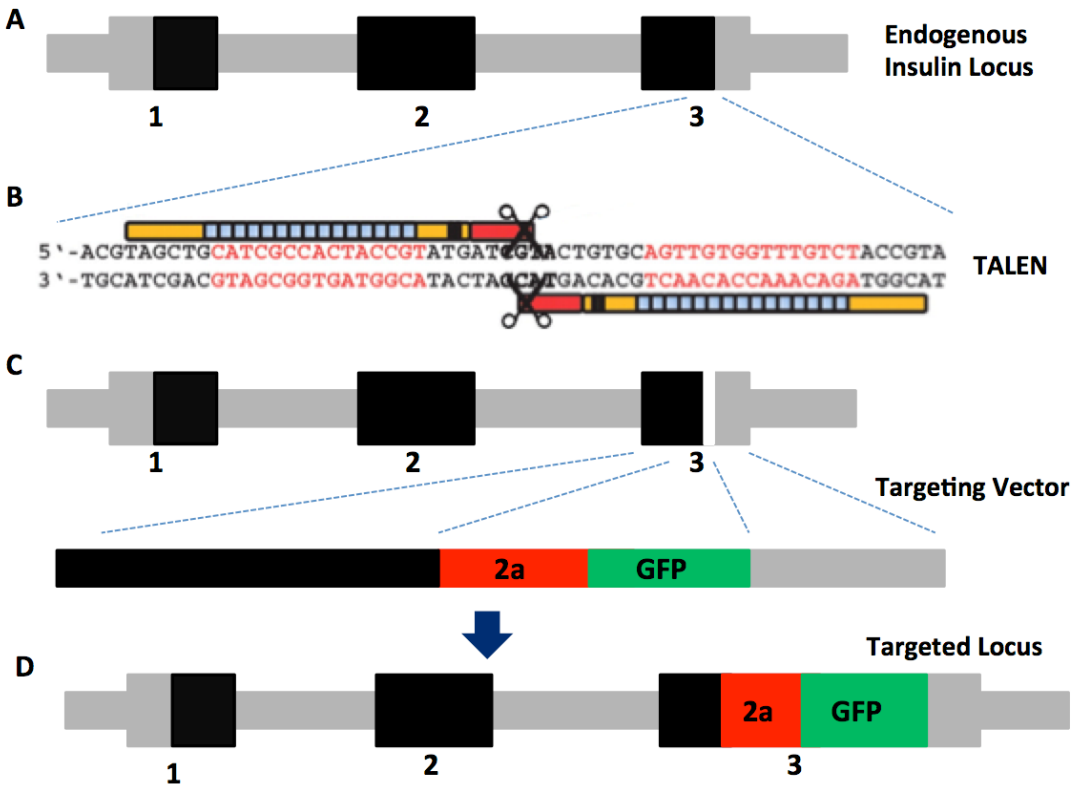


Figure 3: TALEN mediated knocking of 2a-GFP into the exon 3 insulin locus. **(A)** The endogenous insulin locus with the ATG start codon and TAG stop codon located in the first exon and the third exon, respectively (indicated as black rectangles). **(B)** TALENs targeting 3rd exon of insulin gene: the color blue indicates RVDs and the color red indicates Fok1 dimers. **(C)** Targeting vector consists of 500 bp of overlapping homology arm sequences flanking either sides of 2a-GFP. **(D)** End result, a knocking of the 2a-GFP in Insulin exon 3 loci.

Assembling and validating insulin CRISPR/Cas9

For this experiment, the Multiplex CRISPR/Cas9 Assembly System Kit was used to make CRISPR/Cas9 and induce DSB around the exon 3 of insulin gene. Two candidate sgRNAs targeting the exon 3 of insulin gene were designed based on the software developed by Dr. Feng Zhang at MIT (http://www.genome-engineering.org/crispr/) (Appendix A). As previously mentioned, the PAM motif must be present in order for gRNA-Cas9 to bind. Therefore, we selected two regions from insulin exon 3 locus where the PAM motif (5'-NGG-3') is present about 20 nucleotides upstream. We expected Cas9 complex to be localized to these regions and cut the locus with extreme accuracy. **Figure 4** shows the schematics of gRNA-Cas9 complex that gets recruited to the exon 3 insulin locus located directly upstream of the PAM motif. Once the complex gets localized to the target DNA, Cas9 excise the desired region with extreme accuracy resulting in a DSB. Similar to that of the TALEN system, the DSB is repaired by homologous recombination (HR) in the presence of a repair template, either the 2a-GFP or the 2a-GFP-PGKNeo.

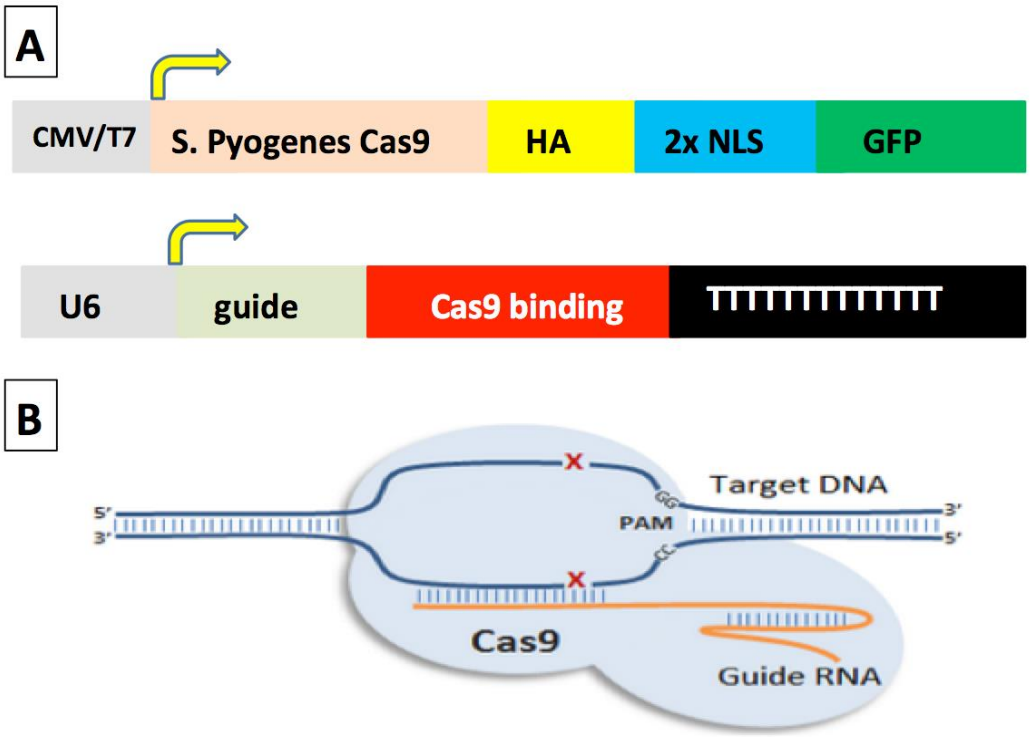


Figure 4: CRISPR-mediated gene knocking of 2a-GFP or 2a-GFP-PGK-Neo into the insulin exon 3 locus. (A) Schematic of CMV promoter for mammalian expression and T7 for in-vitro transcription of Cas9-GFP expression vector: HA (Hemagglutinin epitope) tag, and NLS (nuclear localization signals for nuclear localization). A U6 promoter driven sgRNA expression cassette harboring the guide RNA and Cas9 binding sequence. (B) Schematic of Cas9:sgRNA mediated targeting of intended insulin exon 3 locus.

Preparing targeting vectors for gene targeting in somatic cells and embryos

In order to optimize the targeting efficiency at the insulin locus, five different ways were used to prepare the targeting vector. Different methods yielded different qualities (ex. purity, concentration) of the targeting vector. In this section, all the methods that were used to prepare the targeting vector for SCNTs and microinjections are described in details.

1) *Linearizing targeting vector with one enzyme*

   Once the targeting vector has been made and validated through standard DNA sequencing (Macrogen USA), the targeting vector was maxipreped using the maxiprep kit from QIAGEN, and digested with EcoRV-HF restriction endonucleases from New England Biolabs (NEB) at 37C° for 2 hours. The targeting vector was then subsequently treated with sodium dodecyl sulphate (SDS) and proteinase K at 50C° for 30 minutes and 95 C° for 10 minutes. After all, the treated targeting vector was purified using phenol chloroform (from Thermo Fisher Scientific) extraction and ethanol precipitation.

2) *Linearizing targeting vector with two enzymes*

   The validated and maxipreped targeting vector was double-digested with EcoRV-HF and KpnI restriction endonucleases from New England Biolabs (NEB) at 37C° for 2 hours. The targeting vector was then sequentially treated with SDS and proteinase K at 50 C° for 30 minutes and 95 C° for 10 minutes. After all, the targeting vector was purified using phenol chloroform (from Thermo Fisher Scientific) extraction and ethanol precipitation.

3) *T7 exonuclease treatment of the targeting vector*

Previous studies have shown that the nicking of double-stranded DNA increases on-target activity by reducing off-target activity by 50 to 1500 fold (Zhang, Tee et al. 2015). In order to nick the double stranded targeting vector, we used T7 exonuclease (from NEB) that acts in the 5' to 3' direction, catalyzing the removal of 5' mononucleotides from duplex DNA. The validated targeting vector was PCR amplified with the primers that was designed to amplify the entire homology arm of the targeting vector from the porcine genomic DNA at 63 C° annealing temperature. We then ran our PCR product on 1.2% TBE gel, and gel extracted the region that contains homology arms and the 2a-GFP fragment (~2.3kb band). The gel extracted product was then sequentially treated with SDS and proteinase K at 50 C° for 30 minutes and 95 C° for 10 minutes. After all, the treated targeting vector was purified using phenol chloroform (from Thermo Fisher Scientific) extraction and ethanol precipitation.

4) *Guide RNA site modified targeting vector*

The PAM motif sequences that are present on the targeting vector were modified to avoid any excising event of the targeting vector by CRISPR/Cas9 complex. The PAM motif sequences were modified using the Q5- Site-Directed Mutagenesis Kit from Biolabs. As stated in the protocol, we designed a set of two mutating primers (Appendix A) and exponentially PCR-amplified at the annealing temperature of 65 C°. We then performed KLD (Kinase, Ligase, and DpnI) reaction that is responsible for ligating the mutated product into a commercial PCR 2.1 vector. After all, we transformed the plasmid into the thawed NEB DH5α cells and plated unto kanamycin plates.

5) *Single-stranded targeting vector*

Previous studies and our preliminary data have proven that the use of single-stranded DNA drastically increases targeting efficiency by reducing off-target activities (Yoshimi, Kunihiro et al. 2016). The single-stranded targeting vector was prepared by Long ssDNA Preparation Kit from Diagnocine. The region that contains homology arm and 2A-GFP fragment of targeting vector was restriction enzyme digested with Sma1 and Xba1 at 37 C° for 2 hours, and cloned into pLSODN-4D vector that was provided by Long ssDNA preparation kit. The cloned plasmid was then sequentially digested with Not I at 37 C° for 2 hours and Nb. BsrDI (a nicking endonucleases) at 65C° for another 2 hours. We then ran the digested products on 1.2% TBE gel, and gel extracted the region that contains homology arms and the 2a-GFP fragment (~2.3kb band). The gel extracted product was then sequentially treated with SDS and proteinase K at 50C° for 30 minutes and 95 C° for 10 minutes. After all, the treated targeting vector was purified using phenol chloroform (from Thermo Fisher Scientific) extraction and ethanol precipitation.

Preparing TALENs for gene targeting in somatic cells and embryos

Once the TALENs were assembled and validated through restriction enzyme digestion, the TALENs were in-vitro transcribed with MEGAscript T7 Transcription Kit from ThermoFisher Scientific. Before in-vitro transcription, the TALEN set was linearized with EcoRV restriction enzyme whose site is located right at the downstream of the region to be transcribed. After linearization, the template was sequentially treated with SDS and proteinase K at 50 C° for 30 minutes and 95 C° for

10 minutes, and subsequently was purified using phenol chloroform (from Thermo Fisher Scientific) extraction and ethanol precipitation. The purified TALEN set was thoroughly mixed with the RNA polymerase, ATP, CTP, GTP, UTP solutions, and was treated at 37 C° for 3 hours. Sequentially, TURBO DNase that catalyzes the degrading reaction of DNA was added to the reaction, which was then treated at 37 C° for additional 15 minutes. The transcribed reactions that contain mRNAs of TALEN set were purified with MEGAclear Kit from Life Technologies.

Preparing CRISPR/Cas9 for gene targeting in somatic cells and embryos

As mentioned earlier in this thesis, the main components of CRISPR/Cas gene editing system are guide RNA (gRNA) and Cas9 endonuclease. For both gRNA and Cas9, two different forms of gRNA and Cas9 were made. For gRNAs, we prepared both the in-vitro transcribed mRNA form of gRNA and the PCR- amplified DNA fragment form of gRNA. For Cas9, we commercially purchased both Cas9 plasmid and Cas9 protein. In general, gRNA plasmid was used with Cas9 plasmid, while in-vitro transcribed gRNA was used with Cas9 protein.

Two gRNAs were designed and in-vitro transcribed using the MEGAscript T7 Transcription Kit from ThermoFisher Scientific. Prior to the in-vitro transcription, the mRNA regions were PCR-amplified (Appendix A for primer information), which was then cleaned up using the QIAprep Spin Miniprep Kit. The PCR products, the templates for the in-vitro transcription, were thoroughly mixed with the RNA polymerase, ATP, CTP, GTP, UTP solutions, and was treated at 37 C° for 3 hours. Sequentially, TURBO DNase that catalyzes the degrading reaction of DNA was

added to the reaction, which was then treated at 37 C° for additional 15 minutes. The transcribed mRNAs were purified with MEGAclear Kit from Life Technologies.

Two 20 nucleotide- gRNAs sequences that contain the same 20-nucleotide sequences with those of in-vitro transcribed plasmid gRNAs were cloned in to a vector with U6 promoter. For these constructs, the regions that contain U6 promoter, 20-nucleotide guide sequences, Cas9 binding sites, and poly-T signal were PCR amplified, which were than cleaned up using the QIA prep Spin Miniprep Kit and used for SCNT.

Cell nucleofection

In our lab, we have two devices, Lonza's Nucleofector™ and Neon Transfection system, to transfect foreign DNAs into a cell line. Both devices are based on electroporation, a process which involves creation of small pores in cell membranes by applying high-voltage electrical pulse to introduce foreign DNAs. For our Somatic cell nuclear transfer (SCNT), we transfected our targeting vector and site-specific nucleases (either TALENs or CRISPR/Cas9) into porcine fibroblast cells. The porcine fibroblast cells were cultured in a medium, consisting of DMEM/F12 (1:1) medium and 10% fetal bovine serum (Hyclone). The cultured cells were then maintained in a cell culture incubator at 37°C. For each round of cell transfection, we transfected about 100,000 cells using either Lonza's Nucleofector™ or Neon Transfection system to compare transfection efficiency and cell survival rate at different concentrations of foreign DNA, pulse durations, numbers of pulses and voltage magnitudes.

Gene targeting directly in embryos

Maturing oocytes from sows were purchased from Desoto Biosciences (Seymour, TN). These oocytes were shipped to the lab overnight in their commercial maturation medium #1. Twenty-four hours after being placed in the maturation medium #1 (provided by Desoto), 50 to 75 COCs (cumulus oocyte complexes) were placed in 500 µl of tissue culture medium 199 containing 0.14% PVA, 10 ng/ml epidermal growth factor, 0.57 mM cysteine, 0.5 IU/ml porcine FSH, and 0.5 IU/ml ovine LH. The oocytes were then cultured for additional 20 hours at 38.5°C and 5% CO_2 in air, 100% humidity(Abeydeera, Wang et al. 1998). After incubation, COCs were vortexed in 0.1% hyaluronidase in HEPES-buffered medium containing 0.01% PVA for 4 minutes to remove the cumulus cells following maturation. Groups of 30 to 35 mature, denuded oocytes were then placed in 100µl of a modified Tris-buffered medium (mTBM) and were fertilized according to an established protocol from Abeydeera's lab (Abeydeera and Day 1997). After all, we injected our targeting vector and site-specific nucleases (either TALENs or CRISPR/Cas9). Once the presumptive zygotes are injected, we placed them in Porcine Zygote Medium 3 (PZM3) supplemented with 3 mg/ml fatty acid-free BSA (Yoshioka, Suzuki et al. 2002) and cultured at 38.5°C, 5% CO_2 and 100% humidity.

Screening targeting events

In order to screen targeting events from either SCNT or embryo injection experiments, three sets of primers (Appendix A) that target the region outside of the homology arms of the targeting vector were designed using the BLAST database. When the potentially targeted genomic DNAs were harvested from the cells or the embryos, we

used these screening primers to PCR-amplify the region that is supposedly targeted. Initially, the targeting events were screened by the size analysis of PCR amplicon. We confirmed the targeting events by sequencing the PCR amplicon through standard DNA sequencing service from Macrogen USA.

1) *Screening transfected cells from SCNT*

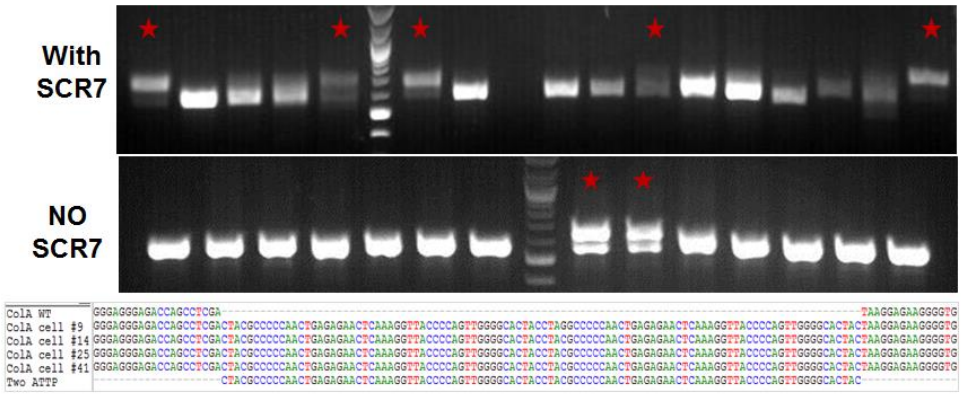
   The transfected porcine fibroblast cells were cultured in a medium, consisting of DMEM/F12 (1:1) medium and 10% fetal bovine serum (Hyclone) at stable 37°C for about 12 to 14 days. The status of cell growth was tracked about every 3 to 4 days. Once the cells propagate enough to be screened, a small subset of each cell colonies were picked up and washed with 1X Phosphate-Buffered Saline (PBS) solution. The colonies were then lysed with cell lysis beffer containing SDS and 0.1 mg/ml Proteinase K at 65°C for an hour. The lysed products were purified using phenol chloroform (from Thermo Fisher Scientific) extraction and ethanol precipitation to be screened with screening primers (Appendix A).

2) *Screening blastocysts*

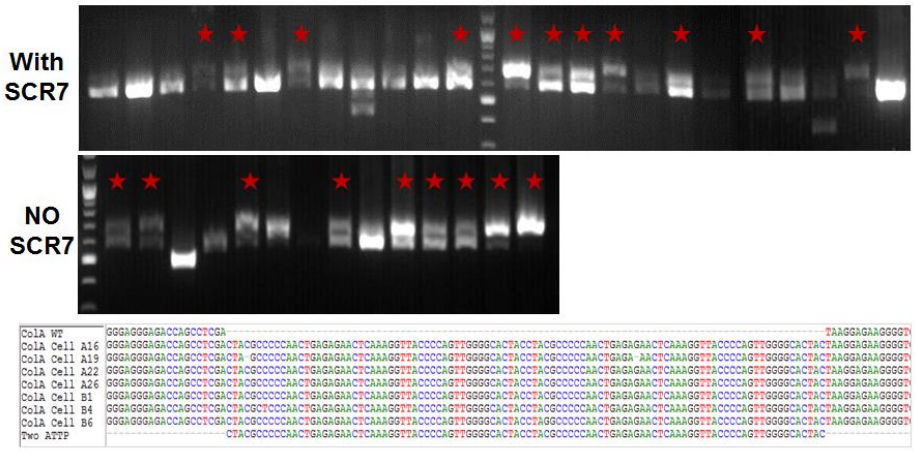
   The injected blastocysts were cultured in Porcine Zygote Medium 3 (PZM3) supplemented with 3 mg/ml fatty acid-free BSA (Yoshioka, Suzuki et al. 2002) at 38.5°C, 5% CO_2 and 100% humidity. After 6 to 7 days, the cultured blastocysts were lysed with BL lysed buffer (non-denaturing condition, 0.1 mg/ml Proteinase K) at 50°C for an hour. The lysed products were directly used for PCR assay.

Chapter 3: Results

A. Cas9 plasmid + sgRNA plasmid + Oligo at 30°C



B. Cas9 plasmid + sgRNA DNA + Oligo at 38.5°C



C. Cas9 protein + sgRNA mRNA + Oligo at 38.5°C

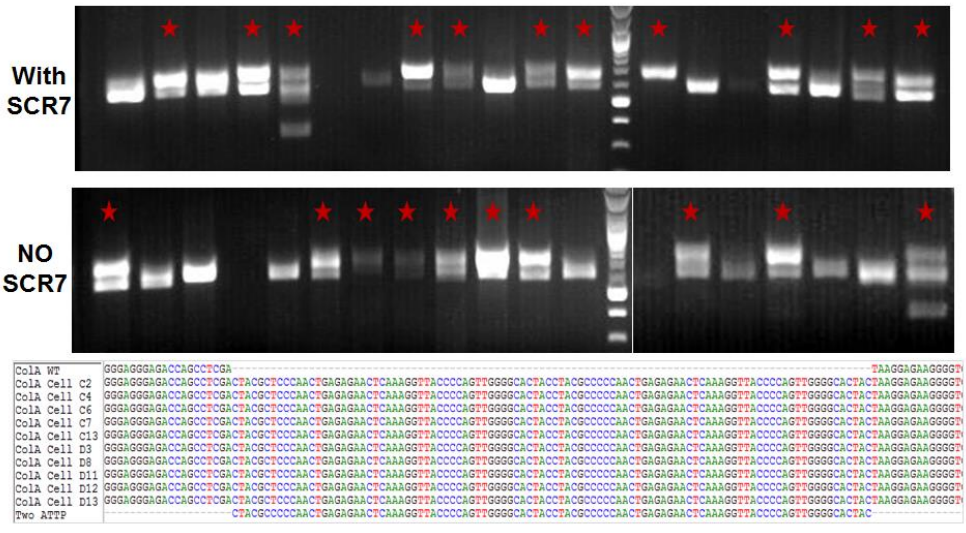


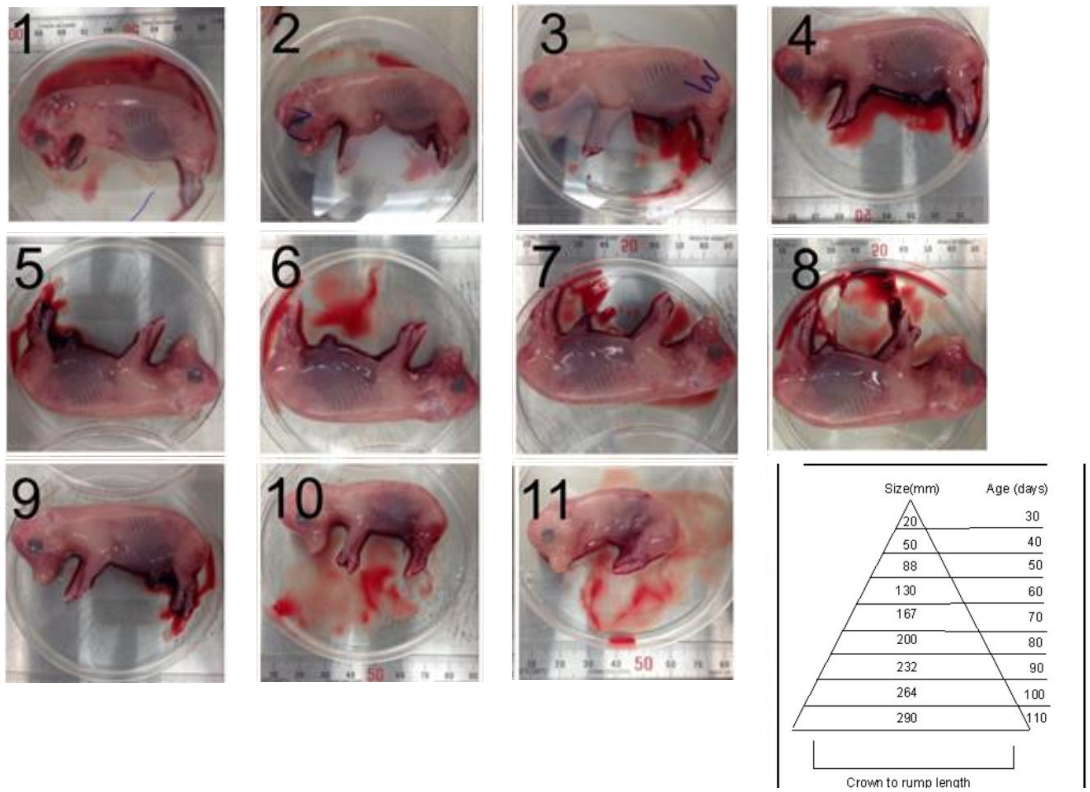
Figure 5: Targeted Knockin of Pseudo-attP Sites Downstream of Porcine COL1A Gene in Somatic Cells Using CRISPR/Cas System.

Knockin of two pseudo attP sites (100 bp) downstream of the porcine COL1A locus. (A–C) Agarose gel electrophoresis analysis of amplicons generated by primers bordering the COL1A targeting site. PCR amplicons of representative colonies derived from plasmid transfection and transient incubation at 30 °C (A) or 38.5 °C (B); and ribonucleoprotein (C) transfection and culture at 38.5 °C. In each panel, representative colonies showing a new band with predicted shift of 100 bp on the agarose gel produced by successful targeting and introgression of the two pseudo attP sites is marked by a red star. In each panel, the top row represents knockin in the presence of SCR7, the middle row without SCR7 and the bottom row represents Sanger sequencing of the amplicon. Multiple sequence alignment of wild type (WT) COL1A gene and two pseudo attP knockin sites are shown. Each nucleotide is color coded.

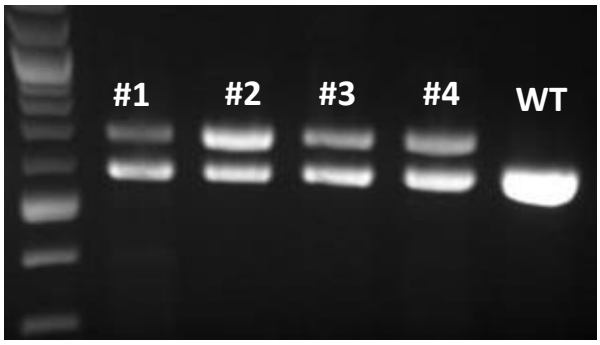
Table 2. Summary of CRISPR/Cas knockin experiments.

Cas9 & sgRNA Combination	Treatment with SCR7	Initial Culture Temperature	Number of Single Cell Sorted Wells	Number of Colonies Formed	% of Colonies Formed/Sorted Wells	Number of Targeted Colonies	Percentage of Targeted Colonies/Colonies Formed
Plasmid	None	30 °C	960	16	1.70%	2	13%
Plasmid	SCR7	30 °C	960	41	4.30%	18	44%
Plasmid	None	38.5 °C	376	14	3.70%	8	57%
Plasmid	SCR7	38.5 °C	376	39	10.40%	17	44%
Ribonucleoprotein	None	38.5 °C	376	58	15.40%	18	31%
Ribonucleoprotein	SCR7	38.5 °C	376	61	16.20%	19	31%

The incubation of plates at 38.5 °C resulted in a greater number of established colonies. When compared to plasmid delivery, nucleofection of Cas9 ribonucleoprotein complex resulted in a fourfold and 1.5-fold increase in the percentage of established colonies in the absence and presence of SCR7 (3.7% vs. 15.4% and 10.4% vs. 16.2%), respectively (Figure 5C and Table 2). Even though we found more colonies with the use of Cas9 ribonucleoproteins, the fraction of targeted colonies were not higher than plasmid transfections (31% vs. 57% or 44%).



(A)



(B)

```
ColA WT          GGGAGACCAGCCTCGA----------------------------------------------------------------------------------------TAAGGAGAAGGGGT
ColA NT fetus #1 GGGAGACCAGCCTCGACTACGCCCCCAACTGAGAGAACTCAAAGGTTACCCCAGTCGGGGCACTACCTACGCCCCCAACTGAGAGAACTCAAAGGTTACCCCAGTTGGGGCACTACTAAGGAGAAGGGGT
ColA NT fetus #2 GGGAGACCAGCCTCGACTACGCCCCCAACTGAGAGAACTCAAAGGTTACCCCAGTTGGGGCACTACCTACGCCCCCAACTGAGAGAACTCAAAGGTTACCCCAGTTGGGGCACTACTAAGGAGAAGGGGT
ColA NT fetus #3 GGGAGACCAGCCTCGACTACGCCCCCAACTGAGAGAACTCAAAGGTTACCCCAGTTGGGGCACTACCTACGCCCCCAACTGAGAGAACTCAAAGGTTACCCCAGTTGGGGCACTACTAAGGAGAAGGGGT
ColA NT fetus #4 GGGAGACCAGCCTCGACTACGCCCCCAACTGAGAGAACTCAAAGGTTACCCCAGTTGGGGCACTACCTACGCCCCCAACTGAGAGAACTCAAAGGTTACCCCAGTTGGGGCACTACTAAGGAGAAGGGGT
Two ATTP          ----------CTACGCCCCCAACTGAGAGAACTCAAAGGTTACCCCAGTTGGGGCACTACCTACGCCCCCAACTGAGAGAACTCAAAGGTTACCCCAGTTGGGGCACTAC----------
```

(C)

Figure 6: Somatic Cell Nuclear Transfer to Generate Clonal COL1A Targeted Fetuses and Fibroblast Lines.

Somatic cell nuclear transfer (SCNT) of porcine COL1A targeted fetal fibroblast cells. (A) Photomicrograph depicts 11 clonal fetuses (numbered 1–11) at Day 45 of

pregnancy along with a chart depicting crown-rump length and age of the fetus measure is shown; (B) Agarose gel electrophoresis image of COL1A amplicons from fetuses (numbered 1–4) on the top showing targeted knockin of the two attP sites and a 100 bp shift in size. Amplicon from wildtype (WT) animal is shown as a reference; (C) Sanger sequencing of the clonal lines confirming the inclusion of the target sites. Multiple sequence alignment of wild type (WT) COL1A gene and two pseudo attP knockin sites are shown. Each nucleotide is color-coded.

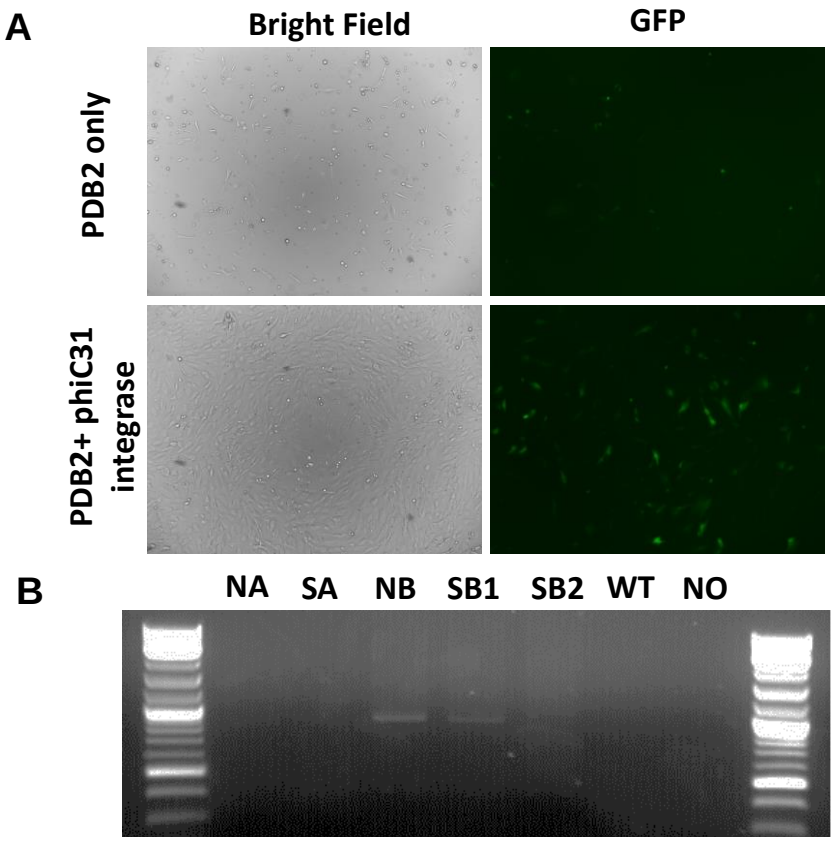
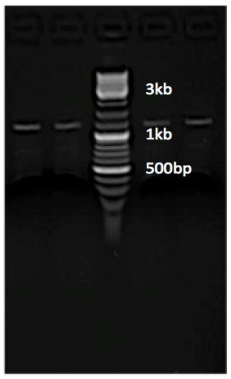


Figure 7: Phic31 Integrase Mediated Integration of Functional GFP Transgene into the attP Modified COL1A Locus.

PhiC31 integrase mediated integration of GFP transgene into COL1A targeted clonal lines. (A) Brightfield and fluorescent micrographs (100× magnification) showing integration and expression of GFP transgene in COL1A knockin lines in the absence (top panel) or presence of integrase (bottom panel); (B) PCR amplicons showing targeted knockin of GFP transgene into the target sites. The cells following nucleofection and G418 selection and sorting were analyzed by PCR. Amplicons that show knockin were identified only in the presence of integrase and not in vector only treatment, wild type parent cells and no-template controls.

A

Amplified Insulin 3rd Exon



B

Sus scrofa EWB tyrosine hydroxylase (TH) gene, exon 14 and partial cds; preproinsulin (INS) gene, complete cds; and insulin-like growth factor 2 preproprotein (IGF2) gene, complete cds, alternatively spliced

PCR product sequence (Ins-homology F primer)

Range 1: 2522 to 2966GenBankGraphicsNext Match Previous Match

Score	Expect	Identities	Gaps	Strand
798 bits(432)	0.0	440/445(99%)	1/445(0%)	Plus/Plus

Figure 8: Assembling and validating insulin locus targeting vectors.

A) An approximately 1.2 KB of genomic fragment containing the arms of homology was amplified from porcine genomic DNA. The gel picture below shows the PCR amplified 1.2 KB band, which was then gel extracted and cloned into PCR 2.1 vector. B) This recombinant DNA was sent our for standard DNA sequencing (Macrogen, Rockville, MD). The DNA sequencing result was then blasted on a database. The BLAST result shown above confirms that the 1.2 KB genomic fragment that was amplified from the PCR reactions correspond to 99% porcine insulin gene.

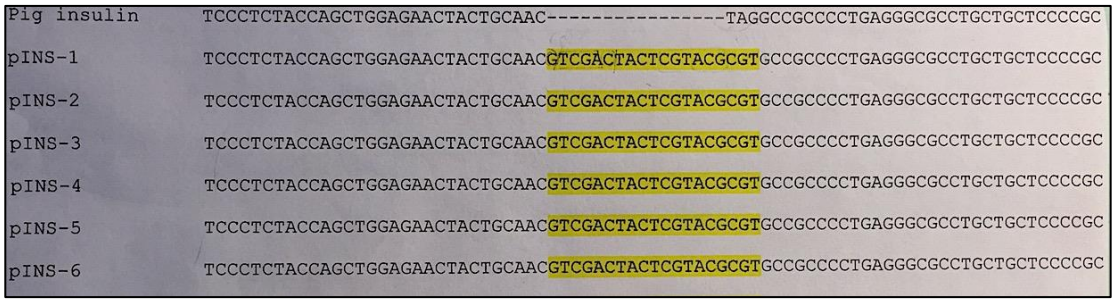


Figure 9: Sequencing Results- pINS homology arms with enzyme sites.

Overlap extension PCR was used to replace the stop codon (TAG) with the 3 enzyme restriction sites (Sal1, Mlu1, and BsiW1) that allow us to clone in either the selection cassettes (2A-GFP-PGK-Neo) or the reporter genes (2A-GFP) by restriction enzyme based cloning. Below sequencing results reveal the sequences of 3 newly generated enzyme sites by overlap PCR amplification. In comparison to the sequences of the wild type pig insulin gene (first lane), 6 bacterial colonies that contain the plasmid from overlap extension PCR exhibited three enzyme sites (Mlu1 (ACGCGT), Sal1 (GTCGAC), and BsiW1 (CGTACG)) in a replacement of TAG, the stop codon. Once the enzyme sites were generated, we used Mlu1 and Sal1 sites and cloned in both the selection cassettes (2A-GFP-PGK-Neo) and the reporter genes (2A-GFP).

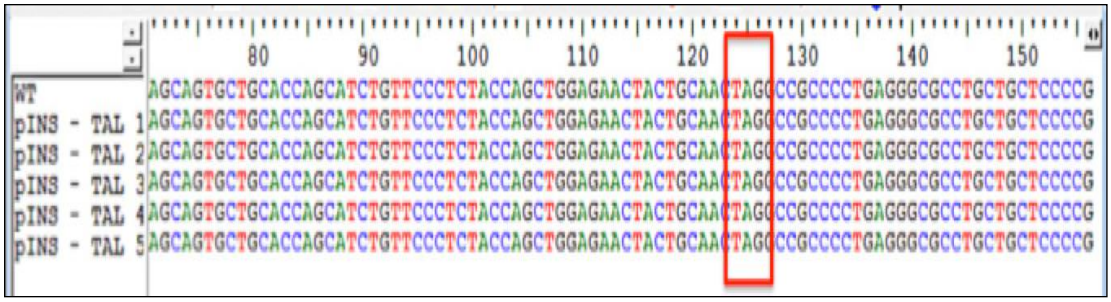


Figure 10: Assembling and validating TALENs.

TALENs, constructed using Golden Gate TALEN and TAL Effector Kit 2.0 from addgene, were in-vitro transcribed and transfected into the porcine fibroblast cells along with the insulin-targeting vector. The targeting event was screened using the primers around the stop codon locus (indicated as a red box). The first row indicates the DNA sequences of wild type (WT) insulin gene. The other rows labeled from pINS–TAL 1 to pINS–TAL 5 all indicate the sequences of genomic DNA harvested from the transfected cell colonies. The sequencing data below show that the transfected cell colonies do not have any gene modification at the target site.

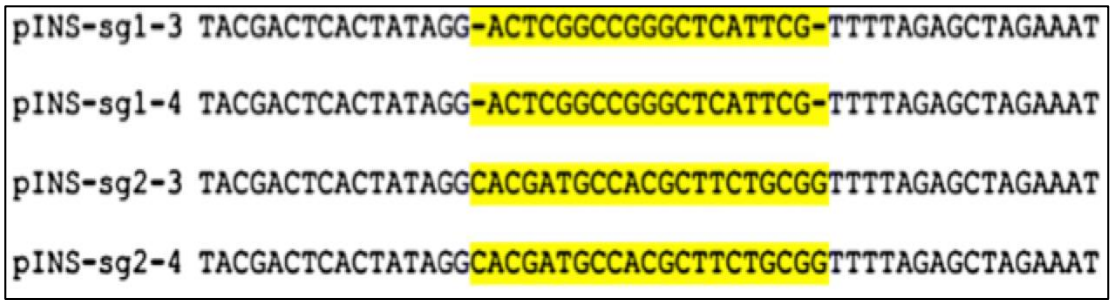


Figure 11: Assembling and validating CRSPR/Cas9.

Two candidate sgRNAs targeting the exon 3 of insulin gene were designed based on the software available from MIT (http://www.genome-engineering.org /crispr/) and commercially synthesized (IDT DNA Technologies, IA). Two complementary sgRNA oligo were annealed to form double-strand DNA and cloned into a vector that consist of T7 promoter and Cas9 binding site. The partial sequencing results of two sgRNAs-containing plasmids are shown above. The highlighted regions indicate the two sgRNA sequences.

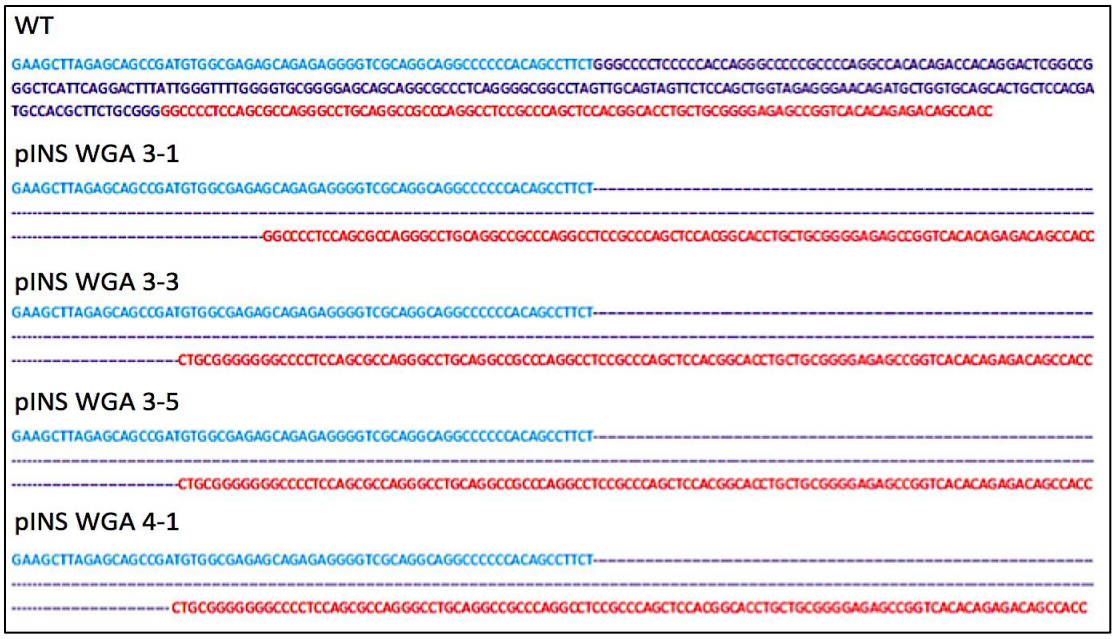


Figure 12: CRISPR/ Cas9 injected embryo WGA sequencing results.

The in-vitro transcribed sgRNA #2 and Cas9 plasmids were injected into one-cell stage blastocysts. The genomic DNAs of the blastocysts were harvested and whole-genome amplified. The whole-genome amplified products were then screened using the screening primers (Appendix A) and cloned into PCR 2.1 vector. The sequences of the resulting plasmid are shown above. In comparison to the wild types sequences (indicated as WT), the harvest genomic DNAs from blastocysts showed huge genomic fragment deletion.

Table 4: Summary of somatic cell gene targeting experiments.

Experiment Date	Targeting vector	ssNuclease	ssNuclease concentration	Transfection system	# of cells	Screening PCR polymerase	Targeting events?
10/27/14 (Experiment 1)	One enzyme linearized pINS-2a GFP & pINS-2aGFP-NeoKan (1.8mg/µl each)	TALEN RNAs	1mg/µl, (For); 1mg/µl (Rev)	Lonza's Nucleofector™	200,000 (100,000 each)	KOD, Taq	No
02/02/15 (Experiment 2)	One enzyme linearized pINS-2a GFP & pINS-2aGFP-NeoKan (1.8mg/µl)	sgRNA/ Cas9 plasmid	sgRNA: 357 ng/µl, Cas9: 1.8 mg/µl,	Lonza's Nucleofector™	200,000 (100,000 each)	KOD, Taq	No
04/08/15 (Experiment 3)	Two enzymes linearized pINS-2a GFP(3mg/µl)	sgRNA/ Cas9 plasmid	sgRNA: 357 ng/µl, Cas9 1.8 ng/µl,	Neon Transfection system	Total of 400,000	KOD, Taq	No
01/08/15 (Experiment 4)	T7 exonuclease-treated pINS-2aGFP (2.9mg/µl),	sgRNA #2 mRNA/ Cas9 protein	sgRNA mRNA: 792 ng/ µl, Cas9: 1.2 mg/ µl	Lonza's Nucleofector™	100,000	KOD, Taq	7
05/31/16 (Experiment 5)	Mutated ssDNA plasmid- pINS – 2AGFP	sgRNA #2, Cas9 protein	sgRNA #2 (1.3 mg/ µl), Cas9 (2mg/ µl)	Lonza's Nucleofector™	100,000	KOD, Taq, GoTaq, Platinum Taq	7-10

Over a 2-year period, a total of 5 somatic cell gene-targeting experiments were performed. An overview of each of these experimental conditions and the number of cells that potentially were targeted is briefly summarized below. For each cell transfection experiment, 1µl of targeting vector DNA and 2 µl of site-specific nucleases (1µl of forward/1µl of reverse TALENs or 1µl of Cas9/ 1µl of sgRNA) were transfected.

Table 4: Summary of embryo injection.

Experiment date	Targeting vector	ssNuclease	ssNuclease concentration	# of blastocysts	PCR polymerase	Targeting events?
06/19/14 (EI 1)	N/A	sgRNA #1 and # 2 mRNAs, Cas9 protein	sgRNA #1 112.5ng/µl, sgRNA #2 301.6ng/µl,, Cas9 protein (25ng/ µl)	30 (15 for sgRNA #1, 15 for sgRNA #2)	KOD, Taq	N/A
01/02/15(EI 2)	Linearized pINS-2a GFP (20 ng/µl)	sgRNA #1 and # 2 mRNAs, Cas9 protein	sgRNA #1 112.5ng/µl, sgRNA #2 301.6ng/µl,, Cas9 protein (25ng/ µl)	32 (15 for sgRNA #1, 15 for sgRNA #2)	KOD, Taq,	N/A
03/16/15(ET3)	Linearized 2 enzyme digested pINS-2aGFP (20ng/µl)	sgRNA #1 and # 2 mRNAs, Cas9 protein	sgRNA #1 112.5ng/µl, sgRNA #2 301.6ng/µl,, Cas9 protein (25ng/ µl)	20 (10 for sgRNA #1, 10 for sgRNA #2)	KOD, Taq	N/A
02/26/14(ET 4)	Mutated pINS-2aGFP (20ng/µl), Non-mutated pINS-2aGFP (20ng/ µl)	sgRNA mRNAs/ Cas9 protein, sgRNA mRNAs/ Cas9 nickase protein	sgRNA: 12.5 ng/ µl, Cas9 plasmid: 25 ng/ µl, sgRNA mRNA: 12.5 ng/ µl, Cas9 nickase: 25 ng/ µl	110	KOD, Taq	10
06/01/16(ET 5)	Mutated ssDNA plasmid (50 ng/ µl), Mutated ssDNA (20 ng/µl)	sgRNA mRNAs/ Cas9 protein	sgRNA mRNAs (12.5ng/µl)/ Cas9 protein (50 ng/ µl)	80	KOD, Taq	9-10

Over a 2-year period, a total of 5 embryo injection experiments were performed. An overview of each of these experimental conditions and the number of cells that potentially were targeted is briefly summarized below. For each embryo injection experiment, 1µl of targeting vector DNA and 2 µl of site-specific nucleases (1µl of Cas9/1µl of sgRNA) were transfected.

Chapter 4: Discussion

The pig is a valuable animal model for translational research and biotechnology. In this thesis study, we sought to optimize procedures for performing gene targeting in pigs for biomedical applications. We have succeeded in demonstrating that it is possible to create site-specific knockins of short DNA sequences in a gene locus of our interest. Over a 2-year period, we tested site-specific nucleases (TALENs and CRISPR/Cas), established targeting vector preparation methods, improved targeting event screening methods, and tested experimental conditions for optimized targeting efficiency. Through direct comparisons of different gene targeting reagents and experimental designs, we have established and optimized protocols to successfully knockin in porcine somatic cells, created a stable platform for future transgene integration, and generated transgenic pigs that will allow us to study biomedical applications, such as type 1 diabetes mellitus (DM).

I. Optimizing methodologies for gene targeting in porcine somatic cells:

In part one of this thesis study, we specifically created site-specific knockins of short DNA sequences (pseudo attP sites (50 bp)) downstream of a ubiquitously expressed *COL1A* locus. In doing so, we were able to ultimately permit subsequent phiC31 integrase-mediated introduction of functional transgenes. As previously reported, we tested whether transient culture in low temperature promotes HR (Doyon, Choi et al. 2010, Miller, Tan et al. 2011). In this study, we also tested the use of SCR7, a DNA ligase IV inhibitor that is known to bias the repair pathway towards HR as opposed to the NHEJ pathway (Chu, Weber et al. 2015, Maruyama, Dougan et

al. 2015). Additionally, we also compared the use of different choice of CRISPR reagents in improving the HR frequency (Park, Park et al. 2016).

In this study, we have not recognized any influence of low temperature on HR outcome. In order to test the influence of temperature on colony formation and HDR outcome, the nucleofected cells (porcine fetal fibroblasts in early passage that were nucleofected with plasmid expressing a Cas9:GFP fusion and an sgRNA that targets the *COL1A* locus) were cultured in 10% fetal calf serum (FCS) supplemented DMEM medium in the presence or absence of DNA ligase IV and NHEJ pathway inhibitor, SCR7 (10 µM/ml) at 30°C for 3 days or 38.5°C for 1 day before sorting. One day after nucleofection, the GFP expressing cells were sorted into a 96 well plate. The sorted cells were then cultured in irradiated mouse embryonic feeder conditioned medium (CM) supplemented with 5 ng/ ml FGF2 at 38.5°C in 5% CO_2 and 5% O_2. The colonies that formed following sorting were further sub-cultured in 48 well plate at 38.5°C until confluence, when the colonies were counted, collected by trypsin, and genomic DNA isolated for genetic screening. As shown in Fig. 5A and Table 2, less than 2 % (1.7%) of the sorted cells established colonies at 30°C, and inclusion of SCR7 resulting in greater than doubling of colonies established (4.3 %). In the established colonies, approximately 13 % of colonies showed a 100 bp shift in the product size suggesting a knockin of two pseudo attP sites in the *COL1A* targeting site (that was verified by DNA sequence analysis of an amplicon of this region). Inclusion of SCR7 at 30°C resulted in greater than a 3 fold increase (13% vs. 44 %) in the percentage of targeted colonies (Table-2). When compared to cultures maintained at 30°C, incubating the plates at 38.5°C resulted in a doubling of the

percentage of colonies in both the presence and absence of SCR7 (10.4% and 3.7%, respectively) (Fig.1B and Table 2), and an increase in targeting efficiencies (44% vs 57 % respectively). In conclusion, we have not identified any advantage of culturing nucleofected cells at low temperatures as previous reported (Doyon, Choi et al. 2010, Miller, Tan et al. 2011). In fact, there was a greater than twofold reduction in the number of colonies formed at 30ºC as compared to 38.5 ºC (the temperature at which porcine somatic cells are routinely cultured, presumably due to a lag in cell cycle).

In regards to SCR 7, our study has agreed with previous report that the use of SCR7 enhances gene targeted colony formation (Chu, Weber et al. 2015, Maruyama, Dougan et al. 2015). The use of SCR 7 in part one of this thesis study improved colony formation efficiency and consequently the frequency of gene targeting events by 3 fold at 30ºC. Similarly, at 38.5 ºC using plasmid transfection, SCR7 increased the number of colonies by 3 fold, however, the percentage of targeted lines was not different. Our studies have revealed that the SCR 7 solution noticeably improved the effectiveness of survival and proliferation of cells. However, the finding that SCR7 clearly and absolutely improves clonogenicity cannot be conclusively determined, because control experiments with just the solvent (DMSO) are lacking.

Finally, the choice of CRISPR reagents, either plasmid- or protein- based delivery into cells has been tested. The use of CRISPR ribonucleoprotein results in greater than 1.5 fold increase in the number of colonies established. Additionally, we observed an obvious improvement in the number of robustly growing colonies and greater than 30% targeting efficiency in the established clones when we delivered CRISPR ribonucleoprotein complex as opposed to Cas9 nuclease expression plasmid.

By delivering CRISPR ribonucleoprotein complex, we have identified two main advantages: Firstly, there was high survival and proliferation of nucleofected cells, making it less labor intensive; Secondly, CRISPR ribonucleoprotein is commercially available, preventing any potential differences in reagent preparation between different investigators and improving the ability of an experiment to be duplicated.

Utilizing the CRISPR/Cas system we were able to knockin pseudo attP sites (50 bp) downstream of a ubiquitously expressed *COL1A* gene in porcine somatic cells for transgene integration in future and consequently established means to generating transgenic porcine models more efficiently. In fact, we have utilized this site to integrate functional GFP transgene in *COL1A* locus mediated by phiC31 integrase. In the future, the pseudo attP sites can be used to achieve routine targeted genetic modifications in pigs. To do this, we expect to use porcine fetal fibroblasts containing pseudo attP sites, nucleofect with donor vector that consist of any desired transgene and consensus attB site for phiC31 intergrase, sort the cells into single cells, identify the percentage of integration of transgene into the attP knockin sites, and finally generate genetically modified pigs. With this process, we expect that the transgene is not only specifically integrated downstream of an ubiquitously expressed *COL1A* gene, but also protected from any potential limitations that are stated above including but are not limited to insertional mutagenesis, lack of control over transgene copy number, positional variegation, random assortment and segregation in subsequent generations.

In summary, through the head to head comparison of different gene targeting reagents and experimental designs, we standardized procedures for routine gene

targeting and demonstrated that it is possible to create site-specific knockins of short DNA sequences in a gene locus of our interest. These experiments were performed in parallel with our second effort of generating insulin-GFP reporter pigs in part two of this thesis study, where we hoped to tag the porcine insulin locus with green fluorescent protein (GFP) to generate insulin reporter pigs.

II. Generating Insulin- GFP reporter pigs: In part two of this thesis study, we sought to generate insulin reporter pigs and utilize these pigs to further test potential for cellular therapy, validate potential medications, and perform reprograming studies. As part of this study, several approaches were tested. 1) Choice of site specific-nuclease (TALENs or CRISPR/ Cas9); 2) choice of starting material (somatic cell vs zygote); 3) choice of CRISPR reagents (plasmids vs ribonucleoprotein); and 4) choice of targeting reagents (double stranded vs single stranded). An ultimate goal was to replace the translation stop codon (TAA) and introduce a sequence for "2a" self- cleaving peptide and GFP transgene in frame with insulin locus. The expectation was that when the insulin gene is transcribed, the GFP transgene is transcribed in cis- and results in co-translation, giving a readout of insulin expression in real time, and in doing so label the insulin producing cells.

1. TALENs vs CRIPSR/cas: Both of TALENs and CRISR/cas site-specific nucleases are known to engineer a DSB at the target site and promote gene targeting or homologous recombination that improve gene-targeting efficiencies by nearly 1000 fold. Except for the site-specific nuclease, these two targeting experiments were performed under the same working conditions (same targeting vector, same cell transfection system, and same screening methods). From the DNA sequencing results

of the harvested genomic DNA samples from TALEN experiment (Table 10), we did not observe any difference between the DNA sequences of the transfected cell colonies to that of WT insulin gene. Although the sequencing result (Figure 10) only represents 5 transfected colonies, these 5 colonies were selected from over 65 colonies (not shown) based on gel electrophoresis analysis with various screening primer combinations (Appendix A, Appendix C). Initially, we ran screening PCR reactions with 65 harvested genomic DNA samples as templates. For each of these reactions, we used a GFP primer along with an insulin gene-specific primer that is located outside of the homology arm to observe the presence of the GFP fragment at the insulin locus. From the 65 transfected colonies that could have potentially undergone targeting events, we did not see any target size band that must be observed if any targeting event has occurred. Additionally, we have run 4 more PCR reactions with different combinations of primers to validate our results. All of our gel electrophoresis analyses have confirmed that the GFP fragment was not integrated at the locus of interest. In order to see if the TALENs have disrupted the insulin gene at all, we PCR-amplified and sequenced selected 5 colonies around cut site (Figure 10). Here, we confirmed that the TALENs were not effective in not just targeting, but also modifying the porcine insulin locus.

Conversely, our results from Experiment #2 with CRISPR/cas seemed more promising than with TALENS. The gel electrophoresis analysis of the harvested genomic DNA samples from Experiment #2 has shown 1-2 colonies (out of 100,000 transfected cells and 78 propagated cell lines) that might have undergone targeting events. When amplified with the screening primers (the same primers that we used to

screen Experiment #1), the harvested genomic DNA from the transfected cell colonies of Experiment #2 showed the target size band that should be observed if any targeting event has occurred. This result suggested that the insulin locus may be comprised of 2A-GFP fragment around the 500-bp homology arms. Despite the fact that the target size bands were too weak and smeared for further analysis, the CRISPR/Cas system was confirmed to be more effective and promising than TALENs in targeting insulin locus.

2) Neon Transfection System vs Lonza's Nucleofector™: In addition to the choice of editors, we also tested the choice of transfection methodologies Neon Transfection System from Thermo Fisher Scientific vs Lonza's Nucleofector™ System, both based on electroporation and apparatus that are widely a process which involves creation of small pores in cell membranes by applying high-voltage electrical pulse to introduce foreign DNAs. In order to optimize electroporation parameters for the cell type, different treatments (varied voltage, time, # of pulse) were given to the total of 400,000 cells (100,000 cells each).

Treatment	Voltage	Time (in nanosec)	# of Pulse
1	1000	20	2
2	1200	30	1
3	1400	20	1
4	1500	20	1

From Experiment #3 (Table 4), genomic DNA from a total of 88 transfected cell colonies were harvested and screened for targeting events. The same screening methods were used to screen colonies from Experiment # 2 (Table 4). Despite all these efforts to increase insulin-targeting events, none of the transfected cells were

found to be targeted. In general, the survival rates from Neon nucleofection system were a lot lower than Lonza system.

Through Experiments #1-3 above, an often-overlooked difficulty in gene targeting was raised, that is to target a locus that potentially lies within a heterochromatin region of the genome in somatic cells. For TALENs, more so than CRISPR/Cas, the target locus must be readily available for them to bind and generate DSB. Unlike CRISPR/Cas system, which unwinds some portions of genomic DNAs to scans for PAM motif site (Zhang, Tee et al. 2015), TALENs neither unwind genomic DNA nor actively scan for the target site. That being said, if the insulin locus lies within a heterochromatin region in porcine somatic cells, then TALENs may not be able to bind or generate any DSB at the locus. We figured that the insulin gene, because it is not necessarily expressed in matured fibroblast cells, might be located at a region that is tightly packed, preventing TALENs to readily access the site. Therefore, we decided to use only the CRISPR/Cas9 gene editing system for the following SCNT experiments. Additionally, because of high viability and clonogenesity, we decided to utilize traditional Lonza's Nucleofector™ system for all of our subsequent experiments in cells.

3) Choice of starting material (somatic cell vs zygote): Given the specific difficulty in accessing loci in potential heterochromatin regions, we considered direct injection of gene-editors into the one-cell stage porcine zygote. Previous studies have already successfully demonstrated the generation of transgenic pigs by injection of editors into zygotes (Whitelaw, Sheets et al. 2016). The genome of the one-cell stage embryo, which is in an open chromatin formation, theoretically is readily accessible

for any site- specific nucleases to trigger modification. As a result, we decided to inject insulin editors into zygotes for our subsequent experiments. In order to test if CRISPR/Cas system works in embryos, Cas9 plasmid and an in-vitro transcribed sgRNA #2 were injected into one-cell stage zygotes (EI #1). Figure 12 shows the DNA sequencing results of PCR-amplified products of whole-genome amplified harvested genomic DNAs from each blastocyst after 6 days of in vitro culture following embryo injections. Here, the region around the cut site was amplified with the screening primers (Appendix A) and got cloned into PCR 2.1 vector to obtain clearer sequencing results. Comparing the wild types sequences (indicated as WT) to that of editor-injected blastocysts, the harvested genomic DNAs from editor-injected blastocysts showed huge genomic fragment deletion. Only the DNA sequencing results from two injected embryos were obtained due to difficulties in obtaining clear DNA sequencing results and/or cloning into the PCR 2.1 vector. Solely from this experiment (ET #1), we could not determine which method (either cells vs embryos) is more effective in generating insulin reporter pigs. Therefore, both cell and embryo targeting were performed side to side under similar experimental conditions. Regardless, the huge genomic fragment deletion observed from the obtained DNA sequencing results (Figure 12) confirmed that the CRISPR/Cas system is very effective in embryos.

Previously, many studies have been done to systematically optimize targeting events in porcine species (Wang, Zhou et al. 2015, Whitelaw, Sheets et al. 2016). These studies have proven that the uses of single stranded DNA (ssDNA) is more effective in facilitating homology directed repair (HDR) driven knockin. In our

studies, we sought to validate this argument by attempting to knock-in the porcine insulin gene with nicked double-stranded DNA and ssDNA.

4. Single vs double stranded targeting vector: Our targeting vector whose 5' end mononucleotides were removed from duplex DNA upon T7 exonuclease treatment was nucleofected into fibroblast cells (Experiment #4). All the experimental conditions of Experiment #4 were similar to that of Experiment #3, except for the use of Scr 7, a DNA ligase IV inhibitor that is known to bias the repair pathway towards HDR as opposed to the non-homologous end joining repair pathway. From our part one of the thesis study, we observed that the use of SCR 7 improve colony formation efficiency and consequently increase the frequency of gene targeting events. However, our control experiments results did not conclusively validate this argument. From the gel electrophoresis analysis of the harvested genomic DNA samples from Experiment #4 (Table 4), we have found 7 colonies (out of 85 propagated cell lines) that might have undergone targeting events. When amplified with the screening primers (the same primers that we used to screen previous SCNT experiments), the harvested genomic DNA from the transfected cell colonies of Experiment #4 showed the target size band that should be observed if any targeting event has occurred. This result suggested that the insulin locus may be comprised of 2A-GFP fragment around the 500-bp homology arms. Comparing data between Experiment #3 and Experiment #4 (Table 4), we have decided that the nicked targeting vector is yielding better targeting efficiency. Although we were able to see some promising gel electrophoresis analysis result (specifically from Experiment #5) showing predicted shift of 1 kb on the agarose gel produced by amplicons that were generated by

screening primers bordering the insulin targeting site, we were not able to further investigate with the experiment. We performed somatic cell nuclear transfer of these potentially insulin targeted fetal fibroblast cells as followed by the same protocols that we used to do part one of this thesis experiment. However, pregnancies of surgically transferred recipient pigs were futile. As a result, we were not able to obtain fetuses that we could have harvested from day 60 pregnant euthanized sow for further investigation. Additionally, direct embryo injections also did not result in successful pregnancies.

Conclusion

In the process of trying to generate insulin reporter pigs, we were able to further test and compare different gene targeting reagents and experimental designs to specifically modify insulin locus in pigs. Even though we were not able to generate insulin reporter pigs, we tested several gene-targeting methodologies that will ultimately help us to generate the pig model of diabetes in the future.

As thoroughly explained in part one of this thesis, we standardized procedures for routine porcine gene targeting methodology and demonstrated that it is possible to create site-specific knockins of short DNA sequences in a gene locus of our interest, specifically porcine *COL1A* locus. In part two of this thesis study, we utilized some of these established and optimized protocol for CRISPR/Cas mediated knockin in porcine somatic cells. Despite our efforts, we were not able to successfully and clearly verify by DNA sequence analysis that we have a definitive clonal targeted porcine fibroblast cells that have undergone gene-targeting events at the insulin locus. We believe that this has to do with the transgene size difference between insulin targeting vector plasmid (2.5 kb) and *COL1A* targeting vector plasmid (200 bp). In regards to *COL1A* targeting vector plasmid, we assembled a total of 200 nt single stranded DNA oligonucleotide containing two pseudo attP sites (50 bp each; 100 bp total) and 100 bp overall homology (50 bp on either side of the attP sites). On the contrary, we assembled insulin targeting vector that consist of a total of 2.5 kb double stranded DNA containing 2A-GFP sequences (1 kb) and 1.5 kb overall homology (700 bp and 800 bp on either side of the GPF sites). Over the course of our part two

of this thesis study, we generated single stranded DNA oligonucleotide and tested its targeting efficiency (Experiment #5 and EI #5).

Regardless, our studies created a stable platform for future transgene integration and stepped forward to perform more efficient porcine transgenesis in the future. In part one of this thesis study, we established and optimized protocol for CRISPR/Cas mediated knockin in porcine somatic cells. In part two of this thesis study, we stepped forward to generate the pig model of diabetes as we further optimize our porcine gene targeting methodology and partially demonstrate that we could modify the porcine insulin locus. We believe that all of our efforts will help us to generate insulin reporter pigs and pig models of diabetes in the future, and further study the dynamics of PBC, its capacity for spontaneous regeneration, the cellular origins of new PBC, and the determinants of PBC (Nir, Melton et al. 2007). All of these studies on reprogramming potentials of porcine PBC may provide an alternative treatment option for human type-1 diabetic patients. The primary sources of islet from porcine tissues might correct insulin dependence upon transplantation. Moreover, the clinical implications that we could have not studied using the mouse model of diabetes could be studied more effectively using our pig model of diabetes (e.g. new diabetic medications could be tested using the pig model).

Appendices

A) Primers used in the study

COL1A Gene knockin primers	COL1A For	AGCCAGGCTGCCTTGTTTG
	COL1A Rev	GCCAACCTCCCCTTTGCACT
pDB3 integration screen primers	EGFPC	CATGGTCCTGCTGGAGTTCGTG
	COL1A For	AGCCAGGCTGCCTTGTTTG
Insulin targeting vector designing primers	pINS Targeting Vector F1	GCCCGTCGGGAGGCGGAGAAC
	pIns Targeting Vector R1	CCACCTCTGGCGGGTGTGCAGTC
	pINS Lower Arm F	GTCGACTACTCGTACGCGTGCCGCCCCT
	pINS Upper Arm R	ACGCGTACGAGTAGTCGACGTTGCAGTA
	2A Sal 1 F	TCGAGGGAAGCGGAGCTACTAACTTC
	2A Sal1 R	TCGACAGGGCCAGGGTTCTCCTCCAC
	2A GFP BsiW1 R	CTAGTCGTACGTTACTTGTACAGCTCGTC
pINS integration screening primers	pINS Targeting Screen F	GCAAGCAGGTCTGTCCCCCTGG
	GFP F	CTATATCATGGCCGACAAGCAGAA
	GFP F2	ACCACTACCTGAGCACCCAG
	GFP R	ACTGGGTGCTCAGGTAGTGGTT
	EGFPN	CGTCGCCGTCCAGCTCGACCAG
	pINS Screen F	GGTGGCTGTCTCTGTGTGAC
	pINS Screen R	GAAGCTTAGAGCAGCCGATGTG
	pINS Screen R2	TGCATGTGGAGGAAGCTTAG
	pINS Target Screen R1	ACTTCTCCCCAGGGGTGTCTGT
	pINS Target Screen R2	TGGACAGCTCGCCACTCATC
	pINS Target Screen R3	CACTCATCCCAGCCGCCTAC
	pINS Target Screen F3	GGAGCGCGGCTTCTTCTACAC
	pINS Target Screen F4	GGCTTCTTCTACACGCCCAAGG
	INS Screen F	CCAGCTGGAGAACTACTGCAACTA
	INS Screen R	CTAGTTGCAGTAGTTCTCCAGCTGG

pINS sgRNA primers	pINS sgRNA 1F	TAGGACTCGGCCGGGCTCATT
	pINS sgRNA 1R	AAACGAATGAGCCCGGCCGAG
	pINS sgRNA 2F	TAGGCACGATGCCACGCTTCTGCG
	pINS sgRNA 2R	AAACCGCAGAAGCGTGGCATCGTG
pINS mutating primers	pINS MUT F	CCCTCAAAAACGTGGCATCGTGGAGCAG
	PINS MUT R	GGCCCCTCCAGCGCCAGG

B) COL1A gene sequence

TTTTAGCACCCAGGCGGCAGGAGGTTTCGGCAAAGTTGGAGGTACTGGCC
ACGACTGCATGCCTGCGCCCGCCAGGTGATACCTCCGCCGGTGACCCAGG
GGCTCTGCTACACAAGGAGTCTGCATGTCTAAGTGGTAGACATGCTCAGC
TTTGTGGATACGCGGACTTTGTTGCTGCTTGCAGTAACTTCGTGCCTAGCA
ACATGCCAATGTAAGTGCCTTCAGCTTGTTTGGGGCAGTCTGGGAAAGAG
TTAGATGGGTGGGCATCCTGGTCTGAAAAGGAAACCTTAAACAAAGTTCT
AGATACAGCTTGGAGGGAAGGCTTTCTGGCATGTGGAGGAAAGCTGGCA
GTGATGAAAAACATGGAAAAGACTTTAGAAATGAAAACAAAAAAGGCTC
GACTCTTGCCAGAAGTTAGGTTACTTTTAACTCAGAGACTGTTAACTTGGT
CTGACTCTTCCCGTTAATGAAGCTCAAAGACTTTTCCTTAAGAATGAGATA
TTAACTGAAGTGATGAAGGCAGGCAAAGGCCTCCGCCTAAAACTGCACC
GATGCCCACCTTGTGAAACTTGGGGCTTAGTTTACAAAGCCCTTTAATTTA
GTAATTAAAAACTCACCCCCTAAAGGCTTTTTCTCTTCAGGGTGTTTTAAG
ACTAGTCTGGATGATGACGACTCTGAGGCGCCCTCTGGAGGGAGGACAAT
GAGAGCACTGGGGCCGAGACTTAGGGCTGATGGAACCCCATTTAAAGACT
GCAGGTTTATACTGAGAATGATCTACATTTCTCATTGAAGCCAATCCCAAT
AGCTAGTAAGTGTATCCTGGGGAGAAAAATGGATACATTCTGAAAAGTAA
AATTAATTTCTCAATTCATTATCTGGTCATGGAAATGGTAACATTATTTAT
CTAAAATCTGCTCTTTTATTGAGCCTGAGAGTTATCTCTTTAACATGGTCA
CAATAGCAGACATCCTGAACATACCAAGTTGGGAAATGTCAACACTGATC
TGAGACTCATGGGACTGGGGTTTTTCCCATGACTTTGCCACTAATTTATCA
TGTGACCTTAGGCAAGTCAAACTGAGCCTTTACTATAAAATGCAGCATAG
AGATTAGATGGGCTCCAAGGCTGCTTTCCAGCTGTAGGTTATAAAACCAG
ATTTCCAGGTCTAGATAGAATGAAGAGAGTCTGGTTTTGTCTTGAAAATG
GCCGACTAATTACCAGACAAAGATATTGGCTTGATGGGGTAAGAGGAGTA
TCTGGCCACACCCAACAATTCAAAAATGTTTGCTCCACCCACATATGCCA
CGGTGTCTACCATTGCATCATTACTTATGTATTTCTGTGTGCGTGTTTTCTC
CCTCCATCCACTTTTAGTATATACACTATGAATACTAAACAATAACTGCTT
TGAACAAATTCACACTACCAAGTTCATTTTTTATGAAATGAGAGTGAGGG
GTAAAGAGTTGGCTTCTAGAGAAAAGATGAAAATAGACCTCGGGAAGTTT
GAAGGTACAAAGGGCAATGGGAAATTCACCTTCTGATCAACATCCAGCCA
CGGTCTGGAATAAAAGTTACCAAACACACATTCTTCAAAATGGCTCAGTC

TGAATAATTAAAAGGCAAATTTCCAAAATCATAAGGACTTCCATTAATCA
AGTCCAGGCATAATTATTCTTCCTACTGATGAACACAATGAAGTAAACAT
ATCATTTTCCTAAATTAAAAGTAATTCTCATAAATTGCCCTTAAATGTCAG
TGCTGGATGTGGTCCACCCCCCTGAATTGTGACTGTTGCAACAGATGTTCT
CACTCCAAATGACACAATTCTTGGCCACTTTAATTAGAGCAATTTTTAGGG
GTGATTCATTTTGTTGCCAGCTTGGCCACACGTATATCCTGTGTTAACCTA
TAATTTTTATACCGTAGGTGAGGAATTCAGTCTTTAAGGACTTATAGCAAT
TATGGCAAAGGGGGAGGATATGACTTTAGTATTTTTCCCATTATTTTTCTA
GATCTTTAATCCACAGTTTTTCCATTTATACAGAGCAATTCTTTATCACAG
ACAAAAGTGATACATACTGGTAAAATGACCTTACCTAATTTAAGTGAACT
ATTTAAAGCAAAATTTCAGATGTCCAAATGAATTCATTTTCACGGTGACTA
ACTGGCAAATTAGGCCTTTTGAAAAAGGACATTAAGAATGGTACAAACTC
AGTAACTTTAACGTTTTATGAAAAACAAAGCCTACTTTAGGTAACATACTT
TCCTAAAATTTCCCCTTTTTAAAATTTTAAGTATCAAAAATTATCTAAACT
ACAAATAATTTTCTCATCAATCTCCTGTTTTTGCTCACATTCAATGATTCAC
CTGTGCAATTTACCATCTTAATTACTAAGCATTAGAAATATCCTTTAGACA
CATTAATATTTCACAGCATTTATATAATTGTAAATAATTAATAATTTAATT
CAAATATATTTACATATTAATATTGAAAAAAAATCACCCCTGCTGATCCCT
GCCATAATTTTGACCTGCATAATTTCTAAGTCATTAGAATATTCTTAATAA
AAAATACAAATGGTTCTTTAAGTCAATTTTATAAACTAATATATCCATTTG
TTAATCACTGCCAGTGGTCAATGAAATAATACTAATAATTCTTTCTTTCTT
TCCCCTTTCCTCTCTAGCTTTACAAGAGGTGAGTCAAATTATATTTAGAAA
TTGTTATTTTAAACTTTGATTGCCTTCGCCATAGTAATGTCTTCTCAGGGA
AGGAATGGGAGGCTATATTAGTATTGACCAACCTGAATTTTCAAAAAAAA
TCATCAAGAGTTATGGCAGCCAAACCTCTCTGGCTGCCTTAGAAAGTACC
CACCCCAATTTCCAGAATGGGCAGGGCCACTGAATAAGACCACCCTAAAA
TGAACACCAATAAGAGAGTAAATAATCCAGTGGAAGTCTGGGTTTTAGCA
TCAGCTACTACAAAGTAGAAGACTTCACCACACTAGAGAATCTTATCGCT
CCCTTGGAATCATTTGCAAGAGTATTATGAATAAATTCTTAGCTTGTGAAT
TGAATTTGAGGCATAATTACAGATAAGTATATATCTTGGTTCTGTGCATAT
TTAATTTAATAGTAAAATATAAAATTATTAATTTGAAATATTAAGATAATA
GTAATACACAATATATATACTCTGTAGATATGTTTGTATACTACACAGAA
ATGGGAGACATTTTAAAATGTACATATACAATTCTCATAATAATCTTTTAT
TTATTCTTTTGTAGGCAACTGCAAGAAAGGTAAGAGTCCATTATTTTTCCA
TAATATCTAAAATAATCAGAGATAACTTAGAATTCAACTAAATTTATAGT
AGACTACGAAAAAATACTTTGTTAAGATTTTGCCAATATGAATATGCATT
AGCTAAACCATGAGATAAAGTAGCTCTGATGATTAAGATAACAAAGCTTT
ATTATCTTCCAGAATCATCTCTAATTATACTTATACAATTGGTGCTTTACA
TGCTAAAGGGAGAAAAAGACTTATTTTTTATTCTAGAGATAGAATGAATA
TTATGCTAATTATGAATAGTAAACAAAAGTTTATTTTAAAGGGCATTGACT
ATGTAAATATATTGTTATGAATGTTAAAAAATGATTTTATCATAGAAATCA
TCTTTCTAAAAATTGTGGTTATAATATTGGCCTAATGCTAAAGACTACAGC
CTCCAATCTCCCAGTTTAAAAAAATTACATGTAATTATAATAAAAACAAA
ATACTTTCATCTATTGTATTGTACCATTTTTATAGCTACTGTCTCATAATAT
TATGGTTAGATCAGCCTTACCAACTAATTATTATCAAGAATGCTTTGCTCA
TTCACTGGAAATTGCTTAGACATTTATTATTGTGCCATTTGTATCTTCCTTG

TAGGGCCCAACTGGAGATAGAGGACCACGCGGAGAAAGGGTGTGTAATT
TTTGAACTATTAAAGGGCTCTGGCCCATATTTTAATAACTACATATTAGAA
TCTACAGGAGACTGACAATTTATAAGAATATTATGTATCCAGATAATTGC
CTACCTTTTAAGCAGATAATGCCTTACAGAAGAAGTGAATGAGCATTATT
ATATATAATCGATATTCTTTTTAGGTCAAGATTATCCTTTAAAAAAGCAGA
ACTATTATTATAGACAGGCATATTCTTTGCATGGTCTATCTTTCTTTGTTTT
TAGTTGACCTATCCAACATATACATAATAGAAATGTATCTACCTACCACC
ACATGGCCCTTCTCTTTTTAATGTGCTTTGTTTGCTTTTCAAGTAAGATCTC
TTATGTAGGCTAGGAATAATAGCGAAACAGCAGTCAACAACATGTTTCTA
TCTAAAGTAATAAATGTACAAGTTAGGCTTACATTTTCTTCATGGTATTCT
TAGATCTCCGTGCTACATTCATGTGGCATGTTCAACCATATGGTTTTCATT
ATATGACACTTTTGTCACTCTATAAAAACCTATATCTTGTATCTATTGAA
AAAGAACAGACTTAAATCTCTTCTACTCAAGCAATCTGTTAGAGGTAGGA
TTATACCAGTAAATTACTTTAACAACATAAATAGGGCTAATGGTAGGAAA
TCTGCAAGATTTTCTACCCTGTGATGAAATAACTATTAGTTATCAATTAAT
TTACTTTTCACACATAACATACCAAAACAATTAAGTAGTCAGAAGCTGAT
AATCAATATATACAGTACATCACTGTGCACAAATTTGTAACTATAAAATTT
AGCTATAATTGTCATCTTTATTTGAATACTGGAGCTTCGATATGAATTAAT
ATCAAATGGCTAAGATATTCCTATTCTAAGAGGAATTATCTTGTCTT
GCATCTTCTTCAGATATGAGTTACTAAACCTCCTTCTCAAATAGACCCATG
TCAATAGTGACAAATCTGACATTTTAAATTTCCACACCACTTACGCATTCT
GACCAACTAATATTACTACATGTATTCTGAGGTCTCTTTTTGGAAGAATAG
GTTTCTACGAGTCTGTCTAATGCTGAAAACTCACTTTTTACCTAACAGGGT
CCACCAGGCCCACCAGGCAGAGATGGTGATGATGGTATCCCAGGCCCTCC
TGGTCCACCTGGTCCTCCTGGCCCCCCTGGTCTTGGCGGGGTAAGGTGTCT
TAAGTATTGCTAACATTTAGCTATCTTCAGCTTCTCATACCTCTATGGAAG
AAGACTTACTGTAGAAATTATTTTTAGTGGTTTTAAGGGAGAGTCTGCTAT
TTGGATATTAGTATAATTATTTTATTTCATTTCTTTGCTTTCTTCATTCAAG
ATTCTGCTTTGCCCGATCTCTTTGTGAGCCCTTGTCAATTACAGGACAGAT
CTTTGTGTTCACTGAAGGTAGCTGCAGTCCTCACGACCTTTATTTAATTT
GGGGATTTAATATCTTTTATCAGTAGGGCACAAATAAGAGGTGTTGAATT
ATTAAAGAGTTTGATTAGATTGAATTATAAATGAAATCCCTGATCTTAAG
CAATTTCACAAACATCCTACTCTTTTTATTCTCCTTGGCTTGAAGTCTGATG
AACCAACATTCAAAGCAGTTTTTGGTTTAATTTGCCTGAAACTAATTTGAG
AAAAGTACATTTCCTTTTTTCATTAATATCTTCTTTTAGCTTTATGTCTTTA
ACAATGGTATGAGTGCCAAATGGCCTCACTGCAGGAAGGAAAACATATTT
GCTTAATTGGTTAGCACTATGAATCAGAAGTCTGATTCTAATACCCAACTG
TACGCCAGTAAAATAAGACCTTTCTCCCCCAAAGATCTTATTATGATTGCT
TATCTACTGAGTGCATTAAAAAGCACCTTCTTTAATAGATCTACCACTATA
AGAAAGATCTTTAACAGTAAAGCCATTACTGTGAACTATTTTTTAAAAGTT
TAATCTTCCAAGGGGCATTTTAATTTAATATTCCTCTAAACTTGAAAAGTC
TTTATATCCTTTTTGAAACTCTAGCAAGAAAAAGGTCTCTTAGTCATTTTG
TGTAGTTCTGGAGGAGCAGGGCAACAGGTAGAAATGCAAAGTTGTCCTCT
TGGTTTCTGCTTAATTTATCAGGGACCCTCACAGACTCTTAAAGTTTAGAA
AGCTCTCTTTTTTCCGTGTACAAGCAGGGAGTGGGGATTTGAGTGAGCGCA
GGAAAAGTTTCAAAAATAAGGTAAGAAAACTTCCTGTTTCAAACTAGGTG

TGGACTAAATTTTAAACATATTTCTCCTGAATTTTTAGGAATTTAGCTCAA
TATGGAATTCTAAACTAATAATTATGTCAAAAAAATATTGCTCTCTCCTTT
TTTTTTTTTTTTTTTTTGTATTTTTAGGGCTGCACCTGCAGTATATGGAGGTT
CCCAGGCTAAGGGTCGAACTGGAGCTGTAGCTGCCAGCCTACACCACAGC
CATAGCAACCACAGGATCCAAGCTGCGTCTGCGACCTACGCCACAGCTCA
CAGCAACAGCAGATCCTTAGCCCACTGGGCGAGGCCAGGGATTGAACCC
ACGTCCTCATGGATGCTAGTCAGATTTGTTTCCACTGAGCCATGACAGGA
ACTCCCCTATTGCTCTCTTTTAAACAATAGAAAAATCTTTAGAAGTAGAAT
AAAGAAATGAACTACATGGCCAATAGTTATTATATTTTTTTTTCTCCCATC
TAGAACTTTGCTGCTCAGTATGATGGAAAAGGAGTTGGAGCTGGCCCTGG
ACCAATGGTATGCTTATCTCTTTTTATTTTAGCCAAAAAGTTACTGAAGAG
AGAATTTACTTTATATATTGCCATTTATTTAGCTATCTAGATTAACACCCA
ATACTTAGATCAGATAAAAACAATTATTTTTGTATTGAAATCCATAAACAT
CTAAGCCTAGGTTGGTGGGTTTAATTTTCCCTATCAATGTGCCTCATTAGG
ATAGATTAGATATTTATATCCAAGAAATAATCACTGGTTATAGTAACCAT
AACTTAGGTAAATCAATTTACTTCAGAAGAGTGATCACTTGGATTATTTTT
TAAATTAACATTTTAACTTTAACATTGCCTAGAATTATAAACAACTCACAG
AGACCCAGAACTAATAAGTTCTATAAACTTGCTGGAAAGGATCAGCAAGT
TCTAGTCTAGCCTGTTTATGTTCTTTGTAGAGAGATTAAGCCCTTTTCCAA
GGTCATACATGTGGTTAATGACAGAACTAGGACTAACCACCAGGCAAGAT
GTGTGATATTCAAGCAACATCATGGCTTGTGTATCTAGTGCTTGAATAAA
ACCTTACTTTGTTCAAATATTTACCAGTTGTCTTCAATGAAACCCAAACTG
CATTATGCATCTGGGATACAGCAGTGATATGGCAGATTCTCACCCTCCCG
CGAGGAATTAACAATATAGGAAATATTTCTTTTTATCCATGTTGATAAACA
TTTCTTTTTTAAAGCCTGATTTATTATAAATCAATATAGGAGCGAATATGA
TAATGATGGTCTATCTTTAACAAGTGAACCTTCTGACAAGTATTAATGAG
ATGTGTTGGGGGAAAAAACTGATAGAGGCAAACAAAATATATACTAGTA
TGTATAGAAAAGAGGAGAAAATTGCTGGAAGTAAAAACGTTAAGTTGAT
TAAACACTGCAGTTTAGCACATACCCACACATAATTTAGTGATTTTAATTA
TTTCCAAACACAGGTATTTATTTGCCTCAAGTTAAATGTCTTTAAAATATA
AATTTTACCTGGACAATGAATCCAGTGCTCTGTCTTCAAAAGATGTTCTGC
TGGAGCTAATGGTCCGCAATAAGCTAATACACTGTAAGGACAAGACAATA
TTTTCTGTAAATTAGAACTTTCACTTGAGAAGTAATATGCCTTTAATTGAT
GATTTCCAATTTCCTTATTTTGCATGGTAATTTTAAACATTCATATCTTCAA
TGAAGAAGTAGAGATCTTCTTTTGTTTATATTGGATATTTGAAATTAGCCA
TTTCAGCTAACACTGGACATTAGTCAGTTTTAAAGCAGCACCTATATCTCA
AGAAGAAGCTAAAAAGAAAAAAAAAAGTAAAGCATTAATAAAAAACAA
AAAAGGCACATAAGCCTGGAAAATTAATATGGATTTCAGATTTCCCTGAC
TTTGGAACATCTCTCTGGGGAGCATAAAGCTGAATGTTTTACTTTTAAATA
TCAGTGAAATATAATTTAAGCTTGCATATCCACACACATTCACAGACATA
TGTAATCAACAGATATGTGTTTCACAAATAGGACAGACAGGCAGCAATAA
AGCATTAAAATAATTTTCCATGTTGGAAAATCAATAACTATAAAACCCCA
CAGGGTTCGTCTCTGAATTAATGAGTAATCATCATCTGTCTCACAGCCTCC
ACAAAATAAACTACCTTTTATGTAAATGAAATTGTTGCAAATACATGGAA
AAATGAATAAATATAATTAAAATTCATGATGTCAAGGAAAATTATTT
TTAATGTATACCTAAAATATTATTTTGATGGAAAAGTGACAGTTTCTTCTA

ACATGTAAGTAATTTATAAAACCAAATAAATAATTCTTGTAATTAGGATG
ATGTGCTTTTTAATAAAATCCACATTATTTCATTTTAGGCATTAACACAAT
GGGGCAGACTGCAGAGTCCCCATGGTTGATGTGTCATGAACAAATTACCC
AAGACTTGTCTGTGCAACATTTGCACCGTTAGAGGAATGGCCCCTACAGA
AGAGTTTGCACAGGCACTAGTTATTGCAAAGAGGACTCAAGAACATGTAT
TTAGCTCCCCAATTAGAAACACTCTGTTTTACTCTGTGAGCTACAGAATGA
AGAACAAAGGGGAGAATAATTGCTGTCATTTGAGCCATATCTTCAATAAT
AGTAGCATATAATAGATTCTTCCAGTAGGATGTTTCAATATTTTATTTTTC
CTTCACCCCTCGCTACATCCACTTTTGTTTGACATTTTTATCAGTCACCAAT
AGCCCTTAAAGGAACTATAAGGTTATAACAGCCCTCCCTTTAATCATTAA
GAAAGACTTTTGATCCTTTTTTATTTATGCCCCTTGTGGCAATAAAAGTAT
CAATTTCTTAACTAAGTATGATATGAAGAGATGATATGTGATCAAGCGAA
TCCACATGGAGATGCTCTGTCTGCTGTATGGACAGCAGTTGTATATCCCAT
ACCGCTCAAAGACTATTTCATTGCATTACAGCAACTTCAATATGCCCAAC
AGTGTATATCTTTAAGGAACTTGTGCCCAACTCATCCAAATGAGTTCTGAT
TTGAATTAGGAGAAAAGGTGAAAGAAAAGGTAGAGAAACAGTGGAGGA
AATTTGGGTCCTTAACTGCTTCCTTGGGAGGAACAAAAACTATGAAATTA
AACCAGGACAATTCTGTTCATTCGTCCATCACAAAATGTCTATGGCACCA
CTAAGCAAATCCTGTCTATCCCAATCCTAGGTCATGCCAACTTGTCAGTGC
TAACTATATCTGTCCTCTTTACAGGGTTTGATGGGACCTAGGGGCCCTCCT
GGGGCAGTTGGAGCCCCTGTAAGTACCAAAATTTGTAATCTCTCTTATGTA
AAAAGACGAAGAATTAAAGAAAAGCTTGAAGTGTGACATACGGTTCTCTT
TTAGGGCCCTCAAGGTTTCCAAGGACCTGCTGGTGAGCCTGGCGAACCTG
GTCAGACTGTGAGTACATTTCTCCCTTTCTAATAAATATTTTTTCAAGAA
GTTTATTAACAGAGTATTAACACAATGAGGGGTCTCTCATTTTCCTAGGGT
CCTGCTGGTGCTCGTGGTCCACCTGGCCCTCCTGGCAAGGCTGGTGAGGA
TGTGAGTATTTACTCTTAAGCATTTCCAAATGCCATTTAAATACTCTTGCC
TCAAATGGGAATTTCTATATTCAAATTAAGTATTCTGCCAAAAGCTGAAT
ATGCCTGACATAATTCTCCCCATATGGAAAATATCACTGTAATGAAATAA
TTTTCTTGTATATAAGGAAATATAATTGTTAATTAAATATATATTTGGATC
CACGTTTTTATGATACGTTTCCCGCCATTGTAAAAACCAAGATGCCTCTAC
TTTCATTTGAGGCTTTACCAAGAAAACATTCATCCTGTGATGGTTTCATGT
TTTAGGGTCACCCTGGAAAACCCGGACGACCTGGTGAGAGAGGAGTTGTT
GGACCACAGGTGAGATACTTTACAACAGTACATAGTCCAAAGATCAGAG
GTATCTCATATACAGGATAGCTGCTAAAACTAGAATCAGTCTAAGTGGCA
ACACAGACATCTGCACACTTAGAACATTTGTTGAAGGCATCTGCCAAATA
AAGGAGCCTAGTTAAGAATGCATATAGTTTTAACTGAATCACTTTGCTGT
AGAGCAGAAATTAACTCAACATTATAAATCAACCATATTTTAATAAAAAT
TTTTAAAAATTATTTTTAAAAGAATGCCTACAGTTTTTATTAATGCAAATA
CTATAATATTAATGACAGCAAGCATACTGTAAACAACTAAACTGTGTGTT
CTACAAATACCATAGTATTAATTATCAAATTAGAGCCACTTATTTGGGGCT
CGGTGCTTAAACACATACTAGTCTTTGATTCCAAGCTTGGACTGTGATGAG
GAGAAACACTTTAGAGGGAAGAGCTCATTGTCTTTTTATTTAGGACAAAA
GCATTATTTGACTTCTTCTGTGTTTCTTTTTAAGGGTGCTCGTGGTTTCCCT
GGAACTCCTGGACTTCCTGGCTTCAAGGGCATTAGGGTGAGCACATTCCT
TACCCACAGGAGAGAACATGTTGATTAGTTTTAGTAAATTCCTGGAATTT

GACTAAATAGAAAGTGAGTTAATGAGTAAAATGGCTGGAAAGTAAAGCA
GGGTGTAGTCTGTTTTTTTGTTTTTGGTTGGTTTTCGGTTTTTTTTTTTTGTT
TTCTGTTTTTTGTTTTTTGAGTCTGTTCATGGTTTCCAAGAAACAAGGATCC
CTACAGAGAAAATTGTATAAAATTGGTATAACTTAAGAGGTATTTAAAGC
CCTCTGAAGATTGTAGAATTATTAAAGATGGAAAGACTGTGAGAGAGCAT
CTTCATCCTTCCCTTTCAAGTTCAGGGAGGTTTAGCAACTTACCTAGGTAG
GGCAACAGAGGGAGTTTGCCCAGCGCTGGGGCACTTAGGGCTCCATTCTA
CTGAAGCAGTCACAAGATTCACCCAGGGGATAGTAACTGCAAAGACAGA
AGATGTTTGAGTTGACCAAGCACTGTAATGGAATAGCATTTTAATAATAA
GGACTTGCCTTTCAGGGTCACAACGGTCTGGATGGATTGAAGGGACAGCC
CGGTGCTCCAGGTGTGAAGGTAAATATTAAATCAAAAGCATTTTTTTAAA
AACAATTGATTATGATTCGCCATCACCATCAGAAAGTATATTTTTACTGAA
GGCTACTCATAATCTGAAGAGTATCTTATTTCCTCTTCAACACTTTAATTC
CCTTAATGTGAGGCCTATGGGATTCTATAACAAAAGGCCAATTATTAATA
ATATTTCTTTAATTCAATATCCATCAAACAATTGAGATAAAAATGTCAATA
GCTTCTTCTTCCATTTCCTTTCCTTCTATAATTTCAGTCAATCTCTGCCACC
AAAATAAAAGTAAATAAACATATAATCATATATTAGAACACTATATTATT
TTAAACTGTAAATTCTCCTTTGCCATGCACTAATTAGATAGGTACATTCAT
GTCACAATACACTTTTCAACCTCTTTCCTGTGATTTATCTGTGCACACTCA
AAAAATTTTAATTAGGTAATTAAAGTCTCAGAAGTGTGTTATCTCTTGGCT
GGGCTCTTCTCTGACAGCGTTTTCAACTCTAAAATGTTCTCTCTCCTATTA
AGGAGATAATGTGATATTAAAGTGAATACCAACGTAATTACAAATTAATG
AGTAACGAATACTAGCGGGACCAGAAATGAACATGAATGTGGAGAATCT
GTTCTAACTTTCCAGCTGCCACACAAATGGATAAGGTCAAACTCATTCTCC
CTAGAGCCTAATATAATAGCTCAGACTACTAATCACAGCATCATAAAATG
TTAAGAGCTGCAAGGAGCTTTGGAGACTCTCTCATTTTGCAGATGAAGGA
AACTGAGGCTTAGAGAGAGCAAGTGACTTGCCTAAAGTCACACCTCCAGT
CCTTAGAAGGGCACAGGCTAGAAATCCTTTCTCCCACTCTAATGCTTATTC
ATAGAGAATACACTTCAGATAAAATCCTAGAAATGAAGAGTCTAATGGCC
TCCCAGTTTTTTTCTATATCTGAATTAGATTACCTTGAAGGAAAGTTCTTTT
TCATGTGCCATTGTATTATATTACCACCTCCACTGTCCTCCACCATCACCA
GAATGGTTTTTACTCAAGACAAGAATAAAGATAGCAAATAAATGACAGTT
CTATAAAATATTATGAAGCTTTAAATTTTACACATATTTTATGAAAATGAA
AATATCGAGTGTTCAAAATAACTTTAAAGAGGGTAAATTGTTATGGTTAT
GTTGAAAGCAGATTATCCTCTGGCATTTCTTTCTATGATATCATTCCTAAT
ATATAAGTCTAGTCTAAATGATATCCAAAATCTATGGGGAAAATTGAGGA
AAGGGTAAATAACCCTCAACCACTCCCTAGTACTGTGATATTTGGTATTAT
CTCAACAAAAACCCACATTTTATAATTTTTCTATAATTTTTAAAAATTTAC
ATGACTAGCATAATGGAGAAACAGTCTATGTAAGAAATGTTACAAAATA
CATAAATGGTTGAAACTAAAGTAACTGGATAAAAATTATCATAAAAAAAT
CAATATAGATAATGAACAAAAACTCAATCTTTCGTTTCCATATAGGGCGA
ACCTGGTGCCCCCGGCGAAAATGGAACTCCAGGTCAAACAGTAAGTATTG
ATTACTTTATTGTAAATCAGAATGTGTACACTCTTTGCAAGCTGGAACTTC
TTTAATGTTTTTGCTAATTACCATTTCCCTTGGCATTGTATTCTTTGATATT
TTTCTAATAGCCTTCTGCATACTTAATTGAAATCCATTACCTTTTAGTTGG
AATTAGAACAAGATCTATTTATTTTCAGTGGATTCTTGCACATGTTGGAAA

TTGGACAAAGCAAATGATGCCTGTAACTTTTTCTTAAATTAGCATTCTGGA
TTTCATTGGAAATATTTCTGCTTCTAGGGAGCTCGCGGGCTTCCTGGTGAG
AGAGGACGTGTCGGTGCTCCTGGCCCAGCTGTGAGTGCTTCTACTTTTGTT
CATTTTTATACTTTTTCAGAATTTTCTAATATGTGCCATTTAACTTCCTACT
TGACTTTTTTTCTTTATTTTCTTCAAAGGGTGCCCGTGGAAATGATGGAAG
TGTGGGTCCTGTGGGTCCTGCTGTAAGTTTTGACACTGAAGAATCTGGAC
GGGGTTGAGAGTGTGGGATGGGAGAAATCTTCTTTATTAATTACCTATGA
AATAATACCATTTTAGACACCTTACCAAATGTTCTGTGAGGTCATCTGAAG
GTTCAAGTTATAAAAACTACTATCCCTGAGGCATGTATTGATATACAAAG
TAAAATCTACCTCATCTTAGAGACCTAGAATGTCTTCCCTTTGATAAGATA
CCAGTGACAAGAAACTAGCTGTCAGTTTTCTCTTTGCCAACATTAAGATTA
AGCCACTCTTAAGCACTCAATCTTCTGTTGTTGTTATGTTGTTAATAGTGG
GAGGAAATTTCTTACCTCCTTCTACTTTGATTTCAGGGTCCCATTGGGTCT
GCTGGCCCTCCAGGCTTCCCAGGTGCTCCTGGCCCCAAGGTAAAAACACT
GGTGACCACTGTCATTATCTTGATAAACTTTTTATTATCATGTGAAAGATA
GAGACTGTGTTCGCAGATAGGTAATTATAGAATAGGCTGCTTAATTAGTC
CCTCCTTCAAATGGATGTAGAATGACCAGGTTTCTCACTCCCAACTTCTGA
ACCTGAGCACTATGAACAAAGATATAGATGGCAACATAGAAATTGATTCT
GTGAGACCCCTCGGGGATATACTCCCAAAACTAGTTGGGACTGACAAGCT
GAGAACCAGAGTCTCTGAGAACCAGGCCAGGAATATAGGAATGCAAACC
CAAATCTTGTGGTTCCGGACCAAATGAGGGAAATTATTACAATTCTGAAG
AAATCTCATAAAACTTGGTTACCTTAAAACAGAGATGTGCTTATAACTGA
TTTAATTCGCTGGTACACCTGAAATTAACTATATTGTAAGTCAACTATATT
TCAATAAAATTAAAAAAAAAAAGATATGCTGTTTTGTTATTTGCTAATGC
CTCAGTGTCATTCCCTCTTTTAGGGTGAACTTGGACCTGTTGGTAACCCTG
GTCCTGCAGGTCCTGCGGGTCCCCGTGGTGAAGTGGGTCTTCCAGGTGTTT
CTGGCCCTGTTGGACCTCCTGTAAGTAGCCATTGTCATTACTCATCTCATT
GAGTGGAAGGAAAGAAAAAAGAATGAAGAAGAATAGAAATACTCCTATG
AATTTGTATTTTTCTCCCTTCCTTTTCTTCATAGGGCAACCCTGGAGCCAAC
GGCCTTCCTGGTGCTAAGGGTGCTGCTGTGAGTATTTGGCTAAAATGTGCT
GCTGTGGTTTTAAAGGTATTTTAATGAATGATGTCCCTCCTGCCAGTTGCC
TCATGATGGACGCATGCTGCATTTCTTCCCCAGGGCCTGCCTGGTGTTGCT
GGGGCTCCTGGCCTCCCTGGGCCTCGAGGTATTCCTGGCCCTGCTGGTGCT
GCTGGTGCTACTGGTGCCAGAGGTCTTGTTGTAAGTGGTCATGACTGATG
CTTTTATCATCCCAAAATATGAGCGCTGTCATCACTCCATCCCCATCTAGA
CTTCCTCTTGTCGTTTTATCATCCTTATTTCTTCCTTAGCATTTTTAGTTTGT
ATTTCTTCTATAAGTTCATGCACATGTATGTATACTCTCTAATACACAGAT
ACACCCTCCCTGCTATCCTAGAACATTTCCCAAAGGTCAGTGAGGCCCAT
TCCTATATCAGTCTTGTCCTTTCAAGACTGTTCCTCTGTAGTCATTTAAATA
TGAACTGAGCATATTTCTTAAAGACCTAGAGCTCTCTTTAATTGATCCATA
AGTAGTCCCTGATATTTGTCTGCCAGAGAGTTGTGACAAATGTGATTTAAC
CACTGTTTTGTATTGGACCCTAGGGTGAGCCTGGTCCAGCTGGTTCCAAAG
GAGAGAGCGGCAACAAGGGCGAGCCTGTAAGTAGTTATATAATGACAGT
TTCATGGAGTTACTAGTTTTTCAATTCAAAAAGTAATATTTCTACTTTTTTA
AGTTATTATTCTTTGTCTAGATGGGTATGAAACCAAATTTTATTTTTCCTCT
CCCTCCTTTTTATTTTTTCCTTCATTAAATAGGGTGCTGCTGGGCCCCAAG

GTCCTCCTGGTCCCAGTGGTGAAGAAGGAAAGAGAGGCCCCAATGGAGA
AGTTGGATCTGCTGGCCCCCCAGGACCTCCTGGGCTGAGGGTAGGTTCCA
AATGCTCTTAACACCCAGACACACAAGAGGTACCCCTTCCCTGGGAATTG
GTCAGAATCACTGAGTGCATTTTCTAGACGAAGAAAGCATCTGCTTTCCA
TCTGCTTTATTAAATCACTGACTCTTAATATGATATGTTATAAAAATCGGT
CTGGAAACAATGTTGACCTACTTTTGCAGAATGTTTTCTCTACACATTCCT
TTGGGGTTGGATGAAGTATTTCGGCTTCGCTTGTAATTGTATCTCCTCCTT
AAGGGAATCAAGATTTCCTGCATGCATGCCATATTTCAATAAATTCTGCTC
CACCTGATTCAAGAATTTTCTCCAATTCTCTTCTGTTTCATGTACTTTCTTA
CAGGGAAATCCTGGTTCTCGTGGTCTCCCTGGAGCTGATGGCAGAGCTGG
TGTCATGGTAAGTTGACCATTAGTCATTTCCTGGGAAGGAACTTGATGCTT
CTGGTGGTTGGTATGTCAGGTATGTAATTGGCTTTAGTAACCTCCCTCTCC
CTTCTTTTTAATAGGGCCCTCCTGGTAGTCGTGGTCCAACTGGCCCTGCTG
GTGTTCGAGGTCCCAATGGAGATTCTGGTCGCCCTGGAGAGCCTGGCCTT
ATGGGACCCCGAGTAAGTTTCATACTGATTCTGAACAAGACACACCCAGC
ACTGTTTCCTTTTTTAAAGAGATGATTAATATTTAAAGACAGAAATTAAAA
AATCAAAAGTTAATTGTTAGACATGCTGACAGTTCCCTTTTTAATTTAACT
TTATCAAAGCCCAGTAGCTACTTTGACAAATTTGGTTAGTTCATAAATTCC
CTATTACTTGATACAATGGCTATGAGGTTTTGGGAAGAACAGATTTATTTT
AATATATCCAAATTAGGGTGGTTCTTCTATCAGCATGAATCTTTTATCTGA
ATTTGGGAGTTTTATATAAGGTGTTCGTGAAACATATTAGGACTATACATT
ATCAGTTTATTAGATTCATAAGTGAATCATTTTTCCAGCAATCACAAAGTG
CCATAATGTATTCAGCATCACACTAGCTATGGAGAAATGACCTTTAGATC
TGTAGACACTCTAACCCATAGCAATAGAGTAATTTTTTGCCTCTGTTATCT
CTTACGGATGAGTATTAACTGTAATTGTATCATAAACAAGTAATAAAGAT
ACCAAATCTAACAACAAAAAAATATACTTTGAAAATTGCTTACTGAAAGT
ATTTGTTTTTGTATTATAAGGTAAATAACATATACATTTTTCCCACATAGT
ATCTTGCTCAAGTTTCTTGAGATATCTTTAAGTGGATGAGTTGCACTAATA
AAAAGAAAGGCATAAAAATATAAAGAAGAAAGATTTCAGAGCTCTTTTC
ATATTTCCCAACTAGTGACTAATACTCCTAATGATCTTCCCTAGGCAACCA
AACAAAAAGTGGGGGAAAAGTGCTTTTGTGAGATTTCAATTAATTGGAGC
TTTGAACATATGTGTTACAGTGGCCAATATGAAAGTGTCCTCATTTGTTTT
ACAGGGTTTCCCTGGATCCCCTGGAAATGTTGGTCCAGCTGGTAAAGAAG
GTCCTGCGGTAAGAGTTGCTCATTTTCCATTATATTTTCAAAGACAAATTG
CATCCGTATCAAGTCTACTCCATAGTTTATTTACATGTGAACACACTGAAA
GTAAGTGTCAGCTACATAGAAATGTGAGAACACAAAGTAATGCATTGTGA
TTACAGACTCCAGAGGAACACAAGAGTAAAAAGAGAAAAGGAAAAAGA
AAAATATGGCAGGGAAAACTGAAGAGGGTGACAAGGGTTGAAAAGAGA
GGAGAAAGAGAGAATTGTACACATGAGATTGACTTGGCTGTGTGTGTGCC
GACATCCTATTTAGAAAAGGACATGATTTTTGACACTTTATACAGGAAGA
TCTATGAACTCAGAAAATGATTGTTTCATCCCATGGCAACATCGCACGCTT
GAGGCCATGAGATCCTTTTGATGCTGAATAAATTCTAAATGAACTCTGGTT
TCAGGGCCTCCCTGGTATTGATGGCAGGCCTGGACCAATTGGCCCAGCTG
GAGCAAGAGGAGAGCCTGGCAACATTGGATTCCCTGGACCCAAAGGCCC
CACTGTAAGAATCACTATAACCTCCTCTCCCTCAGCACTTTCTTGCAACAG
TTCACCAAAACTCTATTATTTCTCTCCAGCCTCAGTCTACTCTAACTCATG

GGGATTCCTTTCAACACAGTGAAACTGCAAGCCACTTACCACATTAAAAG
TGGACTCCCAGTGTATCATTATACCCCTGCTGAAAATCCACATCCTGAGTT
ATATGCTCCTGCAGACACAGGTGTACTCTATTGTTCATGGAGCTGAATGAT
GATGGATCATGCCTTAGATTAAGGAATTAAAGGACTAGGAATCTAAAAAA
AAAAAAAAAAAAAAAAAAAAAAGGCAAGAAGTTAACATTTTTAAATTGT
CATGGTTAATTTGAAATTAAATGTATTTTTAAATGTTTATTATAGGGTGAT
CCTGGCAAAAATGGTGAAAAAGGTCATGCTGGTCTGGCTGGTGCTCGGGT
AGGTATTAACTTCTGTGTGGGTCTATCCAAGGAGTATTCACGTAGGATAC
ACATCTTTTGAGGCCATCTGATATTATTCCATCACTTTCTTTCTAAGATGG
CATTCCCAAGGGTCCTTTTACCATCATAAAGTGTCTTTTTTTTTAACCACAC
TTATAAGGTAGTAAGCTATAACAAATCAGAATATATACTAAATCAAAAAT
ATTTATGACTGCATTTGGTTTTTACTTGATTCATTTTGTGATTAAAAGGCTA
TAATATACTAGTAGCATTCTTTGACCTTAATAAATTGCCCTTCATCCGTGG
AGTTCCATGAGCCTTATCAAATGTAGTAGGTCATATCTGGGCCTTGGCTGG
GGACCCACAATTAGTTTAATTCATGCTAAGATGACAAGCTTTGTTTTGAGG
AAGTAACACATGAGGGAGGCTTTGGGGGCATCTTAAATTATCTGACTCAC
AGTTAACCCTGAGCTTTTTCTGTTAAACATTTTAGGGTGCCCCAGGTCCTG
ATGGAAACAATGGTGCTCAGGGACCTCCTGGACCACAGGTGGGTATTTCT
CCTACTCTTGTTTTTTTCTGCCCTAAAATGAATCTAGTCCCACAAAAATGA
CTGTCTTTTTTTTTTTTTTTTTCCAGTCCAAAAGTTATGTAGCTAGACAGCA
ATGGTGACATGTATCTCTAATTTATGTTATCTCTCTTATCACCTTCAGGGT
GTTCAAGGTGGAAAAGGTGAACAAGGTCCCGCTGGTCCTCCAGGCTTCCA
GGTAAGTCAGTGAAAACACACAGACTATCCATCTGTAAATCACTACACAC
TCTTCCCCACCACACAAAACATGATTTTAGAACTAAAGCAAACTTTTCTGT
AATTCTTTAGTAACACCCACTAGTTAGTGATCATGTTCCTATTAGATAGAG
ACCCAAAAACAAATTATTAGGAAATCCCCCAAATACAGCCATAAAGATTG
AGGTTTGTATTTTTTTTTTTTCAATAGGGTCTCCCTGGCCCCGCAGGTACA
GCTGGTGAAGTTGGCAAACCAGGAGAAAGGGTGAGTAAAACAAGTGATA
GTAAATTCTTCTAAACTTAAGATTTCCTCTCTGTTTAACACTCTACTGCAG
TGCAATTATAATATGCAAAAGGAAACTTCATATTTCATATGTTAATGATA
GTGTTCTACATATGCTGGTACTGATGGAGAGAATGAGCCAAGTTACTCAT
TTTCATTCAAATTACTTTGTCTTAAGGACTGAATGAAACATCACATTATGG
AAGTAAAATATATCATGACCATTCCATTTAATAGCCTTACTTTTTGTTTTG
TTCTTGGTAATGCTAAAATACAAATTATTTCCCTCTCCATAGTGAAAAAGT
AAATGCACTGATTTTCTATCAAACTAGACCCTGAAAAATTGCTTCTAATGT
TTTTCTTGGCATTCAGACATGACACTGCTATTCACAATCAGGGCATGAGTT
CTGAGTCATTTTATCTGTAATTGTGATGAATGTGCTTCAATTCAGTTACAT
TCTGTGGCCTGCTCTCCTTTGTCAACAGTGGGGCACATTAAGGAGACAGC
TGGTCAGTAATAAAGGAGATACACTAGGGTATACAATTAACTAGGTAATG
TGCAGAATATGATCATTTCTATTAAGAAGGCAGTCTGTTATATAACTAAA
AATAGGCAATGTCTGATATTGTTCATTATTTATCATATTATCAAAGGAGAT
AGTCTTTTGTCCAATAAACTAAATAAAGAGAAAATTTAAATTTCTGATTCT
AGGTTCTAGGATAGATCTATCTTTAATATCAGTGGAATTCTCACTTTTTAT
ATCTTTTCATACACCTCTGAAAGTTCTCTTGTTCTTAGAAGCTTGTGTATTT
ATACACTCATGTAGGTAGAGCCTATCTGGTTTTATTTCACTCTTTCAAAAC
TTTTCTAAATATCATAAACATAAGTAGGTTTAGAGAACATACATAATAAC

CTTTACACTTTATTTATTCTTGTGTTTTGCCTAGGGTATCCCTGGTGAATTT
GGTCTCCCTGGTCCTGCTGGTCCAAGAGTAAGTGTTGTTTAACTTTCATAA
ACTTCTGGTGATGTTTTTAAAACAAATAGAGCTTTCTCCTGAAAACTATTG
ATGACCTCTTAATAACAAATTCTTGACACTTTACTGTAGTGAAACATACTC
TGCAGAGAATATTCAATTTCTAAATTGACAATAGACTTTTATTTAATTGAT
CATATTTCTTATTCAAAAACCACACAGTTATAAGAAACATAAATAAACTA
ATATATGTATTATATCCAGTGGTTTTTACAGACAAGCATGATATATTTATG
ACAGTGGCTCGACCTGAGCTAAGGAAAAGAAATATCTTACTAAGATTTAA
AGTGTGTGATTTCCATCTCAGTAAATTATAATTTCAATTTCTGCAGTGTAT
GTTGGTGTTCAAAGTCTAAAAGCTGAAGTCATGAGTGTACTTACATTTTAA
AAATTAAATAAATTGGTGTAGTTTGTCTTTAGCTATTAGAATAACTGGAAT
GATTAGCTGGCCTCTAAAGTATCCTGTTATGTGGCTAATCACCAAGTATCT
AGGATGCCCAAACAAATTTAGTATCATAGAACCATGGTACCATAGAACCA
AGTTGTAAGAAGGGCCTTTGAAAAAGCTGCAATACGGACTTCCATGTTAA
ATATTCTGGACCATAGAATGCATAGTGATATTTAATGCAAATTGGGAGAT
ATTTAGATAGATAGACATATATATATATAAACACACATTTTCATATATATT
TTTTTTCTCTATTAACTATATTATACTGAGACATCACTGTTGCATGCAATTG
TAAAAAGACAGCATACACATGAGAATATCTAATCTTATTTATACCTGGCA
TTGCTAGTCACCTAAAATAATGCAGTTTTTCCTTTTGAATACTATGCCTAA
CTGGCATCCATCACAAAAAGACCAAAAAAAGGAAATGAATTCCAAAAAT
TGAATATTATACTCTAGAAGCAACAAATTTTGGAGTTGTAGCCATAGAAC
ATTAGAGCTGAATGATACTTTAGAGATTATCTCATCTCTTCCCTTCATTTT
ACTAGTAAAAAGTTAAGGGAAAATAATTATAATTTTCTAATAAATGTCAG
AACCAGGATTCAAAAGCTAAACAAGTGTGAACTCGTGAAAAATAGTATGT
CTAAAAGTGATGAACAGTGCAACACTTCTTCTAATCCCTTTTTCACAGGGG
GAGCGTGGTCCCCCAGGTGAAAGTGGTGCTGCTGGTCCTGCTGGTCCTAT
TGGAAGCCGAGGTCCTTCTGGACCCCCGGGGCCTGATGGCAACAAGGTAA
AATTTTATGTTATATATTGCTGGTTTAGCCCAGTATATTCTAGAATAAGTA
AATCCTTTACACTAGCAAATGAGTTGTTTCTTAGATTTTTGTTTTATTTATT
TCCAGTTCTGTGATTATTCCCATCTCAGTAAATTACAGTCTGATTCCAACA
GACCTGAATCATTCCAAAAGAAGCAAAACCCCAGTGTTCTACATACATAA
ATGAATTAGTATGGGTTTTCACTCTTTTCTCATCACATTATTTATACTTTTA
TCTTTTTAAAATTTTTTATTATAGCTGATTTACAATGTTCTGTCAATTTTAT
AATTTGTCTTAATCTAGATAGAAACAACATACTTATTTGGACACTATTTTT
AGAGAATGAACTTTTACCTTCCCCAAATCCTCTATAAAGAGTAAAATTTC
AAAAAGTGATACAATAATAAGAGAGTTTATGATGATAAAAATTAACTGG
AAAGTACATTCATTCTTAAAACATAAGCCTGAGAACTTGTTTCAAAATAG
AATCTGTGGTTTATATCTTAAAATAGATGGAATTGTACATTTCAAAAAAA
ACTGATTACTTCAACCACAAGAAAGAAAGGCAATTTATCTTACTTCTAATT
CTTACCTTTAAGGAAAAAAGTGTAAAGATTAGCTGTAAAAAGCTATTTTG
GTATATGGAAGGATATTGGAAAAAAAAAAGATAAGAGGGAGAGAGGAA
GTAAGAAAGGAAACGTTAGTTCTGTGAGAAAGTACTTTCAAAGGGTTTTA
TCGCAGAAGCTAGTCAACAGGTTTTGTACAGAATTGTAAAAAAGCTTTAT
AGAAAAATGGAAGCTACAGCCCCAACACTTAAAAAGATTAAAATGTAAA
CATCTCTCCAATTGATGAGGTCTCTTTCCATATTTCCTGACCTAACCTTGA
AAAATGTTTGTGGGCAAACATTTTTCACTCTGCCTCCCATTGTCCTACCCT

CTTCTTTTCCACGCCTGCCTTCCTGAAGAGGACTGGGAGCAGGTTATAGA
AAAATCACAGCCACCCACTCCCACTACCTCTTCTCTTTAGTCACCCTGTCA
TTAATAGCATCTTCCTCAGTTGTATTCTCCCTCCTTTCAGCAGGCTTCCTTT
GTGTCTCTAAGCAGGCAAGAATGGTATCTATTAACTGTTTAAATTGGAATT
CTTCAAGAGTTTTACTCTTCATTTTCTTCTTTTACCACTGAAATAATTGATT
TCACATGTGTTTGATTCAAGGGCGAACCTGGTGTGCTTGGTGCTCCAGGC
ACTGCTGGTCCATCTGGTCCTAGTGGACTCCCAGGAGAGAGGGGTGCTGC
TGGCATACCTGGAGGCAAGGGAGAAAAGGTATCCTCTCAGTGCGCGACCT
TAATGCATACTGTTGATGAACCCTAGTAACGAAGGCTGCACGAGGTTAAC
TAGTTTTTTACAATTCTTGGCAGGTGGTCTGGTAGCATTCTGATATCTATC
TATACACATTTCCCTCTGCCACCTAGCACCTACACATTTCTAAACTCACTA
ATCTGGCAGAATTCCTTGCTACCATGGAATTTCACGCAAACAGATGGTGT
TGAGTAATACATGAGGCTCATTTTAATGCCACTAACAATAATGCCTCAGC
CTGCCCTAATTAATGGGAAGAAGCTATATTGAACAGCTGTCAACCGTGCT
GCTGCATTAGTTATGCCGTAAGAGTGATCAGGCACTGAAGCCCATTGTGA
TGTTGCCTTATAATTCTGTCCACATGAATTTGTACCTTGCTTGATTATGCTT
CAGGAGGGTGTATGGAAATTAGAAAAGAATATTAATCTGGAAAGAGCCC
TGAATTATATTTCCCTTTTTTACCTTCATTCAACGTTTTATAGAATGGTAGG
GAATCAAAACATTACTAAAAATCTTAAATAAATAAAGGTATTACAGCATC
TTCTGTAAAAAAAACACTTCAGATTAATTTTGACTCAGAAATTTTGTCAGA
AAACAAAAAGGTGTTCTTGCTTTATACTTTCAGGGTGAAACTGGTCTCAG
AGGTGACGTTGGTAGCCCTGGCAGAGATGGTGCTCGTGTGAGTAGAACTG
TTTTCACTTAGATTGGGTTTTGTTGTTTGTTTTGTTTTGTTTTCATTTCATTC
AATTTGAGTTGAGAAGCTTTCCAAAATCAAACTGACAAAAGAAGGAGTCA
AGTTTAGTAAAAGAGTATTCCAGTACTTTCTGGATCTATGACTAATAATAT
CAATGTTCCAAAAGAAAAATTCAACAGGACTTATCCCCTTGAAAAATCAC
AGCCTGAATTTTTTACATTCCTTAGTGTTAACTATACCACGTAAAACTATA
GTCCAAAACCTCTAGGTTTAATAGAATGTTAAATGATTGTTTTAAGTTACT
ACAGGGCTTTAGAAAAAAAAAAAAAGTAATTCTAACATTTTAGGAAGCC
AGGACCTCCAACAAGGTCATATGTACTCAAATGTATACCCTAAGAATTTA
TCACTCAAAGAGCAGTCCAAAAGATGAACTCTTAATGCCATGGAGATGCC
AACAAAAATCACATAAAGGTATAATAATACTGCCTAGAGTGTTGGCATCA
TGGCTCTTCCAGGCCATAAGCTACTTTGGTTTTTCTCACCTCTCCTTCAGTG
CTGATAAATTTCTGTGGCACATTCTCAATAGTCCTTTCTCTGAATCACACA
GTCCTAACACTTCTAAACAAGGCTAGTCATTTGCTGCTCTCCTGGGAATCT
TACAAGAGCATAGAGTTGAGATCTCCCCAGTTCCTAACCTGCCTATGACTT
AGTCACTTCCATAACCCTCGCCATTAGGGTACTTTCCTCCTTAGAGTGGCT
TCTAATAGTTGATAATTAGAACTGAAACGCATCAAAGGCTTTCTTGATGTC
CACCTAGAGTTCCCAGTAAAAGTGTTCAAACACTTTATAAGCATAGCAAG
GGGAAATGGCAGGGGTCACTTCTAATGGTATGAGGGGGAGTGATATTAAC
GAGATGTTTTCTTTTTGGTACTAGGGTGCTCCTGGTGCTGTAGGTGCCCCT
GGTCCTGCTGGAGCCAATGGGGACCGGGTAAGCATACATTTTCACTGAGT
TGCAGCTCTCTGAGCCCATAGCAGTATCTAAAATTTCCTGTTGCTCTAGCC
CCAAAGAACCCCAGCAACTCATTTTTATGGCTTGGTATAAAGCCCACTTAT
TTAAAAACCTATCTGTTAGATGGGGCATCAGGAAACAAATGCTGTGCGTT
TAGAGTTCTCTTTCCCTTGTTATACATGTGCCAGGCATCTGATCTATACTTT

AATTCCTATCATTTGTTTTAAAATCTTTTGGGGTTGCCCATTAACAATATTC
TAATGCACTGAGCCTTCTTAAAGCCTTCAACTGTAAAAGAAAAAAAATCA
ATAAAATACAGTGCGCCCATTTCACATCGAACAGCTCACTTAAAATAAAA
CTTATTGTCCTGCTGAATAGCATTGAGAACTTTGTCTAGATACTCATGTTG
CAACAGAAGAAAGGCTGGGGGAAAAAATGTAATTGTAATGTATACATGT
AAGGATAACCTGACCCCCTTGCTGTACAGTGGGAAAATTAAAAAATAAAA
TAAAATAAAATAAAACTTATTGTATTGCACAGATAGCAACAAATAGATCA
GCCTCATATCCACCTAAAAAATAAAATATCCTCCTCCTGGTATTGTGGGCA
CTGATTTATAATGTTTTTGCAATTGTATGACCAATACCCCTTCAACCTCCA
AAATGAATACAGTATTACACAGAGTAAGAATCCATTTCATTTTACTCTGTG
AGATGTGAATTAGCTATCCTTCTGGATATAGTCACAAACTCTGACAAGAG
CCCACTGTCTACCTCTGTTTCCTCTTCCTCATCTCTCTTTGAGTATCTATGT
GAGGCATATTAACAAACTCATTTTTTTGTGCCCTGATAGGGTGAAGCTGG
CCCTGCTGGCCCTGCTGGCCCTGCTGGTCCTCGTGGTAGTCCTGTAAGTAG
AAATCTGGGTCATTTTGGACACTTAGAACTTTTATACATTGATGAATCCAA
GATTTCGAAGACATTTTTTTAATACCTTCTCCTACCTAGGGTGAACGTGGT
GAGGTTGGTCCTGCTGGCCCCAATGGATTTGCTGGTCCTGCTGTAAGTATC
ATGTAATGAAGGTTAATCTGAAAACATCCCAAATTAGCAAGTAGAGAGA
GTTGGAGTACCTGAACTATAACTCTTCCTTTCTAACAGGGTGCTGCCGGTC
AACCTGGTGCTAAAGGAGAGAGAGGAACCAAAGGGCCCAAAGGTGAAAA
TGGTCCTGTTGGTCCCACAGGCCCTGTTGGAGCTGCTGGCCCAGCTGTAA
GTTGAATTCACTGGTGGTCGACACAGCCTGTGCAATAATCTATGAAATCTT
ATAACTCAACATGAATCAATCAAGTGCCTACCATACATCAGGGAACATAC
CAGAAAGGATGATTTTTAAAATAAGAGCTAACATTTGCATACCACTTTAC
ACATTATAAAAGTTTCATAAGCATTATCTCACTTTTTTGATAACCCCATAA
AGGTGGTAATATTGAAGTAAGGACATTTATTTAATGCAAATAATTATTTTA
AATTTCTAGATATGCTTCTAATAAACATGGGAAAAGAGTAACAATTATAA
AGATTGTCTTGTGATATGACTCTCTCTTTTTTGTTTCCATTTAGGGTCCAAA
TGGTCCTCCTGGTCCTGCTGGCAGTCGTGGTGATGGCGGCCCCCCTGTGA
GTACTTAAAATGGATTCTGATCATTTTTCTTTCACATCTATTATGGATACCT
TGAAAGTTTAGAAAATTTTGGTAATTTGGATTATAATGAGAGACTTGAGA
TTCAAGCTAACTCTGTTTAGAGAAATTTTTATATTGAAGTTAAGGTCAGGT
TCATGGAATGTTAACTTAATCCCAGCTGGAACTCAGTGCACTGTACACTG
CACTTCACTTGAGGTAGTAATAACCAAAGTTAACCCCAGCTTTAACTATG
AAACAGGTAAAATATTTCTGTGGCACATTTTATCCTTTTGTTGATGCCCCC
TTCTTCTGCCTGATCACTTCCTTTCCCTTGGTGTTCAGTGCTGTCACACTGA
TTCCCTAAAGACCCAAAGTTCCCCAGGAATGAACCCTCTGACTCTTTCCTT
CTCTCACTGAACAGCAGTTAATCTTGGTCTTTTGTTGTACCTTCTCCCATA
AACAAATTACTTATTCAGTTTCCATTCTCTTAGCACCTAGCCAGGTGTTTA
ATTAATAACATATATATATTTATATATATACATTTATATATATAATCTCAT
ATATTAGTATATATATAATCTCATATATTAGTATATATATAAGTATATATT
AATAGTATAGTGTATTATCTCATTGCTTTTACAACTATCAAGTAATATGTT
AGCATTTTAAAAAGAACCATCAGAGCCCTAATGCAAAATGTGAACATTGC
ACCTGAAGGCATGATGGTCAGTAGAGATGGAGGGTCTTTGGGGTCCTGGT
TATATAATACCCTATCTGAAACTGTCACTACCTACTCTTTAGTCCATCCTT
GGTCATGTGTACTGAGGATCTGCCAAAGGGGAAATTTCCTTCTTACCCAA

ATTCCTGAAGTTGGTGCTGACTGTGGAATTTTAATGTGCTTGGTTCTTAGG
GTGCTACTGGTTTCCCTGGTGCTGCTGGACGGATTGGTCCTCCTGGACCTT
CTGTAAGTAAACCACTTCAGAAAAAGGTGTTCATTTACTTACTCTAGCCA
AAAGACCTGGCCTCTAGTAAGAAACTGGTGAGAAACTCTTTATGAAACAT
ATGCTGTTAATATTTTGCTCTTACTTTCCTAATATAAAATCAGTTTTTCTGA
ACTATTAAGGCAAAATCATCACAAATTATATTTTATAATATCAGTCAGGA
GGCTTTTATTCATAAGAATACAAGTCCCTTCATCCCTTTTTAGGGGCATGG
TTTCTTTCTAAAAAATAAAAAACAGAATAGTTTTAGTCACATATCAGATAT
TTTTATATCTAATTACCCTTTATGGCCAACATTTTGCCTCCATTGTTAAAGT
ATAATTGTTCCTGAATTTAAAGGTGGTTTGGCCTCTAATTTAATTCTGATT
CAGACTCTCCTGTCAGGACTCAAGAGAATTTAATTAATTACCAAGGATTA
AGTCTTCCGGTTAAGGTTTCAGGGGAAAAAATAGCAAAGATGTTGATTTC
TTGGGATCCTTTCACAGGTTCATAACAGAAAAATCTTCATTCCCTGTAGGC
ATTTAATTAAGCCTAGTTGAGAGCATATGCGATTCCTCAATTATGAACAA
AATGCCTGTATTGACTTTCTTCAAAAAGGAACAAGCAGAGTCCTCCAAAA
GCCAAAGTCCCATTCATTTACAGCTAACTCTTAGTTAGAACTCAACATAGC
TCTGACTCCAGTATGCCAGGTGACCAGTGGCACAGCATTGCTTTTCTCCAC
AGACATTTGTCATAGCCTCTGCCAAAGTTTTTCACTTTATTACAGTCTCAG
GAGAAGAAAGATAAAAGTTATTTCTTTCTACAAGAAAGATTGCTTATAAA
TCTAAACATTTTCCAAAATGACTAAGATATTATTTTGTTGCTTCAAATTGT
TTGCATCTTAATATTTAAAATCTCAGCTGCCCTCTTCCAGGTCAGCGATGA
TTAACAGAAAGTAAATAACCTTGTACATTTGTTTATAGGGTATCTCTGGGC
CCCCTGGACCCCCTGGTCCTGCTGGGAAAGAAGGACTTCGTGGGCCTCGT
GGTGACCAAGGTCCAGTTGGTCGAACTGGAGAAACAGGTGCATCTGGCCC
CCCTGGCTTTGCTGGTGAGAAAGGTCCCTCTGGAGAGCCTGGTACTGCTG
TAAGTGATTTCAAACTCCTCCTTCTTATTACCTTGTACTGAATTAGAATAA
AGCCCTTATAAAAAAATCTTCAAGTGGCATATATTGATGAGTACTGCAGG
CTCCAAAGTAGAGCTCAGTTTGGCCAGGAAATCTGTCCAACAAATATTTA
GGTGTCATGGCTTTCACAGATGCTCAGAAGGCAGTAGATGGTGATGACAA
TAAGTGAGGGAACTCACTGTTTTATCAATACAACCCTTTAAACTGTTGAG
GATACTTCACTGGTACTGGGATTAGAACAGAATCCCATTGCTGTGGTCAT
CCTACTATAAATTAATCAATCTCAGTAAGCTACCAATTTTTCAAACATGCA
TTGTCCAGTATTATTGTTACTCTGATGTGTGGCTTTACTAGATCTGGTTCA
GTTAGAGGCCCAGGTCAAAATTTGCTTCCCAGCATAAATCCTATAAAAAT
TCTATTCCCTCATTCCAATTTGGCTTCATTAATCCCAACCACATGTTTCTAA
ATGTTGCTCCAGTGGTCTCAAGGAGGAACGAGACACAAAGAAATACATAT
ACCCAGATGTAAACTCATTGTATATCATGTATACAATGATATGTCCTAGG
AGAACATTAGCCTTCCTAAACTAAAGATAAACTCATTCTCAGAATCCCCA
AGCCAATCTGTGTTACCTCCTAGGGTCTCACCAGTAATACTCAGTATTTTT
ATTCTTATTCAGGGACCTCCTGGTACCCCAGGTCCTCAAGGTATTCTTGGT
GCTCCTGGTTTTCTGGGTCTCCCTGGCTCTAGAGGTGAACGTGGTCTACCA
GGTGTTGCTGGATCAGTGGTGAGTGTTTGACACCATTCTTATTCTCATTAA
CATAACAAAAGATTTTAAAAGCTGCCACTTCAGCTGTGACAGATTGCTCA
CTGAATAACTCCACTTAAGATTTTTTATTCATGGCATTTTCTTTATACAGAT
CACATGTCACTTATCTAAGAAGCTTTAATACCACCTTACTTAGACACATGC
TGTAAAGACACAGCTTAACATTATGCAGAATGATTTTTGTTCTTTTTACAT

GCTTAAGAATGATATAAGCCCTAATTGCATTTACTCCTCTGCAGTATGAAA
TGCTGGCATCATTTATCCTGTAGGAAGAATGAAACTCTGGAAGTTCTGAA
TCATTCCATTAAACCTAATATGTGACAGCAACTTAGAATCATCTAATCTCC
ATAAACTCTCGCAACTTTATGAATTCATTTTATTTGGTATATCAGATGAAA
CTAGGCTCACCTTGTACAGAGCAACATTACAAAATGACATAAAGCTGTCC
ATCTAAAGTGTGTAGGAAGAGGGCAAGGATGACACAAATGACACTTCTTA
CTGTTCTCTGGCCCATTCACTTTCACTTCCCTTCGAAAACCAGAATCCCCT
GGCTAGGACTGTTTTGGGAGAATGAAAGTGGCACTCAATCCAATGAGAGG
AATGTTTGTCTTCTCTGCCTCTCTAGGGTGAACCTGGCCCTCTCGGCATTG
CAGGCCCACCTGGGGCCCGTGGTCCCCCTGGTGCTGTGGGTAATCCTGGT
GTCAATGGTGCTCCTGGTGAAGCTGGTCGTGATGTGAGTCCAACATCTGG
TTTGTAAAATAAAACCTAGCAGTACTTCATTGTGTGCAACTTTGTATCCTG
AGCTGAGGTTCCCTTGTTCCAAGTTTCAGAACAAGATGGTCTGTTTCTCCA
TATTATCTACATGTGCTAACCACTCCTCGAATTGAAAATATGAAAAATAA
CTTGAGTTGAGGAATGAGTCCTTAAATAGCACTTTTTAAGTTGACCAAAA
CATTAAAGATACCTCTAAAAAAATCTCAGTTGGCAAATTCCTATTTTTGGA
AAGCCTAAAAGGAGTTTATTAGGTGCTGGGAATAATGAAGGCAGTTAGCT
GAACTGTGGGGGCCAGTGGGCTAGAGCCTCCCTAAGTGCCTTAATTGGTG
TGGTGTCTTCACAGGGCAACCCTGGAAGCGATGGTCCCCCAGGCCGAGAT
GGTCAAGCTGGACACAAGGTCAGTACACCTCTTCATTTTTTTCTCTAACTT
AAAAGGGACTAAAACATAGGCCAGATTGATGCTAAGCTTTATTTTGTCTT
TGGTAGGGCGAGCGTGGTTACCCTGGTAATCCTGGTCCTGCTGGTGCTGC
AGGAGCACCTGGTCCTCAAGGTGCTGTGGGTCCCGCTGGCAAACATGGAA
ACCGTGGTGAACCTGTAAGTTTGTAGATACCAGTCTCTTAATGCTGTCATC
CTCAACAAGCTTTATCTGTGACCTCCCCACACTTGGGAACTGTGGAGGGA
GGATCTTGGGCTTGGCCAATAAGTTGTGTTTTTCTTTTCCAATTCACAGGG
TCCTGCTGGTTCTGTTGGTCCTGCTGGTGCTGTTGGTCCAAGAGGTCCTAG
TGTATGTACACGGGGAAGATTTCTGTACAAATCAGCATTTAGGGAGGATG
AGTCCAGACTATCCATTAAGAAAATAATACATAGACTAGACTTTCTATTT
AAAAAAAAAAAATTAGTCCATGCTGAGAATTGCTACAAATAGCTTTTAAT
CAAATCTGTGTCAAATCTCTTATTATTTGTATTAAATAATCACAAAGATTA
GCCACTTTATAGACAGCTCAACTAAAGCAGGTTAAGGGAATAGATGAGA
GTAGCATCCACATGGCTTTTAGACCTTTGTATATCTTAAATTATCTTTGTA
AATAGATTTAACAATATTTGTGTAGAAGTGAATATCTTTTTTTAGAGGTAA
ATTTATTTCTTTCTCTTTCTTTATAGGGCCCACAAGGTATTCGAGGTGAGA
AGGGAGAGCCTGGTGATAAGGGGCCCAGAGGTCTTCCTGGCTTGAAGGG
ACACAACGGATTGCAAGGTCTTCCTGGTCTTGCTGTAAGTAAAAAATGTG
GACATCCATCATGTATACAGATCTTAAAGTTCTTCAACTCAGAATAATCTT
CCTCAAAATTTCACTACTGTTGTTTTAAAATGTCCTACATCATATTCATTTC
CCATTCTCTGGCTTACACCATGGCTCAGAGATAATGCTTTTTCATCCCAAC
ATGAAAAGGTGAAGGTTAAGGAAGAGAGAAATATACATATATTAAAAAT
CTCTCAGCAACAATGTTCTTCAACCCCATTTAAAATAAATATAATTAGAG
GAATGACTAATACTGCACCGCTGAAATAGCTTGTAGAAAAAAATAATTGA
ATATAATAGACATAATGAGAGAAAAGAGCCCCATTTTACATTTTCAATTT
TCTCAATCCAGAGTTCATCAAACTAAAAGTTTCCATTCAGTTTTTTAAAAT
CTATCTATCGACACATGTTCTGAAAGTGTGATTTTTCCTCTTCTTTCTTTAA

AGGGTCATCATGGTGATCAAGGTGCTCCTGGCCCTGTGGGTCCTGCTGGT
CCTAGGGTAGGTCAATCTAAGAAATGTTTATCTCTGAAACAAAGCCTGCG
GCAGGCAGTAAGCCCCAGGCTCTTCTCCCTTGTGGGATATTCACTGGCTC
ACACAGACTTGGTGTCCTTCTTCCCTAGGGTCCAGCTGGTCCTTCTGGCCC
TGCTGGCAAAGATGGTCGCACTGGACAACCTGGTGCAGTTGGACCTGCTG
GCATTCGTGGCTCTCAAGGAAGCCAAGGTCCTGCTGTAAGTATGATTTGG
GGAAATTAAAGGGTATCACTGGCCTAAGGGAATATTTTCTTTAGACTAAA
CCAAGACAATTATAACACAATAAAGTAATCCCCCCACTTCCGTATGTTTTC
TAAAATATATTAAAATTGTCTGTTTGTAGTGAGCTCTTCAAGAGCATCCCT
GCATACAAGTATGTCTGGGAGCAGGTTGTTATTGACATGAATTTGACTCT
ATTTTACTTACTAGAACATGTTTTTGTGTGAATCTCAACTTAAAAACTGCT
GTCACTGTTGTCTTTACTGTAACAAAATACAAAGCCTCCTCTTTCTCACTT
TCATTTTTTCAGGGTCCTCCTGGTCCTCCTGGCCCTCCTGGACCACCTGGC
CCAAGTGGTGGTGGTTATGATTTTGGATATGAAGGAGACTTCTACAGGGC
TGACCAGCCTCGCTCACCACCTTCTCTCAGACCCAAGGATTATGAAGTTG
ATGCTACTCTGAAATCTCTCAACAACCAGATTGAGACTCTACTTACTCCAG
AAGGCTCTAGGAAGAACCCAGCTCGCACATGCCGTGACTTGAGACTCAGC
CACCCAGAATGGAGTAGTGGTAGGTAGAGATGTCCAACCAGACCAACTTG
TCTCACAAGTTGGGTTTTTCAAAATCAGTTCCCACTTATATTTAGAATAAC
AAGATATTTGGGTAAAAACTGCATTATGTGAAATCATTTCCAGTTAATGA
AGCATTTTCTTCTCCCAACCAGCACCCATTCTCAAGCTTCTCTTGATATAT
ATTTCTGTCCTTTTCATGGTAATATGTACCTTGACTTGATTTTCATAGAAA
AGAGAGAAAAAATTATCTAATACTGAGAATAGCCTTTCCTATTAGTATGG
GGTAAACAAAATAAAAAGTTCAATCAAATATGTTATCTTGGGGAATCAAA
ATCTTTGACTAATAATTATTAGAACTTACTGTTCCACTCAATGAGGTTTTT
TGATTTGGGGTCTTGCTTTCTTAAAAGGTTACTACTGGATTGACCCTAACC
AAGGATGTACTATGGATGCTATCAAAGTATACTGTGATTTCTCTACTGGTG
AAACCTGCATTCGAGCTCAACCTGAAAACATCCCAGCCAAAAACTGGTAC
AGAAACTCCAAGGTCAAGAAGCACGTCTGGTTAGGAGAAACTATCAATG
GTGGTACCCAGGTAAGGAATCCTATAAACGGTCCTTCTCCTACTAAATAA
TATCTTGTTAGAATTGTTTGCTTTTTAATTGTTTACATGTTAGCCCTTTACC
AAAGGTAAAACTTTAAATACATTTAAAGTTATCATTCAAATGTGACTTAC
CTGATAATCACTGGAGAAACTGACTAATTTCATTTAGTGTGAAGGAGTAC
TGGCCAGGATATAGTCCAGTACCCATTTTTAGGATATGTGCTTCTCAACTT
CTGATGAAAACTTTTCACCAGTTTTATGAAATGCCCACCGATGAAAATTA
CAGAAAAATAAAGGGGTCTGAAATCTCTATCTCCTTGCTCTCTTATTTTAA
AAAAGAAACAAACATACAAAGGATTTATCTGAAACAGGTACTTTTTTTGA
GATAACTTGCTTCAGACTGGTAGTTCCTATGTCCCTTCTAGGTAGCTAATG
CTTCAAAAACAAGTGAATTGAGTTTCCTTTAAAAGTACAAGTTCAGATCT
GAGTCCATTTTGCCTCCCTATACCAGACTATTTTTTAAATTTGATTCTTAGT
ACCTGAGTCCTTCTCAACTTAACTGGAATTCCCATCTTATTCTCTGTAGTTT
GAATATAATATGGAAGGAGTTACCACCAAGGAAATGGCTACACAACTTGC
CTTCATGCGCCTGCTGGCCAACCATGCCTCCCAAAACATCACCTACCATTG
CAAGAACAGCATTGCATACATGGATGAAGAGACTGGCAACCTGAAAAAG
GCTGTCATTCTGCAAGGATCCAATGATGTTGAACTTGTTGCCGAGGGCAA
CAGCAGATTCACCTACACTGTTCTTGTAGATGGCTGTTCTGTAAGTAATAC

TGAAATATAGGAATAACATTTATTTGGGAAGTGGGACAGTGGAGCTTAGA
CCATTAAATGAACAGATGAATGAGAGATCTCACTTATTTGATTTTTCTTTC
ATAAATGTATTGTTTAAATTCAAATGGTAAATCCAGAGTATTCTTATCAGA
TCTTACCTTCTGAAATTCTCTGAAATAACTAAAATGTATACTATGTCCATT
ATTAGTCATTTTTCTTTTAACATGAGATTTTTAAGTATGGCCATAACTTAA
AATCACTTGAGGTTAAAAGGGCCCAAGATATAGTTTGTAGTTACATGTAT
TTTTATTTTACTTATTAATATAGTCTACATTTAATTTTTCTTTATACTTTCTC
TATAATATGCTAAAATTGAAGGAAGAAAATTTTAAATTCACTAATAGTAC
TACATATCAACCCTACTCTTCCTACTTTACACATCCTGAAAATTATAATGT
ATAAATAAACCACTTAGAATATGCATTAATTATTTTTAATAAAATTTGACT
TGTAGAAAACCTCTCTCCATAAAGATACCCCTGATTCTACTGTTTCAAAAA
TAAACTTGTTGATCTGAGATACACAGTTTCAGAATGACATGAGCCAAAAA
GTGGTTTTATAATATCTAGTATATATTATATATTACATATCACTACATATT
ACATAATTGTATGTAATATATAATTTATAATGTTATGTAATATAATCAGAT
TCCATAATAATGCTGCTTTATGTATATCTCTTCTTCCTTTTTTTAAACAGAA
AAAAACAAATGAATGGAGAAAAACAATCATTGAATATAAAACAAATAAG
CCATCTCGCCTGCCTATCCTTGATATTGCACCTTTGGACATCGGTGATGCT
GACCAAGAAGTCAGTGTGGACGTTGGCCCAGTCTGTTTCAAATAAATGAA
CTCAACCTAAATTAAAGAAAAAGGAAATCTGAAAAATTTCTCTCTTTGCC
ATTTCTTTTTCTTCTTTTTTAACTGAAAGCTGAATCATTCCATTTCTTCTGC
ACATCTACTTGCTTAAATTGTGGGCAAAAGAGAAGGAGAAGGATTGATCA
GAGCATCGTGCAATACAATTAATTCGTTCCCTGTCCCTCTTCCCCTCCCCA
AAAGATTTGGAATTTTTTTCAACATTCTAACACCTGTTGTGGAAAATGTCA
ACCTTTGTAAGAAAACCAAAAATAAAAATTGAAAAATAAAATAAAAACC
ATGAACATTTGCACCACTTGTGGCTTTTGAATATCTTCCACAGAGGGAAGT
TTAAAACCCAAACTTCCAAAGGTTTAAACTACCTCAAAACACTTTCCTGTG
AGTGTGATCCATACCATTAGGTGCTGACCTATACAGAGATGAGCTGAGGT
ACTTGTTTTTTGTTCAAAATACAAAGGTGCTAATTAATAGTATTTTAGATA
CTTGAAGAATGTTTATGGTGCTAGGAGAATTTGAGAAGAAATACTCTTCT
GTTTTGAGTTGTATTGTGTATTTTTTTTCATCTGATTTAATATCAGTACTTT
ACATCTTATTCCTGAATAAAAGTAAAGTGGATCATTTGCCCAGCATTGTCC
TCTTCTTTAAATTCAGCATTTGTTCTTTGCCAGTCTCATTTTCATCCTCTCC
CATGGTTCCAAAGAAGCCTTGTTTCTCAGACAAGCAGAAAAATTAAATTG
TACCTATTTTGTATATGTGAGATGTTTAAATAAATTGTGAAAAAAATGAA
ATAAAGCATGTTTGGTTTTCCAAAAGAAAA

C) Porcine insulin gene sequence

CCCCCCACCCAGGGACTGCTGTCATGAGCTGCTGGGGCATGTGCCCATGC
TGGCCGACCGGACCTTCGCCCAGTTCTCGCAGGTGTGTGCCCCAGGCCTC
CGGAGCTGGGGGCCGGGGAGGGGCCGGGGTTGGGGGGCTGCCCGCCTTT
CCTGCTTCGGCTTCAGGGAGACGGCATCGTCCTGGAATACAGGGGCGGGC
GGGCGGGGCGGGCGCGCCGGGGCTGGGGAGGGGATGCTCAGACCCGCTC
TGGGAAGAAGAGAGCCTCAGAAGAAATCCCTTCCCAAGGGTCACGCGGT
GGAGCCCAGGGGCCCGCTAGGGGCCGATTCCCACAGCTCGTGCTGCCACC

TGCTGGCGCTCCCAGGAACTGCGGAGGCGGTGGGGGCCCTGGATGGGTCC
GGCAGTGGGCTCGCAGGAGACCCCTGGAGGGGCTGCGGACACCCCAGCT
GCCACTCACAAGGTGCCCAAGCGGCGGTGGCAATGGGCTGAGCCTCTCCC
CCCCTCCTCCTCCGCAGGACATTGGCCTCGCATCCCTGGGGGTCTCGGAC
GAGGAAATTGAGAAGCTGTCCACGGTGGGTTTCTCCCCCTGCAGGGCCCT
GGGTTCCAGCCAGGCCCTCCTGTCCAAGGGGTGTCGTCCTCACGCTGTGA
CCGCCCGGGAGCCTGGATCGGTTCTGCCTGGGTGGGCGGTGCCCGGGCCA
CGGGCAGCAGGGGCAGCGGTGCGGGCCCCAGCCGTGTCTGAGCCCCCTTG
CCGCCTGTCCCCACCAGCTGTACTGGTTCACGGTGGAGTTTGGGCTCTGCA
AACAGAACGGCGAGGTGAAGGCCTACGGGGCTGGGCTGCTGTCCTCCTAC
GGGGAGCTCCTGGTGAGGCCTCCCCCACGCGCTGGGGCCTGGGTCCCCGG
GGGAGGTGACCCCTGCGGTGCCTTGTGGATTCCAGCTCTCGGGAGGCTGG
AGCGAGGGGCTGCCCTCCTGGGGGCACCAAGAAAGCTGGTCTGCGCCCCT
CTCCACACACCTGTGCCTGGGCCCTGGGGGGACCCCTGCTGGGGGATGTG
GGTGCACAGCCAGGGCCACCAGGGAGTCAGGACACGGGGCTCCCTTCCCT
CGGGTCCCTGAGACCCCTGGCCTCCCGCCAGCACTCCCTGTCCGAGGAGC
CCGAGATCCGGGCCTTCGACCCCGACGCGGCGGCCGTGCAGCCCTACCAG
GACCAGACCTACCAGCCCGTCTACTTCGTGTCTGAGAGTTTCAGTGACGC
CAAGGACAAGCTCAGGTGGGCCGGGGCCCGGGGCCCCCAAACTGGAGGA
TCCAGCCTGCAGCCCCGCCTATGAGCCCATTTCCCAGCAGAGGGAGCTGC
TGCGGACCCCACCGTCACAACCCCCCTCCCACAGCTGGAACCCCAGAAAG
CCTGCGGAGGGGGGACCTGCAGGGCTGTGGCCAGGTCAAGGCCAGGTCG
AGGCCAGGCTTTTAGGGGTGAAGTCTGACTTTGTAAGAGGGGGTGCAGGG
TCCTTCCCAGCCTCCTCCCCTCCGAGCAGCTGGGGGCGGGGCGGGGGTGC
GATGAAGGCAGAGATGACGCAGCCACCCGTTCACCCTCAGGAGGCGCCT
CCTGTCCAGCCAGGCTCCTGTTGTCACAGGGGAAACTGAGGCCCCAGGTG
TGTGTGTGGGGGGGTGATTCTCACACACAAGCTTAGGGACAGGGACATAA
CGGCCTCTCCAGGGCACACAGTCTGGAGGCCTTCCTGGAGGAAGAAGGCC
TGGCTGAACCGGGCCTTAGGAATGAGCAGGAGTGGTGGAGTACCTGGGC
AGAGGCTCTCCCAGAAAAAGCCCATCCCTCCCACTGTAAACACCCCCCCA
GCCCTTGGAGTGGGGGGGCCTCCGCCCCCGTCTTATCTGGGTGGGCGGAC
AGAAGGAGAGCCAGGCGGCTCATAAAAGCCCCGGGCGCTCTGAACCTGA
CATGACGGAGGAGCCCCGGCCAGAGCCGCAGACCCTGTCCCAAGAGGCC
CCCGTGCCCCCGAGTCCCCGCTGGGCGACTGAGGGGCCGGGCTGCTCCCT
CCCCAGGAGCTACGCCTCCCGCATCCAGCGCCCCTTCTCGGTGAAGTTCG
ACCCGTACACGCTGGCCGTCGACGTGCTGGACAGCCCCCACGCCATCCGG
CGCTCCCTGGAGGGCGTCCAGGACGAGCTGCACACCCTCGCCCACGCGCT
GAGCGCCATCGGCTAGGCATGGGGGCGCCACCCCATGGCCAGGGTCTCCT
GGCCCTGCCCTCGCCCGCCCAGGGCACCCCAAGAGCCTGCCTGTGACCCC
CCTGCTGGAGGGCGCCCTGCCACCCCTGCCGCTTCCCAGCTCTGCCTGCA
GATGTCTTTGTGCTGCTTCTGCCCGCTGCCCATTTGCAATCTCAATAAAAC
AAAGAAACCCTCCGAGAACCCACCCAGGGTCTGCGTCTGCACGGGCTCAG
GGGCCACGCACACAGCCCGCGGCCCCGGGCCCTCACCGACCCCGCCCGGC
CCAGCTCAGAGGTTCCTTTGCTGCCCACCCTACCCCCCGTCCCCGCAGGCC
TCGCCAGGCGGCCCTGGGTGACCTGCTCCCTAGGGTACCAGCCTGTGCTG
AGCCTCGGGCAGCTCCAGAAGCTGGATGTCCTGGCTGCACTGTCCCACCT

TGGTGGGTGGCAGCATGAAGGTGGGGGCCTGGCTGGCAGGAGCCCCACT
AGGTCTAGGCTTGGGGGGCAGGCTCCCAGCTGTGGTCTTGGAGGTGGGAC
GGTGGGCTGGGCCTCCCCCAAAACCTGATGTGCCCCTAGGAGGAGGCCAG
GGCTCCAGGGCCAGCCTGTGAGAGCAGGTAGGACAGGCAGATGACAGCT
GCCTCTGACCCCCATTCCCCCACCCCCAGAGCCCCCGGGAAGGATCCCCA
AATGTGTGCTGACAGAGAGAGTTCAGCTGAGCTGGCTCCCAGGTCACCTC
TCTGGGTCTTGGTGCCCCCAGCATCTCCCAGGCTGGCCCAGCCCTGGGGG
TGCCCCTCCCACCCTGCCCTGGGCCTTGCGGAGGACACCCTGGGCTCACTT
CCCGCAGGATGTAAGCACCAGCCTATCTTCCAGGCCCTGCGCCCTCCCTG
GGTGCCCCCCTACTTCCCCTTCTACCTCCAGAGCCCTGACTCTAGGCTCTT
AGGATGTCGGTCTTGGAAAACTCCTACTCATCCGTCAAGACCCTCCTGGG
CAAACCCTTCCTTCCCAGCCCCCCACCCTGGATCTGTGCCCTTTCGGCCTT
TGAGGCCACAAATGAGGCTGTTTCCAAAGGTTGGAGGCCCCTGGGAAGG
GCTGACGGCCGGCCTCCTCCCCTCCAACCCCTGGGCCCTGGGCTCTGCCCT
CATCCAGTCTCCTGCCTTGCACACCCTCTCATAGAGGCCCCCAGATCTTCC
CTGGCTGCAGACGGGCCTCAGGACCCCCTGCTGTCCTGGGAAGCCAGGGC
CCAGCTCCTTCCTCCGCGTGGGGTGGGGCCTCCCCACAGGGGCCTGTCCC
GGGGGGGTGGGTACCAAAGGGTCCCCCCCGCCCATGGGACAGCGAAGGG
AAGCAGTATGTCGTGGGGCCGGGTCTGAAAGGGGTCAGCAGCAGGGAGG
GGCAGGGGCTCCAGGAGGCAGGGCCACTGAGCGGTACCTGGTGGGGGGA
GGTGGTGGGGCCACACCCAGGAGTCCTGTGCCCCCCCCACTCCCGCCGTT
GGAGATGAGAAGCAGGGGCCAGCCTGCGGGTCCCTGAGTTNAGCGCCCC
CCCCCCCCACCGCCGCAGCAGCCCGGGGTCTCAGCAGGCTGCTGTGCTGG
GGGCGGGGGCGCTTATGGGGCCGGGAGCAGCCCCCCCCCACGGCTTCGG
AGCATCTCTGGGGCCTCAGGGATGGACCGGGGTCTGCAGGCAGGTGTCCT
CTCGCGCCCCCACTCCCTGGGCTATAACGTGGAAGATGCGGCCCAAGCCC
GGCCGGTTTGGCCTTTGTCCCCAGCCAGTGGGGACAGCCTGGCCCTCAGG
CCGCTCGTTAAGACTCTAATGACCTCAAGGCCCCCAGAGGCGCTGATGAC
CCACGGAGATGATCCCGCAGGCCTGGCAGCAGGGAAATGATCCAGAAAG
TGCCACCTCAGCCCCCAGCCATCTGCCACCCACCTGGAGGCCCTCAGGGG
CCGGGCGCCGGGGGGCAGGCGCTATAAAGCCGGCCGGGCCCAGCCGCCC
CCAGCCCTCTGGGACCAGCTGTGTTCCCAGGCCACCGGCAAGCAGGTCTG
TCCCCCTGGGCTCCCGTCAGCTGGGTCTGGGCTGTCCTGCTGGGGCCAGG
GCATCTCGGCAGGAGGACGTGGGCTCCTCTCTCGGAGCCCTTGGGGGGTG
AGGCTGGTGGGGGCTGCAGGTGCCCCTGGCTGGCCTCAACGCCGCCCGTC
CCCCAGGTCCTCACCCCCCGCCATGGCCCTGTGGACGCGCCTCCTGCCCCT
GCTGGCCCTGCTGGCCCTCTGGGCGCCCGCCCCGGCCCAGGCCTTCGTGA
ACCAGCACCTGTGCGGCTCCCACCTGGTGGAGGCGCTGTACCTGGTGTGC
GGGGAGCGCGGCTTCTTCTACACGCCCAAGGCCCGTCGGGAGGCGGAGA
ACCCTCAGGGTGAGCCGAGGGGGCGTCCCGGGAGCGGTCGGGGGAGTTT
TTAAGGAGGAAATTGGTAAAAGTGACCAACTCCCTGGGAGCTGAGCCCA
GAGACACCCCTCCCACGCCCCGGTCCCGCTCGAGAAGCCCCCCTTCCCTC
CCCTCCTCCCGAGGCGGCTCCAGGGAGGAATCTTACGGAGTCAAGGCCCG
GGTGCCGCTGGTCTCCGAGTGACATGGCCGTGGTGTCCCGTCTGCCGGCC
CACATGCCCGTGGGAGATGCCCCATCCCCCTGGGAGGGGGCCCCGTGCCG
GGCAGGCGGCGGGAGGCCCAGGACCGGTGGCTGCTGCGGCTTCCACTCC

AGGGTGGGCGGGGTGGGGGGTGGCTGTCTCTGTGTGACCGGCTCTCCCCG
CAGCAGGTGCCGTGGAGCTGGGCGGAGGCCTGGGCGGCCTGCAGGCCCT
GGCGCTGGAGGGGCCCCCGCAGAAGCGTGGCATCGTGGAGCAGTGCTGC
ACCAGCATCTGTTCCCTCTACCAGCTGGAGAACTACTGCAACTAGGCCGC
CCCTGAGGGCGCCTGCTGCTCCCCGCACCCCAAAACCCAATAAAGTCCTG
AATGAGCCCGGCCGAGTCCTGTGGTCTGTGTGGCCTGGGGCGGGGGCCCT
GGTGGGGGAGGGGCCAGAAGGCTGTGGGGGGCCTGCCTGCGACCCCTCT
CTGCTCTCGCCACATCGGCTGCTCTAAGCTTCCTCCACATGCATCGGGTGC
CCACAGGCACATGGGCACCGGGGGACCAGGGCCCAGGGCAGGGCCCTTC
AATGTGGCGAGCTCTGGTTTTCAGGGCTCCAGACACCCCCTCCTGGGTGC
CCACTGCTGCACAGGGTCACTCTGAGGGTCACAGGGCACCCACCCAGACT
GCTCTTGGGCACACAAAATAGCCCAGGGGCTTCTTGGGCTGGCTGCNGTC
TGGGAGGTCAGAGAGTGACCCCGCGGGACCAAGACCTGGCCAGCCTGCC
AGTCGCCCAGGCCAAACCAATCTGCACCTTTGCTGAAGGTTCCACCCGGG
CCAGCACTGGGGGCGGCCGGGCCTAGAGCTGGGCGCCCGGGCCCCAGGG
ACTGCACACCCGCCAGAGGTGGGCCTGAGGGGTGGCAGCAGGCTCTCCG
CCTGGGACCCAGCCAGCTGGGCAGCTCACCTCTCAACACGAGGCTCTCAC
CTGTGTCGTCCCCTCCCCACGGCCACACAGACACCCCTGGGGAGAAGTCA
CAGGCCCCCAGCAGGCCCCGCCCCTGGAGAGGAGGCCAGGGCTGGGCAG
GCGGGTGGCCGGCCGGACACTGGACCCGGAAGGGGGGTAGGCGGCTGGG
ATGAGTGGCGAGCTGTCCATGGGAGCACCCAGCGGCCCCATTGGCACCAG
TACAGGCAGGGGCACCTGCAGCAGCTGAGGTACGTGGGGTCCCCGGACT
GGTTGGTGTCCGGCTGCCCTCTGGGAGGCAGCGGGCTGAGCTTGTGGTCC
TGCCAACCAGGGAGACCCGTGACCACCCTGCTGCTTCCCCTCCCCCCCAG
GGCCAGCAGACTCCTTTGGGACTCGGGGCCCCTGAGCCGCCCCCACTCGC
AGGACTCACGGGGTGTGCGGTCCTGGGTGAGTGGGGGCTTGGGAGAGGG
TCACTCTTGTCCGTCGGGTGGGGAAGGCTGAGAGTCATGGTGTGACAGCG
CCCTCGGCCTGCCGGGTGGGGGGTCTCCCTTCTCCCGAGCCCAGATCCAG
CCCTGAGACCCTGTATGCTCCCCACGAGCCTTGCCCGCCGGCCCTTGGGC
CCAGTGACATGCTTTGGCACAGGTGGGGAAGGGGGCTTTCCCGCGGACAA
GGTCACGCAGGGAAGGAAAGGCGGTGCCCCAGGGGACACCTGGGGGCCT
GGGTTCAAACCTGGCCTTTTGCTCAACAGCCCCCTTGGGGCTTTAAGCCCC
AGAAGGCCCTGCTTTCCTGAGATCCCCCGTCACCCGCAGAAGCCAGGCCC
ACAGGCCTTGGCTCAGCCCCTCCACCCAGGCCCACGTTCCGCCCCTTCTGG
GAACTGGAGGACAGCCCGCCCTCGCCCTCGGACCTGGCTTCGTTTGCCCT
GGCATCTGGCAGTGGCCGGCAGCTGCGTTCAGCCCTGGATGACACCCTGG
CGTGAGCGGTGGGTCCCCGTGCTGAGGGCAGCCCCCACACACGTCCTGCT
CACTTGCCTTGTGTCTGCTCCGCATCCCGTCATCACACATGCCATGCTGGG
GCACCGTAGCGCCTTGCCCTGTGTGGCACTGTGGCACTGTGTTCCTGATGG
GAAGACTGAGGCTGGGGTCAGGCCCGCTGCTGCCCACCCTCTAAGGACAT
TCTGCCGGTGCAGCTGCCTCCAGGCTGGCCCCCCGGATTGCATCTGCTTCT
GGCACGGATGAACTGGCACCTCTGCCTGACCATTAGGGCTGTATTTGCCTT
CTCCTGTTGGCAGTAAATATTTACTGTCCTCCCTGTCCTCCAGGCCGGAGC
AGATCCTGAGGGGCATGGGAGGTGGACACAAAGGTGCCCAAGCAGCCCC
CTGCTCTTGAGGGCCCAGTGTCTGGTGGGGGCCAGCCTGGGAAGGAGGAG
CGAGACTAGGAACCAGAGGCCTGTGTTCCTGGAAAAGGCCCCCTGGCAG

AGTTCCGGCTGGTGTGTGTCCAGCTAGGCTGTGAGTCTTCAAACTGGGGA
GCCCGGCCCCTGGACCCAGGCAGGGCTGCACCCCTGGTGCCAGTGCTTCA
CTGGGTGGGCACCTGTCCCCACCAGGCAAGGTGGTCCGAGCGGTCATTCA
CAGACAGAACCAGCAGAGGGCGCCAAAGCCCCACTTTTGACAAACTCCC
CTTCGCCCTGAGCCGAAAGTCCAGGCGGCAGGTGGACCTCTCTGCAGGGC
TCTGCCACCCCTGCTGCCGCTTGCCAGCACTCACAGGGGCTGCGGGGGGT
GCCCAACAGGCCGGCTACCCTGAGCTCTGGAGGCGATGGAGTTTAGGAG
GGAACGAGGGGACTCCTGGGGGTGACTTTCTTCAGCGCCCACATTGCGGC
CCAGCAAACCGAGGCTGGAGGAGGCCGGGCACCTGTGCCCAGCTGGAGC
CTTTGCTGAGGGTCTCCAAGGCCTGGGGAAATTGAGGCTGGGGGCTGGGG
GGTGTCACTGTCGGGCCAGGAGGCCCCTCGCTCTGATTGGAGCCGCCTCG
GCCACTTGAGCCAGGAGGCTCACATGAGGCGGGGGCTGCAGGGACAGGA
CCCTCGGGGCCCGGGAGGCCTTGGAGGGGGTCCAGCTGGGCCAGGGTTC
GTTCTTTCCCGGGTCCATGTCCACCGCCCTCCCGCTGCTGGGAGGAGAGG
AGGTCCAGGGCAGAAAGAATGCGTGGGGATGGGGGGGTGGTCAGGGGTC
TGGGAGCTGTGGAAACAACAAACAGACAGCGAGGTCCTGGGGCGCCCGG
CCCCCCGCCCCCTCCGGCACTGTTGTTTCTGGCCGGGGTGCAGGGACAGC
GAGGCAGATTCCTTCGAAAGTGGAGACTGGCGGGGGGCCCCTCGGGTCCT
CAGCTCACCCCCTGAGCTAGCCCGCCCACTCGGCTCCAACCTCCCGCAGG
CCCCTGGCACGGTCTCCAGGAGTCCACTGAGGGGTCCCCAAAGCTGCCAC
CAGGAGCTGGGCCTGGGTCTGTCACCACCCCACCCCACCCTCCAAGTCTG
AGATATGGGATCTTGAGGGCACCCTTGGACCCCCAGGGGTGGGTGCGGGG
TGCTGCCCCCCAGTCTCCCTTCCCCCCAGAATGCCCAGAGATACAGACCT
GGGTGCAGGCCCCGGAGTGAAGATGACAGGGTGGGACCCAGCCAGGACG
GGTTATGAGATTGGCCAGGGAGGGAGGTGGGCCCCCAAACCTACAGCAT
GGGAGTCCAGAGCCTGTGGGTTTTGGGGGACCCCACCCTCTGCAGCTCAC
GTGGATCAGGCTGCTGAACCATCAGAACCAAGTGGGCCTGGGGGGGTGGG
GGGGACGGGGCGGGGGACATTCCACCGGCCTCACTCTGACCCCCTCCTCA
GCCCATCTGTGTTTTTGGAAAACAGAGCCGGCGCCCAGGCAGCCCTGCAG
CTTGCGGCCGCATCCACACTGTGTCTCAAACTACACGCCTCCCTTCTTCAA
AGCCCAGCCTCGCTTCAGAACCTGGCCCCGTTCCCACACAGCCTCCCTCG
GGGCACCCTGCCTGCCCCTGCTGCACCCAGGGCCCACTTGCTGGGGACCT
TGGAAGGACCCTTCCCTCTCTGGGCCTCGGGTTCCAGGCTGTGGGGCATG
TCCTCGTGGTAGCTCCATGCGCAACCAGAACGGGGCTGGGATGAGTGAAG
CTGGGGGCTCAGGGTCGAGCCTGAGGCTCGGAACTCATGTTCTGGCCTCC
TTGGCCCTCCCCGGTCCTCCCCGGTCCTCCCCGGTCCTCCCTGGCCCGCCC
TGGCACCCACCCGAGGTCGCACAAGCTCGGCCTGTCCAAAGAAAGCCTCA
AGGAGCAGGGGGAGAGCGCGCGTAGCAGGTGGAGGCCTAGCAGATGGGT
CTGTCCTTGGTGGCCTGCTGGGGGCGCTTCCTTGAGCAGGTCCTGCTGTGG
GGACAGAGACGGCAGAGACGCCAGGCTCTGTGTGGGAGGAAGAGATATC
CGGGGACCAGGCGGGGCTGAGGGCCCCCTCCCTCCGGCTGTGCAGCTTAG
AGGAGGTCAGATGGGCCGACAGGGAGGCTTCCTGGGGGAGGCGTCCTGT
GTCCCCATGCAGGCTCCCTGACCTGAAGGTCAATTCTGCATCTGGCTGTCA
TGGCTGAGGGGGCAAAGCAAGCCCCAGGATGGCCATGTTCAGATAGCGC
CTGTGCCACCTCCCCTGAGCCACCCTGGGAGCCGTGTTCATGCAGAGATG
GCGTCCTCCAGGCCCTGGCTGCCAAGGAGGCCCCGGCCAAGTTCTGGAGG

ACACAGTGGGTCACGTCCTATGGGCGTCTGTCCTCCTGGAATGCCCCCAG
CCTGTCGGAGCCAGCGAGAGTGTCCCTTCTGGGCCACTGCAGGCCTGAAG
GGTGAGAGTTGGCGGGCCCGGAAGCCAGGGTTTAAGTCCTTGCTCTGTTG
CCCATGGCTGTGTGGCCTTGGGAATGTTACTTAGCCTCTCTGGGCCTTGGT
CCCCTCATCCATCAAAAGGGAACCAGGAGACTCCCCCTGCCTGGGTGGTT
GATATGTGAGTGGGGACCCCTGCCAGGCCCGCGGCCTTGGGGTCAGCAGG
TCCTTCTCATCAGACAAGCAGGTTTCCAACTTCCTTTGTGACCCCGGCCAT
GACATGCAGGTGAGGAAGCCTGAGGGAGACCTGCAGCCCCGGGCGCTAC
TTGCAGGGCCTCCCTGTGTGTCTCCAGCAAGCTGCTTCCCCTCCCTGGGCC
GCAGAGCCGTGGGGCAGCCCCCGGGCCCGTGCTGCTCCAGGCAGCATGG
ATGGGGCAGCTGATGCATCTTCCGTGGGGAGAGGAGCCGACTCCGCGAG
GACCGGGGCAGCCCATGGGCATGGGGTGGGGTTGCCAGGCCCCTGGCCA
CTGGGAAGGCCTGCCTGGTCTCCCTCTCCCCAAATTCCAGGGCCCTCTCAG
GTGGCAACTTTTACCTGTGGGGGATGGTGGGGGATGGCAGGTGCCTAAGG
ACCAGCTACCACCCCCATAGGCTGCCTTTCCATGTGGCCCCAGGAAAGAA
CGTTCTGGTACCCCTGGTCCTGGGAGTCAGGGGCAAAAACACCTTTGCTT
CGGGAGAGTATTACGCTGAACGTGTGTTCACGTGTGTCTGACGGACGCCG
ACCCACGGGCATCCGGGCCCCGACACATTGGGCCTGTGAGGTGGGAGGG
GCGGGGAGGCCCAGAGGTGGCTGGGCCCACCTGGGACGTGAGCCGGGTC
AGCCTCCAGGCCTCCTCCTGCCCAGTCCTCTACCCGTCTGGAAACACTGCC
ATGTGAGCTGCACAGCATGAGCCCTCGGCCCCACTGCTGTGGCCTTGCGG
ACATTGAGGTGTGTGCCGCCCAGGGCGACCACACCCTGGCCTCTCAGGCT
GCCCGTACAGAGGCGGCTGGGTCGTAGGAGGTGCGGGGCTCTGGGGACC
GCTGGTGAGTTCAGGACGGGGGTCATGCCACCTCCTCTCTGAAGGTTTGG
TGAGGTGGCCCTTCTCTTATCGTGATGACAATACTGATTTCTGGAAGAGCC
AGGTGTTTTCTGAGGCTGTGGTTGCACTTCTCCACGTGGCCACAAGGTGCC
GGGCTCGGGTCAGATTTGAGAAGCCCTGCGGGAGCGGGTGTCATGCGCCA
GATTCAGCTTGCCTCCTGCGGGTCTGGGGTCAGGACGTGGTCCCCAGCAG
TCTGCTCCAGAGCCTGTCAGTGATGTGTGGGATTTTACCGCTAGAACACA
GTTTCCTCTGATTCTCAGAAACCAGCAGATGCTTTAGGAGGGCGTGCAGG
TTTCACCTGTGCTGCAGGCAGCCACCCTGCCACACCCCTGGGCTCGGGAG
CCCCCAAAGACAGGCATCCCTTAGAGATCAGTTCCTCATCATCAGTTCCG
CAGTGCTGGGGTGGGGGCAGATGAGAACCTCAGGGCTGGGCGCAGAGGT
GGGGAGCCCGCCTGGACCCCGACACTGCAGGGGGGCCTCCCCCTTGTAGG
AAGAACAATGTCGCTTTGCCACCCAGCCCTCTCCCCAGGGTGCCCCGAAC
TGTTGCTCCTAAGACCTCTGGGCTGTGTGCTGTAATTCTATAAGTGGCCAC
CAGGTGTCAGCAGGAGGCCACTTAAGCATCCATGTGGCGGAAACCTGGA
GCTGGGGGTTCCTAAGGGTCCCTCGAGTGTCTCCTGAATAAATAGGCGCT
GACCTGATCCCCAGGAAGGGATAACCCTCTCCCAGGCCTAAGAGGCAGTG
GGGCAATGAGGTTTATGTGTCCACTGTACCCCCAAATTGTCTCTTCCTTCC
CTCTACCCTGTGTCCCCACCGTGGACGATACACGGAGTGCGAGGCTGCGG
GTCACAGCCCTCACAGCCCCAAAGCTGCAGGTCCTGCCTCAGGGGCACCG
CAGCTTGGCTGGTCCCCCTTGGGTCCTCCCCACCCTGACCCGTCCTCTGCT
CCCCTCCCTTTGCTTAAATGCTCTGCGTTTCAAGGTTCTGATGGAATAAAA
TAGCCCTGCACTGGTGTGTTCCTCTTTGGGGCTGTGCCAGAAGTGGGAATT
CAGACCAGGGCAGAGCTCAGATTCCACATACTGTGTTAGGGATGGCAGGT

GCCACATTTCCAGGAGTTTCATTGGTGGTTTGTAAATGCTACTTCCGTTTC
AGCCCCTCAGCTGCCCACCTCCTCAATTTAGGGACCCCCCCCTTTGGCGGG
TTGCCCATGGAACCACATCATCTGGCGTGGGGTGAGCCCTTTATCCTCCCT
GGCCCCACTGGAGGTTTGGGAAGTCCCAGCTAAATTTCTCCGTAGGGACC
TGGAAGGAGCCCTTGTGACATCTGGGCACAGATAAGAGGTAGGGGGCAC
AGGCCGTGAACACTTGAAGCTGCAGAGCCCAGAGCAGAGCCAGCAGGAG
CAAGTGACTGCTCCCCACCCCAAGAACTGTGGGCTGCGTCACACACTCCC
CACTGTGTGCCCTGGACCTGACAGGGCCTTTAGCCTCCCTGCATCCCTCCC
CACCCAAGAACCCAGTGAGGCACCCCACTTGCCCCTCCTTAGTGTTGTTAT
GGCTCTGGGGCATCTGCATTTTGTTTAGGACACCCCCAGCTAGATTTAAGT
CCCCCCAAGTGTGACTCTTTCCTCCACTGAAAACCCTGTCCTCCCACCAGA
GGGCCCTATCCCTTTAGCTGAGCAAGGAAATTCAGGAGGGGCCTTGAATG
ACAAAGGAAGAGGGGGAGAGTTAAACCCCAACACTGGCTGGCAAGCTGG
GTGGGGTGGACACCCCAGGGTGCAGGGGTGCAGTGAAGGTAGCGGCTGG
TGGCCTTCTGGAAACTACATGTGACTTTGCCATTAGGTGAGTCTTTGCTTT
GCCCCTGCTCTATCTGCAGGCTTATGGAAGAAGTTTAAATTCCCAGGGAC
ACTTGGTCTAACCAGGCAGCGCTTGTATCTGGGCCCTTCCCCAGCTGCTGA
CCACTCTGAGTCTGCGCCTTAGTTGGAGTTTTGGCCAAGCTCAAGAGGCT
GTGGACCCCAGTCATCCCACCCAGGGGTGCCTGTGGGCAGGACGCTGCTG
CCTGCCATTTGCTGCAGTATTGTCACTGTCCGGCACCACACACATGGTGCA
GGGGGTGGTATCAGGTGCCACTGGGGAAGGGAGAAAACTCCCAGGTGAG
TCCCCTGCCTCTGGAAGCAAGATGGACATGACCGCACTGTGTTGCAGCTG
CATTGGGAGGCCCCGAAGAAAGATTTTTCTGATCTTTCTCGAACCCTGCTT
TTCCCCATCATGCCCCGCCCCCATTTTACCCGTGCCACGCCCACTGGTGTG
CCGGGGTGTCAAGTGACTGACAAGTGTCAATCTACTGAGGCCCTGCCCAC
TCTCCACCCCCCCACATAGTCCCACCTCCCAGCTGGCAGGGAGAACTTCC
AGCTAATGCCCATGCCCACAAATGTCTTTCTGTCAGCCTAGAGCTGGACC
AAATCTCCACCCTGTAACATGCTGTGCCCTGGCGTGGGAAGGTGCCAGAG
CCAGTTGCCCCAGCAGCCCCAGAACCACTAAGTTGGCACAAAGCTACCCA
AATTTGGAGGGGCTTGGGGAAGGGCATGGAGGGGATGAGGAGGTGAGGG
GCAAAACTAATTTCAGTTAGCATTTGAGCAGGTGCCACGCTCAGCGTGGA
GAGGCTCTCTTGCTTCTAGGGACCCATTATGATGCACACGCTAAAAGCGC
CCTTCACCATCTCTCCAGCCTCAGCTTTGTCCCCCTCCTCCTCCTCAGCGG
CAACCCGGCTGGAGGGTCTGGCCACTACAGCCAGAGCGCCCCCTACTTTG
GTGGCGACTGCTACTATTGGCCCAACCAGCGGATCACCGGCCAGGCAGTT
TCGGCAGAGAGTCTGGGGCACCAGTGACTCCCCCGTCCTCTTTATCCACC
ACCCAGGAGCTTCAGGGACTACACAGCGACTAGAGGGCAGGTAACTGGT
CTGCCCTCCCTAGGGCTGCCCCCTCAGAGTGTGTGAGAAAAGCTGCATTG
AGTGTTTGGGTGCAGGTGGGCTGGGGGCTTGGGGCAGCCAACAGGAACG
GCGGGACCTCTGCTTCCAGAGGACCCCAGATCCTGGCAAGCTTCGACTTT
GGAGGGGACAGGAAAGACAGGTGGAGAGGGGACACTTCCCTCTTCTGTA
CAGACGCCCACCCGGAGCCACAGAGGCTTTTGCAAGGAAAATAGGTTTTC
CCTCACTAATGCAGCAGGCAAAATGGGAGGGGCAGGGGTGGAGGGTAGT
GCCCCCGCCCCCAGCAGGAGGGCACAGCTGTTTCTGCAAATGTAAAAAAG
CAGGGTTTTTCTGTGTGAGAAGTTCCCTCTTGCTGCATGTCCCCACCCCCG
CCACCAAAGACAAACAGGACACTGTGCAGAGGGGCCAGAGCCCCGAGAT

TTTGGAGTTGTTTTTATATGCATATATACCATTTTGAAAGCAAAGCTTCCC
TCTCCCCTACTCCCTACATGTCCCCCTTCACCAAAAAATCCCACCACGTAA
CTGGAAAGGGGAGTGAGAAGGACGACGAAGGGGCACTGTCCCCTCCCGT
CCCACAGCGGGACTTAAAACGTACAGCTTTTCGCCTCCGGACAGTGTGCC
GCCCCCTGGCCCCCGTCACGCTCCCCTGCCCGGGGGGCTGAGTGTGGGGC
CAGGGCCTGTCTCCAGGCATGCATTATTTTGTGCATGAAGGTTTTGTCCCG
CCCACCCAGGCTGGTGTTGGGGGGAAGGGGTTCATTGCTCCAAAGAAGCC
CATCTCCCCCCTCAGCCACCTTCAGCCGCCTTCGCAAGGCAGAGCTGTGTC
CTCTGCTGTGTGCCTGGCCCCCTCCTTGCTTCTATTCAAGGTGGAAGTGTT
GGGGGGAGGAGAAGAGTTTTTATATTGTGTCTGTGATCCCCCGAGGCAGG
GCATTTGTGTGCGGCCCCCCAGCCCCCAGGCCCAGGCAGATGGGCCAGCC
TGCCCGACAGAAGGGTCTCCTGCTGCTTGGCTGCAGGGAAACCCAGCTCT
GGGTGAACCGTGGGCACCTTCCTTCCTCCATGCCCTGTATTTAAAGAAGG
AGAGCTGGGGGGCCAGAGGCACAGGGAGGGGAGCCACGGCCCCAGGTCT
GACAAGATGACCTGCGGGCCTCTCCACCCAAGAGTCGGGGTGGGGGGGC
GGATTTGGTTTGAAAAGAGAACAAATAGGAACACACTCTTTATTTTCCCC
AGGGGCCGAAGAGTCACCCCTGAACTTGAGGACGAGCAGCCGGATTCCA
GCCCCCAGCCCCAGGGCCCCACATCTCCTCGGGCTCAGCCGCGCGCCCCA
GCTGCCCCCCCAGCCTGAGCTGCAGCAGGCCAGGGCTGCCCGAGACCCCAG
CCCCCAGGTGAGCTGCTGCAGCCTGTGGCCCAGGAGATCTCCGCCGGCTC
AGAACTGAGGCGGGCAGCCCACCCAGCCCACAGCGGTGAGTGTCTCCAG
ACCCCAGGGCAGGGCCCGGTGTCCCCCGGCACAGAGAGCTGTGCTGCAG
GCCCAGACCTCCCAGGCCGTTTTAGTTCCCATCTCCCCTTGGGGGAGGGG
TGGGGCTCAGAGGGGCTGGGGTGCATCCGCAAAGCTGGGGTGCAGGGCT
CCAAGTGCCTCTCTCCCAGGCGGCTGGCTCGGAGGGGGGCTGAGCAGGG
ACAGTCGGGGTCCTCAGCGGGGAGCTCCTGGTGGAAGAAGGAGGGGGAG
GGGGTTGGTGGGGGGGAGGCAGGGGGTCCGGGTGAGGGCGGGAAGGGG
AGAGAAGAGGGTTCCAGGGACACTCCTAAGCTGGGGAGACTCCGGGACT
CCCAGGGAGGCTGGGCCCAGCTGAGGCAGGTGAGACCCAGTCCTCCCTGC
CCTGGGCTGCTGGCTGTGGGGCCTCACACCAGTGCTGTGGCCTCTCTGTGT
CCTTATCTGTCAAAGTGGCCGGGGTCACAAGCCTGGCTGCTGATCCCCGC
GAGCAGCACGTGACAATGACAAGGCCTAGAGGCTCGGGGGCCTTACGAG
TCCCTGTGTCATGATTCCACTGTGTTGAGCCGGCAGGGCTTCCTGTGTGTG
AGTGGGGGGCGGGGGGGGGCGACTAGAGAGGTCCAGGGGCCCAGATTCC
TCTCGGACACCTGGTGGAGCAAGCTCTGTGGTGTGTGCCCCGCCCATCCC
CTGATAAGGTATTGGGCTGGGAGCGCAGGCTGAGCTGCCCCAGAGGGTCC
CGGAGAGCTCTGGAGGCGGAGGCGTGGGCCGTGTCCGTGCACCTGCCCCG
CAGCGACAGAGTCCGCTCTTCAGGGCCTGCAGTGAGGGGCCCAGCAGGT
GTCCCCTCGCAGGCCGGTGTCGCCTCCGAGAGAGATGCCAAGAACGTTGC
TTAAAGGTGTCTTAGTCCAGACTCTACATTAGATGGTGCAATGGCGACGC
CTGTGTCCTCCAGGAGAGCAGGGCTCAGCCCCTCTTCTGGGGCCCCCACC
CCGCCCATACATACTTCCGACTGACTTCTGGTTTATCGATCACATGGCGTG
GGTGTCTTGTGCCGTCGGAGTCCCCGTGTCCGGCGGGGGGCCGGGGTCCC
TTGTTTCCAGCCAGGGTGGGAACAGGGTCCTCACCCAGGCAGCAGAGCCC
TGCAGGGGTGAGGGAGGGACGATCAGGGGACTCTGAGAACGGAGCTGGG
TTTGTCAAATGCTGGGCTTGGTCCCAGGGAAGGGGGGAGCGGCAGCTCGG

AGCCAGAGCCGGGGGAGGAAACCGGGGAGGGGAGAGGCAGGACGAGCA
GAGAGGGAACTGGCGGAGCAGGCGCGAGGCGGCCGCAGGCCTGGTGCAA
GGGCCAGGCTCGCGGGGTGAGTCCGAATCTGGGGGGACGCTTGTGTTGTT
GGAATAGGGTTTGTGGATGGGGGGGTAGTGCTGGAGCAGGGCCGGGGTG
GGGGGGTGGGCCAGGGGATGGTGCTTGGGGCTGAGGTTCCTGCCAGCCA
GGCCGCAGCGGAGACCCCCCAACCCCATAACATGCTTGGGTTGCTGGCAG
GTCCCTTCCTACTGAGGGGCTGGAAGCTGGGGAGGCCCGGGCCACTGGTC
CTGGCTAGGGGTGGCCAGGGCCCTTTCTGAGCCCCGCAGTGACCTGGATC
TGCCCCTGACAACCCTGCCCTGTTAGCCACACTCGCGACTAATAAGGCGA
GAGGTCAGCGGGCAGCCCCACGGGGAGAAAGTGCCTCCGTGCCCCCCAC
CCCTGGCTCTGATGGCCCAGCCTGGCACCCCAAGGTGGCCTCGGCCTTCC
TACCTCCAAGGTCCAGGCGCATGTCCAAGCACCAGCAGAAGCTTCTCCAG
GGTTGGTGCCTGCTCAGGGCAGAAAGCAGGGGTGAGGCTCCCCAAAGGG
CCACTGGCACCAATGCCCCCAGGCAGCCCCAGCGAAGGGGACAGCCCAC
CCCCAGCCCGGGGACGCAGGCCTGAGGGGACATGGGGAACCCAGAGCAG
GGCCAAGGGGAGCAGAGCCCCTCCTCCGGGACTTGAAATCTTTCCCGGGG
GGCCCAGGGAGCTGGGGTCTGCAGAGGGCACTTTCAAAATACGGCCCAC
CCCCAAATTGCCACGTGGCCACAGAGCAGGGAGTCGCTGCAAAGTGGCCT
GGCTTCAGCGCAGGAGGTCCCCTCCTGGGCCTCCCTCCTATAGGCACAGG
CCTGAGTGCAAGGACGGGGACTGTGGAAGGCTGTTGGGGAGGGAGGGGG
GCGGCGATGGGTGGGGGGAGACAGGCCATATGTTTTTGGAAAGGGACAC
CGGGGTGACTGATGCCGACTGGAATTTGAAGGGGGAACATTCCCCACGTG
CCTGGGGCTCTGGGTGGAAAATGGACGGTTTTTCATGGGGGGGGGGGTCT
CCCTGTCCTGCCAAGCGAATGTGAAAGGCCTGCCTCATCCCTGCCCGGTC
CCCACACCTGTCCCATGGCATTAACCCCCGCCTTCCCGGAGAGTCAGGCT
TTAAGGGGACCCCCTGCCAGCGGGGTCCCTCTGTGGGGTGACAGTGGGGA
CAAGCTGTGTCCAGTGTACTGGCCATAGGGACCTGCCGCTCTGCTCCCCA
GCAGTGCCCCCTCCCTGGCAGCCCCTCAACCCACAGCTGCCCCTGGGGCG
GCTGGCCCCTGCCCCCAAGGCTGCCCAGTGCCGAGGGGCTGGCCAGTTCC
TGGCAGCAGTGCCTTGGGGCTGGGGAGGGTTCAGCAGTTGCCTTAAGGAA
CCAGGTTTTCGCAGCCGGGAGAGGGAGGCTTTTCGAGAAAATTCCCTTGG
AGAAAACGAAAGGAGACGCTTGACCCCCCGGCCAAGGGCCGGTCCTGCC
CCCATCTCCCCCCCCCCATTTTCTCCTCTCCTTTCCCAGTCCTTCCACTCTC
TCAATTCCCCAAGCAAAACTGGTTTCGCCCTCCTCCGGGGACTGTTGAAG
TCCCCGAGAGCGCCCGCGGAGCGCCGGGGCGAGCGGGGGTGGCCGCCGG
GGGTGCTCCCGGGCCCCCGGACCGAGCCAGGGACGAGCCTGCCCGCGGC
GGCAGCCGGGCCGCGGCTTCGCCTAGGCTCACAGCGCGGGAGCGCGTGG
GGCGCGGCGGCGGCGGGGAGTCCGCGGGCCTCCTCGGAGGCGGCCGACC
GGGGAGCCTGGGGACCCCGGAGCGCCCGGGGAGCAGCGCCCCGACGCGC
CCCGGGCCGCTCTCGGCTTCCTCCCTTCCAGCCGGCGCCCGCGCGGCCGG
GCTTCGGCACCGGGGCGCTCTCAGTGGCAGGAGAAGCGTGCGCCCCGCG
GGGTGGGGGACCCGCAGGAAGCCCCGCACCGCCTGGAGCCGCCGCCGCG
CGGCCAGCGCTCGCGTCCCCCGGGGAGGGCGCCACTGCTCCGCGCGCGCG
TCCCCCGACGCCCCGCGCGCTTCCCCGGCCGGCCCGGGATCCTAACCTCT
CTCTCGGTCGCAGCCCCGCATCCCCAGGGCTCCAGGCCCCCGGCGACTTG
CCCGCTCCTCCCAATTGCAGACACGACTTTTTCTGGGACCTCCCAAAGGA

CAGCCTGGCTCCAGGGTCCCCCAGATACATTCACCATTTCTCCAGATCAC
AAGTGGGTTTTTCGGGCACTAACTTCCAGAGACCTCAAAGCACATGAGCC
CCTACTGGCTTTCCCAGGTTTCCACTAGTGGCCTCGGTCCCCACCTCACTG
GGGATTGTCTCCCAGGCTCTTCGCGGTGTGATCCCACCCATTCGCGCCCAG
GTCCCGCAGTGCCAATCCCTCCTCTAGAAAACTTAAACACTGACTCCTGG
TCTCGGGGTGAGGCTGCCCAATGTGCCTGACTCCCCAGAAGGTATACCAG
TGTTTTTCTGGCATTTGGGCACCGTTCCCCCAAAACACGTGAAGCTCTTTT
CCCGCGTCCCCATAATTTTGGACGCCAGGGGCACCCAAGCTTAGCGCCCC
TGTTTGGCTCCCCCACACCGCGAAGCCCTGCTCCCTGGGGTTCACGACAG
TTTGGGACTTTATCTGCCAAGTTCCACAAACTGATTGGCCCCAAGCTGGG
GTCCCTAAATTGTACACAAAGAACCCCAGCCCCCCCCCCCAACTCCAGTA
CAGGAAGCGATGGCCCCAGGGACCCTCGGAGTTGGAACGTGGCTTCCTAA
GCCTTCACCAAAATTGAGGCTTTCCGCGCATGGCGCGCTGATGCCCTTGCT
GAATCAGAAGCACTCTGCCCTCTGATTCCTGCTTTCCACAACCCTGAGAG
CATGATTTCTGGTCCCCCAAACTCACTGAGCAAAAATCTTTTTGTGGGGGC
TGCAAAGATAGGAGGCATTTCTCTCCGGAGCTCTCCAAACTCCCTTGCCT
ATAATCAAGTTCCCTAAAACTTAGACAGAGCTTCCCAGGCCCCAGAGGCA
CACAGAGCCATTATTGGAGCTGCGTTTAATGATGACAGGGACCATGGGTC
ATGCAGCTCCCCCAAGTCACAAATGCCCCAGGTATCCTTGGCTCCAGCCA
AGCCCAAAGCAAACTCTTGCACAGATCCCATATCTTGTTATGTCAAGCGC
TTTGCGTGTCCCAGTAAACAAATAGTCTGAGTGTTTTCTCCACCTCATAAC
ATTCGGAATATTAAAAAATTCCCTGGGCCCCCGGAGCTGACAGACAAGAA
TCCGGGCTTCCTAAAATTCAGAACTGATTCCCAAATCCCAGGCCAACGCC
AGACCCTCTCCCAATCTGGAGCCCCTCCGACTGGACACACTGGACTCCTA
AGTATTACGCGCTGTCCTCCAGGCACCCCAAATGCATTCAAAGTGACGCT
TTGGTCACAGAAAGGCACTGATTTCTTGGGCTCCAAAGCAGCCCATGCAC
CCCCGAGTCACCCCAAACTTAGTCAGCATTTCCCGGGTCTCCCTCCGCACT
GCAAACTCCCAACTGCGGACACCGGTTCTTCAGGACCCACCGCCTAGACG
GTCTTAATCCCTTTTCCCCCAGACCTAGATTCGGATCCTCAGTTCTTCCCG
AGGACTTCTTCAAGAACGTCTTTGGACCCCNCCAGATTCAAAAACTATTTT
TCTGGGGCCTCCAAATTGAGGTGCTGCCTGCCAGTCCTCCAAAATAAACT
GAGGGGTTTTTTGTTTGTTTGTTTTTTTGTTTGTTTGTTTTTTTTTTACCTTC
CACGAAACAATCCAACTTTTTTGGACCATTGATTTATGGGTCCCCTGACTT
TATGACCCTTGCCCCAAGTCCCCCTAAATGTAGGCCATTTTCCACGGGCCT
CCCAAAATGAAATTGCCCAGATCCCGCCGAAAAAAATATCCCCGGGTCCT
GGAAATCCCAGGTATTACAGGCCTGCGGCTGACACCCCTCCTTGCTACTA
ACCAGGTTCCCTGAAGTTTAGAGATCACTACCTAATGAACAAATCCACGG
ATCCCCAAAACTGGGGCCCTATCTTACTAGGGTTCCCTAAATGCAGACAG
CGCCCGGGAAAATAGGGGCGTTTTTTTTCCTGTTTGCCAAAAATAAACTA
ATTGAAACCAATTTTTAGAATTAAAATCTAAAATGACCTTGATTTTCTGCG
TTCTCCAAATGTACTTTTCACAGCCCAGGTTGCCCCCAGTTTAGACGGTGT
TGCTTGAATCTCTAAAGCACCCTGAGGATTTTTCCCGAGGAAGCCACCAC
AACTACGGAATTTACTGTCCTTCGGGGCCACAAGCCTCCAGGCCACCAAC
TTGGATTTCTAAACCGTGGAAATCAGCCTCCACTTCCCTCCGCCACCCCGA
GGGTCTGCTCAGACCCCCCAAACGTGCCCGCTGTTCTTCTCCCCCCAAATT
TTATTTAGAGAATATGCCTCTCTCGGGTTCTGCCAAGTTTCCCGCTGAGAC

TTCCTCGGTCATCCCCAAATCCTCTTCCCCACAGTCCGGGAGCCCCCACAA
GCTTACCGACCCACATGCTGGGGTCCCCCAACTTAAACGCGATCCCCTGT
CCCCCAGATTCACCGAGTGATTTCCCTGGTCCTCAGACTGGGACTCTTTTA
CTGGAGTCTCGAATTTAGCCATTAATCACAGTTCTCCACTCCGACGCAGG
CTCCCTTGGGTCCCCACGTCGGGGACATGGGTTCTCTTGCCTGCAAATCAG
GCTGCTCTGACTTGCATTCAGGCCTTTGGGCATTGTTCCCCGCCCGCCGCG
GTCTCGGTTCTCCCCCCATCCCGCGCACGACGGGCACTGGGTCTGGGCCT
CTTGGTGTCTCCTACAAGTCCCCGGAGCTCCTCGGACTTGGGAACTGTCTC
TTGCGTTCCCCAAATACACTCGGCCCGGCAGTGTGTCCGCCAGGACGTAG
GCAGAGCTTCTCCCGCGTCCAGGAAAACGACTGGGCATTGCCCCCAGTTT
CCCCCAAATTTGGGCATTGTCCCTGGGTCTTCCAACGGACTGGGCGTTGCC
CCCGGACACTGCGGACTGCCCCCGGGGTCTCGCTCACCTTCAGCGCGTCC
ACCGCCCGCTGCAGAGCGCTCGCTCTCCGTCTCTCGGCTCCCAGCGCGCTT
GGGGACGCAGCCTCCGGGCCTCCAGCCTTGCGGTGAGCTCCCCGTCGCCT
CGCGTGTCCCGGCCCGGCTCCCAAACCCACTCGCCGCCGTCCCGCTGGGG
CTGGCACTGGCCTCCGGCGACTGCCGGGGACACGGGAGCGGAGCGCGGG
AGCCTGCTGCAGGCCAGCCCGTCGGCCGGGCCGCGCGCCCTGAAACGCGC
GCGGCTTTCGTTTGCTCTTTGCAAAGGTCACAACCGTGGGGAAAACGCCT
CGGCGGCCCCCAAGCGGGGCAGGCAGGGCGTTGGGAAGGAGGGACACGC
GGGAGAGGAGCACCCCGCTGGGGCGGCGCAGCGCGGCGCCTCCAGCCGC
CGGGCGGAGGATCCCGGGAGGCGCGCGCGGAGCGCGGGCGAAGTGATTG
ATGGCGGAGCGAGGGGGCCAGCGGATCGCGGGCTTCCGCCGGCGGCGGC
CCCTTCCCCTCGGAGGGACTCGGGCGGCCGGGGTTTCTGGGGGCGGGCGG
GGCGCGGGGGCTTGTGCGTGGTCTCCACTTGGTAAAAATCACAACGACTT
TTTACGTCGCCCCAACTCTCCAGGAGATGGTTTCCCCAGACCCCCAAATTA
TCGTGGTGGCCCCCGGGGCTGAACCCGCGTCTACGCAAGGCCAACGCGCT
GAGGACGGGGGAACCATTATCCGGATATTTTGGGTGGGCCCCCAAAGCGA
GCTGCTTAGACGCGCCCCGGTGAGCTCGGTCCTGCAGGTAGGCTTGGAGC
GAGGTTCCCCGCCCTGCTCCTCTCTCTTCGGGCAGGCGCGGCCAGGCCGG
CCGGCCCTCCCCACGTACGGCACCTGGCGGCCGCCGAGACGACTCCCCGG
TTCCCGCGCGGCACCGGGGGGCGCTCGGGCTCTGGCTGCGGCTCGAGGCG
CTGCGCCTGCTCGGGCAGGTGGAGGCTTCACGCCGGGCCCGCGCCCAGGG
ACGACCCCTTACCCCGCAGGTCCCAGCGGGACTCGGGGCCCCCGGATCCA
GCGTCTAGCCACCTGTGCCCGCACCGCCGCGAGGGCTTGTGACACCTACC
ACCCTGGCCGCCCCGCGTCCCCCCGCGCACGAATGTAGGGATCCTGACAC
CCCGGAACCTAAGACGGGGCCCCCATACACTTTCGTACAGCGATTCGGGA
TTTCTCTCGAACTCTGCAGATCCCTGCCCCCAATCCAATCTCGCACCCGCG
CACCCCACAGCCCCCTCCCCGCCGCCCCCCCGGACCCCAAGCCCACCCCC
CCCCGTCTCGAGTTGCGGGGGCGGTCCGGGGGCGGGGCGAGGCCCCGCG
CGCCCATTGGCGCGGGCGCACGGCCAGCGGGGCCTGACGGGCGCCGCAG
GGTGGCGCGGCTATAAGAGCCGGCGTTGGCCCCGGAGTTCGCCTGCTCTC
CGGCGGAGCTGCGTGAGCAGGCAGCCCGGCCCCCCTTCGGCCGCCCCCGC
CTCCATGTGGCCCGACTGCCCGCCCGCGGCTCGGCCCGCCAGCGCCTCCA
CCCGGCTGGCGGCCCCGCGTACGACTGACCGTCCGTCGCTACGCTGCTGA
CTCCCATCTCGGGCGCCGTCGGCGGGAGCCGCGCTCCGCCGGCCTGCGGA
TCCCCCGCCGCCTCCTCCTCATCTACCTCAACGCCCGTCCCGCTTCGCCCG

AGGAGGCGGTCCCCCCCGCAGGCAGTCCGGCTCGCAGGCCGCCGGCGTTG
TCATCCCCCGCGCTCCCCCCCAGCCCTCCCCGGCGCGCAGCCTGGCCGCTC
CCCCTCCTCGCTGCGTCCCGAGCGGCCCCGGCGCGCCGCCACCGCCCCCC
CACCCCCCGAGGCCCGGGCTCGCGACGGCCGAGGGCTCCGTCGGCCCAA
ACCGAGCTGGGCGCCCGCGGCCCGGGTGACGCCTCCTGCCGCCCCCCTCA
GCACCTTCCCCGGCCCCCGACGTTGCGCCCTCTCCCTCGCTTCTCTGCGCT
CCCCAGCCCCTCTCCGGGCCTCTGGTCCCGGCCCTCTCTTCTTCCGCAGCC
TTCCCTGTGCTCCCTCCCGCCCCCCCCAGCTCCTTGCCTCCAACTCCCTCCC
CTTCCACGCCCGCCCTCTCCGCTTCGCCGTCCCAAAGTGGATTAATTATAC
GCTTTCTGTTTCTCTCTCCGCTCTCCTCTCCCGCTGTGAGCCTGCCCGCCTC
TCGCTGTCCTCTCTCTCTCTCTCTCGCCCTCTCTTCGGCCCCCCCCGTTTCA
CGTTCACTCTGTGTCTCTCACTATCTCTGCCCCCCTCTATCCTTGATACAAC
AGCTGACCTCACTTCCCGATCCCCTTCCCCCCCCCAAAAGTACAACATCTGG
CCCGCCCCAGCCCGCAGACGGCCCGTCCTCCCCAAACAATCAGACGAGTT
TCCCACCCCCCCCAAAAAAAGCCATCCCCCCGTTCTGCCCCGTCGCACATTC
GGCCCCAGCGACTCGGCCAGAGCGGCGCTGGCAGAGGAGTGTCCGGCAG
GAGGGCCAACGCCCGCTGTTCGGTTTGCGATACGCAGCAGGGAGGCGGG
CGGCCCCTGCGCCGGCTTCCAGGTAAGCGGCGTGCGGGCCGGGCCGGGCC
GGGAGGGGCGGCGCGGGCCCCGGGCAGAGCGGGGTGGGGGGCTCTTGGT
CACATTCCCTGTGGCGAACCCTGCGCGCTCCCCAGCGCCGCGGAGGAGGG
GGTGGGGGGGGGCCGGGCAGGGACTTTGGGAAGTGCCAAGGGAGGTGTGA
GCTGCTGCGAGGCGACTTCCTGGTCGCTTTGCGGGTGCAGGGGGGTGTTG
CTGAAAGTTTGTGATTTGTGCCGTGCGGGGGCGGGGGGCTGTGTAATGAG
GAGGGCGCAAAGCTCAGAAACCCACCCTGGTATGTTGACTCGGTGCCAGC
AAGACGGAGAGGAAAACTGGGTGGGCGGGCCANGGATGGAGCGAGGGG
GGCTGAGTTGGCAAAAGAGTCCTGCCAGACGCCACATTTGCAGCGCCCCC
CCCACCCCACGGCCTCCCTTTCCCCGTCCCCAACATTTCCAACCAGGAGTC
TGGGGAGAGGGCCTGGTCCAGGGCTGGGAAGGGCAGGACCGTGCAACCC
CAGCGAGGGTAACTGCCCCTGCCCCAGTGGGCAGATGGAAGTTTCCATAC
AAGGAGGTGGAAAGCAGTCCCCTCCCTGCCCCCCCCCATTGCCCTGTGCT
GAGATGGGGGGGGGCTGCGTGCAGGACAAGCCCCCCGCCCCCATGGTAC
TTTGCTAGGAGATGGTGGGGCTAGCGGTTGGGGGTGCAGAAGAGAAGCG
AATGGCTGGGAGGAGGGAGGGGGCGGGAGCAAAGGGGGCGGGGGGAGT
GGTCAGTGCGAGAGGGGTGGGGGTAGGGTGGAGCCGGGGCTGGGAGGAG
CCGGCTCAGACATAAAAGCTGAGGCACTGACCAGCCTGCAAACTGGACA
TTAGCTTCTCCTCTGAGGCAGGCTCCCAGCTTCCTCCTCCTCCTCCTCCTCC
TCCTCCTCCTCCTCCTCCTCCAGCCCCAGCGAGCCTCCTGTCCAGCTGCAG
CAGGTAACCGAGGCTGTTGAGCACAGACCTGCTGCCCTCCTGCTCACCGC
TCGGGCTGCTGAGCTGCAGCCTCGGCCTCAGCCCCCTCACCCCCCACCCC
CCCACCCCCCCCCACCCAACGCCCGCCTCTGCCCTCCCTTGCCCCCCCCCAAC
CGGTGTGCTCACCTGGCCCGCACCTCTGCATGCTTCCTGTCCAGTCTTCTC
TCTGCTGCTGCGTTGGCATGGACCCCACGCTCCGTGGCTGGGGATGCGAA
AGGATGGAGTCTTGCTGCTCGGACGGCCCCGAATCACTGGCTGACCACAT
TTGTTCTTAGAGTTTGGAGGGGACAGCAGTGGGGTGGCGGAAGACGGGCT
GGGGGGGGAGGAGGTGAGGAAGGTGGCCCTCCGAAGGCCTTCGAGGAGCT
GGCCCCGCGCGTCAGGGGCTTACGGGAGATTGTGCCTCCTCCAGCCCCCT

CCTCTCCCACACGATTTATTGCATCAAAGTCCCCATGGGCCAAGCCCGCA
GCAGCCGGGAAGGGTCACGCTCTTTGGTGCCCCTGTGCTCACTGGGTGTG
CAGCCTGGGAGGGACACACGTGGGCAGGCCCTTCTTCCAGTCGCATCAGG
GTGCCCGCAGGCTCAGGGGCCGTGACTGTGTCCCTCAGAGGGGCTCTCAC
CTGAGTCTTTGCTAGGAGTCTGCAGGTGGGGCCTGCCTGCACAGGGGGCT
TCCGCCTTTGAGCAGCACAGGGGTGGCTTGCTTTGTCTAAAGAGAAGCCC
AAGTGCCTTCCCAGGCCCTCAGCCTCTTAAGGGGCCTGGTTCGCTCTCTAG
CTCCAGGCCTGCAGCCACACAGCCTTGCGAGGGGGTGTGTCCCAGGCGGG
GCCCTGAGGGACCTGGGGCCTGGCCTCCAGGTAGCCTGACCTGTTCTCTC
CAGCTCCCGGGGTGCTATTTGGCACCCCAAGTGGAGAGGGGGCGCTGGTG
GTCCTGGAGGGACCTGACTCTGGGTGTGGTTTGAGGCCGGGAAAGGAGGC
TTTTGGTTTTGGTGTGTTTGGGATGCCCAGCGCAGACTGCAATCGAGAGA
GGGGGCAATTCGGACCCCCTGGGAGTGTGGCTGGGACCCATGAGCTCTCT
GCCGGGGCCTCGACCCAAGCTTCGCTCAGCCATAGACTCGGGGGATAGGC
AGTGGGGGAGGCGGGGAGGAGAGACTAGTCCAGCAGGACCCCCCTGTTG
CTTTGCGGGGTGGGGATGGGGTGGGGGTGGGGGGCTGGTCCTTCTGGAAG
CCTTGCTCAGCTGGTCCTCTGGGATGTACATGCCCCGTGGTGGGGTGGGG
CCCTCTGCCATCCTGGCTGGTGCTGGGAGGCCCGAGGTGTGGCCACTGCC
CTGTGCTGGCCCCGCCCCCATGACATCATAGGACGGAGGCCAGGTCCAGT
GTCCAGCACTGGCCCCCGCAGCTCCGGGTGTGGCCGAGCCCCTGCAGTAA
TGACGATGGGGAGATGGGGACAGCGACCCCCATGTGAAGTTCCCGGCCC
GGTCCCCACCAGAGACGCGCTTCACCCACGGCGGCTCTTCGCTCTGGGGT
GGGATTCCATTCTGTGTCCCGCCTGCCTCGGCCACGAAAGCGCCTTTCACT
TCTGTGGCTGAAATGATTGACGAAGGGCGCAGCCGTGCTCGAAAGCCGCA
GCTGGACTTTCGGGGAGGAGCCGGCGGGGCTTCGTGTCTCCGACTGGAGC
AGGTCCTCATCCCTTAGAGGCCCCCTCCTCTTGAGACATGCTGGCTGGGC
GGGTGGCAGGAGAGGCCTGGCCCTGGACGTGGCGGGCGGCGGGCTCACG
CCTCTTGTCTTTGCCCCCAGATTCCAATGGGGATCCCGATGAGGAAGCCG
CTGCTCGTGCTGCTCGTCTTCTTGGCCTTGGCCTCGTGCTGCTATGCTGCTT
ACCGCCCCAGTGAGACTCTGTGCGGCGGGGAGCTGGTGGACACCCTCCAG
TTTGTCTGCGGGGACCGCGGCTTCTACTTCAGTAAGTAGCTCAGCGGGGC
ACGGGGGCGGGGCGGACACAGCAGGTGCTCCATCGGTGCTGCCCCGGTA
CCTGTGCGGGTCCTTCGGGATGGATGGTGTGGGGGACGGGGGGCGGGGG
GCGGCCAAGGGAGGACCTCTCCTCCGAGGGTCTGAGACTTCAGACCGGG
GGCGCCCTGGCCGTGCGCATTGATTGGCACCTGCCATGTGCCTGGCTGGG
GCTCACACCCCCTGACGTTCCTGCAGCGTGACTCGAAACGGGAAACCGAA
GGGACGGGTGGCACGGGGTGGGGAGGCAGACCGTGAGTGGCAGGCGTGC
GAGGGGTTCTTTCGGGCGGGGTGGCCCAGGCAGGCCCCACAGGATGACA
GCCTGTCCCCTCCTGCTCCTCCTTGACCTGCCCACAGCCAGGGCTGCAGGC
ACTGACATTCACCCATGGTATTGTGGTGCCTGACGTCTTGGCAGTGGGCAT
GGGTTCATGGACTGTTGGATTGAAAGTGGAATAAGATGGGTTGAAAACCA
ATAAGAATAAAGGCGCGTGTGGCTGGCGGCATCTGCGAGAGGTGACCGC
TGCCCTCCCTGGGGTTGGGCTTTGGGTGGGTTCCCATGGGTGGGGGCGGG
CCGCCATGCAGGGTGCCCGCCTGCTGGCCTCAGAGTGCTTTGCCGTCCTC
ATCTTTCTCTCTGGCCCCCGTCCCGCTCCTGAGGCTGGCTGGCTGGGCCCG
CGGAGACCTCCGCTCCCGCCTCGTCTGTGCCCAGGGAGCAGGGTGGACCC

TCCCTTGGGCTCTTGCCTGCACCTCCCAGCAGGCTGGGCCTCAGTGTCCTT
ACCTGTAGGATGGGTCAGGGGCGTCCTGGAGAGAGTCCTCGGGACAATG
GGGAGGCTGGGGGCAGGCCCAGCCTGACCCTGAAGGTGGGAGTGTGTGC
TCCCCCTGGGCTCAGCCAGCCGCGCTTGGGGCCGGGAGGGGGTGGGGGA
CGTGGCTGGGGCAAGTTGTCAAGGGCCGCGAGGCTCACCCCCGCCCATCG
CTCCCCATGTGGCAGCCTCTTCTGCAGCCTCTACTTACCCACCCTCTGAAA
TGGGCTGAAAACACCCATCTTGGCATGCCAAAGCTTCTCTGTAAAAAGCG
TTGCTGCTTCTTGATGCTTCTGAGGCCCCTGCCTGCCCTGGCCTCTGAGCC
CTCTCTCTCCTGCCTCGTTTGGGGGCAGGGAGTGGCACCATAGAATCTGG
CGCTGGGCCTGGGGAGCGGCCCCCTCGTGCCAGGCTTCCCCGAAAGGAGG
GCTGGGCTGAGCTCCCGACCCTCTGGACCCCTTACCAGGACCCCTTACCA
GGGGCTTCCCCCCCCCCCCCCCCCCGGTGGCGGCGGGCTGGGCTGGGGCC
TTTTCCTTGCAGCCGAGTCGGAGCTGTCGGAGGCGAGGGCGAGGACGGGA
AGAGAGGAGGGCGTGGTTTCTGCTGGTCCTCACTCCTCTCCTCCCGTCTTC
CTCCTCCTCCTCCCATTCCCACCTGTGTCTCCGGGTCCCGGGGCCGCAGGC
TGCCCAGGCGCCTGCTGATCCATTGGGGACCGCACTCGGGTCCCCGCTGG
CCTTCGGGTCAGGGCCACGGTCCACCTATTTTCCAAACAGCCTTGGGTCG
AGGCCCAAGAGGCTGGGCCCGGTTTAAGGACGGGGAGGGAGGCGCCAAG
AGGCCAGGGGCTGGTCCCGAGCACGCCCGCACCCGCTCACCCCCGCTGTC
CCCTCTCCTTCCCCGGGGGGCCCCTGTGCACCCCACTCTCACTTCTTCTGC
TCGAGGCCACGAGGCTGGCTGTCCCCGCAAGGTGACCGGGCGTCCTGTCT
GGAGGGCGGGGGCCGGGGCGGCTGGGGGCACCGTCCGTGCCCGGGGCCC
CTGTGCTGACGTGCCCTCCCCTTGGTCCTGTGGGACTTCCAGGCAGGCCG
GCAAGCCGCGTGAACCGCCGCAGCCGTGGCATCGTGGAAGAGTGCTGCTT
CCGTAGCTGCGACCTGGCCCTGCTGGAGACCTACTGCGCCACCCCCGCCA
AGTCCGAGAGGGACGTGTCGACCCCTCCGACCGTGCTTCCGGTAAGGCAG
CCCCTCTCTCGGCAGCGCCCCCCCCCNGGGGGGGGGCTGTCTCCTCTGAG
CCGGGGGACCGGGGCGCAGCCGGCTCTTGGGCTTCAAGTGCTGCCAGAG
GGGCCTTCCCCGCTGGGGACCCTGGCCAGAAGCCAGGGCAGTCTTCGCTC
TGTCGCAGGGCAGGCAGGCAGGAGGACCCCGCAGAGGTTGTTGTTCTGGG
ACAGGGGCTGGGGGGCCAGGCCCCCCCCTGACGGGCCCTTCCCCTCTCAG
GACAACTTCCCCAGATACCCCGTGGGCAAGTTCTTCCGCTATGACACCTG
GAAGCAGTCCGCCCAACGCCTGCGCAGGGGCCTGCCGGCCCTCCTGCGCG
CCCGCCGGGGTCGCACGCTCGCCAAGGAGCTGGAGGCGGTCAGAGAGGC
CAAGCGTCACCGACCCCTGACCGCCCGTCCCACCCGAGACCCCGCCGCCC
ACGGGGGCGCCTCTCCCGAGGCGTCCGGCCATCGGAAGTGAGCCAAATTG
TCGTAATTCTGCGGTGCCACCATCCACCTCGTGACCTCCTCTCGACCGGGA
CCGCTTCCATCAGGTCCCCCTTCTGAGATCTCTGTACCCTTCTGTCTGCGG
GCATCTCCGCCCCGGGCCCCGTGCCCCAACCTCCCCATGTCAGGCTAGTCT
CTCCTCGGCCCCTTCCATCGGGCCGAGGGCATCCAAACCACAAACCCAAT
TGGCTTGGTCTGTATCTCCCCCCAAATTATGCCCCCAATTATCCCCAAGTT
ACATACCAAAAATTGAACCCCTCAACCACACCCACATACAATCAGCCCCC
GTAAAACGAATTGGCATCTTTAAAACACCAGAAAAGCGAATTAGCTTTAA
AAAAAAAATAAACCCAAAATATCAATTAGCTGAAAAAAAAATACTAAAA
ATAAATTGGCTTAAAAACAATTGGCAAAATAAAAGAATTTGGCCCCCCCC
TTCCTTCTCTTTCTTTTCGGACCTTGAGTTAAATTGGCTGTGACCCATCATC

CAAGAGAAAGGAAGGGACCAAAATTTGCAGGTAGGCTTGTCGCCGCTCA
CAGCCATCTCCCTCCTCCTGCCACACCCTCGCCGGCCACTGGCGGTGTGGC
ACCAAGGACCCAGTCCCGTCCTCTCTCTAGTCCCATGACCGAGACCGCGG
TGGAGTTGGCTGGGAGACCCCGTGAGATCAGAGGAGGGGAGCACGGAAC
CAGAAACCCAAACCTGCACAGGTACAACATGACTGGCCCCCCGCACAGC
CCAAGACCTCTCATCTCAGTCTCCACTTAAAAAGCACCTGTACCCACACG
CATCCCTGCTGAAACACACACACACACACACACACACACGCACGCACGC
ACACACGCGCGCACGCACGCGCACACACACACTCATGCGTATACACACAC
ACACACGCACGCACGCGCACCCACACACACACATGCATTCACACACACAC
ACACTCGTGCATACACACGTGCGCGCGCACACACACACACACACACACTN
TTTTTTTNTTGGGGATCCCTGAGCCCAGCTATGTAGATTAGACCCCGGTCC
GTCCCAAATTCTTCTCAAAGCTGTCCCGAGATGAGAGATGAGGTTTTCGT
GTCCTGTGCTCTCCTCGCTTCCCCTGGGATGTGCCCTAGGGTGGGAGAGG
GTGTGTCCCAGGGCTCAGCAGGCGGTCCCATCTTCCCGAGACGGGAGAGA
TCCCCTCCTTCTCGGCGCCTGTCCCCACGGCCCCCACAGACACCCCCCCCC
CCGGCATGGCACCCATGCACCTGCCATCGTGCCCAGTAGGGGATGGGTTT
GGCGAGACTGGAGATGGCTGTAGCCAGTGAGACATGCCCTGCCACGTAG
CCTGACCCCCTGGGTGTGCTCTGTGAGATCTGGGGACCCCCAGCACACCT
AGGGATCATCTTTGCCAGCCTCCTGGGGAGCCTCTCAGAAATGGGGGCCC
CCAGAAGCTGGCAAAGGTGATGGGGAGCGTGGGAAGTCTGGCGGTTGGC
GGGGTGGGTGGGGGGCAGTGCGGGCTGGGTGGGGGGTGCTCCGGGGTCG
GAAGTGGTCCAGCAAGGTTTTGGACACAAAGTCAGGAGGAAGGAGTGAC
GAGGAGACTTGCAGAATTACAGGTAGAATCAGGAACCCACATCGACGCC
AATTGATCTATCCCCCCCTTTGATTGTTTTCTCCTGGGGCTTTTTTCCTTTTT
TTTTTTTTTTTTTTTTTTTTAATCCCTCCTTAGCTTTTTACGCGCTCAACACCA
AATTAAACGTACTCCCCACCCCACGTAACAGGGGGGCGGTGACCCGAAG
GACGAGGAGCACACGAAGCCACCATCCGTCACCTTGGCGGCACCAGCCG
CTGTCCTGCCCTCCGCCCATTTATCGCCCTTGATTGATTTTTGTTTTGCTCT
GTCCCTGTCGCTTGGGTAGAGTGGAGAAGGGAGCCTCTGTGGGGGTGCCA
GCCCACTGCGCCCCCCAGAGAATGTCAGGGGAATGGAGACGCTGCCTGCC
CAGCCTGGCTCCCGGGGATGGTGGCTGGTCCTCGGGGGTCCTGAGGGCGA
GGAGGGGGCCTGAGGTGTCTCCCCAGGTGTCAGAAGGTGCTCGGTGGCCA
CCAGGGTGGGGCCCCCGGCTGGCCCACCGGCCGGGGGGGGAAGGGGCTGC
AGAGGCTTGAGTGAGGGGCTCCGTCGGGCGGAAGCAGGTGACCACCTTC
CGTGCACTTGGGAGTCCTCTCCGCACCCCTCGCTGACACCGCGCCCGCCTC
TCAGCGCCCCGTCTGGCCCCCAGAAAGCTCGGAGCTCCGCGGGGCTTCCT
GGGGTCCAGGCGGGGTGAGGTGGGGCAGTAGGAGAGGTCCAGAGGGATG
AGCCCTGCTGAGCAAGAGAGCTGCGAGCGTGTGCGACAGCTGGGGTCAC
CCGGGCCCCGAGGCCCTCGGGTCTGGCGGTACCGACCCAGCCACCGTCTC
GGCGCCTAGGGCTGCGGCAGGGACCCCGGGCGGCTGGGAGGTTTACCTC
ACCCCCGCCTCTGCCCCCAGCGCAGCCCCCCTGCACAGCAGCCCGACTAG
CCAGCTAGAGGCCCGAAGCTTCTGGGCCCCGGTGACGGGGCTGGCACGA
CCCTGGGGGCGGTCCGTGCCAGCCCACCCGAGGCACAGAGGACCCTTGGC
CGGCGCCCGGGGAGTGGGCCATTCGGACCATTGGACAGAAGCCTAAAGA
ACCAAATTGTCCCCATTGGAGACCAGACAGATTGGCCTGAGATC

References

1. Abeydeera, L. R. and B. N. Day (1997). "Fertilization and subsequent development in vitro of pig oocytes inseminated in a modified tris-buffered medium with frozen-thawed ejaculated spermatozoa." Biol Reprod **57**(4): 729-734.
2. Abeydeera, L. R., W. H. Wang, R. S. Prather and B. N. Day (1998). "Maturation in vitro of pig oocytes in protein-free culture media: fertilization and subsequent embryo development in vitro." Biol Reprod **58**(5): 1316-1320.
3. Aigner, B., S. Renner, B. Kessler, N. Klymiuk, M. Kurome, A. Wunsch and E. Wolf (2010). "Transgenic pigs as models for translational biomedical research." J Mol Med (Berl) **88**(7): 653-664.
4. American Diabetes Association. (2013). Diagnosis and classification of diabetes mellitus, American Diabetes Association.
5. Blundell, T. L., J. F. Cutfield, S. M. Cutfield, E. J. Dodson, G. G. Dodson, D. C. Hodgkin, D. A. Mercola and M. Vijayan (1971). "Atomic positions in rhombohedral 2-zinc insulin crystals." Nature **231**(5304): 506-511.
6. Bonner-Weir, S. (2000). "Life and death of the pancreatic beta cells." Trends Endocrinol Metab **11**(9): 375-378.
7. Brown, W. R., N. C. Lee, Z. Xu and M. C. Smith (2011). "Serine recombinases as tools for genome engineering." Methods **53**(4): 372-379.
8. Butler, A. E., J. Janson, S. Bonner-Weir, R. Ritzel, R. A. Rizza and P. C. Butler (2003). "Beta-cell deficit and increased beta-cell apoptosis in humans with type 2 diabetes." Diabetes **52**(1): 102-110.
9. Campbell, J. D., M. S. Sansom and F. M. Ashcroft (2003). "Potassium channel regulation." EMBO Rep **4**(11): 1038-1042.
10. Carlson, D. F., W. Tan, S. G. Lillico, D. Stverakova, C. Proudfoot, M. Christian, D. F. Voytas, C. R. Long, C. B. Whitelaw and S. C. Fahrenkrug (2012). "Efficient TALEN-mediated gene knockout in livestock." Proc Natl Acad Sci U S A **109**(43): 17382-17387.
11. Cermak, T., E. L. Doyle, M. Christian, L. Wang, Y. Zhang, C. Schmidt, J. A. Baller, N. V. Somia, A. J. Bogdanove and D. F. Voytas (2011). "Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting." Nucleic Acids Res **39**(12): e82.
12. Cheng, J., Y. Wang, Z. Zhang, Y. Jin, Q. Li, R. Want, Y. Wang, X. Li, Q. Xiong, M. Zhang, R. Li and Y. Dai (2015). "Dominant-negative inhibition of glucose-dependent insulinotropic polypeptide impairs function of beta cells in transgenic pigs." J Biomed Res **29**(6): 512-514.
13. Chu, V. T., T. Weber, B. Wefers, W. Wurst, S. Sander, K. Rajewsky and R. Kuhn (2015). "Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells." Nat Biotechnol **33**(5): 543-548.

14. Cong, L., F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini and F. Zhang (2013). "Multiplex genome engineering using CRISPR/Cas systems." Science **339**(6121): 819-823.
15. Dimitriadis, G., P. Mitrou, V. Lambadiari, E. Maratou and S. A. Raptis (2011). "Insulin effects in muscle and adipose tissue." Diabetes Res Clin Pract **93 Suppl 1**: S52-59.
16. Doyon, Y., V. M. Choi, D. F. Xia, T. D. Vo, P. D. Gregory and M. C. Holmes (2010). "Transient cold shock enhances zinc-finger nuclease-mediated gene disruption." Nat Methods **7**(6): 459-460.
17. Ducluzeau, P. H., N. Perretti, M. Laville, F. Andreelli, N. Vega, J. P. Riou and H. Vidal (2001). "Regulation by insulin of gene expression in human skeletal muscle and adipose tissue. Evidence for specific defects in type 2 diabetes." Diabetes **50**(5): 1134-1142.
18. Eventov-Friedman, S., D. Tchorsh, H. Katchman, E. Shezen, A. Aronovich, G. Hecht, B. Dekel, G. Rechavi, B. R. Blazar, I. Feine, O. Tal, E. Freud and Y. Reisner (2006). "Embryonic pig pancreatic tissue transplantation for the treatment of diabetes." PLoS Med **3**(7): e215.
19. Ferrer, J., W. E. Scott, 3rd, B. P. Weegman, T. M. Suszynski, D. E. Sutherland, B. J. Hering and K. K. Papas (2008). "Pig pancreas anatomy: implications for pancreas procurement, preservation, and islet isolation." Transplantation **86**(11): 1503-1510.
20. Garrels, W., S. Holler, U. Taylor, D. Herrmann, H. Niemann, Z. Ivics and W. A. Kues (2014). "Assessment of fetal cell chimerism in transgenic pig lines generated by Sleeping beauty transposition." PLoS One **9**(5): e96673.
21. Gilbert, S. F. (2000). Developmental biology. Sunderland, Mass., Sinauer Associates.
22. Gittes, G. K. (2009). "Developmental biology of the pancreas: a comprehensive review." Dev Biol **326**(1): 4-35.
23. Gosmain, Y., C. Cheyssac, M. Heddad Masson, C. Dibner and J. Philippe (2011). "Glucagon gene expression in the endocrine pancreas: the role of the transcription factor Pax6 in alpha-cell differentiation, glucagon biosynthesis and secretion." Diabetes Obes Metab **13 Suppl 1**: 31-38.
24. Gosmanov, N. R., A. R. Gosmanov and J. E. Gerich (2000). Glucagon Physiology. Endotext. L. J. De Groot, P. Beck-Peccoz, G. Chrousos et al. South Dartmouth (MA).
25. Groth, A. C., E. C. Olivares, B. Thyagarajan and M. P. Calos (2000). "A phage integrase directs efficient site-specific integration in human cells." Proc Natl Acad Sci U S A **97**(11): 5995-6000.
26. Hammer, R. E., V. G. Pursel, C. E. Rexroad, Jr., R. J. Wall, D. J. Bolt, K. M. Ebert, R. D. Palmiter and R. L. Brinster (1985). "Production of transgenic rabbits, sheep and pigs by microinjection." Nature **315**(6021): 680-683.
27. Han, X. and B. E. Tuch (2001). "Cloning and characterization of porcine insulin gene." Comp Biochem Physiol B Biochem Mol Biol **129**(1): 87-95.
28. Hebrok, M., S. K. Kim and D. A. Melton (1998). "Notochord repression of endodermal Sonic hedgehog permits pancreas development." Genes Dev **12**(11): 1705-1713.

29. Inoue, H. (2015). "Central insulin-mediated regulation of hepatic glucose production [Review]." Endocr J.
30. Jeong, Y. H., C. H. Park, G. H. Jang, Y. I. Jeong, I. S. Hwang, Y. W. Jeong, Y. K. Kim, T. Shin, N. H. Kim, S. H. Hyun, E. B. Jeung and W. S. Hwang (2013). "Production of multiple transgenic Yucatan miniature pigs expressing human complement regulatory factors, human CD55, CD59, and H-transferase genes." PLoS One **8**(5): e63241.
31. Jinek, M., A. East, A. Cheng, S. Lin, E. Ma and J. Doudna (2013). "RNA-programmed genome editing in human cells." Elife **2**: e00471.
32. Keravala, A., J. L. Portlock, J. A. Nash, D. G. Vitrant, P. D. Robbins and M. P. Calos (2006). "PhiC31 integrase mediates integration in cultured synovial cells and enhances gene expression in rabbit joints." J Gene Med **8**(8): 1008-1017.
33. Kim, W., Y. K. Shin, B. J. Kim and J. M. Egan (2010). "Notch signaling in pancreatic endocrine cell and diabetes." Biochem Biophys Res Commun **392**(3): 247-251.
34. Kiselyov, V. V., S. Versteyhe, L. Gauguin and P. De Meyts (2009). "Harmonic oscillator model of the insulin and IGF1 receptors' allosteric binding and activation." Mol Syst Biol **5**: 243.
35. Larsen, M. O., B. Rolin, M. Wilken, R. D. Carr, O. Svendsen and P. Bollen (2001). "Parameters of glucose and lipid metabolism in the male Gottingen minipig: influence of age, body weight, and breeding family." Comp Med **51**(5): 436-442.
36. Lehninger, A. L., D. L. Nelson and M. M. Cox (2013). Lehninger principles of biochemistry. New York, W.H. Freeman.
37. Liu, M., I. Hodish, L. Haataja, R. Lara-Lemus, G. Rajpal, J. Wright and P. Arvan (2010). "Proinsulin misfolding and diabetes: mutant INS gene-induced diabetes of youth." Trends Endocrinol Metab **21**(11): 652-659.
38. Luo, Y., Y. Wang, J. Liu, H. Lan, M. Shao, Y. Yu, F. Quan and Y. Zhang (2015). "Production of transgenic cattle highly expressing human serum albumin in milk by phiC31 integrase-mediated gene delivery." Transgenic Res **24**(5): 875-883.
39. Maruyama, T., S. K. Dougan, M. C. Truttmann, A. M. Bilate, J. R. Ingram and H. L. Ploegh (2015). "Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining." Nat Biotechnol **33**(5): 538-542.
40. Mayo, K. E., L. J. Miller, D. Bataille, S. Dalle, B. Goke, B. Thorens and D. J. Drucker (2003). "International Union of Pharmacology. XXXV. The glucagon receptor family." Pharmacol Rev **55**(1): 167-194.
41. Miller, J. C., S. Tan, G. Qiao, K. A. Barlow, J. Wang, D. F. Xia, X. Meng, D. E. Paschon, E. Leung, S. J. Hinkley, G. P. Dulay, K. L. Hua, I. Ankoudinova, G. J. Cost, F. D. Urnov, H. S. Zhang, M. C. Holmes, L. Zhang, P. D. Gregory and E. J. Rebar (2011). "A TALE nuclease architecture for efficient genome editing." Nat Biotechnol **29**(2): 143-148.
42. Moehle, E. A., J. M. Rock, Y. L. Lee, Y. Jouvenot, R. C. DeKelver, P. D. Gregory, F. D. Urnov and M. C. Holmes (2007). "Targeted gene addition into

a specified location in the human genome using designed zinc finger nucleases." Proc Natl Acad Sci U S A **104**(9): 3055-3060.
43. Morel, P., D. B. Kaufmann, A. J. Matas, P. Tzardis, M. J. Field, J. K. Lloveras and D. E. Sutherland (1991). "Total pancreatectomy in the pig for islet transplantation. Technical alternatives." Transplantation **52**(1): 11-15.
44. Murakami, T., S. Hitomi, A. Ohtsuka, T. Taguchi and T. Fujita (1997). "Pancreatic insulo-acinar portal systems in humans, rats, and some other mammals: scanning electron microscopy of vascular casts." Microsc Res Tech **37**(5-6): 478-488.
45. Nir, T., D. A. Melton and Y. Dor (2007). "Recovery from diabetes in mice by beta cell regeneration." J Clin Invest **117**(9): 2553-2561.
46. Nussey, S. and S. Whitehead (2001). Endocrinology: An Integrated Approach. Oxford.
47. Pandol, S. J. (2010). The Exocrine Pancreas. San Rafael (CA).
48. Park, K. E., C. H. Park, A. Powell, J. Martin, D. M. Donovan and B. P. Telugu (2016). "Targeted Gene Knockin in Porcine Somatic Cells Using CRISPR/Cas Ribonucleoproteins." Int J Mol Sci **17**(6).
49. Park, K. E. and B. P. Telugu (2013). "Role of stem cells in large animal genetic engineering in the TALENs-CRISPR era." Reprod Fertil Dev **26**(1): 65-73.
50. Quesada, I., E. Tuduri, C. Ripoll and A. Nadal (2008). "Physiology of the pancreatic alpha-cell and glucagon secretion: role in glucose homeostasis and diabetes." J Endocrinol **199**(1): 5-19.
51. Raducanu, A. and H. Lickert (2012). "Understanding pancreas development for beta-cell repair and replacement therapies." Curr Diab Rep **12**(5): 481-489.
52. Renner, S., A. Blutke, E. Streckel, R. Wanke and E. Wolf (2015). "Incretin actions and consequences of incretin-based therapies: lessons from complementary animal models." J Pathol.
53. Renner, S., C. Fehlings, N. Herbach, A. Hofmann, D. C. von Waldthausen, B. Kessler, K. Ulrichs, I. Chodnevskaja, V. Moskalenko, W. Amselgruber, B. Goke, A. Pfeifer, R. Wanke and E. Wolf (2010). "Glucose intolerance and reduced proliferation of pancreatic beta-cells in transgenic pigs with impaired glucose-dependent insulinotropic polypeptide function." Diabetes **59**(5): 1228-1238.
54. Richter, B. and G. Neises (2003). "'Human' insulin versus animal insulin in people with diabetes mellitus." Cochrane Database Syst Rev(3): CD003816.
55. Rogers, C. S., D. A. Stoltz, D. K. Meyerholz, L. S. Ostedgaard, T. Rokhlina, P. J. Taft, M. P. Rogan, A. A. Pezzulo, P. H. Karp, O. A. Itani, A. C. Kabel, C. L. Wohlford-Lenane, G. J. Davis, R. A. Hanfland, T. L. Smith, M. Samuel, D. Wax, C. N. Murphy, A. Rieke, K. Whitworth, A. Uc, T. D. Starner, K. A. Brogden, J. Shilyansky, P. B. McCray, Jr., J. Zabner, R. S. Prather and M. J. Welsh (2008). "Disruption of the CFTR gene produces a model of cystic fibrosis in newborn pigs." Science **321**(5897): 1837-1841.
56. Sharma, P., F. Yan, V. A. Doronina, H. Escuin-Ordinas, M. D. Ryan and J. D. Brown (2012). "2A peptides provide distinct solutions to driving stop-carry on translational recoding." Nucleic Acids Res **40**(7): 3143-3151.

57. Smith, M. C., W. R. Brown, A. R. McEwan and P. A. Rowley (2010). "Site-specific recombination by phiC31 integrase and other large serine recombinases." Biochem Soc Trans **38**(2): 388-394.
58. Steiner, D. J., A. Kim, K. Miller and M. Hara (2010). "Pancreatic islet plasticity: interspecies comparison of islet architecture and composition." Islets **2**(3): 135-145.
59. Swindle, M. M., A. Makin, A. J. Herron, F. J. Clubb, Jr. and K. S. Frazier (2012). "Swine as models in biomedical research and toxicology testing." Vet Pathol **49**(2): 344-356.
60. Thakur, S., A. Jhobta, D. Sharma and C. S. Thakur (2014). "MR in complete dorsal pancreatic agenesis: Case report and review of literature." Indian J Radiol Imaging **24**(2): 156-159.
61. Wang, X., J. Zhou, C. Cao, J. Huang, T. Hai, Y. Wang, Q. Zheng, H. Zhang, G. Qin, X. Miao, H. Wang, S. Cao, Q. Zhou and J. Zhao (2015). "Efficient CRISPR/Cas9-mediated biallelic gene disruption and site-specific knockin after rapid selection of highly active sgRNAs in pigs." Sci Rep **5**: 13348.
62. Wang, Y., F. Wang, R. Wang, P. Zhao and Q. Xia (2015). "2A self-cleaving peptide-based multi-gene expression system in the silkworm Bombyx mori." Sci Rep **5**: 16273.
63. Wernersson, R., M. H. Schierup, F. G. Jorgensen, J. Gorodkin, F. Panitz, H. H. Staerfeldt, O. F. Christensen, T. Mailund, H. Hornshoj, A. Klein, J. Wang, B. Liu, S. Hu, W. Dong, W. Li, G. K. Wong, J. Yu, J. Wang, C. Bendixen, M. Fredholm, S. Brunak, H. Yang and L. Bolund (2005). "Pigs in sequence space: a 0.66X coverage pig genome survey based on shotgun sequencing." BMC Genomics **6**: 70.
64. Whitelaw, C. B., T. P. Sheets, S. G. Lillico and B. P. Telugu (2015). "Engineering large animal models of human disease." J Pathol.
65. Whitelaw, C. B., T. P. Sheets, S. G. Lillico and B. P. Telugu (2016). "Engineering large animal models of human disease." J Pathol **238**(2): 247-256.
66. Wolf, E., C. Braun-Reichhart, E. Streckel and S. Renner (2014). "Genetically engineered pig models for diabetes research." Transgenic Res **23**(1): 27-38.
67. Yang, D., C. E. Wang, B. Zhao, W. Li, Z. Ouyang, Z. Liu, H. Yang, P. Fan, A. O'Neill, W. Gu, H. Yi, S. Li, L. Lai and X. J. Li (2010). "Expression of Huntington's disease protein results in apoptotic neurons in the brains of cloned transgenic pigs." Hum Mol Genet **19**(20): 3983-3994.
68. Yoshimi, K., Y. Kunihiro, T. Kaneko, H. Nagahora, B. Voigt and T. Mashimo (2016). "ssODN-mediated knock-in with CRISPR-Cas for large genomic regions in zygotes." Nat Commun **7**: 10431.
69. Yoshioka, K., C. Suzuki, A. Tanaka, I. M. Anas and S. Iwamura (2002). "Birth of piglets derived from porcine zygotes cultured in a chemically defined medium." Biol Reprod **66**(1): 112-119.
70. Zhang, X. H., L. Y. Tee, X. G. Wang, Q. S. Huang and S. H. Yang (2015). "Off-target Effects in CRISPR/Cas9-mediated Genome Engineering." Mol Ther Nucleic Acids **4**: e264.

71. Zhou, X., J. Xin, N. Fan, Q. Zou, J. Huang, Z. Ouyang, Y. Zhao, B. Zhao, Z. Liu, S. Lai, X. Yi, L. Guo, M. A. Esteban, Y. Zeng, H. Yang and L. Lai (2015). "Generation of CRISPR/Cas9-mediated gene-targeted pigs via somatic cell nuclear transfer." Cell Mol Life Sci **72**(6): 1175-1184.